2018

Electroencephalography (EEG) Based Analysis of Human Brain Activities for Audio Stimuli

Ali, Md. Sujan

University of Rajshahi

UNIVERSITY OF RAJSHAHI

DOCTORAL THESIS

Electroencephalography (EEG) Based Analysis of Human Brain Activities for Audio Stimuli

Author:
MD. SUJAN ALI
Signal Processing & Computational Neuroscience LAB
Department of Computer Science & Engineering
Roll No.: 13713
Registration No.: 10033
Session: 2013-2014

Supervisor:
DR. MD. KHADEMUL ISLAM MOLLA
Co-supervisor:
DR. MD. EKRAMUL HAMID

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh.

JUNE 2018

University of Rajshahi

Department of Computer Science and Engineering

Rajshahi, Bangladesh

Certificate

This is certify that the Dissertation entitled **"Electroencephalography (EEG) Based Analysis of Human Brain Activities for Audio Stimuli",** submitted by **Md. Sujan Ali** is a record of original work carried out by him, in the partial fulfilment of the requirement for the Degree of Doctor of Philosophy at the Department of Computer Science and Engineering, University of Rajshahi. This work is done during July 2013 - June 2018, under our guidance. The contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree.

Prof. Dr. Md. Khademul Islam Molla	Prof. Dr. Ekramul Hamid
Supervisor	**Co-supervisor**

Date:

Declaration of Authorship

I, MD. SUJAN ALI, declare that this thesis titled, "Electroencephalography (EEG) Based Analysis of Human Brain Activities for Audio Stimuli" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"If the brain were so simple that we could understand it, we would be so simple we couldn't."

Emerson M. Pugh

Abstract

A brain-computer interface (BCI) uses signals originated from the brain to direct some peripheral devices. This thesis deals with multichannel electroencephalography (EEG) based BCI implementation. Synchronization of neural activity between different parts of human brain has great significance in coordination of cognitive activities. The time-frequency (TF) representation can be used to measure the synchronization between different channels of EEG. The technique, called synchrosqueezing transform (SST) is one of the techniques that operates with the continuous wavelet transform (CWT) and produces impressively localized time-frequency representations of nonlinear and nonstationary signals. The SST based method is proposed to effectively measure the synchronization in TF domain. Due to its data adaptability and frequency reassignment properties, the SST produces a well-defined TF representation. The marginal time coherence for different channel pairs is used to quantify synchronization. The experiment is performed with both synthetic and real EEG data. The results show that the marginal time coherence based on the proposed SST exhibits very clear discrimination between two types of motor imagery (MI) movement.

This research also presents a novel method for the selection of effective spatial filter pair and discriminative features in EEG based MI classification. Usually, the spatial filter pair is selected manually. However, the manual selection of CSP filters does not confirm that the approach will achieve the best accuracy. In the proposed method, the analyzing EEG data is divided into training and test sets. The training set is used to select appropriate spatial filters with dominant features. To accomplish such features, the EEG of training set is segmented again into two subsets termed as training subset and test subset. The features of both subsets are extracted using common spatial pattern (CSP). The mutual information between the features of training subset and class levels of the training subset is calculated. Then features of training subset are ranked on the basis of values of the mutual information. Besides, the features of test subset are also ranked according to the order of the training subset features. The initial classification performance using training and test subsets is obtained by using linear discriminant analysis (LDA). Then a grid search method selects the effective number of spatial filter pairs as well

as the discriminative features by which the maximum accuracy score is yielded. Thus the selected spatial filter and features are used in actual classification accuracy of the test set of EEG. In this research, the binary classification performance of the proposed approach is evaluated to classify MI data where the datasets are widely used as the publicly available dataset from BCI competition III. The approach achieves more increased mean accuracies than different existing methods of MI tasks. Finally, the method is verified to classify audio stimuli based EEG. In auditory EEG, the proposed approach produces superior classification accuracy compared to prevailing methods.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor Professor Dr. Md. Khademul Islam Molla for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. It has been honor to be his first Ph.D student. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D study. Besides my supervisor, profound gratitude goes to my co-supervisor Professor Dr. Md. Ekramul Hamid, who has been a truly dedicated mentor. I appreciate all his contributions of time, ideas, comments and critical assessments to make my Ph.D.

I express my heart-felt gratitude to Professor Dr. Somlal Das, Ex-chairman and Professor Dr. Bimal Kumar Pramanik Chairman of the Department of Computer Science and Engineering for always being helpful and motivating. I would like to acknowledge the entire faculty member of the Department of Computer Science and Engineering for their kind help at various phases of this research.

I would like to extend my sincere word of thanks to my fellow lab mates, Jannat, Rabiul, Sanjoy, Shimul, Talha, Mussa for always being there and bearing with me the good and bad times during my wonderful days of Ph.D.

I acknowledge the Information and Communication Technology (ICT) division of the ministry of Post, Telecommunication and Information Technology, Bangladesh for providing me funding and fellowship to pursue research at University of Rajshahi.

Finally, I would like to acknowledge the sacrifice of my parents, maternal parents, family members, my sisters and brothers. I do not imagine a life without their love and blessings.

Dedication

This thesis is dedicated to my beloved kids Janan and Manan whom I can't force myself to stop loving. This work is also dedicated to the people with locked-in syndrome who encouraged me to pursue the research.

Contents

List of Figures

List of Tables

List of Abbreviations

AEPs	Auditory Evoked Potentials
AER	Auditory Evoked Response
ALN	Adaptive Logic Network
ALS	Amyotrophic Lateral Sclerosis
AM	Amplitude Modulation
ARTMAP	Adaptive Resonance Theory MAP
ASA	Auditory Selective Approach
ASSR	Auditory Steady State Response
BCI	Brain Computer Interface
BLRMN	Bayesian Logistic Regression Neural Network
BMI	Brain Machine Interface
CLIS	Completely Locked-in-State
CSP	Common Spatial Pattern
CWT	Continuous Wavelet Transform
DNA	Deoxyribonucleic Acid
DOC	Disorder of Consciousness
ECG	Electrocardiography
ECoG	Electrocorticography
EEG	Electroencephalography
EMCS	Emerged from Minimally Conscious State
EMG	Electromyography
EP	Evoked Potentials
ERD	Event Related Desynchronisation

ERP	Event Related Potential
ERS	Event Related Synchronisation
FIRNN	Finite Impulse Response Neural Network
fMRI	functional Magnetic Resonance Imaging
GCS	Glasgow Coma State
GDNN	Gamma Dynamic Neural Network
GMM	Gaussian Mixture Models
GS-CSP	Grid Search Common Spatial Pattern
HCI	Human Computer Interface
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IFCN	International Federation of Clinical Neurophysiology
IOHMM	Input Output Hidden Markov Model
kNN	k Nearest Neighbor
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode
LiS	Locked-in-Syndrome
LVQ	Learning Vector Quantization
MCS	Minimally Conscious State
MEG	Magnetoencephalography
MI	Motor Imagery
MI-BCI	Motor Imagery Brain Computer Interface
MLP	Multilayer Perception
MMN	Mismatch Negativity
MTG	Middle Temporal Gyrus
NIRS	Near Infrared Spectroscopy
NN	Neural Network
PCA	Principal Component Analysis
PeGNC	Probability estimating Guarded Neural Network
PET	Positron Emission Tomography

pSTS	posterior Superior Temporal Sulcus
RBF	Radial Basis Function
SFPF	Spatial Filter Pair and Features
SNR	Signal to Noise Ratio
SOM	Self Organizing Map
SSEP	Steady State Evoked Potentials
SSEPs	Somatosensory Evoked Potentials
SST	Synchrosqueezing Transform
SSVEP	Steady State Visual Evoked Potentials
STFT	Short Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TDNN	Time Delay Neural Network
TF	Time Frequency
TFR	Time Frequency Representation
UWS	Unresponsive Wakefulness Syndrome
VEP	Visual Evoked Potential
VS	Vegetative State
WOSA	Welchs Overlapped Segment Averaging
WT	Wavelet Transform

List of Symbols

$s(\mathrm{t})$: Non-stationary signal

$\delta(\mathrm{t})$: Synthetic signal

$*$: Complex conjugate

$w(\mathrm{t})$: Window function

$\Phi(\mathrm{t})$: Mother wavelet

Γ : Synchrosqueezing transform operator

$\bigotimes$: Convolution operator

σ : Standard deviation

λ : Eigenvalue

$\overrightarrow{\nu}$: Eigenvector

Σ : Matrix of eigenvectors

Λ : Feature vectors

m : Spatial filter pair

θ : LDA accuracy score

Θ : LDA test operator

κ : Discriminant hyperplane coefficient

Chapter 1

Introduction

1.1 Brain Computer Interface

Brain-computer interface (BCI) is a connection between a brain and an external device that facilitates signals from the brain to direct some peripheral action, which controls a cursor or a prosthetic limb. This type of interface enables a direct communications pathway between the brain and the device to be controlled. The cursor is controlled by the brain signal which is transmitted directly from the brain to the mechanism directing the cursor, rather than taking the normal route through the body's neuromuscular system from the brain to the finger on a mouse. By capturing signals from an array of neurons of human brain and to translate it into action using computer chips and programs, BCI can enable a person suffering from paralysis to write a book or control a motorized wheelchair or prosthetic limb through thought alone. Present brain-interface devices need deliberate conscious thought; some future applications, such as prosthetic control, are likely to work effortlessly. For the development of BCI technology development of electrode devices and surgical methods are important that are minimally invasive. In the existing BCI model, the brain accepts an implanted mechanical device and controls the device as a natural part of its representation of the body. Recently, the current research is focused on the potential on non-invasive BCI.

BCI is the most recent development of Human Computer Interface (HCI). Without use of traditional input devices it reads the waves produced from the brain at different locations in the human head, translates these signals into actions, and

commands to control the computer. The BCI system can be employed in many applications especially for disabled persons such as [168]: (1) new ways for gamers to play games using their heads, (2) social interactions; enabling social applications to capture feelings and emotions, (3) helping partially or fully-disabled people to interact with different computational devices, and (4) helping understanding more about brain activities and human neural networks. The BCI applications depend on the basic understanding of how the brain works. Moreover, it utilizes the brain and its nervous system functions where the humans central nervous system consists of the spinal cord and the brain. The function of the BCI is to process and integrate incoming sensory stimuli received via peripheral nerves and to give impulses back to actuators, e.g. to muscles or glands which cause automatic or voluntary action. Additionally, the central nervous system, mainly the brain, is responsible for higher integrative abilities such as thinking, learning, production, and understanding of speech, memory, emotion etc. Moreover, respiration and the cardio-vascular system are controlled by the central nervous system. The activities of central nervous system are measured by the BCI system and translate the data into an output, suitable for a computer to use as an input signal. Both human and animals are facilitated by the brain computer interface system. A Monkey in 2008 [189] was able to move a screen cursor as well as controlling a robot arm. With this study it is possible to know how animals can think and discover their brains as well. Moreover, BCI system is used with different human patients capturing their brain signals. This system goes beyond a communication tool for people are not able to communicate. Not only that the BCI is gaining more attention from healthy people for other purposes such as rehabilitation or hands-free gaming. Nevertheless, one of the big challenges is that the BCI tools are not available and yet need to deal with them.

Even though laymen often think of BCI as being a very new field, the concept and term Brain-Computer Interface was introduced by Jacques J. Vidal in 1973 already [238]. The BCI system reads signal from brain, analyzes and interprets and translates the interpretation into actions [62, 246]. Suppose a player who

is sitting in front of a pinball machine with the task to control the flippers of the pinball machine with his thoughts [129] can be an instance of a BCI system. The brain activity of the player can be measured by the Electroencephalography (EEG). To do that the player need to wear an EEG cap with electrodes, which are measuring the current on the scalp of the subject, induced by the brain activities under the scalp. The measured EEG data is amplified and feed into a computer. The computer receives and interprets the continuous data stream for each EEG channel.

Assume that the computer is capable to translate the subjects imagined left or right hand movement into a signal that says left or right whenever the subject imagines the respective hand movement. The pinball machine is connected by the computer and moves the respective flippers whenever it receives the left or right signals. In this way the subject is capable to play pinball using BCI. In this example the BCI system functions by three steps. i) The brain activity is measured by the EEG system, the signal acquisition. This step collects the brain signals and translates it into a data stream appropriate for the computer. ii) The second step is the signal processing step in which the measured EEG data is translated into the output signal. iii) This step is the pinball machine that translates the output of the signal processing into an action. Also this step is called the feedback. A feedback can come in many colors and shapes: it can be the moving flipper of the pinball machine, or could it be a prosthesis [172], a virtual keyboard [71] or even the steering wheel of a car [258]. The most significant feature of a feedback is that it translates the output signal into some kind of action. BCI researchers are concerned in how the brain reacts to certain stimuli, and the feedback part is replaced by a stimulus presentation. In that situation, the BCI system is not used to translate thoughts into action, but rather as a sophisticated measuring system. With an experiment a researcher can try to calculate how well a subject can memorize can be an example for such setup. Suppose wearing an EEG cap a subject is sitting in front of a computer monitor and presented with Chinese symblols. The subject's task is to memorize the symbols as good as possible while

the EEG system is recording the brain activity.

To find out the correctly memorized symbols the subject can perform a pen and paper test without EEG after some days. Now the researcher can try to correlate the correctly and incorrectly memorized symbols with certain patterns in the EEG signals that occurred while the subject was trying to memorize that specific symbol. The success of the learning can be predicted on the basis of correlation and thus provides additional feedback to the stimulus. Actually it is needed to realize that a BCI system is not a mind reading device and cannot read thoughts. The system can only classify very specific patterns of brain activity that the subject voluntarily or involuntarily produces. With the attention and motor imagery based BCI the most useful patterns for the BCI can be produced.

1.1.1 Historical Views

Communication ability of human being is the distinguishing features of them with other members of the species. Humans can express themselves and share thoughts, emotions and experiences with other human beings and can communicate in various ways throughout time. In the early age of human evolution they used to communicate by drawing figures on the walls and by telling stories. Written books started playing vital role in communication through the introduction of printing press. At the present time, all types of communication totally depend on the advancement in technology. With the advancement of the internet communication between computer systems has reached very advanced levels. Recently the research is going to establish an effective communication between humans and machines. There are various diseases that can severely damage and destroy the neuromuscular pathways which link the brain to muscles. The brain cannot communicate with external environment in the absence of these links. People can be affected and lose their motor abilities with diseases amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury and multiple sclerosis such as eye movement, arm and leg movement. Even more, the patients lose all voluntary muscle control and become locked in their bodies. In the case of not all

pathways have been destroyed, people can try to make more use of the remaining intact pathways. Largely paralyzed patients can still use eye movements to answer simple yes/no questions can be an example.To bypass the breaks in the damaged pathway through other neighboring intact pathways can be an alternative way. Patients with spinal cord injury are able to restore movement and control to paralyzed muscles through electromygraphic (EMG) activity from neighboring muscles. However, in the case of locked in patients where all their neuromuscular junctions are destroyed the mentioned solutions are not applicable.

A person is cognitively intact but the body is paralyzed and the paralyzed means that any voluntary control of muscles is lost. In this case, the people cannot move their arms, legs, or faces, and depend on an artificial respirator. In this condition the only effective way to communicate with the environment is with a device that can read brain signals and convert them into control and communication signals. Such a device is called a brain computer interface (BCI). Before few decades, controlling devices with brain waves was considered pure science fiction, as wild and fantastic as warp drive and transporters. In 1929, the German scientist Hans Berger recorded the electrical brain activity from the human scalp. The necessary technologies for measuring and processing brain signals as well as our understanding of brain function were still too limited. The intuition was that EEG signals must carry necessary information and thoughts. However, decoding them was thought to be impossible because of some considerations. The first one is that the EEG is a very complex signal and hence its content reflects not only the activity of certain limited cells but the electrical activity of trillions of synapses in the cortex. Another one is that EEG-based BCIs would require that the EEG is analyzed in real time and at that time, that was either not possible or extremely expensive and hence very unpractical. Therefore, possible applications of an EEG BCI were highly under-estimated.

Currently, neuroscience research over the last decades has led to a much better understanding of the brain and hence the situation has changed. With the development of signal processing algorithms complex real-time processing of brain signals

does not require expensive or bulky equipment anymore. Even though the EEG is truly a complex signal, its main characteristics are very well understood today. Nowadays, researchers know not only the sites and origin of the EEG rhythms and potentials but also their relationship with specific features of the brains state and function. It is proved that EEG signals are highly correlated with imagined movements and mental activities, and hence BCIs are no longer thought to be fictional. The developments of BCI research were very slow for many more years. But, BCI research advanced rapidly after that, particularly during the last few years. BCI research is proving that BCIs can work with patients who need a BCI to communicate. Nowadays, BCI researchers have used many different kinds of BCIs with several different patients. Besides, BCIs are moving beyond communication tools for people who cannot otherwise communicate. Moreover, BCIs are gaining attention for healthy users and new goals such as rehabilitation or hands-free gaming.

However, there are still many practical challenges before a typical person can use a BCI without expert help. There is a long way to go from providing communication for some specific patients, to provide a range of functions for any user without help.

1.1.2 Necessity of Brain-Computer Interfaces

It is very difficult to use the conventional technologies such as a mouse or keyboard to operate a personal computer with the people of physical disabilities. To establish an alternative communication an input device or software are designed so that the disable people can use a computer. Some examples of such devices are single switches of various forms, head mice, speech recognition systems, eye gaze trackers and even software for touch screen devices. Locked-in syndrome (LiS) is the most severe form of motor disability. The term, thought up by Plum and Posner in 1996, talk about a state in which a person is almost completely paralyzed yet remains cognitively aware [120].

In the case of classical LiS, vertical eye movements, including blinking, are

possible. In partial LiS, some additional residual muscle movement such as in a finger, toe or head has been recovered. On the other hand, in total LiS, the person has lost control of even eye movements [11]. Study of cognitive consciousness can be achieved by manual cognitive assessment and is sometimes proved by neuroimaging techniques such as magnetic resonance imaging (MRI) or electroencephalography (EEG). Usually LiS are causes due to a lesion in the pons in the brainstem and neuro-degenerative diseases such as in amyotrophic lateral sclerosis (ALS) [153]. However, a large proportion of people who find themselves in such a state of being are willing and able to continue living for many years with a good quality of life [59]. But empowering people to communicate is essential to maintain a good quality of life. In the classical LiS, the most common and possibly the most efficient means of communication is via eye blinks or eye gaze to a care giver or other human being. One can either communicate 'yes' or 'no', or spell words by indicating the desired letters on an alphabet board. The development of the personal computer and the Internet also improves the quality of life of people who would otherwise be even more isolated from the rest of the world. Eye gaze trackers and eye blink technology allow some people access to computers, and persons with partial LiS can use a single switch device [174]. On the other hand, for people with total LiS, it is difficult to communicate using any overt muscle movement. For locked-in patients, the development of brain-computer interfaces (BCI) is only the way, a direct communication pathway between the brain and an external device that records neural processes. A brain-computer interface (BCI) is a method that aims to extract a user's intention whilst bypassing the normal modality of physical movement by measuring and analysing brain signals.

1.1.3 Attention Based BCI

The BCI system based on attention is work with different type of stimuli. The stimuli are visual [71, 44, 2], auditory[122, 217], or tactile [173]. The visual stimulus is widely used in the BCI. Each of the stimuli is associated with a particular action, like the movement of a part of prosthesis, or the selection of a letter from

the alphabet. The subject focuses his attention on the desired stimulus and ignores the others. During this attention the brain patterns are generated and recorded by devices. The BCI system explores and interprets the pattern and executes the desired action [62, 246].

In most commonly used visual attention based BCI two different brain patterns can be used: event-related potentials (ERP) and steady-state visually evoked potentials (SSVEP). In the case of ERP based BCI, the stimuli are presented successively and for a very short time. In the BCI the duration of a stimuli presentation is usually a few milliseconds and the time between two stimuli about 100ms. During the stimulation the subject focuses on particular stimulus and the subject yields a specific brain pattern that is different from the brain patterns that arise from stimuli the subject is not focusing his attention on. The noticeable pattern is an event-related positivity in the centro-parietal areas around 300ms after the presentation of the stimulus and this is called P300. The P300 is recognized by the ERP based BCI system and hence the targets (the stimuli the subject was attending to) and the non-targets (the stimuli the subject was not attending to) are distinguished. Unlike ERP based BCI, in SSVEP based BCI the stimuli are presented continuously, all at the same time, and flickering with different frequencies between 6-30Hz. The target is selected by the subject by focusing on a particular stimulus. The steady-state visually evoked potentials (SSVEP) are produced by the flickering frequency of the stimulus with the same frequency in the visual cortex. The flickering of the SSVEP depends on the concentration on a stimulus flickering. For instance, if the subject concentrates on a stimulus flickering with 30Hz, the SSVEP will also flicker with 30Hz. The stimulus the subject was attending to can be detected by the BCI system by matching the frequency of the SSVEP with the ones of the stimuli. The classification performance of the visual attention based BCI is very high. The Visual attention based BCI is very dependable and allows in a speller based paradigm for a typing speed of roughly one sentence per hour. The different visual paradigms need the subject to have control over his eye movements in order to work accurately, which is not always

a given for patients, especially the ones suffering from locked-in syndrome. The gaze independent visual spellers [234], or the use of auditory- or tactile spellers can be a way out which requires no voluntary muscle activity. Moreover, people dislike the constant flickering on the screen and hesitate to use over time. Hence the visual speller faces this type of practical problem.

1.1.4 Motor-Imagery Brain Computer Interfaces

Motor imagery is a thought in which a subject can perform a movement without actually performing the movement and without even tensing the muscles. In the motor imagery, one can mentally rehearse the movement of the affected body parts, without ever actually attempting to perform the movement. It requires the sensible activation of brain regions that are also involved in movement preparation and execution, accompanied by a voluntary inhibition of the actual movement.There are two distinct types of motor imagery: kinesthetic motor imagery and visual motor imagery. During kinesthetic motor imagery the subject is imagining the feeling associated with performing a movement. During visual motor imagery the subject is imagining the movement itself.

The input of BCI system depends on its capacity to perceive mental states that are willingly produced by the user. The motor imagery brain computer interfaces (MI-BCIs) are a subset of the BCIs that are based on the voluntary potential. Usually a reference EEG electrode is attached to an electro physiologically neutral part of the body, such as the ear lobe. The EEG measures the voltage potential difference between each electrode in the EEG montage and the reference electrode. In the BCI system, the standard international 10-20 system of electrode placement is used to measure the potential difference. The EEG signal is generated with the firing of a large numbers of neurons at a specific location of the brain. The neuron oscillations are categorized by the amplitude in different frequency bands. These motor neurons are connected with the sensation of different body parts. During motor imagination, the firing of a large numbers of neurons in synchrony produce oscillation in the 7–13 Hz range α rhythm and the 15–30 Hz range β rhythm [198].

An event-related desynchronization (ERD) reduces the amplitude of the related frequency bands in which the synchrony is interrupted by the actual movement of particular body parts. An increase in the synchrony is referred to as event-related synchronisation (ERS). The ERD and ERS can be localized according to certain body parts that are being imagined. The mental states are classified on the basis of the localization. The motor imagery based BCI is the common choice for mental-state based BCIs. Although research of BCI is growing rapidly, still the performance of MI-BCI is not an acceptable level [90, 73]. The reasons for this are under vigorous exploration and several approach have been made into discovering why this may be the case. By using inaccurate mental strategy one cannot control the MI-BCI. Besides, the inter- and intra-subject variation may affect the ability to operate an MI-BCI. An up-to-date review of the current literature on performance variation in BCI is presented in [89].

1.1.5 Brain and Brain Data Acquisition

This section briefly describes the anatomical and physiological structure of human brain. The human brain generates electrical activity that can be recorded on the scalp. This section focuses on the origin and generation of the electrical activity. We should know the basics of brain cell, the neuron in order to properly understand the creation of local current flows within the brain.

The neuron: Neurons also called nerve cells are the necessary units of the brain and nervous system. The nerve cell receives sensory input from the external world and sends motor commands to our muscles. Neurons transmit electrical signals and pass messages to each other over long distances. On account of the flow of ion-based electrical currents, the biometric potentials are observed within the body. The electrical potentials recorded on the scalp are the summation of electrical potentials of many neurons of human brain. During creation the human brain is made up of approximately 100-billion neurons at an average density of 10^4 neurons per cubic mm[181]. With the increase of age of the brain the number

of neurons decreases. The three main parts of a neuron are dendrites, an axon and a cell body which can be represented as the branches, roots and trunk of a tree, respectively. A dendrite is where a neuron receives input from other cells and allows communication with other neurons. The axon is the output structure of the neuron. The neuron carries the electrochemical message (action potential) throughout the entire axon. The cell body is the main part where the necessary components of the cell, such as the nucleus, neuron's deoxyribonucleic acid (DNA) and proteins are housed. There are different types of neurons and their various structures depend on their function. Examples of such type of cells are sensory, motor and cortical pyramidal cell neurons. The most dominant neuron cell is the pyramidal neuron cell of the cerebral cortex, mainly in the cortical peaks and valleys that are parallel to the scalp. The electrical activity recorded by the EEG is mostly depended on the pyramidal neuron cell. The long straight dendrite of the neuron extends up and down towards the surface of the brain. The electrical potentials are accumulated by parallel dendrites in the cerebral cortex. Additionally, there are many neighbouring neurons that will have the same presynaptic sources causing a synchrony of potentials that can be readily picked up on the scalp. The local current flow occurs when neurons are activated by means of an electrochemical concentration gradient.

The brain and its functions: The three parts of a human brain are the brain stem, the cerebellum and the cerebrum. Also, the cerebrum is divided into two parts left hemisphere and right hemisphere. The surface of the cerebrum is called the cortex and is divided into four zones called lobes as shown in Figure 1.1. The frontal lobe of the cerebrum is used for cognitive functions such as speech, movements and other executive functions. The temporal lobe deals with auditory, visual and language functions. It is also used for emotional processing. The parietal lobe is used for reading/writing, language comprehension, attention and spatial awareness. Lastly, the occipital lobe is involved in visual processing [87].

The brain activity is recorded by using electrical fields, blood pressure, or

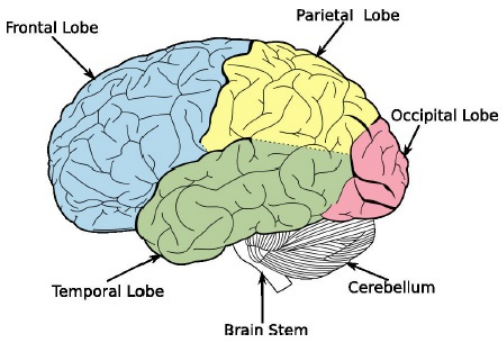


Figure 1.1: Lateral view of the brain lobes. Derived from[87].

magnetic fields. Recording and collecting the data is called acquisition. The brain data can be recorded with one of the two methods called invasive or non-invasive. In invasive method electrodes are placed on the brain tissues. On the other hand, in non-invasive method, the measurement device is placed on the scalp. Based on the electrode placement a BCI is called invasive or noninvasive [159].

More accurate data can be recorded using the invasive systems but need a surgical operation to be placed on the user and have to be removed or replaced after some time due to the rejection phenomenon. The invasive BCIs are mainly used in electrocorticography [139]. The non-invasive BCIs are used in various methods: EEG [177], near-infrared spectroscopy [32], magnetoencephalography [95], and functional magnetic resonance imaging [185].

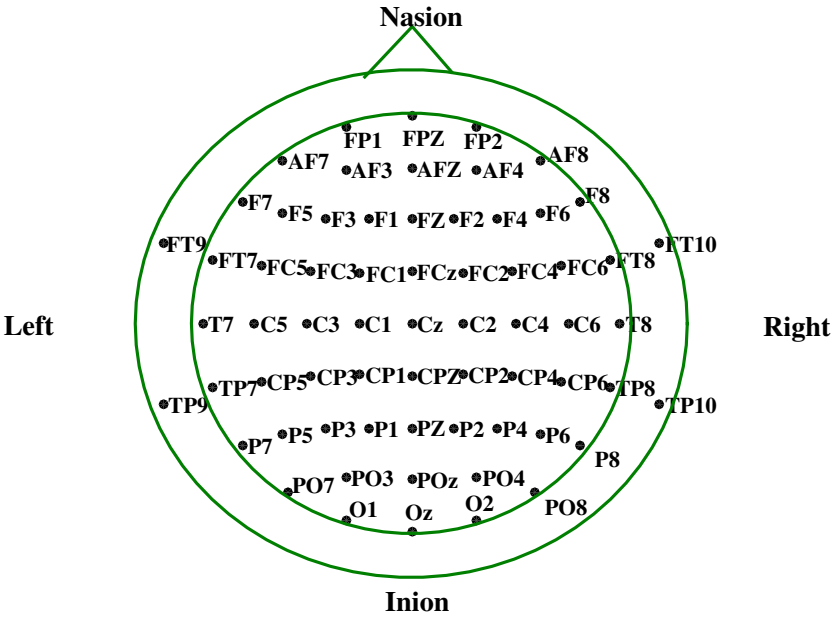
The functionality of the methods varies with cost, efficiency and mobility. Since the use of EEG is easy and low cost compared to magnetoencephalography or functional magnetic resonance imaging devices, most of the BCI systems use it. To collect data from the brain electrodes are placed on the scalp which emits the electrical potentials [177]. The EEG are often used in non-invasive medical applications of BCIs. Two amplifiers named BrainAmp and the g.USBamp are usually used to amplify the recording with EEG electrodes. Recently, new non-medical uses of EEG-based BCIs have appeared in fields like entertainment and video-games. The Emotiv EPOC, a lost cost EEG hardware is produced. A medical and a non-medical EEG device in a specific application is compared by Duvinage et al. [157]. To record brain data the EEG electrodes are placed on the

surface of the head, often attached to a cap. Gel or salted water is used to enhance the conductivity between the head and the electrodes. Electrodes are commonly placed using the 10-20 system as shown in Figure 1.2.

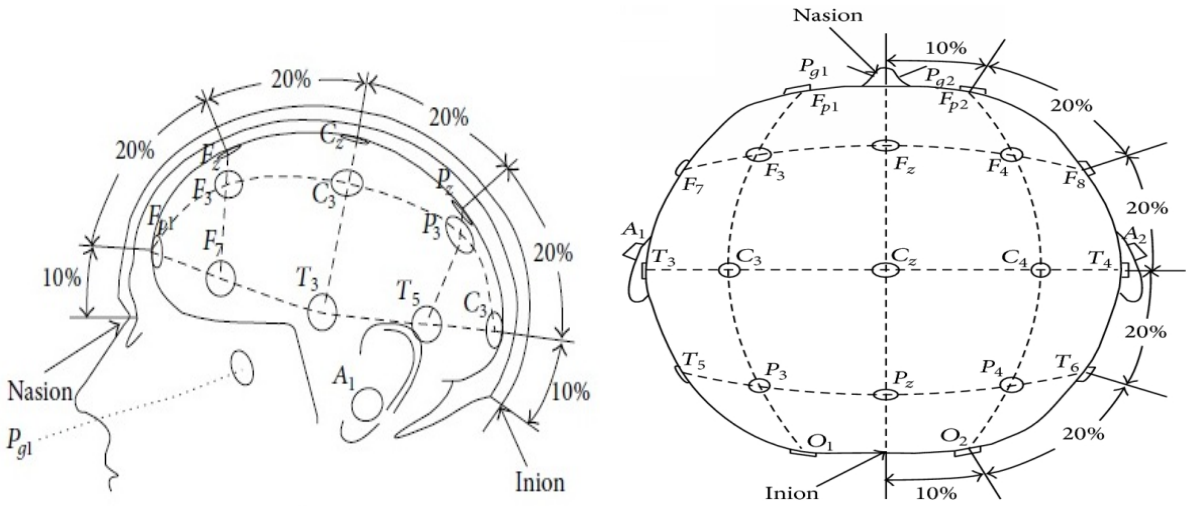
Electroencephalography has also a very good temporal resolution (milliseconds) compared to other methods such as functional magnetic resonance imaging or positron emission tomography (seconds). Unlike functional magnetic resonance imaging, EEG has a bad spatial resolution because of the nature of electric waves and their diffusion. Surface brain activity is easier to detect than internal activity. Since the low signal-to-noise ratio induced by muscular activity EEG requires a non-trivial data analysis.

1.1.5.1 Signal Aquisition

The recorded signals of the brain activity should be in a form that can be successfully interpreted by the computer. The success of a BCI also depends on the signal acquisition. The signal acquisition has to be accurate, efficient, in real-time. Moreover, the acquisition should be out of health risk. There are two methods to accomplish this; invasive and non-invasive BCI. In invasive BCI, electrodes are placed within the scull and acquire signals directly. On the other hand, in non-invasive method, signals are recorded placing the electrodes on the scalp of brain. Electroencephalography (EEG) and Magnetoencephalography (MEG) are non-invasive measurement methods. Since Electrocorticography (ECoG) and microelectrode arrays require surgery to place the electrodes on or in the brain they are invasive methods. The Positron Emission Tomography (PET) uses positron-emitting radionuclide to measures chemical processes. In this process the positron-emitting radionuclide is injected into the body of the subject. The blood flow of human body is connected to brain activity. The changes of the blood flow are measured by Near-Infrared Spectroscopy (NIRS) and Functional Magnetic Resonance Imaging (fMRI) methods. All of the mentioned methods merely depend on the spatial and temporal resolution of the measurement. The measurements also vary in large scale with the area of the brain that can be measured. The temporal



(a)



(b) (c)

Figure 1.2: (a)The electrodes map of 10-20 EEG system (b) Side view and (c) Top view.

resolution for microelectrode arrays, ECoG, EEG, and MEG is in the millisecond range. The temporal resolution of the methods NIRS and fMRI is in the range of seconds and the resolution vary between 10-100 seconds for the PET. Since microelectrode arrays can measure up to single neurons, it shows the finest spatial resolution followed by ECoG, fMRI and PET which have a spatial resolution in the millimeter range. NIRS and MEG have a resolution of circa 5 millimeters, and EEG 1 centimeter and more. During the recording of the brain signal, the methods EEG, PET, fMRI and NIRS usually cover the whole area of the brain followed by MEG being able to cover large parts of the brain. The ECoG can cover small cortical areas, and micro arrays which cover a few hundred neurons. To acquire the brain signals and to provide the recorded data in a computer a signal acquisition system is required. The system needs to provide the data in a format so that it allows both online and offline processing. The signal acquisition consists of the hardware (e.g. EEG cap and -amplifier) and the software to measure and transport the data respectively.

1.1.5.2 Signal Processing

The brain signal contains the patterns of subjects intend. The function of the signal processing step of the BCI is to detect and classify the patterns. The signal processing step receives the raw data from the signal acquisition and translates it into actionable output signals. Usually the functions of the step are: preprocessing, feature extraction, and detection or classification. The raw data is filtered, subsampled, and cleaned to improve the signal to noise ratio in the preprocessing. The feature extraction transforms the high-dimensional data into much lower dimensional feature vectors. The feature vectors represent the characteristics and properties of the respective brain patterns appropriately and allow for detection or classification of them. After the preprocessing and feature extraction, the signal processing uses the lower dimensional features to detect or classify certain brain states. The preprocessed data is then divided into two sets: training set and test set. A classifier is trained using training set with their known labels, and the

classifier learns the brain patterns that predict the labels best. After the classifier has been trained, it can decide, given test set, whether it belongs to one class or the other.

1.1.6 Applications of BCI

The main application for BCI is to create a smart environment so that people especially people of severely physically disabled can communicate with the outside world. The key example, found in BCI publications, are patients suffering from ALS [41] a neurodegenerative disease that slowly paralyzes the patient until he is completely locked-in in his paralyzed body. In the last stage of this disease, the patient is awake, but unable to twitch a single muscle in his body (locked-in syndrome). Spellers can be implemented with the BCI [71, 15] that allows patients to communicate without moving a muscle. Those spellers are probably the most prominent application for BCI. The BCI can be used by the stroke patient to visualize the current brain state in order to learn how to suppress unwanted patterns [52]. Patients with spinal cord injuries can use BCI to control a prosthesis [14], a wheelchair [82], or even a telepresense device [69]. In non-invasive BCI, the data transfer rate is very low, bits per minute and presently somewhere in the low two-digit range. This is the main challenge for complex applications using BCI. A hand prosthesis that is controlled via BCI can be an example which will typically not allow the patient to perform low-level movements like individual finger control, but only a very small set of high level operations like grasp and open hand. Moreover, a patient that is still able to voluntarily twitch a few muscles in his body will often be faster and more accurate using those muscles to control a device than using BCI [155]. Beyond the assistive technology, the applications for BCI can also be used in many other fields [21]. The BCI technology can also be used as a measurement device to measure mental states like attention [218, 98], or workload [124, 170, 236] in order to predict and possibly prevent human failure in critical environments. This technology is used for quality assessment, by measuring subconscious perception of noise in auditoryor visual

signals [200]. The BCI can also be used for entertainment and gaming [130, 179].

1.2 Background Study

Brain-computer interfaces (BCI) are a topic of research and development that has seen increasing interest in the last 20–30 years. The goal of BCI research is to develop systems that decode useful information from ongoing brain activity in real time. In most cases, that information is encoded voluntarily by the user (for example, by performing a voluntary mental operation to produce a measurable signal that can then be used for controlling some device, or by selectively attending to one of a set of stimuli in order to encode a choice). The result, according to the definition of [246], is a system that can replace, restore, enhance, supplement, or improve conventional central nervous system outputs. One of the most commonly considered goals is the development of communication systems for people who are locked-in by a paralyzing disease or accident [11, 133]. Within this field, there has been a recent increase in interest in BCI systems that are based on purely non-visual input. This is motivated by the desire to reach users in the most severely paralyzed states, for whom spatial vision may become extremely limited by the inability to open, direct, or focus the eyes voluntarily, by the inability to make saccades to integrate multiple fixations into a visual scene,and by the frequent infections that result from the lack of blinking (are view of some of these problems,in the particular case of amyotrophic lateral sclerosis,is provided by [226]). Therefore,BCI systems that rely on visual stimuli may work less well that expected [29, 234], or not at all,for the users who need BCI most. Several recent approaches have presented multiple types of auditory stimuli, and required users to make a voluntary choice by covertly shifting their attention to one stimulus type while ignoring the others. Auditory BCI systems can also be divided according to whether they use a streaming or sequential technique. In streaming, streams of auditory stimuli are presented simultaneously or in rapid alternation, and the BCI system exploits the fact that the brain produces a different response

to every stimulus in the attended stream when contrasted with every stimulus in the unattended stream [114, 121, 103] In sequential presentation, relatively infrequent target stimuli are presented among more-frequent non-targets, and the BCI use the difference in brain responses between targets and non-targets [91, 12] Streaming techniques are better suited for interfaces that provide a simple binary choice (such as ayes vs.no decision), whereas sequential techniques are better suited for higher-capacity systems that allow the user to select a letter. Both approaches are useful as assistive communication tools: although a simple yes-no interface has avery limited range of expression relative to a speller, it is easier to learn because the user does not have to keep the assignment between stimuli and letters(or groups of letters) in memory. Therefore,it is a useful first step in establishing communication with severely impaired users, and may also be all that is needed to support some important practical tasks.

One of the most important factors necessary for materializing a successful EEG-based BCI system is the selection of appropriate mental tasks that can elicit distinct task-specific brain activity patterns. To translate the acquired neural signals into appropriate commands, various experimental paradigms and tasks have been introduced, including visual attention tasks such as the P300 speller [71, 132, 220] steady state neural responses elicited while one is gazing a certain visual stimulus flickering with a specific frequency (steady state visual evoked potential: SSVEP)[135, 144, 161]; mental tasks associated with motor imagery[56, 111, 197] or mental calculation [116, 193] and so on. Most of the mental tasks and paradigms listed above use visual stimuli, visual feedback, or both and are thereby applicable only to patients whose visual function is not impaired. In practice, however, some patients with severe neurological disorders, such as ALS and completely locked-in state (CLIS), often have difficulty controlling their voluntary extraocular movements or fixing their gaze on specific visual stimuli. Even for those who have normal visual function, gazing at stimuli for a long time can easily cause fatigue or loss of concentration. Moreover, EEG signals recorded at frontal electrodes can be contaminated by electrooculogram (EOG) elicited by

eye-blinking and eyeball movements. A recent experimental study demonstrated that the performance of the P300-based speller paradigm can be substantially influenced by eye gaze[29] which strongly suggests that the use of visual stimuli or cues might not be appropriate for those who have difficulty in gazing at specific target stimuli. In other mental task paradigms that do not directly use visual stimuli, visual cues or feedbacks are generally provided to the participants so as to instruct or assist them in performing the given mental tasks [108]. Even in such cases, the recorded signals can be contaminated by unwanted visual evoked responses. Therefore, developing new BCI paradigms that are not dependent on visual stimuli remains one of the challenging issues in modern BCI research [178].

To overcome the limitations of conventional BCI paradigms, some researchers have turned to auditory stimuli [101, 114, 147, 122, 217] as an alternative to visual stimuli. Most of the previous studies used auditory oddball paradigms, which share most of the basic concepts with conventional visual BCI paradigms. Two of the earliest studies [101, 114] independently introduced an auditory BCI paradigm in which the authors attempted to discriminate attended brain responses from unattended ones when two simultaneous auditory oddball streams were presented to subjects. In a study by [101], deviant sounds were generated alternatively at either a right or left sound source, and subjects were asked to concentrate on one of the two sound sources. They extracted the feature vectors from the changes in the amplitude of the averaged event-related potential (ERP). [114] used a similar paradigm,where the subjects were instructed to concentrate their attention on one of two oddball audio streams with different frequencies presented alternately with a short inter-stimulus interval. They used the peak amplitudes of P300 and mismatch negativity (MMN) as the feature vectors to classify the subjects selective attention. Recently, [94] refined the auditory oddball paradigm and evaluated various auditory stimuli with different volumes, pitches, or directions.

Another group of researchers attempted to modify the P300 speller paradigm, which is a well-established protocol in BCI research [61] into an auditory version [122]. Instead of presenting matrix-type visual stimuli, they presented different en-

vironmental sounds to participants and detected which sound the participants were attending to. Spatial hearing was also adopted as a new auditory BCI paradigm [217] which used eight speakers spatially distributed around a participant and detected a single sound source that the participant concentrated on.

Apart from the oddball paradigms or modified oddball paradigms (P300 speller), [147] investigated whether the auditory steady-state response (ASSR) is modulated by auditory selective attention (ASA) to a specific sound stream and discussed the possibility of using the ASSR as a new BCI paradigm. They provided eight participants with two amplitude-modulated (AM) sound streams (1 kHz and 2.5 kHz) with different modulation frequencies (38 Hz and 42 Hz) to both ears simultaneously (1 kHz tone with a 38Hz modulation frequency for the left ear and 2.5 kHz tone with a 42Hz modulation frequency for the right ear). The participants were then asked to either concentrate their attention on the stimulus from the left ear or ignore both auditory stimuli according to the instructions appearing on a monitor. In six out of eight participants, the spectral density of alpha rhythm was inversely proportional to that of the modulation frequency for the left ear (38 Hz), providing evidence that selective attention can modulate ASSR. They also showed, using the self organizing map (SOM) method, that the attended and ignored conditions could be clearly classified into two clusters, demonstrating the possibility of using ASSR modulated by auditory selective attention as a new BCI paradigm.

1.3 Motivation

It is very difficult to use the conventional technologies such as a mouse or keyboard to operate a personal computer by the people of physical disabilities. To establish an alternative communication an input device or software is designed so that the disable people can use a computer. Some examples of such devices are single switches of various forms, head mice, speech recognition systems, eye gaze trackers and even software for touch screen devices. Locked-in syndrome (LiS) is the most

severe form of motor disability. The term, thought up by Plum and Posner in 1996, talk about a state in which a person is almost completely paralyzed yet remains cognitively aware [120].

In the case of classical LiS, vertical eye movements, including blinking, are possible. In partial LiS, some additional residual muscle movement such as in a finger, toe or head has been recovered. On the other hand, in total LiS, the person has lost control of even eye movements [11]. Study of cognitive consciousness can be achieved by manual cognitive assessment and is sometimes proved by neuroimaging techniques such as magnetic resonance imaging (MRI) or electroencephalography (EEG). Usually LiS are causes due to a lesion in the pons in the brainstem and neuro-degenerative diseases such as in amyotrophic lateral sclerosis (ALS) [153]. However, a large proportion of people who find themselves in such a state of being are willing and able to continue living for many years with a good quality of life [59]. But empowering people to communicate is essential to maintain a good quality of life. In the classical LiS, the most common and possibly the most efficient means of communication is via eye blinks or eye gaze to a care giver or other human being. One can either communicate 'yes' or 'no', or spell words by indicating the desired letters on an alphabet board. The development of the personal computer and the Internet also improves the quality of life of people who would otherwise be even more isolated from the rest of the world. Eye gaze trackers and eye blink technology allow some people access to computers, and persons with partial LiS can use a single switch device [174]. On the other hand, for people with total LiS, it is difficult to communicate using any overt muscle movement. For locked-in patients, the development of brain-computer interfaces (BCI) is only the way, a direct communication pathway between the brain and an external device that records neural processes. A brain-computer interface (BCI) is a method that aims to extract a user's intention whilst bypassing the normal modality of physical movement by measuring and analysing brain signals

Chapter 2

Brain Computer Interface with EEG

2.1 Electroencephalography

Electroencephalography (EEG) is a method used to measure the electrical activity along the scalp produced by billions of nerve cells, called neurons. The nerve cells communicate with each other through electrical impulses. Each neurons is connected to thousands of other neurons and the electrical activity is generated with the firing of neurons within the brain[177]. It is a combined electrical activity of thousands of the neurons. To record the electrical activity electrodes are placed on the scalp. The EEG is the measure of the brain's spontaneous electrical activity over a short period of time. In neurological perspective, the leading application of EEG is in the case of epilepsy, as epileptic activity can create clear abnormalities on a standard EEG study[119]. In general, a small amount of electrical potentials (normally less that 300 μV) are generated by the brain. Besides, to diagnose sleep disorders, depth of anesthesia, coma, encephalopathies, and brain death the EEG is used. Furthermore, EEG is used to diagnosis tumor, stroke and other focal brain disorders. The EEG method has multiple advantages over other methods used to study brain function. The key advantages are significantly lower hardware cost, higher temporal and spatial resolution etc.

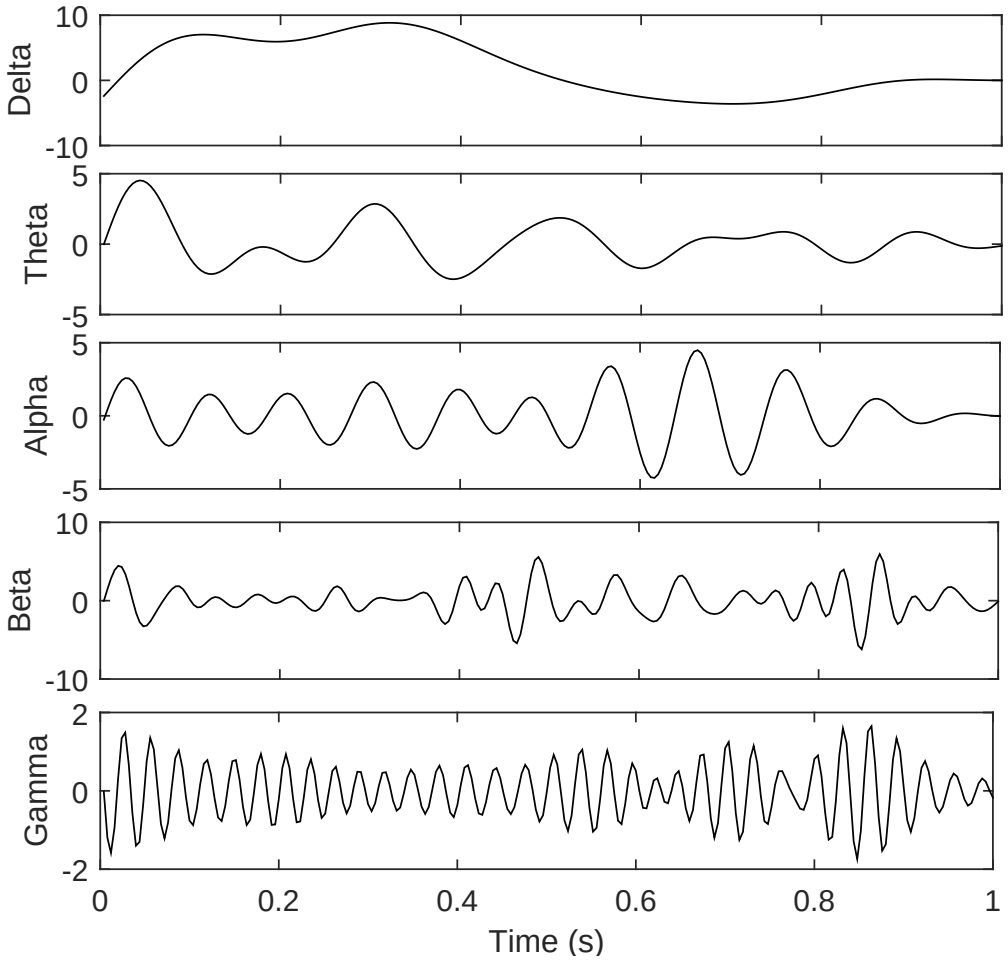


Figure 2.1: Brain rhythms.

2.1.1 EEG Rhythms

The rhythms of brain activity depend on the subject's level of consciousness. For example, various sleep stages can be detected in EEG recording. Throughout the walking state, different rhythmic waves can be sensed. In EEG, the observed rhythmic components are delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–32 Hz), and gamma (above 32 Hz) as shown in Figure 2.1.

Delta rhythm: In adults, delta waves are recorded during deep sleep. Infants (the age around 2 months) show irregular delta activity of 2–3.5 Hz in the waking state. The rhythms have large amplitudes (75–200 μV) and show strong coherence all over the scalp.

Theta rhythm: The frequency range of the theta wave is 4–8 Hz and have a high amplitude and characteristic sawtooth shape. The rhythms can be serve as

a gating mechanism in the information transfer between the brain structures [51]. Usually, activity in the theta band may occur in emotional or some cognitive states; it can be also connected with the slowing of alpha rhythms caused by pathology activity in the theta band may occur in emotional or some cognitive states; it can be also connected with the slowing of alpha rhythms caused by pathology.

Alpha rhythm: The rhythms are predominant during wakefulness and are most pronounced in the posterior regions of the head. Generally, the waves are recorded during the eyes are closed and the subject is in a relaxed state. They are blocked or attenuated by attention (especially visual) and by mental effort. Mu rhythms have a frequency band similar to alpha, but their topography and physiological significance are different. The frequency and amplitude of the Mu rhythm as around 10 Hz and below below 50 μV respectively

Beta rhythm: The brain rhythms of the frequency range from 12–32 Hz are beta rhythms. The amplitudes of the rhythms are usually less than 30 μV. The waves are generally observed over the frontal and central region.

Gamma rhythm: It is difficult to record gamma rhythm by scalp electrodes and their frequency does not exceed 45 Hz. Gamma activity is connected with the recognition of sensory stimuli [76] and the onset of voluntary movements.

2.1.2 Short History of EEG

The brain potentials are first investigated by Richard Caton (1842–1926) who worked on the open brains of cats and rabbits. He observed electric currents by means of a galvanometer, where a beam of light reflected from its mirror was projected onto a scale placed on a nearby wall. He observed that "feeble currents of varying directions pass through the multiplier when the electrodes are placed at two points of the external surface, or one electrode on the gray matter and one on the surface of skull." This statement can be considered as a discovery of electroencephalographic activity. Besides, the spontaneous activity of the brains of rabbits and dogs are investigated by Adolf Beck (1863–1939). In 1890, he first discovered the rhythmical oscillations of brain's electrical activity. The first dis-

covery of so-called *alpha blocking*, the disappearance of these oscillations when the eyes were stimulated with light is observed by him. A co-worker Napoleon Cybulski (1854–1919) of Adolf Beck presented the electroencephalogram in a graphical form by applying a galvanometer with a photographic attachment and was the first to observe epileptic EEG activity in a dog elicited by an electric stimulation[25]. Hans Berger, in 1929 recorded the first electroencephalogram from the surface of the human scalp [13]. The field of the clinical electroencephalography is first introduced in 1935. F. Gibbs and H. Davis showed association of 3/sec spike-wave complexes in EEG with epileptic absences and A. L. Loomis et al. studied human sleep patterns. Moreover, in 1935, the first electroencephalograph (Grass Model I) started the era of contemporary EEG recording.

2.1.3 EEG Recording

The EEG, spontaneous neuronal activity in the brain is observed by electrodes placing them on scalp. The electrical potential of neurons changes over time. The changes are observed over time between a single electrode and a reference electrode [127].To record the potentials, the placement of electrodes is extremely a challenging task. The mapping of functions onto different regions of the brain and electrode placement is very difficult. However, the International Federation in Electroencephalography and Clinical Neurophysiology implemented the standard 10–20 electrode placement system that is widely recognized. With the recording of the EEG, the potential change due to eye movement and muscle activity are also needed to observe. The EEG recording system consists of electrodes, amplifiers, filters and recording device. The electrodes consist of Ag-AgCl disks; diameter varies in between 1 mm and 3 mm. To plug into an amplifier, long flexible leads of the electrodes are used. It is necessary to have a low impedance contact at the electrode-skin interface. To obtain low-impedance and keep the electrodes in place, conductive electrode paste is used. The mechanical stability of the electrodes is also prerequisite for proper recording and contact cement can be used to fix small patches of gauze. For the recording of the EEG, different types of recording

devices are used. A multichannel pen recorder is the most common device which is essential part of commercially offered EEG instruments. To monitor the EEG signals, a visual output device such as an oscilloscope or video display is used at the time of recording. Now a day, computers are used to record the signals and can be used as main device. To convert the recorded analog signal to digital signal an analog-to-digital (A/D) converter is used. To save each sample of the signal in the computers memory, the A/D converter needs to interface to a computer system.

2.2 Event-Related Potentials (ERPs)

Event-Related potentials (ERPs) are very small voltages produced in the nerve cells in response to a specific sensory, cognitive, or motor event [17]. The potentials that are time locked to sensory, motor or cognitive events that provide safe and noninvasive approach to study psychophysiological correlates of mental processes. The ERPs are the sum of electrical potential produced in the nerve cells when a large number of similarly oriented cortical pyramidal neurons fire in synchrony while processing information. The ERPs are of two types. The sensory generated within the first 100 milliseconds after stimulus. It depends largely on the physical parameters of the stimulus. The cognitive, produced in the later parts of the stimulus, in which the subject evaluates the stimulus. An evoked potential is an electrical potential observed from the nervous system of a human or other animal due to the stimulus, as distinct from spontaneous potentials as detected by electroencephalography (EEG). A smart wheelchair is controlled by P300-based BCI. The BCI helps disable people with a virtual keyboard for spelling word and interacting with computers. Auditory evoked potentials (AEPs) are a subclass of event-related potentials (ERP)s. AEPs (and ERPs) are very small electrical voltage potentials originating from the brain recorded from the scalp in response to an auditory stimulus, such as different tones, speech sounds, etc. The functions of a patient's spinal cord are observed by the Somatosensory Evoked Potentials

(SSEPs). To record the SSEPs, peripheral nerves such as tibial nerve or median nerve are stimulated with an electrical stimulus. The response is then observed from the patient's scalp. The potential caused by a visual stimulus is a visual evoked potential. An example of the stimulus is an alternating checkerboard pattern on a computer screen. To record the response, electrodes are typically placed on the back of once head and are observed as a reading on an electroencephalogram (EEG). The potentials are usually produced from the occipital cortex, the area of the brain involved in receiving and interpreting visual signals. Steady State Visually Evoked Potentials (SSVEP) are signals that are produced due to the visual stimulation at specific frequencies. Light-emitting diode (LED) is generally used in a typical SSVEP-based BCI system for flickering. In SSVEPs, the brain produces electrical activity at the same frequency of the visual stimulus by which the retina is excited. The stimulus SSVEP BCIs are useful in different applications, when

i) Large number of BCI commands is necessary

ii) High reliability of recognition is necessary

iii) Self-paced performance is required.

2.2.1 Different ERP Waveforms

P50 wave: P50 is a brain response elicited by audio stimulus, occurs approximately 50 milliseconds after the stimulus. The response is described as P50 wave since it is in the positive direction. The P50 waveform is used to measure sensory gating which filters redundant or unnecessary stimuli in the brain.

N100 wave: The N100 is a negative-going evoked potential occurring at around 100 milliseconds after the onset of a stimulus. This waveform is elicited by auditory stimulus and the amplitude is influenced by interstimulus interval, stimulus intensity, arousal level, subjects attention etc. Usually, the waveform is distributed over the fronto-central region of the scalp.

P200 wave: The P200 waveform is a component of ERP recorded at the human scalp. It is a positive going potentials that varying between about 150 and 275

millisecond after the onset of some external stimulus. The P200 response, in general, distributed around the centro-frontal and the parieto-occipital areas of the scalp.

N200 wave: The N200 waveform, the component of ERP is a negative-going wave that peaks 200-350 millisecond post-stimulus. It is originated mainly over anterior scalp sites.

P300 wave: The P300 wave is a positive-going event-related potential. This ERP component is elicited in the process of decision making. The neural origin of P300 wave is unknown.

N400 wave: The N400 is a component of ERP recorded on human scalp. It is a negative-going potential and peaks around 400 millisecond post-stimulus onsets in adults. The waveform is elicited by potentially meaningful stimuli such as auditory, visual, signs etc. Generally, the waveform is recorded over centro-parietal electrode sites of human brain.

P600 wave: The P600 is a language-relevant ERP and can be elicited by both visual and auditory stimulus. It is a positive-going wave with an onset around 500 milliseconds. The wave peaks around 600 milliseconds after the onset of the stimulus and continues up to several hundred milliseconds.

2.3 BCIs Based on Rhythmic Activity

The BCI technology uses brain signals to control external devices over the last two decades. The rhythmic activities of human brain are recorded over the sensory motor cortex. The activities are modulated by actual movement or imagery movement. The brain rhythms as shown in Figure 2.1 can be used in the BCI [254]. BCIs based on the rhythmic activities have been widely used in people with amyotrophic lateral sclerosis (ALS) [134]. The BCIs have also been used in people with severe damage from spinal cord injuries[249] and in paralyzed people due to stroke[31].

2.4 EEG Artifacts

During the recording of EEG, it is desired to record the cerebral activity of the brain. But the electrical activities arising from outside of the brain are also recorded at the time of the EEG recording. The recorded activities that are not produced in the cerebral are called artifact. The International Federation of Clinical Neurophysiology (IFCN) defines artifact as the recorded activities from extra-cerebral sources. The artifacts are classified as physiologic and extra-physiologic artifacts. Physiologic artifacts are generated from the subject and the activities arise from the body of the subject such as eye movement and muscle activities. The activities recorded from outside the body such as equipment and environment are known as extra-physiologic artifacts. The recorded EEG signals are contaminated by these artifacts that creates great problem for clinical and experimental electroencephalography. Due to the contamination it is very difficult to classify the EEG signal. So really it is a big challenge to deal with artifacts. To deal with the artifact, it needs to know the presence of artifact. The type and source of artifact is also need to determine. Since the rhythmic and frequency parameters of the artifact are very similar to the pure EEG it is quite difficult to separate artifact from pure EEG.

2.4.1 Physiologic Artifacts

2.4.1.1 Electrooculogram (EOG)

The electrical activity generated by eye movement is the electrooculogramn (EOG). The EOG has a huge effect on EEG recording. During EEG recording the voltage difference varies by the eye movement. The EOG is observed typically on the electrodes which are adjacent to eye and the eye movement directed electrodes. The EOG can also be generated by eye blinking, which is produced by the muscle movement of eye lid. This type of EOG may arise only at the wake stage and it generates a different waveform. The EOG due to eye blink has a high frequency.

In practice, EOG signals are recorded by placing the reference electrodes near the eye.

2.4.1.2 Electromyogram (EMG)

Action potentials are produced in the skeletal muscles. The electrical activities of the skeletal muscles are called Electromyogram (EMG). To observe the myoelectric signals, usually a needle electrode is used. The surface EMG, in practice, is measured with the assistance of surface electrode on the skin. Since the EMG has a broad frequency distribution ranging from 0–200 Hz they affects all rhythmic components of EEG.

2.4.1.3 Electrocardiogram (ECG)

Electrocardiogram (ECG) is the recorded electrical activity of heart. The amplitude of the cardiac activity is lower than the EEG signals. The existence of ECG artifact can be identified by the repetitive and regularly occurring waveform pattern of subjects heartbeats. It can be difficult to identify the ECG artifact if the artifact is hardly visible in the background EEG signals.

2.4.1.4 Glossokinetic Artifact

The glossokinetic artifact is produced by the tongue. It has extensive potential field that drops from frontal to occipital areas. In this case the potentials are less steep compared to eye movement artifacts. Although the frequency of the potentials alters it is generally in the delta range. The similar artifacts can be produced by chewing and sucking which are discovered in young patients.

2.4.1.5 Respiration Artifacts

There are two type of artifacts produced during respiration. The first one is slow and rhythmic activity. They are synchronous with the body movements of respiration and mechanically affecting the impedance of one electrode. The second

type can be slow or sharp waves. They occur synchronously with inhalation or exhalation and involve those electrodes on which the patient is lying.

2.4.1.6 Skin Artifacts

The skin artifacts are produced due to biological defects which may alter impedance. During sweating sodium chloride and lactic acid reacting with metals of the electrodes may produce huge slow baseline sways. Important irregularity also can be observed when a collection is under or in the skin. Irregularity occurs due to skull defects. In this case, amplitudes are greater in derivations from electrodes overlying or adjacent to skull defects.

2.4.2 Extraphysiologic Artifacts

2.4.2.1 Electrodes

The electrode artifact is produced due to electrode popping. Because of unexpected impedance variation the artifact appears as single or multiple sharp waveforms. The artifact can be detected by abrupt vertical transient and its distribution.

2.4.2.2 Alternating Current (60 Hz) Artifact

At the time of EEG recording, if the impedance between the electrodes and the ground of the amplifier develops large amount then a problem arises. Due to this problem, the ground becomes an active electrode and yields the 60-Hz artifact. The artifact produces at exactly 60 Hz frequency. However, this type of artifact can be removed with adequate grounding on the patient.

2.4.2.3 Movements in the Environment

Due to capacitive and electrostatic properties, the movement of other persons around the subject can produce artifacts. Infusion motor artifact arises for more use of automatic electric infusion pumps.

2.5 Types of Brain Computer Interface

Brain computer interface system can be divided into three groups based on the technique that the electrical activities are recorded from neuron cells.

2.5.1 Invasive BCI

To record electrical activities of brain, EEG is the most extensively used methods. The electrical activity of groups of neurons is measured by the implanted electrodes which is the rapidly growing part of BCI research [137, 78, 105, 176, 219]. At present the invasive BCI system is implemented for primates only [137, 176]. The use of the system on human is described by some recent results [105, 219]. Using the invasive BCI system it is possible to record signals with a much quality. Moreover, the system provides much better spatial resolution compared to non-invasive techniques. The activity of single neurons can be recorded by the implanted electrodes techniques. On the other hand, a non-invasive system can only measure the resulting activity of thousands of neurons. Therefore, the performance of the invasive BCI system is much better than the non-invasive BCI technique with respect to information transfer rate and classification accuracy. However, some research showed that both the techniques can reach similar information transfer rates [249, 247].

There is a serious drawback of the invasive BCI system. Since the technique is invasive it must require that the subject has to go a surgery operation. Furthermore, the subject needs to replace the implanted electrodes regularly because the electrodes have a limited lifetime. And hence, the subject has to go further surgery. So, certainly the implanted electrodes system has a serious effect on the health of the subjects. In this method, special devices called invasive BCI devices are needed to record the brain signals. The invasive BCI devices that are used to capture signal from the single area of brain cells is called single unit and from multiple areas is called multi-units [248]. As shown in Figure 2.2 the devices are inserted directly into the human brain by a critical surgery.

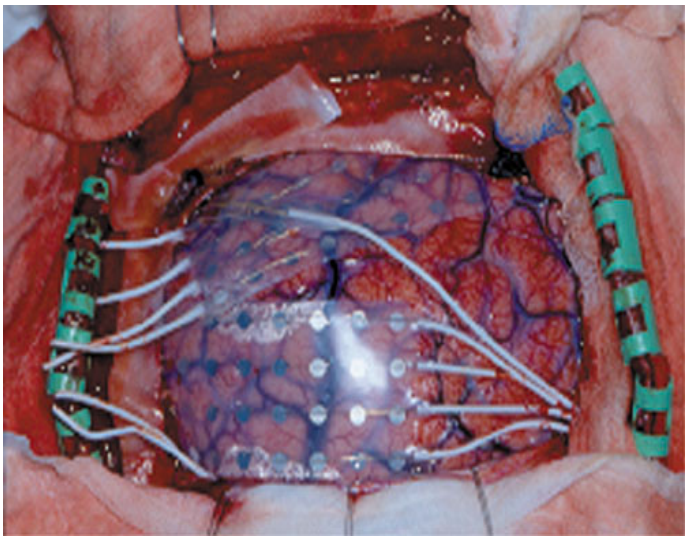


Figure 2.2: Invasive BCI.

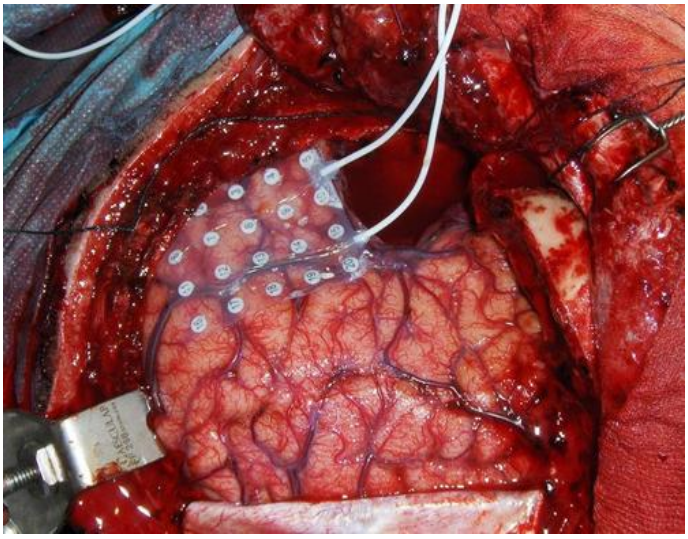


Figure 2.3: Partially invasive BCI.

2.5.2 Partially Invasive BCI

In partially invasive BCI, the signal recording devices are inserted in the skull on the uppermost part of human brain as shown in Figure 2.3. In this system, the quality of the recorded signals is better compared to non-invasive BCI and has a lower risk than the invasive BCI system. The electrical activity of the brain is observed from the beneath the skull. In this system, the electrodes are set in a thin plastic pad that is placed above the cortex. The advantages of the system over non-invasive system are higher spatial resolution, enhanced signal-to-noise

ratio, broader frequency range and reduced training requirements. Moreover, the benefits of the system over invasive BCI system are minor technical difficulty, lower clinical risk and superior long-term stability.

2.5.3 Non-Invasive BCI

The non-invasive BCI system is the easiest and the safest way to record the electrical activity of brain. In this system the electrodes are placed over the scalp which observes minute differences in the voltage between neurons. Due to the skull the device measures weaker brain signals compared to other BCI devices. Besides, some unwanted signals can be mixed with the EEG because of the gap between neurons and electrodes. Therefore, the signal is then amplified and filtered. However, compared to the disadvantages of the other system, the non-invasive system is more acceptable. The electrodes of this system are portable and easy to place on the scalp. Moreover, recently up to 256 electrodes are use on the human scalp to record brain signals that increases temporal resolution.

2.6 Coherence Based BCI

Electroencephalography (EEG) is a cost effective and easier way to implement brain computer interface (BCI). It is captured by spatially distributed EEG sensors of the scalp. The connectivity of different parts of brain is an interesting study to the BCI research community. Recently, the measure of coherence between the signals obtained by different sensors is quantified by coherence analysis [88, 143, 27]. It is usually implemented by spectral estimation with Fourier or wavelet [166] transform. The neural activity data is usually nonstationary in nature and hence it is a great challenge to implement coherence based analysis. The short time Fourier transform (STFT) is considered to solve such problem, it is not entirely resolved due to the following reasons: i) within each short-time period the stationarity of neural data cannot be assured, ii) the resolution of time frequency representation is restricted by Heisenberg uncertainty principle [156]. Although

wavelet transform is considered as data adaptive signal analysis method, it uses basis function called mother wavelet for signal decomposition and faces time-frequency resolution problem i.e. lower frequency resolution at high frequencies and higher at low frequencies. Wavelet analysis also depends on the selection of mother wavelet. The arbitrary selection of mother wavelet without matching with the analyzing signal is the cause of erroneous and non-reversible decomposition.

The synchrosqueezing transform (SST) [53, 54] is one of the techniques based on the continuous wavelet transform (CWT) that generates highly localized time-frequency representations of nonlinear and nonstationary signals. It overcomes the limitations of linear projection based time-frequency algorithms, such as the short-time Fourier transforms (STFT) and continuous wavelet transforms. The synchrosqueezing transform reassigns the energies of STFT and CWT such that the resulting values of coefficients are concentrated around the instantaneous frequency curves of the modulated oscillations [1]. The frequency reassignment method in time-frequency representation [123, 7, 8] develops the meaningful localization of signal components in time-frequency space [35].

Neural synchronization is usually measured using time frequency (TF) coherence as the synchronization described in different frequency bands and to vary over time. It is necessary to smooth the cross and auto spectra between the signals to minimize noise effects to estimate the TF coherence [232, 146]. In any arbitrary TF representation method, the smoothing of cross and auto spectra is usually executed via one of the following method; (i) smoothing of periodigrams via ensemble averaging based on the Welchs overlapped segment averaging (WOSA) method [37], (ii) smoothing in one or both of the time and frequency domains [232, 160, 233, 45] and (iii) smoothing of both cross and auto spectra via averaging a set of spectra estimated by multiple orthogonal taper functions [252, 230]. In the above methods the identical smoothing operators are used for the cross and auto spectra to estimate TF coherence. The identical smoothing operator indicates that TF coherence satisfies the Cauchy-Schwarz inequality [160] and is hence bounded within [0,1]. The estimator with identical operator fails when

smoothing coefficient becomes 1. The better temporal resolution can be achieved by selectively smoothing the auto spectra. With the use of non-identical smoothing operators, the bias of the estimator cannot go to 1 while the cross spectrum is not smoothed, resulting in an estimator with enhanced temporal resolution [160]. Hence, the use of non-identical smoothing operators may assist to identify weak correlations between signals and represent better TF coherence.

In this research, the TF representation of EEG signals is implemented by SST and then TF coherence between two noisy signals using non-identical smoothing operators. The similar analysis is performed with the short-time Fourier transform in place of SST. After validating the TF coherence paradigm with synthetic signals, the method is applied to real EEG. It is clearly observed that in both synthetic and real data, the SST based TF coherence performs better than STFT based method.

2.6.1 Methods

The neural synchronization is varying over both time and frequency. The time-frequency representation (TFR) of any signal describes the energy as a function of both time and frequency. It maps a one dimensional signal of time domain into a two dimensional function of time and frequency. The value of the TFR space provides idea as to which spectral components are present at what time. The TFR is useful to analyze and synthesize non-stationary or time-varying signals.

2.6.1.1 Short Time Fourier Transform

Short-Time Fourier Transform (STFT) is a time-frequency analysis technique suited to non-stationary signals. The STFT provides information about changes in frequency over time. It represents a sort of compromise between the time and frequency of a signal. Also, it gives some information about both when and at what frequencies a signal event occurs. During STFT, the signal is separated into small portions, where these portions of the signal can be assumed to be stationary. For this purpose, a window function w is chosen. The width of this window must

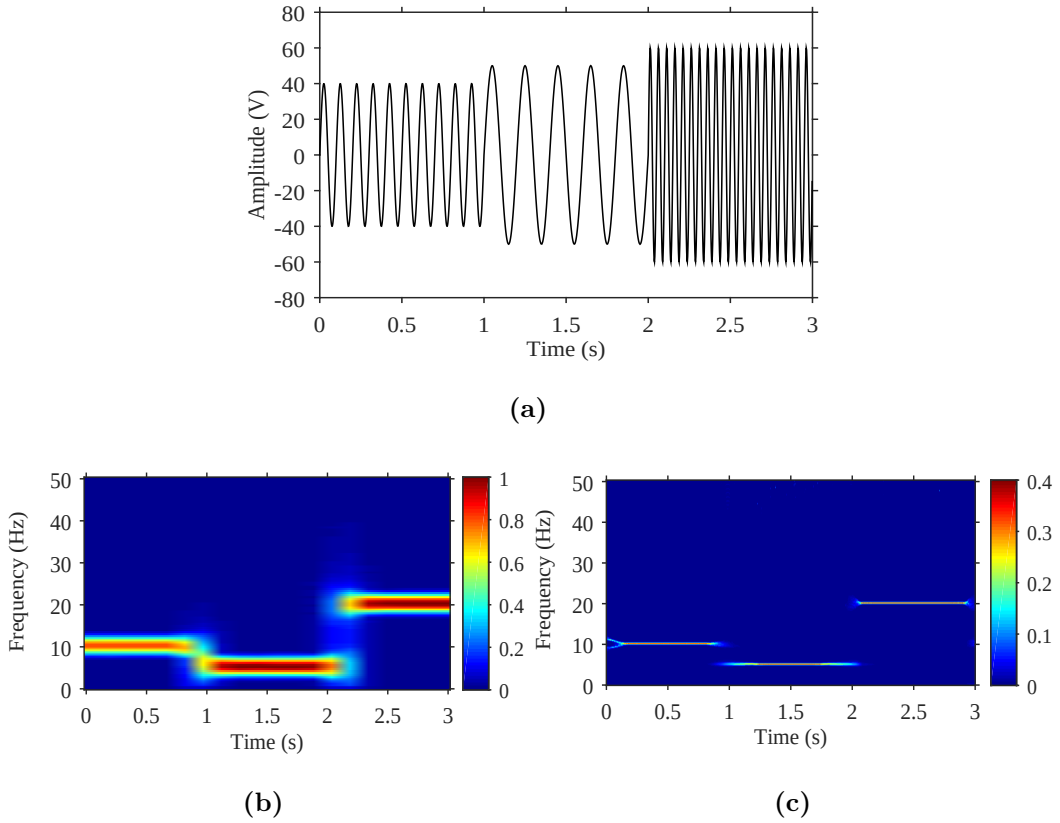


Figure 2.4: Synthetic signal δ(t) and its TF representation (a) the synthetic signals δ(t) with three components, TF represention using (b) STFT and (c) SST.

be equal to the portion of the signal where its stationarity is valid. The STFT for a non-stationary signal s(t) is defined as

$$\Psi(t, f) = \int_{-\infty}^{\infty} [s(t).w^*(t - t')].e^{-2\pi ft} dt \tag{2.1}$$

where * is the complex conjugate, w(t) is the window function. The STFT of the signal is the Fourier transform of the signal multiplied by a window function. To illustrate the time-frequency representation we create a noise free synthetic signal $\delta(t)$ obtained by concatenating three sinusoids s1(t), s2(t) and s3(t) of frequencies 10Hz, 5Hz and 20Hz respectively as $\delta(t) = [s_1(t)s_2(t)s_3(t)]$. The sampling frequency is 500Hz. The synthetic signal and TF representation using STFT are presented in Figure 2.4(a) and Figure 2.4(b) respectively. In the STFT a Hamming window of length 256 and 50 percent overlap is used. The STFT is able to recognize the three components but with a poor resolution.

2.6.1.2 Synchrosqueezing Transform

Synchrosqueezing Transform (SST) is a method functional to the Continuous Wavelet Transform (CWT). The SST is used to localize the frequency components of non-stationary signals in TF space. The CWT is capable to generate a high resolution TF representation. The SST based TF representation taking the synthetic signal δ(t) is shown in Figure 2.4(c). A bump mother wavelet and the discretization with the proper scales of CWT are used to implement SST. It is noted that STFT based TF space has very poor frequency resolution and inherently it suffers from lower time resolution due to the use of window function. The CWT algorithm recognizes oscillatory components of a signal through a series of time-frequency filters known as wavelets. To separate a continuous time function into wavelets the CWT is used. A mother wavelet Φ(t) is a finite oscillatory function which is convolved with a signal s(t) in the following form

$$Z(p,q) = \frac{1}{|p|^{1/2}} \int_{-\infty}^{\infty} \Phi(\frac{t-q}{p}) s(t) dt \tag{2.2}$$

where Z(p,q) is the wavelet coefficients for each scale-time pair and the instantaneous frequency can be estimated as

$$\omega_s(p,q) = -iZ(p,q)^{-1} \frac{\partial Z(p,q)}{\partial q} \tag{2.3}$$

The TF representation maps the information from the time-scale plane to the time-frequency plane. In the synchrosqueezing operation, every point (q,p) is converted to $(q,\omega_s(p,q))$ [53]. Because p and q are discrete values, we can have a scaling step $\Delta p_k = p_{k-1} - p_k$; for any p_k , where $\omega_s(p,q)$ is computed. During mapping from the time-scale plane to the time-frequency plane $(q,p) \rightarrow (q, w_{inst}(p,q))$, the SST $\Gamma(\omega_l, q)$ is calculated [1] only at the centers ω_l of the frequency range $[\omega_l - \Delta\omega/2, \omega_l + \Delta\omega/2]$, with $\Delta\omega = \omega_l - \omega_{l-1}$:

$$\Gamma(\omega_l, q) = \sum_{p_k : |\omega_s(p_k,q) - \omega_l| \leq \Delta\omega/2} Z(p_k, q) p^{-3/2} \Delta p_k \tag{2.4}$$

The equation (2.4) shows that the TF representation of the signal is synchrosqueezed along the frequency (or scale) axis only [140]. In the SST, the

coefficients of the CWT are reallocated to get a concentrated image over the time-frequency plane, from which the instantaneous frequencies are then extracted [250].

2.6.1.3 Coherence Analysis

Coherence is a mechanism of efficient interaction between two groups. In the field of neuroscience, coherence reports systematic consistency between two neuronal populations. Neuronal synchronization refers to the existence of a more or less consistency among oscillatory modulations in neural activity in different neurons. Synchronization plays a significant role in the coordination of spiking activity across the neuronal groups [231, 212, 79].

2.6.1.4 Frequency Coherence

Frequency coherence is a frequently used method to measure consistency between neural signals. The main advantages of frequency coherence are that it is very implicit, firm and pretty robust alongside noise and allows an easy overview over relevant coherent frequencies in the data [150]. The frequency coherence reflects the consistency of cross-spectral densities between two signals, normalized by their auto-spectra densities. Considering two stationary random processes, x and y as the functions of frequency, the standard coherence function of x and y is dened as [214]:

$$|C_{x,y}(f)| = \frac{|J_{x,y}(f)|}{\sqrt{J_{x,x}(f)J_{y,y}(f)}} \tag{2.5}$$

where $J_{x,x}(f)$and $J_{y,y}(f)$ are the auto spectral density functions of x and y, respectively, at frequency, f; and $J_{x,y}(f)$ is the cross spectral density between the two processes. The standard coherence function is not enough for EEG like non-stationary signals.

2.6.1.5 Time-Frequency Coherence

Usually, the coherence analysis estimates the correlation between two signals in the frequency domain, and therefore it is only suitable for stationary signals. However, the temporal information of EEG like non-stationary signals cannot be discovered by the conventional coherence analysis [166].A time-frequency extension approach measures the linear correlation between the two processes in time-frequency plane [245]. In brain computer interfacing motor imagery paradigm, the synchronization of neural activity has been measured using the TF coherence.The TF coherence function is defined as

$$|C_{x,y}(t,f)| = \frac{|J_{x,y}(t,f)|}{\sqrt{J_{x,x}(t,f)J_{y,y}(t,f)}} \tag{2.6}$$

where $t = 1, 2, ..., T$; signal partitioned into T segments and $f = 1, 2, ..., F$; is the discrete frequency.The cross and auto spectral densities are measured as

$$J_{x,y}(t,f) = X(t,f)Y^*(t,f) \tag{2.7}$$

$$J_{x,x}(t,f) = |X^2(t,f)| \tag{2.8}$$

$$J_{y,y}(t,f) = |Y^2(t,f)| \tag{2.9}$$

where $X(t,f)$ and $Y(t,f)$ are the TF transform coefficients of x and y respectively and $Y^*(t,f)$ is the complex conjugate of $Y(t,f)$ The definitions of the TF coherence are straightforward and they follow an approach comparable to the Fourier analysis. In the Fourier analysis, the spectra and the frequency coherence can be estimated based on periodogram method in which the signal segments are averaged to find the estimation. However, the time-frequency coherence faces difficulties during averaging as there are two dimensions of both time and frequency. In this research the TF transform coefficients followed TF coherence is computed using SST that outperforms STFT.

2.6.1.6 Smoothing Effects on TF Coherence

Smoothing operator is a convolution operator that is used to remove detail and noise. In the case of smoothing cross and auto spectral densities, the operators can be identical or non-identical. Unlike identical operators, the use of non-identical operators result the time-frequency coherency is not bound to [0, 1] and hence enhanced temporal resolution [160].To improve the consistency of TF coherence it is required to smooth both in time and frequency. The smoothing operators can be an average over a set of orthogonal-based spectral estimates such as multi-tapper approaches [252, 230, 26]. Usually, it is used to smooth the cross as well as auto spectral densities. The non-identical smoothing operators are one dimensional function of time [252], or two dimensional in both time and frequency [232, 233, 26].Now, standard magnitude squared TF coherence can be found as [160]

$$|C_{x,y}(t,f)|^2 = \frac{|J_{x,y}(t,f) \otimes w[\phi]|^2}{(J_{x,x}(t,f) \otimes w[\varphi])(J_{y,y}(t,f) \otimes w[\varphi])} \tag{2.10}$$

where $\otimes$ represents convolution operator, $w[\phi]$ and $w[\varphi]$ are two non-identical ($w[\phi] \neq w[\varphi]$) smoothing windows of cross spectral density and auto spectral density respectively. The effects of smoothing in TF coherence are illustrated in Figure 2.5, where two synthetic signals $x1 = [\sin(2\pi f_1 t)\sin(2\pi f_2 t)], x2 = [\sin(2\pi f_1 t)\sin(2\pi f_2 t)]$ with $f_1 = 5$Hz and $f_2 = 10$Hz and their TF coherence are presented. There are different time duration of individual sinusoids to generate x_1 and x_2 as shown in Figures 2.5(a) and 2.5(d) respectively. It is observed that smoothing operation makes the TF coherence more representative and informative for both STFT and SST based method as illustrated in Figures 2.5(e) and 2.5(f). On the other hand, a noticeable amount of irrelevant coherence is introduced when the smoothing operation is not performed as shown in Figures 2.5(b) and 2.5(c). Hence the use of non-identical smoothing operator improves the measure of time-frequency coherence. In this paper two 2-D Gaussian smoothing windows of different lengths are used. The window length is represented as $w = [hd]$ where h specifies the height of the kernel in hertz and d specifies the width of the kernel in seconds.

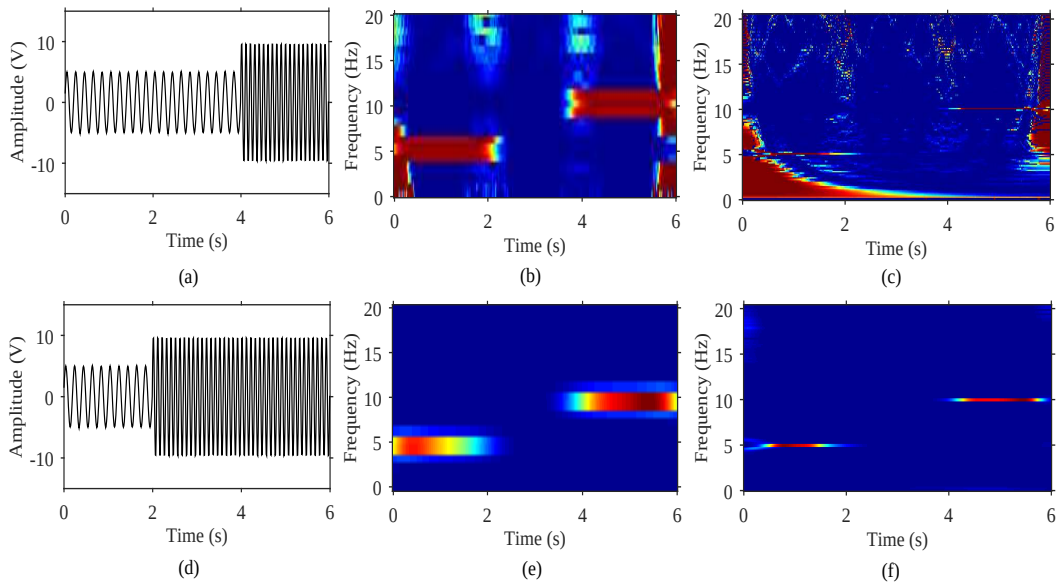


Figure 2.5: The effect of smoothing operation on TF coherence between synthetic signals x_1 (a) and x_2 (d). (b) STFT and (c) SST based TF coherence without smoothing, (e) STFT and (f) SST based TF coherence with smoothing.

2.6.1.7 Proposed Algorithm for TF Coherence

The proposed algorithm of measuring time-frequency coherence between two signals based on SST can be summarized with the following steps:

i) Select two signals or signals from two channels of EEG

ii) Apply the SST on individual signal to obtain the time-frequency coefficients

iii) Compute cross and auto spectral densities based on the SST coefficients

iv) Smooth the cross and auto spectral densities using two appropriate non-identical (different window length) smoothing operators

v) Finally compute time-frequency coherence using the smoothed cross and auto spectral densities using equation (2.10).

The performance of the proposed SST based time-frequency coherence is evaluated with both synthetic signals and real EEG data. The results are compared with the STFT based time-frequency coherence. The time-frequency coherence between channels of left and right hemisphere of human brain is studied in this experiment. Motor imagery classification between left hand and foot movement is also observed. It is observed that SST based method is more efficient than

STFT for localization of frequency components with higher resolution in coherence domain. In addition marginal time coherences are calculated from both SST and STFT based coherences. It is found that the SST based marginal time coherences exhibits very clear discrimination between left hand and foot movement data whereas the STFT based marginal time coherences are unable to do this.

Chapter 3

Stimuli Based Brain Computer Interface

3.1 Introduction

A brain-computer interface (BCIs) is a direct communication method that establishes connection between the brain and an external device [196, 248]. In BCIs, generally used scalp recorded EEG features are P300 potential [71], steady state visual evoked potentials (SSVEPs)[161], N200 [106, 256] and alpha/beta rhythms [194]. P300 potentials can be stimulated by an oddball paradigm and exactly the subject is required to focus on visual/auditory target stimuli. Actually the stimuli are hidden as rare occurrences among a series of more common non target stimuli. In this situation, P300 like event-related potential (ERP) may occur about 300 ms after the target stimulus across the parieto-central area of the skull [199]. P300, by which the target stimuli can be recognized, has been generally used in visual/auditory/audiovisual BCIs. Usually, in BCIs to proficiently control the external devices P300 potentials are used. As an example, some visual BCI spellers based on P300 have been described [61, 131]. A hybrid BCI system combining P300 and SSVEP is proposed in [142]. Concerning auditory BCIs, a spelling system in which the letters in a 5×5 matrix were coded with acoustically presented numbers is proposed in [81]. An auditory spatial ERP paradigm involving spatially distributed auditory cues is proposed in [217]. Actually, these BCIs are naturally based on auditory ERPs, including P300. But, existing results shows

the performance of auditory-only BCI systems are not as good as the visual BCIs. In recent years, numerous audiovisual BCIs have been developed. The usefulness of a P300-based BCI has been evaluated in [220] that could use four different types of visual, auditory, or audiovisual stimuli.

Some offline investigation proved that the level of classification accuracy of auditory mode did not grasp the same level of classification accuracy as the visual or audiovisual mode. Hence, the performance of the audiovisual mode was not significantly better than that of the visual mode. For an instance, the proposed offline audiovisual P300-speller in [12] found that audiovisual stimulation amplified the average strength of the response compared to that of either visual-only or auditory-only stimulation. During the reception of spatial, temporal and semantic congruent audiovisual stimuli brains possibly will integrate the auditory and visual features of these stimuli [55]. Neural responses in heteromodal brain areas, including the posterior superior temporal sulcus/middle temporal gyrus (pSTS/ MTG), are more noticeable for congruent audiovisual stimuli than for corresponding visual-only and auditory-only stimuli [34, 3, 100]. Audiovisual integration methods were evaluated by comparing the ERPs elicited by audiovisual stimuli with the sum of the ERPs elicited by visual-only and auditory-only stimuli [60, 75, 165]. Both early (at approximately 40 ms for sites Fz, Pz) and late (after 100 ms for sites Fz, Cz and Pz) ERP enhancement effects of audiovisual integration are described in [229]. Also, the multisensory integration effects at approximately 100 ms (frontal positivity), 160 ms (centro-medial positivity), 250 ms (centro-medial negativity), and 300–500ms (centro-medial positivity) are perceived in [227]. According to the interpretations in [227], it is estimated that the P300 could be enhanced by combing the paradigm eliciting P300 and that for audiovisual integration. These improved ERPs in the audiovisual situation may be used to advance novel audiovisual BCIs that have not been necessarily measured in existing audiovisual BCI studies [220, 12]. A BCI system can detect intention-specific changes in EEG signals and thus allow patients who have lost movement ability after an injury or disease to convey their intent to the ex-

ternal world [196, 248]. Awareness detection is a potential application of BCIs in patients with disorders of consciousness (DOC), such as vegetative state (VS) also called unresponsive wakefulness syndrome (UWS), minimally conscious state (MCS), emerged from MCS (EMCS) and locked-in syndrome (LIS).

Several patients with DOC may be diagnosed with VS, in which they may awaken but show no awareness of themselves or their environment [112]. Some patients may improve to MCS, in which they demonstrate inconsistent but reproducible signs of awareness [84]. Furthermore, EMCS is characterized by reliable and consistent demonstration of functional interactive communication or functional use of two different objects [85]. In recent times, to subdivide MCS patients into patients responding to commands (MCS+) and patients showing only non-reflex behavior (MCS−) is proposed in [30]. At present, the clinical diagnosis of patients with DOC is mostly based on behavioral observation scales, including the Glasgow Coma Scale (GCS) and the JFK Coma Recovery Scale-Revised (JFK CRS-Revised). For the reason that these patients cannot provide adequate behavioral responses, misdiagnosis rates in VS and MCS patients range from 37 to 43% [5, 216]. More than a few BCI paradigms have been presented for patients with DOC. In [152], a four-choice auditory oddball BCI on 16 healthy subjects and 18 patients with DOC (13 MCS, 3 VS and 2 LIS) is tested. Their outcomes indicated that 13 healthy subjects, 1 MCS patient and 1 LIS patient were able to communicate using the BCI.

In [49] an MCS patient who could use a motor imagery (MI)-based BCI system with 80% online accuracy is testified. A visual hybrid BCI combining P300 and SSVEP is established to detect awareness in eight patients with DOC (4 VS, 3 MCS, and 1 LIS) and successfully demonstrated command following in three patients (1 VS, 1 MCS, and 1 LIS) in [188]. Although the training for MI-based BCI is a heavy burden, P300-based BCIs may be more suitable for patients with DOC the BCI-based awareness detection in patients with DOC is still in its infancy. Since the patients cognitive ability is considerably lower than that of healthy individuals the performance of the BCIs designed for these patients is generally

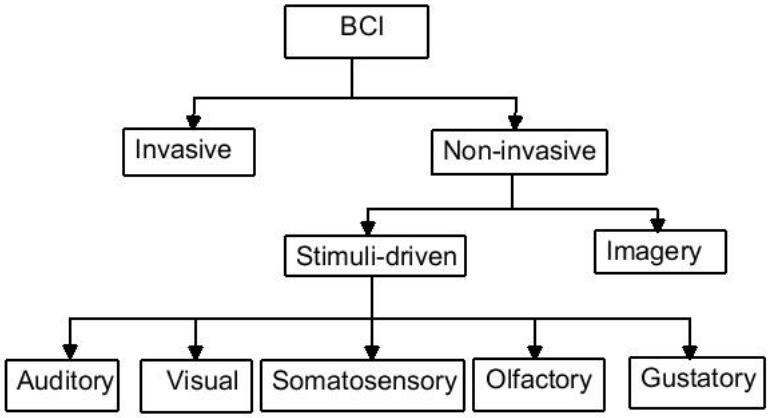


Figure 3.1: Based on implementation BCIs are two types: invasive and non-invasive. Non-invasive BCI can also be divided into imagery and stimuli driven, depending on whether the sensory stimulus is used or not. Moreover, there are also five types of the stimuli driven BCI: visual; auditory; somatosensory; olfactory; and gustatory.

poor. Moreover, most of these patients lack control of gaze movement. To develop novel BCIs to improve sensitivity in awareness detection can be one promising solution.

The brain computer interface (BCI) or brain machine interface (BMI) facilitates any computerized device to be functioned without muscle activity [246]. The interface empowers the ALS patients to communicate freely. On the basis of the classification of feature values extracted from brain activity, a state of-the-art BCI is implemented. However, their classification performance is presently unsatisfactory for a wide range of applications. In order to develop a quality interface, there is a need to improve the classification accuracy. To attain this development, some research projects based on an auditory BCI are designed applying auditory evoked responses (AER) generated by corresponding sound stimuli [217].The different types of BCIs are shown in Figure 3.1 where auditory BCI is grouped under noninvasive BCI and stimulusdriven BCI.

Cranial surgery is mandatory in an invasive BCI to implant the electrodes on/in the surface of the cortex. The invasive BCI measures brain activity in a high signal to noise ratio (SNR) condition with higher infection and side effect complication risk. Unlike invasive BCI, in noninvasive BCI electrodes are placed

on the surface of the human scalp. Hence, the SNR is lower than with an invasive BCI system. Consequently, the real-world applications with a noninvasive BCI are more challenging. The stimulus-driven BCI method can be a solution to his problem that applies brain responses to artificial sensory stimuli to produce commands. As an instance, a user can concentrate on only one stimulus from several stimuli offered in sequence.

The BCI method categorizes the responses and converts them into computerized commands. The stimulusdriven BCI method focuses on a visual modality, which uses evoked responses to visual stimuli such as flashing or switching ON/OFF of a signal. Still, visual BCI involves eye movement for pick out commands and is indeed challenging for advanced ALS patients. Since auditory BCI requires no muscle activity, it can be one of the best solutions to the problem. Moreover, the auditory BCI method has the benefits of *i*) not necessarily need to fix viewing direction and *ii*) easy to set stimulation devices. However, the classification accuracy with the auditory BCI is still not good for real-world applications. The recent auditory BCI research demonstrated consistent classification accuracy but it needs further improvement. A comparative study between auditory and visual BCI spellers is carried out in [40]. The study testified that with the same experimental conditions the accuracy of an auditory BCI speller is lower than that of a visual BCI speller. Because of the dependence of the auditory stimuli the waveform of AER fluctuates. For instance, EEG responses differ for sound stimuli from the right or the left [83, 211], or depending whether sound images are virtual or real sound sources [180].

3.2 Auditory Stimuli

3.2.1 The Auditory ERP

The scalp potentials stimulated by auditory stimuli are auditory ERPs and seen by averaging EEG samples time-locked to those stimuli. The ERP components are categorized into exogenous and endogenous in [61]. Since the exogenous ERPs

are generally determined by the external stimulus characteristics, they are relatively stable in terms of latency and amplitude. On the other hand, endogenous ERPs commonly depend on the subject's intention. Hence, they exhibit unlimited inconsistency in both latency and amplitude, usually in accordance with the subject's varying internal state and behavior. According to the signal latency the exogenous ERP components are grouped into early, middle and late. The early components are supposed to stem from brainstem responses occurring within the first 10–12 ms from stimulus onset whereas the middle components manifest in the first 50 ms after stimulus onset and are supposed to reflect early primary auditory cortex processing of the stimulus. Besides, the late components include the relatively large N1 and P2 waves. N1 indicates a negativity that generally peaks at about 100 ms from stimulus onset and P2 indicates a positivity that peaks at around 180–200 ms from stimulus onset. N1 is preceded by a small positivity, denoted P1, which peaks at around 50 ms. These late components are supposed to stem from auditory cortex as well. The N2 and P3 (P300) components, with a latency of 300–450 ms are considered to be endogenous components stimulated as a response to rare or subjectively significant stimuli.

3.2.2 Auditory ERP Based BCI

An auditory ERP based binary BCI [102] is grounded on attention modulation of ERP signals generated by attending to right ear versus left ear auditory stimuli. The stimuli presented to the subject consisted of a series of rapid beeps concurrently presented to both ears but the phase and frequency are differing. Before each trial a visual queue was presented where an arrow pointing either to the right or the left. The pointer indicates which side the subject should attend to. A random number of frequency deviant stimuli were presented (1650Hz to the right and 880Hz to the left) in each series of beeps in order to keep the subject alert. The subject was asked to count how many deviants there were in the series presented to the ear he/she was ordered to attend to. The ERPs generated by attending to each side were calculated (averaged) and used to train a support vector machine

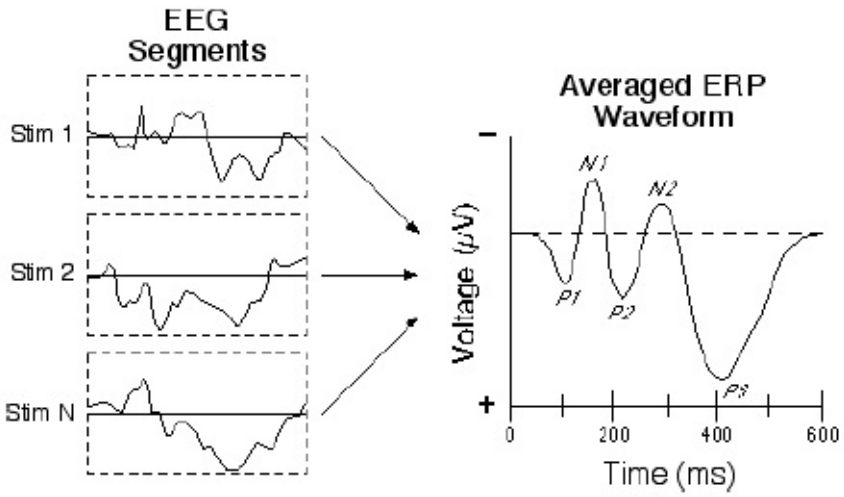


Figure 3.2: ERP is produced by time-locking and averaging of the different epochs of a raw EEG recording.This figure is taken from http://erpinfo.org/.

(SVM) which classifies by creating a separating hyper-plane between data points of each class by maximizing the margins between the data sets. It is established that it is possible to classify single trial ERPs based on that classification. This paradigm requires a two hour training session as well as offline machine learning.

3.3 Steady State Visual Evoked Potentials

In the perspective of EEG signals, an Evoked Potential (EP) is an electrical potential elicited by the presentation of a stimulus that can be measured from the nervous system. EP can be recorded with an electrode positioned on the surface of the scalp particularly in the case of non-invasive EEG recordings. The Visual Evoked Potentials (VEP), are EP elicited by a visual stimulation [184]. The amplitude of the EP and VEP is very low with respect the spontaneous ongoing brain activity that is the key concern in BCI research. The amplitude ranging from less than a microvolt to several microvolts, compared to tens of microvolts for EEG, millivolts for EMG, and often close to a volt for ECG. Signal averaging is essentially required to resolve these low-amplitude potentials against the background of ongoing EEG, ECG, EMG, and other biological signals and ambient noise.

Since most of the noise occurs randomly and the signal is time-locked to the stimulus, permitting the noise to be averaged out with averaging of repeated responses as used for ERP as shown in Figure 3.2. Steady-state VEP (SSVEP) is a particular case of VEP, where the same stimulus is repetitively presented at a frequency at least higher than 3.5Hz, but more commonly higher than 6Hz. Usually, this kind of stimulus is referred as Repetitive Visual Stimulus (RVS). A SSVEP response acquired using the time-locking averaging technique is shown in Figure 3.3. From the time when the existence of the response can be detected investigating the power spectrum in the frequency domain of the recorded signal, summing and averaging different signal epochs corresponding to different presentations of the same stimulus is not necessary in the SSVEP context. Typically Light Emitting Diodes (LED) generated the stimuli and shapes on a computer monitor [261]. Though the SSVEP can be perceived for higher frequencies [201] up to 100 Hz, the device flickers at frequencies ranging between 6Hz and 40Hz.

3.4 Visual Stimuli

In order to stimulate VEP and SSVEP responses in the subjects brain activity, a visual stimulation has to be needed. In SSVEP based BCIs environment, different visual stimuli have been used by means of different stimulator devices with different performances [261] and occasionally with different results [39, 255]. The stimulus properties can significantly effects the VEP waveforms and also influence the SSVEP response amplitude and frequency distribution.

3.4.1 Stimuli for VEP Experiments

VEP experiments are performed usually for clinical applications. The VEP stimulation are classified into luminance and pattern as described in [184]. A constant flash of light is produced with the luminance stimulation. On the other hand, pattern stimulation may be either presented in a pattern-reversal or onset offset fashion.

Pattern Stimulus

Pattern stimulus is a black and white checkerboard as suggested in [184] in which every checks should be a square and where there should be an equal number of light and dark checks as shown in Figure 3.4. The dimensions of the checks should be defined as the visual angles subtended by the sides of a single check and the visual angle should be measured in degrees and minutes of arc subtended at the subjects eye. The same holds for the definition of the whole checkerboard size, named stimulus field size that should be expressed in degrees of visual angle, with an indication of the field shape. If the field is a rectangular field $\alpha^\circ \times \beta^\circ$ large or a circular field of δ° diameter or radius.

Pattern stimulus luminance is a photometric measure of the luminous intensity per unit area of light travelling in a given direction. The luminance defines the amount of light that passes through, is emitted or reflected from a particular area, and falls within a given solid angle. It should be measured in candelas per square meter (cd/m^2) and the luminance of the white checks should be at least $80cd/m^2$.

The surrounding of the stimulus should be homogenously lit, with an average luminance equal to or below the average stimulus luminance. Usually quiet room lighting with no bright sources visible to the subject has to be used and if a computer monitor is used for presentation, a dark background has to be used. The fixation point should be located at the corner of four checks when placed at the center of the field. The pattern reversal stimulus consists of black and white checks and its phase has changed abruptly from black to white and white to black. Since there must be no overall change in the luminance of the screen, an equal number of light and dark elements has to be displayed. Also the pattern stimulus should be defined in terms of the visual angle of each check.

Flash Stimuli

Though there is no pattern in this situation, the flash stimulus is defined as the pattern onset/offset stimulus. Therefore, on a darker background a patch of solid color, uniformly lightened, is turned on and then off. VEP should be stimulated by a flash that subtends a visual field of at least 20° [184]. The strength of the flash

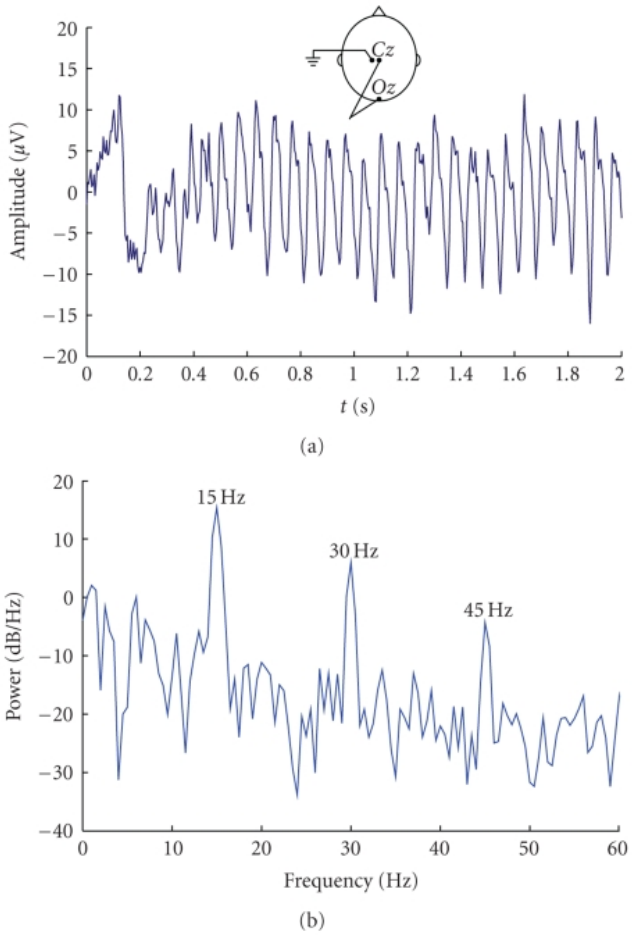


Figure 3.3: Waveform of an EEG signal picked up during visual light stimulation with a frequency of 15 Hz and its frequency spectrum acquired from Cz and Oz locations. (a) The SSVEP waveform give a picture of the time locked average of 10 realizations. (b) Frequency information of the same recording shown in (a). This figure is taken from [261].

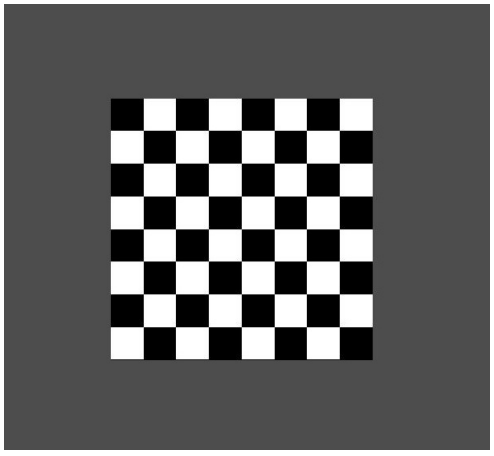


Figure 3.4: The checkerboad pattern stimulus.

stimulus are measured in photopic candelas seconds per squared meter (cds/m^2). The background on which the flash is presented should be measured in candelas per squared meter (cd/m^2) and the flash should have a stimulus strength from 1.5 to 3 cds/m^2 with a background from 15 to 30 cd/m^2. Besides, the stimulus should be presented less than 1.5 times per second in order to elicit VEPs and avoid the elicitation of SSVEP responses.

3.4.2 Stimuli for SSVEP BCI

The stimuli used to stimulate VEPs are same as the stimuli used to elicit SSVEP responses. But the change is the presentation rate that has to be faster than several hertz that creates the steady-state. Generally the flash and the pattern reversal stimuli are used in SSVEP based BCIs [261]. Moreover, the VEP and SSVEP responses are elicited by other kinds of stimuli that have been considered in [237]. Figure 3.5 shows various types of visual stimuli and stimulation devices. Stimulation by means of two different lights placed on a pair of goggles is represented in the upper box (A) where stimuli flicker directly in front of the eyes. Light-Emitting Diode (LED), the familiar stimulation device for SSVEP based BCIs is represented in the second box (B). Different types of stimuli that are fit to be seen on a computer monitor are represented in the last box (C). The flash stimuli (c1), pattern reversal stimuli (c2) and flickering images (c3) which are used to study the SSVEP response change are also represented. Moreover, the stimuli as (c4) to (c7) are hardly used for SSVEP based BCIs but in the fields of physiology and neuroscience to investigate particular relations between the SSVEP response and other perceptual or cognitive functions of the brain.

3.4.3 Stimulation Devices

Light Emitting Diode

LED or LED array lights are preferred to provide flash stimuli with respect to other kind of lamps because of their relatively low latency and fast reaction to onsets and offsets.These lights received excessive attention from electronic research

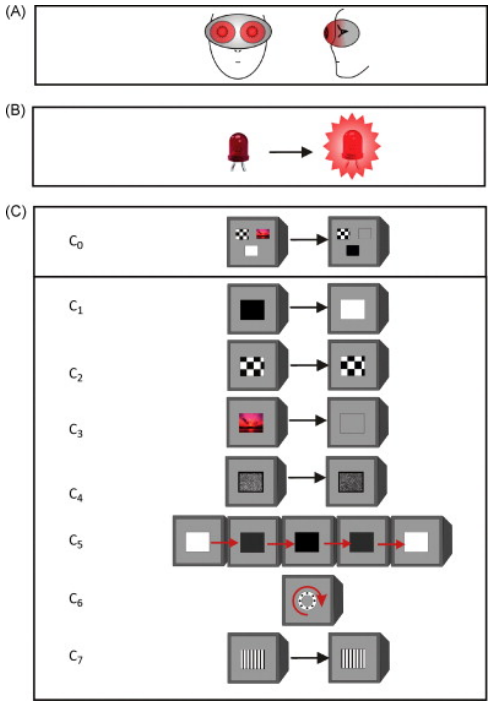


Figure 3.5: Visual stimuli and stimulation devices used in VEP and SSVEP BCIs. (A) flickering light mounted on goggles; (B) Light-Emitting Diode (LED) and (C) flickering images on a computer screen: (c0) combination of images that can be used for binocular rivalry paradigms, (c1) simple square, (c2) checkerboard, (c3) image, (c4) Gaussian field, (c5) sinusoidally modulated square, (c6) rotating or moving stimuli, and (c7) moving vertical or horizontal gratings. This figure is taken from [237].

end industry in recent years because of their high luminous efficiency, reaching for red-orange LEDs, peacks of almost 100 $lmW^{-}1$. LEDs are widely available on the market, relatively cheap and also power LED exists reaching several watts. Since the LED generally very small with respect to the optimal area to elicit VEP and SSVEP responses (several degree of visual angle), they are often used in arrays, or otherwise they are placed behind a diffusive patch as shown.

Computer Displays

To deliver SSVEP producing stimuli, another commonly used device is a regular computer monitor. The wide diffusion are the main advantages of this kind of devices, but also the fact that lot of the stimulus characteristics can be easily controlled by software. In fact, it seems to be much more flexible compared to

LED stimulator devices but it has also a major drawback; every computer monitor is able to update the image displayed on its screen at a certain frequency that is commonly set between 60 Hz and 85 Hz. This means that an upper limit to the displayable flickering frequency exists, given by the half of the screen refresh frequency for flash stimuli and by the screen refresh frequency for pattern reversal stimuli.

Chapter 4

Feature Extraction and Dimensionality Reduction

Usually, the sources of EEG data are presented as a large matrix. To handle effectively, it is needed to reduce the large volume to data into summarized matrix that are very close to the original data. The summarized matrices are generally contained in a small number of rows and columns, and therefore can be used much more efficiently than the original large matrix. Dimensionality reduction is a technique that is used to narrow the large matrix. After signal preprocessing, the information needed for classification into individual states or for cluster analysis is extracted. This process is called feature extraction, and represents an important step in EEG data processing. Feature extraction can be defined as automated recognition of various descriptive features of signals. Each segment obtained by signal segmentation can be represented by its extracted features. A good feature should remain unchanged if variations take place within a class, and it should reveal important differences when discriminating between patterns of different classes. Many variables were tested for the purposes of the thesis. It is helpful to describe subtle changes in the psychophysiological state of the brain during long-term EEG measurements. Feature extraction can be used as a preprocessor for applications including visualization, classification, detection, and verification. Besides, extracting informative and discriminative features from EEG signals are often of crucial importance for representing and classifying patterns of brain activations. Electroencephalography (EEG), a non-invasive tool to measure the electrical activity of the brain is the most studied potential non-invasive inter-

face for BCI implementation, among the brain activities monitoring techniques, mainly due to its low cost, ease of use, non-invasive nature, and comparatively easily recording brain signals. A brain computer interface (BCI) aims at translating brain signals into commands and it will be useful for people with severe motor impairments in order to restore communication and movement, on the other hand, it could be a new interface for healthy people. EEG features extraction plays an important role in conjunction with methods that evaluate this feature in different scenarios such as detection or classification of EEG signals.

4.1 Statistical Background

Before giving a description of dimensionality reduction and feature extraction techniques, this chapter first introduces mathematical concepts that will be used. It covers mean, standard deviation, variance, covariance, eigenvectors and eigenvalues.

Mean: The mean of a data sample is simply the average of the data points contained the data sample. It does not give much information about the data but only the center or middle point of the data sample. The mean of a data set X can be calculated according to the following formula:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Here, the symbol $\overline{X}$ (called "X bar" to indicate the mean of the data set X, n is used to refer to the number of elements in X. The above formula tells "Add up all the numbers and then divide by how many there are".

Standard Deviation: Standard Deviation (SD) is a measurement of the dispersion of a set of data from its mean. It also defined as the average distance from the mean of the data set to a point. If the data points are further from the mean, there is higher deviation within the data set. Actually the SD of a data set tells how spread out the data is. The SD is calculated as to compute the squares of

the distance from each data point to the mean of the set, add them all up, divide by $n-1$, and take the positive square root.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{(n-1)}} \tag{4.1}$$

where σ represents the standard deviation of a sample. Here, the data set is considered as a sample of data set, i.e., a subset of the real-world data. In equation (4.1) the summation is divided by $(n-1)$ because it turns out that this gives more closer result to the standard deviation that would result if one had used the entire data set. However, to calculate the standard deviation for an entire data set it should be divided by n instead of $(n-1)$.

Variance: The definition of variance is the average of the squared differences from the mean. A value of zero means that there is no variability i.e., all the numbers in the data set are the same. Actually, it measures how far a set of numbers are spread out from their average value. Basically, it is almost identical to the standard deviation. The formula is this:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{(n-1)} \tag{4.2}$$

Covariance: It can only possible to calculate the standard deviation for each dimension of the data set independently of the other dimensions since it only operate on one dimension. However many data sets have more than one dimension, and the aim of the covariance is to find out how much the dimensions vary from the mean with respect to each other. It is similar to variance, but where variance tells how a single variable varies, covariance tells how two variables vary together. Covariance is always measured between two dimensions. The covariance between one dimension and itself it is called variance.

If a data set is the three dimensional set (x, y, z), then it is possible to measure the covariance between the x and y dimensions, the x and z dimensions, and the y and z dimensions. The covariance between x and x , y and y and z and z will give the variance of the x, y and z dimensions respectively. The formula for covariance

is very similar to the formula for variance as in equation (4.3).

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)} \tag{4.3}$$

The value of covariance can be positive or negative even the sign is more important than its exact value. The positive value indicates that both dimensions increase together. On the other hand, the negative value indicates as one dimension increases, the other decreases. Moreover, if the covariance is zero, it indicates that the two dimensions are independent of each other.

Covariance Matrix: The covariance is always measured between two dimensions. If a data set has more than two dimensions, there is more than one covariance measurement that can be calculated. To calculate all the possible covariance values between all the different dimensions and put them in a matrix is a very useful way. This matrix is called covariance matrix. In the covariance matrix, the variances appear along the diagonal and covariances appear in the off-diagonal elements. If a data set is three dimensional x, y and z then the covariance matrix has three rows and three columns as shown below:

$$C = \begin{bmatrix} var(x) & cov(x,y) & cov(x,z) \\ cov(y,x) & var(y) & cov(y,z) \\ cov(z,z) & cov(z,y) & var(z) \end{bmatrix}$$

Eigenvectors and Eigenvalues: An eigenvector of a linear transformation is a non-zero vector that only changes by an overall scale when that linear transformation is applied to it.In general, a matrix acts on a vector by changing both its magnitude and its direction. However, a matrix may act on certain vectors by changing only their magnitude, and leaving their direction unchanged (or, possibly, reversing it). These vectors are the eigenvectors of the matrix. A matrix acts on an eigenvector by multiplying its magnitude by a factor, which is positive if its

direction is unchanged and negative if its direction is reversed. This factor is the eigenvalue associated with that eigenvector. Eigenvectors give you the direction of spread of data, while eigenvalue is the intensity of spread in a particular direction or of that respective eigenvector.

Let A be an $n \times n$ square matrix. The number λ is an eigenvalue of A if there exists a non-zero vector $\overrightarrow{v}$ such that

$$A\overrightarrow{v} = \lambda\overrightarrow{v}$$

In this case, the vector $\overrightarrow{v}$ is called an eigenvector of A corresponding to eigenvalue λ. To calculate the eigenvalues and eigenvectors the above condition can be rewritten as

$$(A - \lambda I)\overrightarrow{v} = 0$$

where I is the $n \times n$ identity matrix. Now, in order for a non-zero vector $\overrightarrow{v}$ to satisfy the equation, $(A - \lambda I)$ must not be invertible. That is the determinant of $(A-\lambda I)$ must equal to 0. Here, $p(\lambda) = det(A-\lambda I)$ is the characteristic polynomial of A. Now the roots of the characteristic polynomial of A are the eigenvalues of A. To find the eigenvectors $\overrightarrow{v}$ corresponding to the eigenvalues λ, the system of linear equations given by $(A - \lambda I)\overrightarrow{v} = 0$ can be solved.

4.2 Feature Extraction Techniques

Feature extraction is a key step of a pattern recognition system. Mainly it transforms input parameter vector into a feature vector and/or reduce its dimensionality. To find an informative and small but discriminative subset of features, one of the most important task is in any pattern recognition system. Techniques that can perform the transformation from high-dimensional to discriminative low-dimensional feature space are very important because of the curse of dimensionality. The purpose of feature extraction step is to identify the pattern of brain activity then subsequently the features are used as an input of a classifier. The system performance actually depends on both the extracted features and employed

classification algorithm. The extracted feature space represented as NXD data matrix where N and D defines number of samples and features respectively.

The dimensionality reduction based on the domain knowledge can be performed to explore more important information which may result in a lower dimensional vector, usually the remaining dimensionality is still too high for learning [239]. But the domain knowledge can not perform directly without explicit background or meaning of training data. Dimensionality reduces number of columns, but maps vectors to subspace. During preprocessing the dimension of EEG data is calculated as number of channels X number of trials. During EEG recording, for more than hundred EEG channels, a recording session spanning through hundreds of trial each spanning through several seconds or minutes or even hours with sample frequency of 100Hz or more. Therefore, the amount of produced data may be of the order of ten or even hundreds of gigabytes. Since it is almost impossible even to load the huge amount of data into the main memory of most computers, some kind of data reduction is necessary. As mentioned in [136, 206, 154], dimensionality reduction can be done by selecting appropriate channels or time epochs or trials. EEG signal processing is subject to several important constraints. The salient features in the EEG processing [190] are *i*) the number of signals to be processed is high, and often tightly interdependent, *ii*) signals are unique, in the sense that the circumstance under which they were obtained are normally nor repeatable, *iii*) medical signals are very noisy and *iv*) in some cases information about the signals is required in real time in order to take crucial decision.

A number of techniques have been proposed for dimensionality reduction and feature extraction, such as Singular Value Decomposition (SVD), Principle Component Analysis (PCA), Independent Component Analysis (ICA), and Common Spatial Pattern (CSP).

4.2.1 Singular Value Decomposition (SVD)

The Singular Value Decomposition (SVD) is a method that is used to transform correlated variables into a set of uncorrelated ones that better expose the various

relationships among the original data items. At the same time, it is a method for identifying and ordering the dimensions along which data points exhibit the most variation. This ties in to the third way of viewing SVD, which is that once we have identified where the most variation is, it is possible to find the best approximation of the original data points using fewer dimensions. The SVD technique is widely used to decompose a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix. The decomposition of a matrix is often called factorization. Hence, SVD can be seen as a method for data reduction. The decomposition of a matrix is also useful when the matrix is not of full rank. That is, the rows or columns of the matrix are linearly dependent. The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal. Using the SVD, one can determine the dimension of the matrix range or more-often called the rank. The rank of a matrix is equal to the number of linear independent rows or columns. This is often referred to as a minimum spanning vector simply a basis. The SVD can also quantify the sensitivity of a linear system to numerical error or obtain a matrix inverse.

According to the theory of linear algebra a rectangular matrix A can be broken down into the product of three matrices − an orthogonal matrix U, a diagonal marix S, and the transpose of an orthogonal matrix V. The theory is represented as

$$A_{ij} = U_{ii} S_{ij} V_{jj}^T \tag{4.4}$$

where $U^T U = I, V^T V = I$; the columns of U are orthonormal eigenvectors of AA^T, the columns of V are orthonormal eigenvectors of $A^T A$, and S is a diagonal matrix containing the square roots of eigenvalues from U or V in descending order.

To find the SVD, USV^T consider a rectangular matirx A, where

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

Calculation of U: The transpose of A is

$$A^T = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix}$$

Therefore,

$$AA^T = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}$$

The eigenvectors are defined by the equation $A\overrightarrow{v} = \lambda\overrightarrow{v}$ where λ and $\overrightarrow{v}$ represents eigenvalues and their corresponding eigenvectors respectively. Applying this the eigenvalues and their corresponding eigenvectors of AA^T can be calculated as

$$\begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This can be written as

$$11x_1 + x_2 = \lambda x_1$$
$$\Rightarrow (11 - \lambda)x_1 + x_2 = 0$$

and

$$x_1 + 11x_2 = \lambda x_2$$
$$\Rightarrow x_1 + (11 - \lambda)x_2 = 0$$

The values of λ can be calculated by setting the determinant of the coefficient matrix to zero,

$$\begin{vmatrix} (11 - \lambda) & 1 \\ 1 & (11 - \lambda) \end{vmatrix} = 0$$
$$\Rightarrow (11 - \lambda)(11 - \lambda) - 1 \cdot 1 = 0$$
$$\Rightarrow (\lambda - 10)(\lambda - 12) = 0$$

So, the two eigenvalues are

$$\lambda = 10, \lambda = 12$$

The eigenvectors for the eigenvalues can be calculated. For $\lambda = 10$

$$(11 - 10)x_1 + x_2 = 0$$
$$\Rightarrow x_1 = -x_2$$

which is true for many values, so $x_1 = 1$ and $x_2 = -1$ can be picked. Thus $(1, -1)$ is the eigenvector corresponding to the eigenvalue $\lambda = 10$. Similarly, for $\lambda = 12$

$$(11 - 12)x_1 + x_2 = 0$$
$$\Rightarrow x_1 = x_2$$

In this case, $x_1 = 1$ and $x_2 = 1$ can be picked and $(1, 1)$ is the eigenvector corresponding to the eigenvalue $\lambda = 12$. These eigenvectors become column vectors in a matrix ordered by the size of the corresponding eigenvalue. In other words, the eigenvector of the largest eigenvalue is column one, the eigenvector of the next largest eigenvalue is column two, and so forth and so on until we have the eigenvector of the smallest eigenvalue as the last column of our matrix. In the matrix below, the eigenvector for $\lambda = 12$ is column one, and the eigenvector for $\lambda = 10$ is column two.

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

By applying the Gram-Schmidt orthonormalization process the matrix of column vectors can convert into an orthogonal matrix.

$$\overrightarrow{u_1} = \frac{\overrightarrow{v}}{|\overrightarrow{v}|} = \frac{[1,1]}{\sqrt{1^2+1^2}} = \frac{[1,1]}{\sqrt{2}} = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] = [0.7071, 0.7071]$$

Compute

$$\overrightarrow{w_2} = \overrightarrow{v_2} - \overrightarrow{u_1} \cdot \overrightarrow{v_2} * \overrightarrow{u_1} = [1, -1] - [0.7071, 0.7071] \cdot [1, -1] * [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] =$$
$$[1, -1] - 0 * [0.7071, 0.7071] = [1, -1] - [0, 0] = [1, -1]$$

and normalize

$$\overrightarrow{u_2} = \frac{\overrightarrow{w_2}}{|\overrightarrow{w_2}|} = [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}] = [0.7071, -0.7071]$$

to give

$$U = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$$

Calculation of V: The V is calculated based on $A^T A$, so

$$A^T A = \begin{bmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix}$$

Now the eigenvalues and corresponding eigenvectors of the $A^T A$ can be calculated as

$$\begin{bmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which represents the system of equations

$$10x_1 + 2x_3 = \lambda x_1$$
$$\Rightarrow (10 - \lambda)x_1 + 2x_3 = 0$$
$$10x_2 + 4x_3 = \lambda x_2$$
$$\Rightarrow (10 - \lambda)x_2 + 4x_3 = 0$$
$$2x_1 + 4x_2 + 2x_3 = \lambda x_3 \Rightarrow 2x_1 + 4x_2 + (2 - \lambda)x_3 = 0$$

As above, setting the determinant of the coefficient matrix to zero the eigenvalues, λ can be calculated

$$\begin{vmatrix} (10-\lambda) & 0 & 2 \\ 0 & (10-\lambda) & 4 \\ 2 & 4 & (2-\lambda) \end{vmatrix} = 0$$

$$\Rightarrow (10-\lambda)\begin{vmatrix} (10-\lambda) & 4 \\ 4 & (2-\lambda) \end{vmatrix} + 2\begin{vmatrix} 0 & (10-\lambda) \\ 2 & 4 \end{vmatrix} = 0$$

$$\Rightarrow (10-\lambda)[(10-\lambda)(2-\lambda)-16] + 2[0-(20-2\lambda)] = 0$$
$$\Rightarrow \lambda(\lambda-10)(\lambda-12) = 0$$

Therefore, the eigenvalues $\lambda = 0, \lambda = 10$ and $\lambda = 12$ are the eigenvalues for $A^T A$. Now the corresponding eigenvectors for the eigenvalues can be calculated. For $\lambda = 12$

$$(10-12)x_1 + 2x_3 = 0$$
$$-2x_1 + 2x_3 = 0$$
$$x_1 = 1, \text{ and } x_3 = 1$$
$$(10-12)x_2 + 4x_3 = 0$$
$$-2x_2 + 4x_3 = 0$$
$$x_2 = 2x_3$$
$$x_2 = 2$$

So, $\overrightarrow{v}_1 = [1, 2, 1]$ for $\lambda = 12$. Similarly, for $\lambda = 10$ and $\lambda = 0$ the $\overrightarrow{v}_2$ and $\overrightarrow{v}_3$ can be calculated as $\overrightarrow{v}_2 = [2, -1, 0]$ and $\overrightarrow{v}_3 = [1, 2, -5]$. In the matrix below, the eigenvector for $\lambda = 12$ is column one,the eigenvector for $\lambda = 10$ is column two and the eigenvector for $\lambda = 0$ is column three.

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 0 & -5 \end{bmatrix}$$

By applying the Gram-Schmidt orthonormalization process the matrix of column vectors can convert into an orthogonal matrix.

$$\overrightarrow{u_1} = \frac{\overrightarrow{v_1}}{|\overrightarrow{v_1}|} = \frac{[1,2,1]}{\sqrt{1^2+2^2+1^2}} = \frac{[1,2,1]}{\sqrt{6}} = [\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}] = [0.4082, 0.8165, 0.4082]$$

$$\overrightarrow{w_2} = \overrightarrow{v_2} - \overrightarrow{u_1} \cdot \overrightarrow{v_2} * \overrightarrow{u_1} =$$

$$[2,-1,0] - [0.4082, 0.8165, 0.4082] \cdot [2,-1,0] * [0.4082, 0.8165, 0.4082] = [2,-1,0]$$

$$\overrightarrow{u_2} = \frac{\overrightarrow{w_2}}{|\overrightarrow{w_2}|} = [\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}}, 0] = [0.8944, -0.4472, 0]$$

$$\overrightarrow{w_3} = \overrightarrow{v_3} - \overrightarrow{u_1} \cdot \overrightarrow{v_3} * \overrightarrow{u_1} - \overrightarrow{u_2} \cdot \overrightarrow{v_3} * \overrightarrow{u_2} = [\frac{-2}{3}, \frac{-4}{3}, \frac{10}{3}] = [-0.6667, -1.3333, 3.3333]$$

$$\overrightarrow{u_3} = \frac{\overrightarrow{w_3}}{|\overrightarrow{w_3}|} = [\frac{1}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{-5}{\sqrt{30}}] = [0.1826, 0.3651, -0.9129]$$

Now $\overrightarrow{u_1}$, $\overrightarrow{u_2}$ and $\overrightarrow{u_3}$ make the V as

$$V = \begin{bmatrix} 0.4082 & 0.8944 & 0.1826 \\ 0.8165 & -0.4472 & 0.3651 \\ 0.4082 & 0 & -0.9129 \end{bmatrix}$$

And its transpose

$$V^T = \begin{bmatrix} 0.4082 & 0.8165 & 0.4082 \\ 0.8944 & -0.4472 & 0 \\ 0.1826 & 0.3651 & -0.9129 \end{bmatrix}$$

Calculation of S: The diagonal matrix S can be formed by taking the square roots of the non-zero eigenvalues of U and V and populate the diagonal with them, putting the largest eigenvalue in s_{11}, the second largest in s_{22}, the third largest in s_{33} and so on until the smallest value ends up in s_{ii}. For the purpose of multiplication between U and V, a zero column vector is added to S so that it is of the proper dimension.

$$S = \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix}$$

Now, the equation (4.4) gives

$$A_{ij} = U_{ii}S_{ij}V_{jj}^T = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix} \times \begin{bmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{bmatrix} \times \begin{bmatrix} 0.4082 & 0.8165 & 0.4082 \\ 0.8944 & -0.4472 & 0 \\ 0.1826 & 0.3651 & -0.9129 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$$

4.2.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is very useful multivariate statistical technique which is greatly used by brain computer interface. It transforms a number of correlated variables into a number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The objectives of PCA are to discover or to reduce the dimensionality of data set as well as to identify new meaningful underlying variables. The new variables called principal components are calculated by the PCA which are obtained as linear combinations of original variables. PCA analyzes a data set representing observations described by several dependent variables, which are, in general, inter-correlated. The goals of the PCA are to extract the most important information from the data set; keeping only the important information compress the size of the data set; simplify the description of the data set; and analyze the structure of the observation and the variables.

The basic steps for performing a principal component analysis are as follows:
i) Ignoring the class labels, take the whole dataset consisting of d -dimensional samples.
ii) Compute the means for every dimension (d -dimensional) of the whole dataset.
iii) Compute the covariance matrix of the whole data set.
iv) Apply SVD and compute eigenvectors $e_1, e_2, e_3, ..., e_d$ and corresponding eigenvalues $\lambda_1, \lambda_2, \lambda_3, ..., \lambda_d$.

v) Arrange the eigenvectors in descending order and choose k of them with the largest eigenvalues to form a $d \times k$ dimensional matrix.
vi) Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation:

$$y = w^T \times x$$

where x is a $d \times 1$-dimensional vector representing one sample, and y is the transformed $k \times 1$-dimensional sample in the new subspace.

4.2.3 Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is a statistical technique very useful in systems involving multivariable data. The computational technique is used to separate a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. Suppose, two signals denoted by s_1 and s_2 are produced by two sources and the these are linearly mixed according to the following manner:

$$x_1 = a_{11}s_1 + a_{12}s_2$$
$$x_2 = a_{21}s_1 + a_{22}s_2$$

where x_1 and x_2 are two mixed signals, a_{11}, a_{12}, a_{21} and a_{22} are some parameters that depend on the distance from the sources of origin. Now, it would be worthy to separate the two original signals s_1 and s_2 only using the mixed signals. The ICA technique was originally developed to solve the problem. Recently, the ICA technique is used in BCI application. The EEG signals are recordings of electrical potentials in many different locations on the scalp. These potentials are generated by mixing some underlying components of brain activity. It is important to find the original components of brain activity, but one can only observe mixtures of the components. The ICA technique can provide interesting information on brain activity by giving access to its independent components. ICA can olso be used in pattern recognition, signal enhancement and data reduction.

The main purpose of ICA is to find new components that are mutually independent in complete statistical sense. It is assumed that we observe p linear mixtures of $x_1, x_2, x_3, ..., x_p$ and they can be modeled as linear combinations of n random variables $s_1, s_2, s_3, ..., s_n$ as follows:

$$x_i = c_{i1}s_1 + c_{i2}s_2 + ... + c_{in}s_n, i = 1, 2, ..., n \tag{4.5}$$

where each $c_{ij}(j = 1, 2, 3, ..., n)$ in equation (4.5) is an unknown real coefficient and s_j is statistically mutually independent and can be denoted as the Independent Components (ICs). Equation (4.5) can be written in matrix terms as

$$x = Cs$$

$$or, x = \sum_{i=1}^{n} c_i s_i \tag{4.6}$$

where $x = (x_1, x_2, x_3, ..., x_p)^T, s = (s_1, s_2, s_3, ..., s_n)^T$ and C is the $p \times n$ mixing matrix that contains all c_{ij}. The equation (4.6) is called independent component analysis where only measures variable x is available and the objective is to determine both the matrix C and the independent components.

4.2.4 Common Spatial Pattern (CSP)

Common Spatial Pattern (CSP) is a feature extraction technique used in signal processing for separating a multivariate signal into additive sub-components. The technique used to design spatial filters such that the variance of the filtered data from one class is maximized while the variance of the filtered data from the other class is minimized. Thus, the resulting feature vectors increase the discriminability between the two classes by means of minimize the intra class variance and maximize the inter class variance [205]. This property builds CSP as one of the most effective spatial filters for EEG signal processing. The method of CSP was first introduced to EEG analysis for detection of abnormal EEG [126] and effectively applied on movement-related EEG for the classification purpose [22, 63]. The target of the CSP is to project the multichannel EEG data into low dimensional spatial subspace with a projection matrix using linear transformation [242].

For details explanation of the CSP algorithm, assume the original EEG data matrix E_k^i from trial i for class k. The dimension of each E_k^i is $D \times T$, where D is the number of channels and T is the number of samples per channel. For the explanation, the EEG data of a single trial ($i = 1$) is represented as $E_{k\epsilon(h,f)}$ where h denotes hand and f denotes foot movement. The normalized spatial covariance of the EEG for hand movement, C_h and for the foot movement, C_f can be calculated as:

$$C_h = \frac{E_h E_h^{'}}{tr(E_h E_h^{'})}, C_f = \frac{E_f E_f^{'}}{tr(E_f E_f^{'})} \tag{4.7}$$

where E_h and E_f represent the original EEG matrices for hand and foot movement respectively, $(\cdot)^{'}$ is the transpose operator and $tr(\cdot)$ represents the sum of the diagonal elements of any given matrix. The composite spatial covariance, C is the sum of the averaged normalized spatial covariance $\bar{C}_h$ and $\bar{C}_f$. The $\bar{C}_h$ and $\bar{C}_f$ are estimated by averaging over all the trials of each class. The composite spatial covariance, C is calculated as

$$C = \bar{C}_h + \bar{C}_f = \Sigma\lambda\Sigma^{'} \tag{4.8}$$

where Σ is the matrix of eigenvectors and λ is the diagonal matrix of eigenvalues. The averaged normalized spatial covariance $\bar{C}_h$ and $\bar{C}_f$ are transformed as

$$J_h = X\bar{C}_h X^{'}, J_f = X\bar{C}_f X^{'} \tag{4.9}$$

where $X = \frac{\Sigma^{'}}{\sqrt{\lambda}}$ is the whitening transformation matrix. J_h and J_f share common eigenvectors and the sum of corresponding eigenvalues for the two matrices will always be one. If $J_h = Y\Lambda_h Y^{'}$ and $J_f = Y\Lambda_f Y^{'}$ then $\Lambda_h + \Lambda_f = I$, where I is the identity matrix. Since the sum of two corresponding eigenvalues is always one, a high eigenvalue for J_h means that a high variance for EEG in hand movement and a low variance for the EEG in foot movement (low eigenvalue for J_f) and vice versa. The classification operation is done based on this property. The projection of whitened EEG onto the eigenvectors Y corresponding to the largest Λ_h and

Λ_f will give feature vectors that significantly enhance the discrimination ability. The goal of the CSP is to find B spatial filters to create a projection matrix W of dimension $D \times B$ (each column is a spatial filter). The projection matrix W is represented as

$$W = Y^{'}X$$

The projection matrix W linearly transforms the original EEG into uncorrelated components according to:

$$Z = W^T E \tag{4.10}$$

The original EEG, E can be reconstructed by $E = W^{-1}Z$ where W^{-1} is the inverse matrix of W. The columns of W^{-1} are spatial patterns that describe the variance of the EEG. The first and last columns contain the most discriminatory spatial patterns that explain the high variance of one class and the low variance of the other.

4.2.5 The Composite CSP (CCSP)

The CSP algorithm does not consider the inter-subject information. It only exploits covariance matrices on a subject-by-subject basis. The Composite CSP (CCSP) algorithm, a modification of the CSP algorithm is made by Kang et al [113] in which they consider subject-to-subject transfer and exploit a linear combination of covariance matrices. In the CCSP, covariance matrices, M_c can be measured as a weighted sum of the covariance matrices of the other subjects, $M_c = \sum_{s \epsilon B} U_c^S V_c^S V_{t,c}^{-1}$ where B is a set of subjects, U_c^S is the spatial covariance matrix for class c and subject s , V_c^S is the number of EEG trials used to calculate U_c^S , and the total number of EEG trials for class c is $V_{t,c}$.

4.2.6 Spatially Regularized CSP (SRCSP)

The CSP algorithm overlooks the spatial position of EEG electrodes though it is used to learn spatial filters. The main target of the Spatially Regularized CSP

(SRCSP) is to consider this spatial information [149]. The aim of this algorithm is to get spatially smooth filters so that adjacent electrodes can ensure comparable weights. The spatial smoothness of the filters φ can be achieved by using Laplacian penalty term and the regularization matrix R [149, 251].

$$R = D_N - N, N(p,q) = exp(-\frac{\|\omega_p - \omega_q\|^2}{2d^2}) \tag{4.11}$$

where ω_p is the vector containing the 3D coordinates of the p^{th} electrode, D_N is the diagonal matrix and d is the distance between two electrodes. Once the weights of adjacent electrodes are very close to each other, the penalty term $P(\varphi) = \varphi^T R\varphi$ will be of small value since $\varphi^T(D_N - N)\varphi = \sum_{p,q} N(p,q)(\varphi_p - \varphi_q)^2$.

4.2.7 Regularized CSP with Generic Learning (GLRCSP)

The Regularized CSP with Generic Learning (GLRCSP) [151] algorithm uses data from other subjects and regularizes the estimation of the covariance matrix. The identity matrix and generic covariance matrix are shrunk by the algorithm. Like CCSP, in this algorithm the covariance matrix M_c is calculated using the covariance matrices of other subjects.

$$M = H_M \sum_{s\epsilon B} U_c^S, H_M = \frac{1}{(1-\beta)G_c + \beta \sum_{s\epsilon B} G_{U_c^S}} \tag{4.12}$$

where G_c is the number of trials used to calculate the covariance matrix c and β is the regularization term.

4.2.8 CSP with Tikhonov Regularization (TRCSP)

The CSP with Tikhonov Regularization (TRCSP) [149] works grounded on the regularization of the CSP with quadratic penalties. The Tikhonov Regularization is made up of penalizing explanation with large weights. When the penalty term $P(\varphi) = \|\varphi\|^2 = \varphi^T\varphi = \varphi^T I\varphi$, the TRCSP is obtained by using regularization matrix $R = I$.

Chapter 5

Dominant Feature Selection

5.1 Background on Feature Selection

A feature of EEG dataset is an individual measurable property of the dataset being observed. The classification performance of the EEG dataset is observed using a set of features. Recently, the domain of features has extended from ten to hundred of features in the applications of machine learning. The dimensionality of the data involved in machine learning and data mining tasks increases day by day. Data with extremely high dimensionality poses challenges to existing learning task due to the curse of dimensionality. With the presence of a large number of irrelevant features, a learning model tends to overfit and becomes less comprehensible. Some dimensionality reduction techniques have been employed in the machine learning research area in order to resolve the problem of the curse of dimensionality. Feature selection is a widely used and effective technique to identify relevant features for dimensionality reduction [145, 92]. The objectives of the feature selection methods are to subset the relevant features from the original features which lead to better classification performance with reduced computational cost. The training set can be labeled, unlabeled or partially labeled and accordingly the feature selection methods can be categorized into supervised [224, 244], unsupervised [68, 164] and semi-supervised feature selection [253, 260].

5.1.1 Supervised Feature Selection

Supervised feature selection methods can also be categorized into three models namely filter models, wrapper models and embedded models. The filter model relies on measures of the general characteristics of the training data such as distance, consistency, dependency, information, and correlation. Examples of the filter model algorithms are Relief [209], Fisher score [66] and Information Gain based methods [191]. The wrapper model requires a predetermined learning algorithm to determine the quality of selected features. A major disadvantage of the model is that to run with a large number of features of these methods are expensive. The embedded model was suggested to overcome the shortcomings. In the model features subsets with a given cardinality are selected with the incorporation of the statistical criteria like filter model. Moreover, the subset with the highest classification accuracy is selected in the model [145]. Hence, the model attains both comparable accuracy to the wrapper and comparable efficiency to the filter model. Both the model fitting and feature selection operations are performed simultaneously in the embedded model [38].

5.1.2 Unsupervised Feature Selection

Unsupervised feature selection method selects features considering only cluster information [67] and it does not account class labels. Supervised feature selection approach measures the relevance of features by the guidance of class labels and hence needs enough labeled data, which is expensive in both time and effort. Unlike supervised technique, unsupervised feature selection method considers the unlabeled data although it is not easy to find out the relevance of features. Usually a huge dimensional data set contains only a small labeled sample size.

5.1.3 Semi-Supervised Feature Selection

Supervised feature selection works based on class label and unsupervised feature selection method do not consider any class label. However, usually in real-word

applications, a huge dimensional data set contains a small number of labeled samples and a large number of unlabeled samples. Therefore, in this situation both the methods are not appropriate feature selection methods. The supervised methods use only a small number of labeled samples that restricts the method to provide enough correlated features information while unsupervised methods totally ignore class labels that limits to provide useful information to separate different classes. Hence, it is needs to develop a method that feats both the labeled and unlabeled samples and that is the semi-supervised methods.

5.2 Importance of Feature Selection

The feature selection process selects a subset of original features according to certain measures. Usually the data set practically exists hundreds, even thousands, features, but many of them are either redundant or irrelevant. The huge amount of features enhance the over fitting and computational problem. Dataset always contains more or wrong information than is needed to build a feature selection model. The feature selections not only improve the quality of the model but also make the process of modeling more efficient. In EEG classification problems the feature selections are important because it

i) empowers the machine learning algorithm to train faster;

ii) simplifies a model complexity and makes it easier to understand;

iii) improves the accuracy of a model if the right subset is chosen; and

iv) moderates over fitting.

5.3 Challenges of Feature Selection

The conventional feature selection methods face some challenges to handle the big dataset. Most of the methods accomplished their task only considering the correlation among features. However, they do not consider intrinsic structures among features. In practical applications, features show different types of structures such as disjoint groups, spatial or temporal smoothness, overlap groups, trees

and graphs [141]. The ignorance of the structured features may deteriorate the performance of classification. Linked data becomes pervasive in many areas such as social and biological networks. Because they are correlated with each other by different types of links, the linked data are distinct from traditional attribute-value data. The feature selection from the linked data may face challenges e.g., relation exploitation among data instances, evaluation of the relevance features without class labels.

5.4 Feature Selection Methods

In classification problems, to find the smallest classification error, the feature selection methods select a subset of size m from a given set of Y features. Suppose, the features set Y with cardinality d and the desired number of features is m in the selected subset X, $X \subseteq Y$. Let $J(X)$ be the feature selection criterion function for the feature subset X and the function can be selected as $J = (1 - A_e)$ where A_e denotes the classification error. Hence, the higher values of J shows a better feature subset. The most straightforward feature selection method examines all $\binom{d}{m}$ possible subsets of size of m and selects the subset with the largest value of $J(\cdot)$. The non-exhaustive sequential feature selection methods cannot be guaranteed to produce the optimal subset [48]. The optimal feature selection method which avoids the exhaustive search is based on the branch and bound algorithm. This method avoids an exhaustive search by using intermediate results for obtaining bounds on the final criterion value. The key property of this method is the monotonicity of the criterion function $J(\cdot)$. This property gives $J(X_1) < J(X_2)$ where X_1, X_2 are two feature subsets and $X_1 = X_2$.

Since the feature selection operation is done in off-line, the execution time is not a big concern for the moderate size of feature sets. However, recently more than thousands of features are involving in data mining and classification applications. In this situation, the computational time of a feature selection algorithm is very important. Some of the feature selection methods which have been proposed in

[109] are listed below:

i) Exhaustive Search

ii) Branch-and-Bound Search

iii) Best Individual Features

iv) Sequential Forward Selection (SFS)

v) Sequential Backward Selection (SBS)

vi) Plus l-take away r Selection

vii) Sequential Forward Floating Search (SFFS) and

viii) Sequential Backward Floating Search (SBFS)

The Exhaustive Search method evaluate all of the $\binom{d}{m}$ possible subsets and assured to discover the optimal subset but the method is not practicable for even moderately large values of m and d. The Branch-and-Bound Search method guaranteed to find the optimal subset provided the criterion function satisfies the monotonicity property. In this method, a portion of all possible feature subsets need to be computed to find the optimal subset. One of the simple feature selection methods is the Best Individual Features method which evaluates all of the m features individually and selects best m features. The simple sequential methods are SFS and SBF which adds and deletes respectively one feature at a time. In SFS, once a feature is selected, it cannot be rejected. This method is computationally smart since to select a subset of size 2, it observes only $(d-1)$ possible subsets. On the other hand, in SBS, once a feature is deleted, it cannot be brought back into the optimal subset. The SBS requires more computation than SFS. The Plus l-take away r Selection, SFFS and SBFS [202] are sophisticated methods for feature selection. In these methods the selection process repeats as long as they find progresses compared to previous feature sets of the same size. Also, these methods perform better than the SFS and SBS for the case of large feature selection problem. In the Plus l-take away r Selection method, the value of l and r are determined automatically and updated dynamically.

5.5 Proposed Feature Selection Method

Common Spatial Pattern (CSP) is an algorithm commonly used in BCI systems to preprocess the EEG signals [204, 63, 171]. The algorithm finds optimal spatial filters that are functional in discriminating two classes of EEG signals in MI based BCI. It maximizes the ratio of average variances that belong to two different classes. A computed CSP spatial filter projects the multi-dimensional EEG time domain signal to a one-dimensional time domain signal. The CSP handles two classes at the same time and simultaneously diagonalizes the covariance matrices of both classes [70]. Also, it is proved that the algorithm is efficient in BCI competitions [19, 20]. CSP optimizes the ratio of average variances of two classes, consequently requires only one average covariance matrix for each class. It creates difficulty during handling EEG like non-stationary signals as the covariance matrix of an EEG signal may change over time due to artefact [213]. The fitness function constraint is another drawback of the CSP. CSP does not allow different types of fitness function, which may be more useful in different situation [72]. Over fitting occurs during the optimization of the ratio of average variances of the two classes due to its outlier sensitivity [207].

The most challenging task of CSP based BCI is to select the spatial filters for features extraction. In all of the existing works the spatial filters are selected manually. For example, in [149, 240], 3 pairs of filters are used as recommended in [22], only single filter pair is used in [225, 257], in [6] experimentation is prepared by using 4 pairs of filters and in [113] the classification performance with 1, 2, and 3 pairs of filters is compared. However, the manual selection of CSP filters does not confirm that the approach will be yielded best accuracy. In this research,a novel approach is proposed that selects the best CSP filter pair and corresponding most discriminative features and hence to improve classification performance. In this approach, grid search method is used to select the filter pair and features searching best linear discriminant analysis scores. A common problem of the CSP is that the relation between the CSP filter optimization and the classification per-

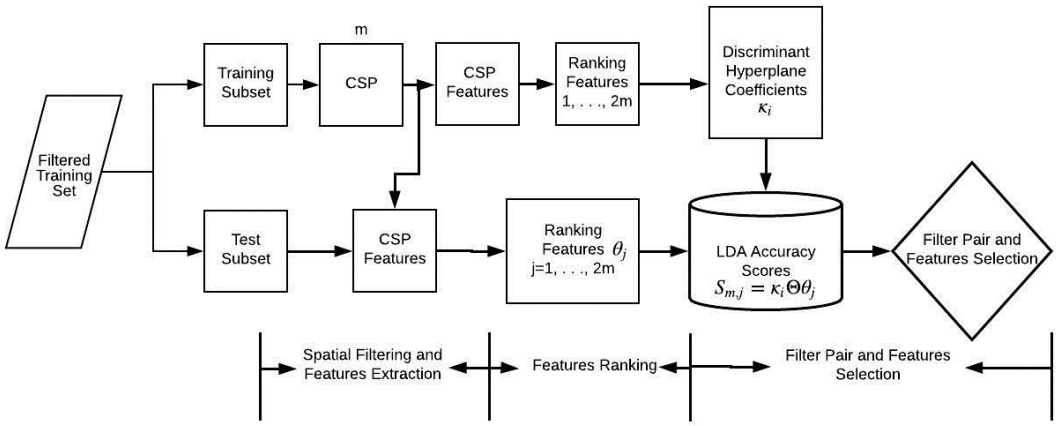


Figure 5.1: Proposed GS-CSP based spatial filter pair and features selection.

formance is not direct enough. The selected spatial filters do not certainly map the signals into spaces where the best classification accuracy is found by the corresponding extracted features [72]. The proposed CSP based Grid Search (GS-CSP) method searches the best combination of spatial filter pair and features (SFPF) that provide maximum classification accuracy.In the pre-processing stage, a number of channels that roughly cover the motor cortex of the brain are selected. The training and test sets of the raw EEG are separated and individually filtered by a band pass filter. Figure 5.1 shows the block diagram of the proposed GS-CSP approach. For the selection of suitable SFPF, the training set EEG signal is used. The GS-CSP comprises three stages and a detail of each is described below.

Spatial Filtering and Features Extraction: The CSP algorithm is greatly successful in calculating spatial filters. In this stage, the filtered training set is divided into two subsets: training subset and test subset. To calculate spatial filters, the CSP algorithm is used on the training subset. The filtering is accomplished by linearly transforming the EEG measurements by using equation (4.10). Using the calculated spatial filters, CSP features are then extracted from each of the trial of both training and test subset of EEG. It is done by projecting the EEG data into the CSP filters. The CSP features are generated as the log variance of the projected signals.

Features Ranking: In this stage the extracted training subset features are ranked on the basis of values of mutual information with training subset EEG

class. The test subset features are ranked in view of the ranking of training subset features. Mutual information is a measure of the amount of information by which one random variable tells about another random variable. The Mutual Information (MI) between two random variables $\mathbf{A}$ and $\mathbf{B}$ is [47]

$$MI(\mathbf{A};\mathbf{B}) = H(\mathbf{B}) - H(\mathbf{B}|\mathbf{A}) \tag{5.1}$$

where $\mathbf{A} = A_1, A_2, , A_d$ is a d-dimensional random variable. The entropy of $\mathbf{A}$ is

$$H(\mathbf{A}) = -\sum_{a\in\mathbf{A}} p(a) log_2 p(a) \tag{5.2}$$

The conditional entropy of random variables $\mathbf{A}$ and $\mathbf{B}$ is

$$H(\mathbf{B}|\mathbf{A}) = -\sum_{a\in\mathbf{A}}\sum_{b\in\mathbf{B}} p(a,b) log_2 p(b|a) \tag{5.3}$$

where $p(\cdot)$ are probability mass functions.

In the EEG classification case, the CSP features are continuous variables and class levels are discrete values. Hence, the MI between the CSP features $\mathbf{X}$ and class $\mathbf{X}_C$ is calculated as [6]

$$MuInfo(\mathbf{X};\mathbf{X}_C) = H(\mathbf{X}_C) - H(\mathbf{X}_C|\mathbf{X}) \tag{5.4}$$

where $\phi \in \mathbf{X}_C = 1, ..., N_\phi$ and the conditional entropy is

$$H(\mathbf{X}_C|\mathbf{X}) = -\int_{\mathbf{X}} \sum_{\phi=1}^{N_\phi} p(\phi|x) log_2 p(\phi|x) dx \tag{5.5}$$

where N_ϕ is the number of classes.

Filter Pair and Features Selection: In the third stage, the GS-CSP based SFPF selection is implemented with LDA. During LDA training the discriminant hyperplane coefficients κ_i are computed for the GS training ranked features $i(i = 1, , 2m)$ where m is the CSP filter pair. The LDA accuracy score of the GS testing ranked feature $\theta_j(j = 1, ..., 2m)$ is now measured by the computed coefficients. For every value of κ_i and their corresponding GS testing features θ_j, LDA accuracy scores are measured as $S_{m,j} = \kappa_i \Theta \theta_j$ where the symbol Θ represents LDA test operator. In this approach, a grid of accuracy scores is formed from where

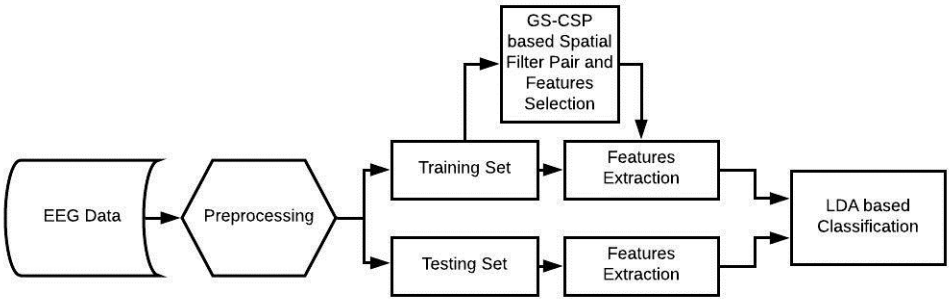


Figure 5.2: Block diagram of motor imagery based BCI with the proposed approach.

the appropriate SFPF by which the maximum accuracy score yielded are selected. The SFPF selection method is summarized as:

1. Learning spatial filters by the GS training EEG using CSP algorithm

2. Extracting CSP features by the spatial filters from both training subset and test subset data. For m pairs of spatial filters $2m$ numbers of CSP features are extracted

3. Calculating mutual information of the extracted training subset features with their corresponding EEG class

4. Ranking the extracted training subset features on the basis of values of the mutual information. Also the extracted test subset features are ranked according to the order of training subset features

5. Computing discriminant hyperplane coefficients (bias and slope of the discriminant hyperplane) for the ranked training subset features

6. Measuring the LDA accuracy score $S_{m,j}$ of ranked test subset feature θ_j based on the computed coefficients κ_i

7. Finally, selecting the best SFPF that produce maximum LDA accuracy score.

To implement motor imagery based BCI, the work flow of this research is presented with the block diagram as shown in Figure 5.2. The GS-CSP approach is employed on the training set EEG data to select the best spatial filter pairs and a number of dominating features. The CSP features of training set are finally extracted, using the selected filter pairs. Besides, the selected features are used in the LDA classifier to classify the testing set.

Chapter 6

Classification

6.1 EEG Signal Classification

In the field of biomedical exploration, the classification of EEG signal plays very significant role. For the diagnosis of brain diseases EEG signal classification is very important. At the same time, it is essential for the understanding of cognitive process. EEG segments can be distinguished by an efficient classification algorithm. The recorded EEG signals are usually huge volume of data. To handle the data is a big challenge. The data representation is the main problem for further analysis, such as classification. Initially, it is essential to extract dominant features from raw EEG signals. Then the extracted features can be used for classification.

The classification task happens through everyday life, and essentially means conclusions being prepared based on currently existing data. Some instances of classification tasks are the automatic processes used for categorization letters on the basis of machine read postcodes, assigning persons to credit status on the basis of financial and other personal information, and the earliest diagnosis of a patient's disease in order to select immediate treatment while awaiting conclusive test results [28]. In machine learning and pattern recognition, classification is a method for assigning a given piece of input data into one of a known number of categories [65, 28]. Assigning an e-mail to a spam or non-spam division or giving a diagnosis to a patient based on gender, blood pressure or presence or absence of definite symptoms, etc. can be an instance of the classification. The piece of input data and the categories are formally known as an instance and classes

respectively. The instance is described by a vector of features, which together constitute a description of all known characteristics of the instance.

The main purpose of the classification is to allocate class labels to the features extracted from the observations of a set of data in a particular problem. Classifier is an algorithm that implements classification, specifically in a real execution. Also a classifier refers to the mathematical meaning, implemented by a classification algorithm that maps input data to a group. The aim of a classifier is to identify the class of a feature vector thanks to training sets. The training sets are composed of feature vectors labeled with their original classes. The recorded brain activity through EEG leads to the acquisition of a huge amount of data. It is necessary to work with a smaller number of values which describe some relevant properties of the signals and that produced the best possible classification performance. The smaller numbers of values are known as features. Features are generally gathered into a vector known as a feature vector [148] that transforms one or several signals into a feature vector. The feature vector contained a set of all features used to describe a pattern, is a reduced dimensional representation of that pattern. Signal classification means to examine different characteristic features of a signal, and based on those characteristic features; decide to which grouping or class the signal belongs. And the classification outcome can be mapped back into the physical world to reveal information about the physical process that created the signal.

6.2 Types of Classification

Classification is mainly of two types: supervised and unsupervised. Observations of a set of data are associated with class labels in supervised classification. On the other hand, in unsupervised classification, observations are not labeled to a known class [110]. Brief descriptions of the supervised and unsupervised classifications are given below.

6.2.1 Supervised Classification

Supervised classification is widely used in the biomedical research. Almost all of the classification algorithms deal with a group of data that has some information about the dataset. In other words, the classifiers are trained by the class label information given within the dataset. This is the supervised learning, in which a supervisor instructs the classifier during the construction of the classification model. Supervised procedure assumes that a set of training data (the training set) has been provided, consisting of a set of instances that have been properly labeled by hand with the correct output [65, 28].

There are pairs of examples in the given training dataset in the supervised classification method. Mathematically the pairs of examples are expressed as $X = \{(a_1, b_1), (a_2, b_2), , ..., (a_N, b_N)\}$. Here, $a_1, a_2, ..., a_N$ are the observations and $b_1, b_2, ..., b_N$ are the class labels of the observations. If the problem is filtering spam, then a_i is some representation of an email and b_i is either spam or non-spam. The observations can be any vector, whose elements are selected from a set of features. Usually we have real valued observations and it is easy to assume $a \epsilon A$. Moreover, any type of representation can be chosen for the class labels. And, they are usually represented as real numbers that is $b \epsilon Y$. The main target of the supervised classification is to find the transformation between the feature space A and the class label space B, that is $f : A \rightarrow B$.

If the class information contains a limited number of elements, that is $b \epsilon \{1, 2, ..., L\}$ then the problem is considered as a classification task. The classes are divided into two groups, such as the target and non-target classes for the case of a binary classification problem. For simplicity these classes can be represented as $B = \{-1, +1\}$ where the negativity represents the non-target case. The classification algorithms are subjects to the type of label output, on whether learning is supervised or unsupervised, and on whether the algorithm is statistical or non statistical in nature. Moreover, the statistical algorithms can be classified as generative or discriminative. The examples of the supervised classification algorithms are lin-

ear discriminant analysis (LDA), support vector machine (SVM), Decision trees, Naive Bayes classifer, Logistic regression, k-nearest neighbor (kNN) algorithms, kernel estimation, neural networks (NN), linear regression, Gaussian process regression, Kalman filters etc. The dataset in a supervised classification procedure is divided into two sets of training and testing. A classifier is constructed using the training set and the performance of the classifier is evaluated using the testing set. This evaluation is sometimes repeated for different parameters of the classifier constructed. After the optimization of the classifier parameters, the classifier is ready to assign class labels to the features with unknown class labels. The main target of the learning procedure is to maximize this test accuracy on the testing set. In this research, the training set is used to train the classifier and the testing set is used to compute the performance of the classifier.

6.2.2 Unsupervised Classification

The data is grouped into classes based on some measures of inherent ability in the unsupervised classification method. In the unsupervised method training data has not been hand-labeled, and attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances [65, 28]. In unsupervised approach, any information about the class labels of the measurements is not available even for a small set of data. Examples of unsupervised classification algorithms are K-means clustering, Hierarchical clustering, Principal Component Analysis (PCA), Kernel Principal Component Analysis (Kernel PCA), Hidden Markov Model (HMM), Independent Component Analysis (ICA), Categorical mixture model, etc. Recently, the semi-supervised method is introduced in which a combination of labeled and unlabeled data is used. Usually a small set of labeled data combined with a large amount of unlabeled data is used. Actually it is a combination of supervised and unsupervised methods.

6.3 Classification Methods

The widely used classification algorithms in BCI research are divided into five different groups: linear classifiers, neural networks, nonlinear bayesian classifiers, nearest neighbor classifiers and combinations of classifiers. Brief description of some well-known classifiers is given below.

6.3.1 Linear Classifiers

Linear functions to differentiate classes are used in linear classifier. Linear classifiers are undoubtedly the most popular algorithms for BCI research. Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM) are the two main classifiers that have been used for BCI design.

6.3.1.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA), also known as Fishers linear discriminant analysis is a technique used to find a linear combination of features that separates two or more classes of data. It is typically used as a dimensionality reduction step before classification [115]. It reduces dimensionality but at the same time preserves as much of the class discriminatory information as possible. The goal of the LDA is to use a separating hyperplane that maximally separate the data representing the different classes. The hyperplane is found by selecting the projection, where the same classes are projected very close to each other and the distance between two classes means is as maximum as possible [96].

Let us assume that we have K classes, each containing N observations x_i. The within-class scatter, $\tilde{S}_w$ for all K classes can be calculated as:

$$\tilde{S}_w = \sum_{k=1}^{K} f_k S_w^k \tag{6.1}$$

where the within-class covariance matrix S_w^k and the fraction of data f_k are

calculated according to the following formulas:

$$S_w^k = \sum_{i=1}^{N_k} (x_i^k - \mu^k)(x_i^k - \mu^k)^T \tag{6.2}$$

$$f_k = \frac{N_k}{\sum_{j=1}^{K} N_j} \tag{6.3}$$

where N_k is the number of observations of k^{th} class and μ^k indicates mean of the all observations x_i for k^{th} class. The between class scatter $\tilde{S}_b$ for all K classes is calculated as:

$$\tilde{S}_b = \sum_{k=1}^{K} f_k S_b^k \tag{6.4}$$

where the between class covariance matrix, S_b^k can be estimated as

$$S_b^k = \sum_{k=1}^{K} (\mu^k - \mu)(\mu^k - \mu)^T \tag{6.5}$$

where μ indicates the mean of the all observations x_i for all classes. The main objective of LDA is to find a projection matrix that maximizes the ratio of the determinant of $\tilde{S}_b$ to the determinant of $\tilde{S}_w$. The projections that providing the best class separation are eigenvectors with the highest eigenvalues of matrix M:

$$M = \frac{\tilde{S}_b}{\tilde{S}_w} \tag{6.6}$$

Since the matrix M is asymmetric, the calculation of eigenvectors can be difficult. This difficulty can be minimized by using generalized eigenvalue problem [126]. Now, the aim of the LDA is to seek $(K-1)$ projections $[y_1, y_2, y_3, ..., y_{K-1}]$ by means of $(K-1)$ projection vectors. The transformed data set y is obtained as a linear combination of all input features x with weights W.

$$y = x^T W \tag{6.7}$$

where $W = [w_1, w_2, w_3, ..., w_H]$ is a matrix form with the H eigenvectors of matrix M associated with the highest eigenvalues. The LDA reduces the original feature space dimension to H. The LDA performs well when the discriminatory information of data depends on the mean of the data. But it does not work for the variance depended discriminatory informative data.

6.3.1.2 Support Vector Machine

Support Vector Machine (SVM) was introduced by Boser, Guyon, and Vapnik in COLT-92[24]. SVMs are a set of related supervised learning methods used for classification and regression. SVM is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. The SVM used a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. Now the SVM is an active part of the machine learning research around the world. SVM becomes famous as it gives accuracy comparable to sophisticated neural networks. It is used for numerous applications particularly in the field of pattern classification and regression based applications and increased reputation due to many promising features such as better empirical performance. SVMs were developed to solve the classification problem, but recently they have been extended to solve regression problems [235].

In SVM, there are many linear hyper planes that can separate the data. However, only one of these achieves maximum separation. The selected hyperplane might end up closer to one set of datasets compared to others and we do not want this happen and thus we see that the concept of maximum margin classifier or hyper plane as an apparent solution. The Maximum margin can be calculated according to following expression (6.8) [33, 50].

$$margin = \arg\min_{x \epsilon D} d(x) = \arg\min_{x \epsilon D} \frac{|\mathbf{x} \cdot \mathbf{w} + b|}{\sqrt{\sum_{i=1}^{d} w_i^2}} \tag{6.8}$$

The illustration as shown in Figure 6.1, is the maximum linear classifier with the maximum range. In this situation it is an example of a simple linear SVM classifier. The hyper plane with the maximum margin provides better separation performance. Another reason to search maximum margin is that even if we have ve made a small error in the location of the boundary this gives us least chance of causing a misclassification. The other advantage would be avoiding local minima

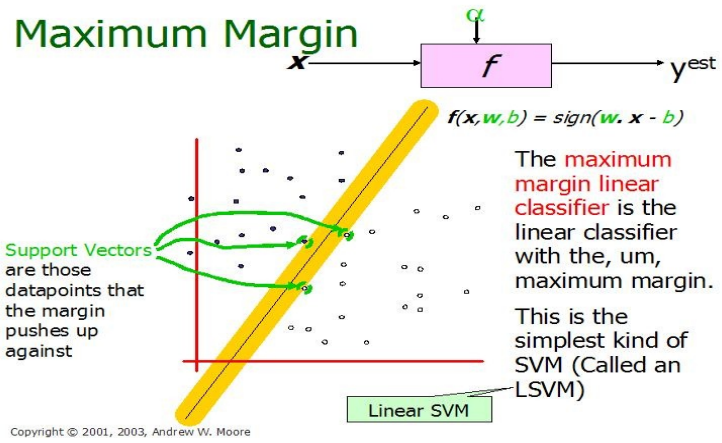


Figure 6.1: Linear SVM classifier. [167]

and better classification. The main target of SVM is separating the data with hyper plane and extends this to non-linear boundaries using kernel trick [50, 163]. To correctly classify all the data using SVM we have the following mathematical calculations,

$$\text{i) If } Y_i = +1; wx_i + b \geq 1$$

$$\text{ii) If } Y_i = -1; wx_i + b \leq 1$$

$$\text{iii) For all } i; y_i(wx_i + b) \geq 1$$

Where x is a vector point and w is weight vector. To separate the data, i) should always be greater than zero. Among all possible hyper planes, SVM selects the one where the distance of hyper plane is as large as possible if the training data is good and every test vector is located in radius r from training vector, the chosen hyper plane is located at the farthest possible from the data and the desired hyper plane which maximizes the margin also bisects the lines between closest points on convex hull of the two datasets.

Although Neural Networks are easier to use, SVM is a useful technique for data classification and sometimes unsatisfactory results are obtained. A classification task usually involves with training and testing data which consist of some data instances [64]. In the SVM process the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes [50].

In supervised learning like SVM, known labels are indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. Identification is a step in SVM in which the know classes are closely connected. It is called feature selection or feature extraction. Feature selection and SVM classification together have a use even when prediction of unknown samples is not necessary. They can be used to identify key sets which are involved in whatever processes distinguish the classes [50].

6.3.2 Neural Networks

The mostly used classifiers in BCI research is the linear classifiers and Neural Networks are one of them [104, 4]. Neural Networks is an assembly of several artificial neurons which enables to produce nonlinear decision boundaries [16]. MultiLayer Perceptron (MLP) is the most widely used neural network and brief description of it is as follows.

An MultiLayer Perceptron (MLP) is composed of several layers of neurons: an input layer, possibly one or several hidden layers, and an output layer [16]. In MLP neuron's input is connected with the output of the previous layer's neurons whereas the neurons of the output layer determine the class of the input feature vector. Also, in MLP when composed of enough neurons and layers, they can approximate any continuous function.

Neural network can classify any number of classes, that is why it is very flexible classifiers and can adapt to a great variety of problems. Therefore, MLP, which are commonly used in classification, have been applied to almost all BCI problems such as binary [186] or multiclass [4], synchronous [192] or asynchronous [42] BCI. However, the MLP classifiers are sensitive to overtraining, especially with the noisy and non-stationary data as EEG. Consequently, careful architecture selection and regularization is required [110]. A MultiLayer perceptron without hidden layers is known as a perceptron and wonderfully a perceptron is equivalent to LDA used for BCI applications [46, 243].

The Gaussian classifier [57, 162] based on the neural network is one of the classifier used in the field of BCI. The classifier deserves a explicit attention as it has been specifically created for BCI. Each unit of this classifier is a Gaussian discriminant function representing a class prototype. It outperforms MLP on BCI data and can perform efficient rejection of uncertain samples [57]. For this advantage, this classifier has been applied with success to motor imagery [221] and mental task classification [57], particularly during asynchronous experiments [57, 43]. In addition to the Gaussian classifier, the following neural networks have been useful to BCI applications, in a more marginal way.

Learning Vector Quantization (LVQ) Neural Network [125, 195];

Adaptive Resonance Theory MAP (ARTMAP) Neural Network [36, 187];

Dynamic Neural Networks such as the Finite Impulse Response Neural Network (FIRNN) [97], Time-Delay Neural Network (TDNN) or Gamma Dynamic Neural Network (GDNN) [9];

Radial Basis Function (RBF) Neural Network [107];

Bayesian Logistic Regression Neural Network (BLRNN) [193];

Adaptive Logic Network (ALN) [128];

Probability estimating Guarded Neural Classifier (PeGNC) [74].

6.3.3 Nonlinear Bayesian Classifiers

The most commonly used Bayesian classifiers for BCI are Bayes quadratic and Hidden Markov Model (HMM). Besides, Bayesian Graphical Network (BGN) has been used for BCI but it is not common and not promising to implement real-time BCI due to slowness [228, 208]. These types of classifiers yield nonlinear decision boundaries. Additionally, the classifiers are generative, which enables them to perform more efficient rejection of uncertain samples than discriminative classifiers. However, although these classifiers are employed in BCI applications they are not as widespread as linear classifiers or Neural Networks.

6.3.3.1 Bayes Quadratic

The aims of Bayesian classifiers are to assign a feature vector the class it belongs to with the highest probability [65, 80]. The so-called a posteriori probability calculated by the Bayes rule is that a feature vector belongs to a given class [80]. The class of this feature vector can be measured using the MAP (Maximum A Posteriori) rule and these probabilities. Bayes quadratic consists in assuming a different normal distribution of data. This leads to quadratic decision boundaries, which explains the name of the classifier. Although this types of classifier is not broadly employed for BCI research, it has been applied with success to motor imagery [138, 221] and mental task classification [117, 10].

6.3.3.2 Hidden Markov Model

Hidden Markov Models (HMM) are widely used dynamic classifiers particularly popular in the field of speech recognition [203]. An HMM is a kind of probabilistic automaton that can provide the probability of observing a given sequence of feature vectors [203]. Each state of the automaton can modelize the probability of observing a given feature vector. These probabilities are usually Gaussian Mixture Models (GMM) for the BCI. For the classification of time series HMM are perfectly appropriate algorithms [203]. Since EEG components used to drive BCI have specific time courses, HMM have been applied to the classification of temporal sequences of BCI features [182, 43, 183] and even to the classification of raw EEG [222]. Although HMM are not greatly employed in the BCI applications these studies discovered that they were promising classifiers for BCI systems. The Input-Output HMM (IOHMM) is another kind of HMM which has been used to design BCI [42]. IOHMM is not a generative classifier but a discriminative one. The key benefit of this classifier is that one IOHMM can discriminate several classes, whereas one HMM per class is needed to achieve the same operation.

6.3.4 Nearest Neighbor Classifiers

The nearest neighbor classifiers are relatively simple and discriminative nonlinear. They assign a feature vector to a class according to its nearest neighbor(s). This neighbor can be a feature vector from the training set as in the case of k Nearest Neighbors (kNN), or a class prototype as in Mahalanobis distance.

6.3.4.1 k Nearest Neighbor

The main target of the k Nearest Nneighbor (kNN) technique is to assign to an unseen point the dominant class among its k nearest neighbors within the training set [65]. These nearest neighbors are typically achieved using a metric distance for BCI. kNN algorithm can approximate any function which enables it to produce nonlinear decision boundaries with a sufficiently high value of k and enough training samples. The algorithms are not very popular in the BCI community, probably because they are known to be very sensitive to the curse-of-dimensionality [77], which made them fail in several BCI experiments [18, 169, 215]. In BCI system, the kNN algorithm is very efficient with low-dimensional feature vectors [23].

6.3.4.2 Mahalanobis Distance

The classifier based on Mahalanobis distance adopt a Gaussian distribution $N(\mu_c; Mc)$ for each prototype of the class c. At that time, a feature vector x is assigned to the class that corresponds to the nearest prototype, according to the so-called Mahalanobis distance $d_c(x)$ [43]:

$$d_c(x) = \sqrt{(x - \mu_c)M_c^{-1}(x - \mu_c)^T} \tag{6.9}$$

These type of classifiers are very simple but robust, which even showed better performance for multiclass [215] or asynchronous BCI systems [43]. Although it performs better, it is still hardly used in the BCI literature.

Chapter 7

Experimental Results and Discussion

7.1 Coherence Based BCI

The performance of the proposed SST based time-frequency coherence as in section 2.6.1.7 is evaluated with both synthetic signals and real EEG data. The results are compared to the STFT based time-frequency coherence. Hamming window of length 100 is used in the case of STFT. Spectral coefficients are then smoothed for the estimation of TF coherence using non-identical smoothing windows. Gaussian smoothing windows of length w [2 1] and w [10 1] are used for smoothing the cross and auto spectral densities respectively. A bump mother wavelet is used and the discretization of the scales of CWT is set to 32 to implement the SST. In the SST, the length of the Gaussian smoothing windows are w [3 1] and w [50 1], i.e., cross spectral density is smoothed over a TF area of 3Hz by 1s and the auto spectral densities over a TF area of 50Hz by 1s.

Synthetic Data: Three non-stationary synthetic signals X, Y and Z are generated by summing up three sinusoids of frequencies 5Hz, 6Hz and 10Hz with sampling frequency 100Hz. Each of the synthetic signals is composed of these three sinusoids with different time alignment as illustrated in Figure 7.1. Finally, the individual synthetic signals X, Y and Z are contaminated with Gaussian noise of levels 5dB, 0dB and -5dB respectively. The time-frequency coherences of each pair of synthetic signals using STFT and SST are shown in Figure 7.2 and Figure

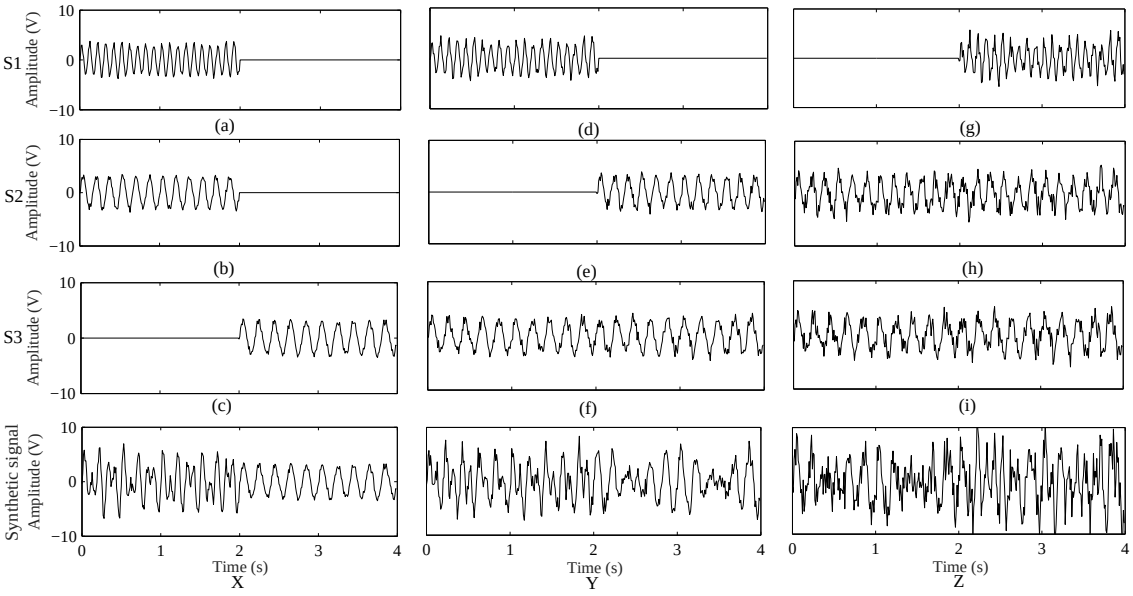


Figure 7.1: Creation of three-channel non-stationary [X,Y,Z] signal. The first three rows (S1, S2 and S3) contain three sinusoids of different frequencies. All of the sinusoids in S1, S2 and S3 comprise 10 Hz, 6 Hz and 5 Hz frequency components respectively. There are different time alignment of the sinusoids to generate the synthetic signals X, Y and Z. 5dB, 0dB and -5dB noises are added to sinusoids (a) to (c), (d) to (f) and (g) to (i) respectively. The fourth row (synthetic signal) is the sum of the three sinusoids; X=(a)+(b)+(c), Y=(d)+(e)+(f) and Z=(g)+(h)+(i).

7.4 respectively. In Figure 7.2, the coherence between signals Y and Z (5Hz and 6Hz frequency) is overlapped each other, whereas, in Figure 7.4, the coherence between the same pair of signals is well separated. The phenomenon has clearly illustrated in Figure 7.3 which represents the marginal frequency coherences of two methods (STFT and SST). The marginal frequency coherence is defined as, $\tilde{C}_{x,y}(f) = \sum_{t=1}^{T} |C_{x,y}(t,f)|^2$, for $f = 1, 2, , F$. With STFT, the coherence values of closer frequencies are overlapped and that with SST sharply represents the coherence of individual frequency components. It is perceived that the STFT based time-frequency coherence displays poor resolution than SST based technique.

Real Data: The real EEG data is collected from the publicly available Brain Computer Interface (BCI) Competition IV dataset *calib_ds*1*a*. The data is used to calculate the performance of the proposed method. It is recorded from healthy subjects. In the whole session, motor imagery is accomplished without feedback.

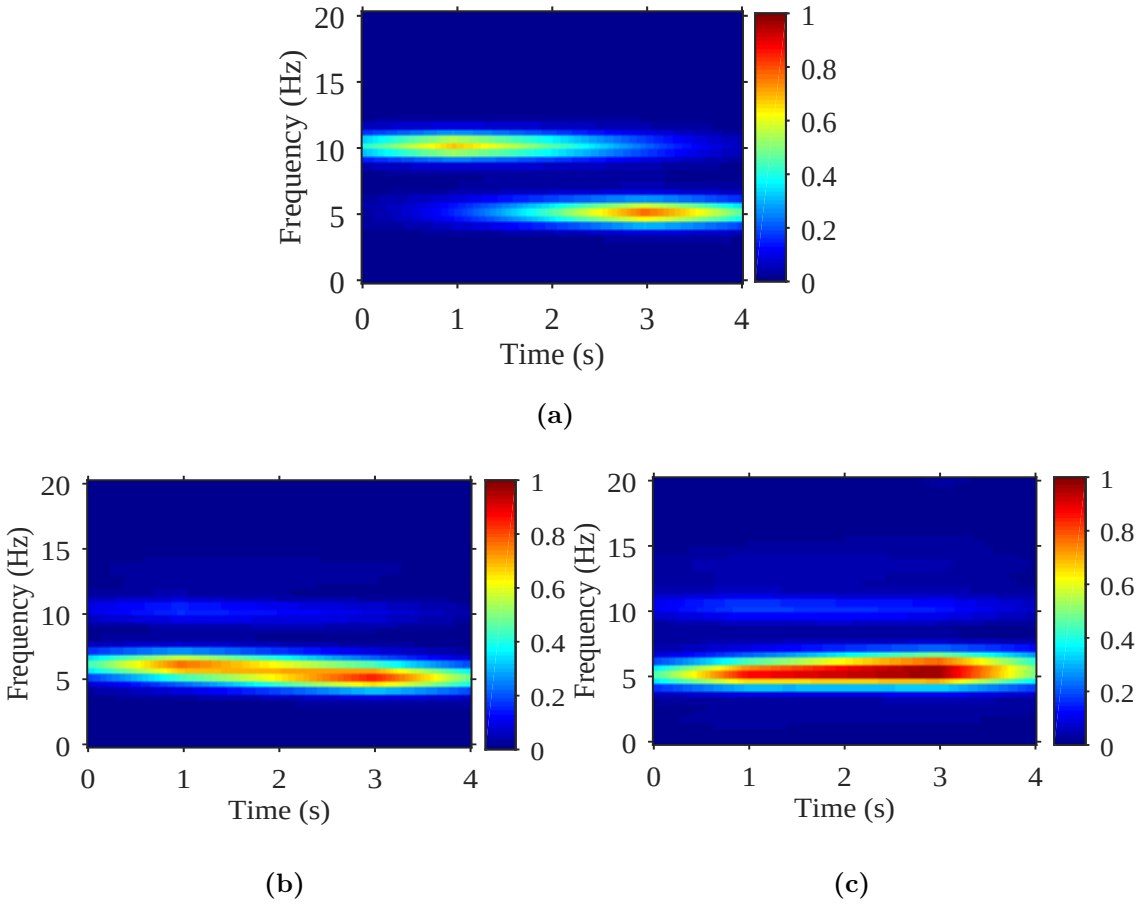


Figure 7.2: STFT based TF coherence between the synthetic signal (a) X and Y, (b) X and Z, and (c) Y and Z.

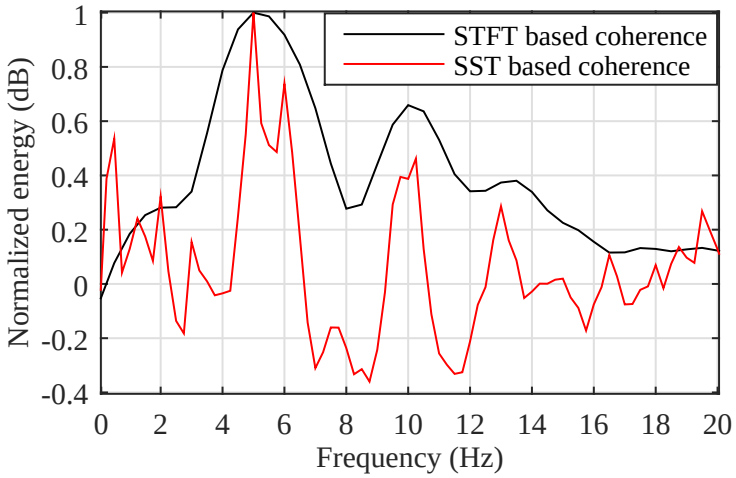


Figure 7.3: Marginal frequency coherences of STFT (black line) and SST (red line) based coherence between the synthetic signal Y and Z.

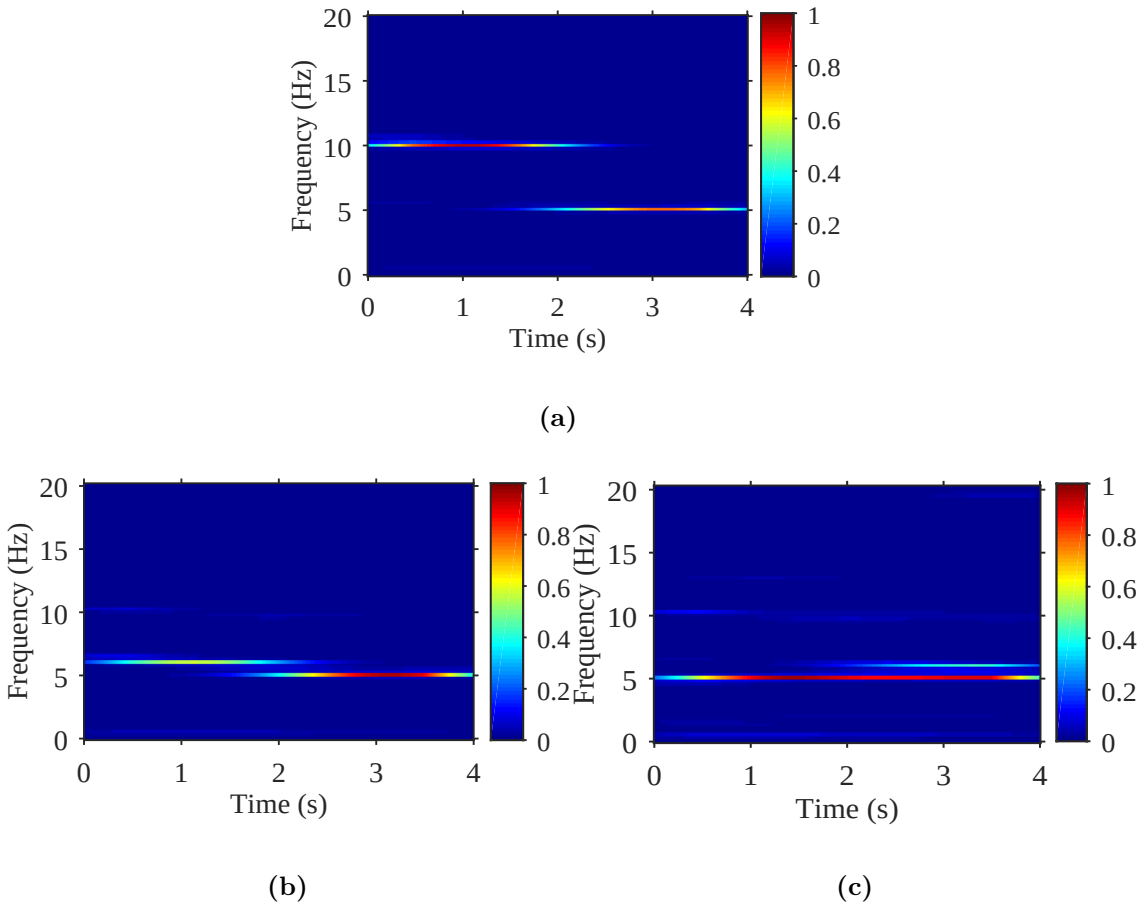


Figure 7.4: SST based TF coherence between the synthetic signal (a) X and Y, (b) X and Z, and (c) Y and Z.

For each subject, two classes of motor imagery which are selected from left hand, right hand, and foot movement. The calibration dataset *calib_ds*1*a* are continuous signals of left hand and foot movement. The data contains 59 EEG channels, total 200 trials with four second each. The sampling rate of the data is 100 Hz. As a pre-processing, the data offset has been removed from the EEG signals. Then the signal is passed through a 4th order Butterworth band pass filter of the range between 8Hz and 12Hz to obtain alpha frequency band as the band contains complex patterns of intermittent synchronization [160]. Two channels T7 and T8 are chosen to measure the inter-channel coherence in this experiment. The raw EEG signal (first row), the filtered alpha component (second row) and the spectrum of alpha(third row) of channels T7 and T8 of left hand movement are shown in Figure 7.5. Similarly, the foot movement data of channel T7 and T8 are illustrated

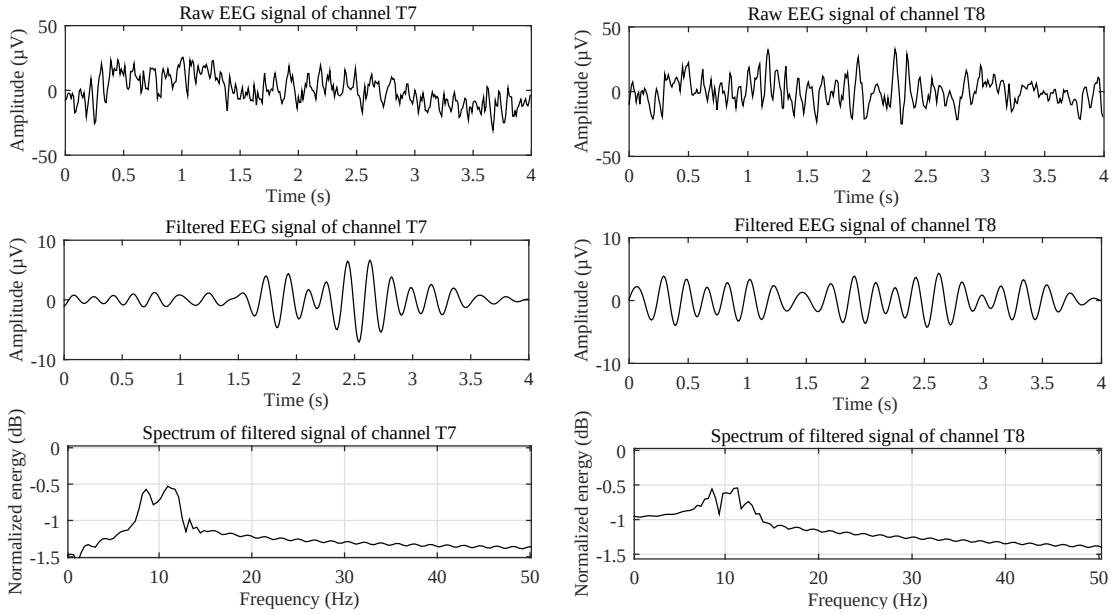


Figure 7.5: Left hand movement data: first row is the raw EEG signals, second and third row are the filtered EEG signals and spectrums of the filtered component respectively.

in Figure 7.6.

The STFT based time-frequency coherences between channel T7 and T8 for left hand and foot movement motor imagery are shown in Figure 7.7(a) and Figure 7.7(b) respectively. The time-frequency coherences based on SST between channel T7 and T8 of left hand and foot movement are represented in Figure 7.7(c) and Figure 7.7(d) respectively. Unlike STFT, the SST based TF coherence illustrates sharp localization of very narrow band frequency components as shown in Figure 7.7.

BCI Interpretation: The time-frequency coherence between channels of left and right hemisphere of human brain is studied in this experiment. Motor imagery classification between left hand and foot movement is also observed. Motor imagery is useful to control sensorimotor rhythms [158] and the rhythms are more activated in central part of brain [93]. Hence, out of the 59 EEG channels, three channels from left hemisphere ("T7", "FC5" and "CP5") and three channels from right hemisphere ("T8", "FC6" and "CP6") are selected for coherence analysis. The spatial distribution of the channels on the scalp in 10/20 EEG system is

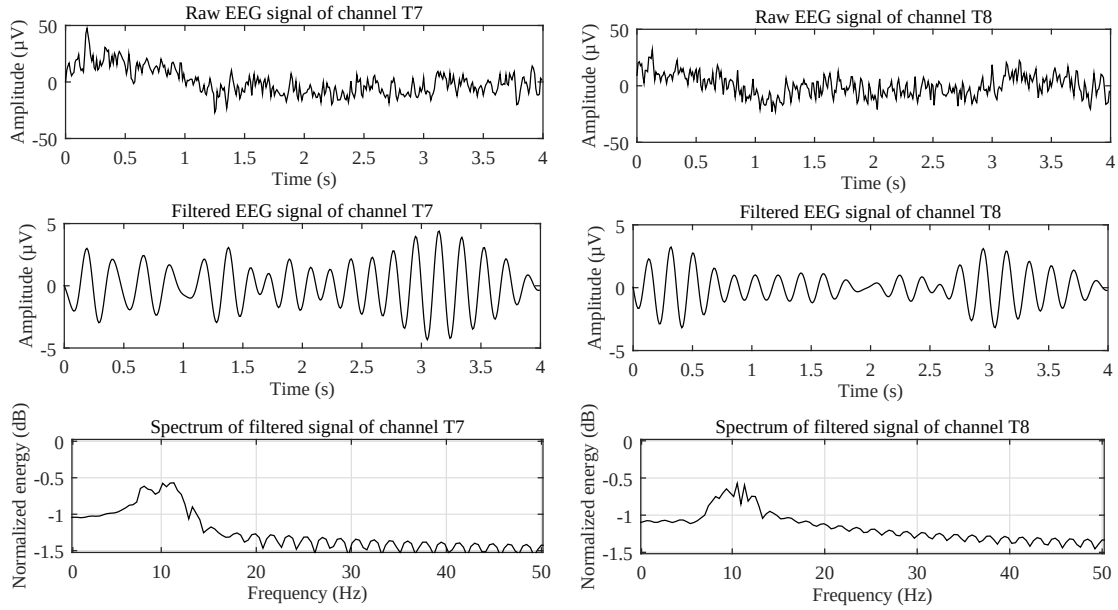


Figure 7.6: Foot movement data: first row is the raw EEG signals, second and third row are the filtered EEG signals and spectrums of the filtered component respectively.

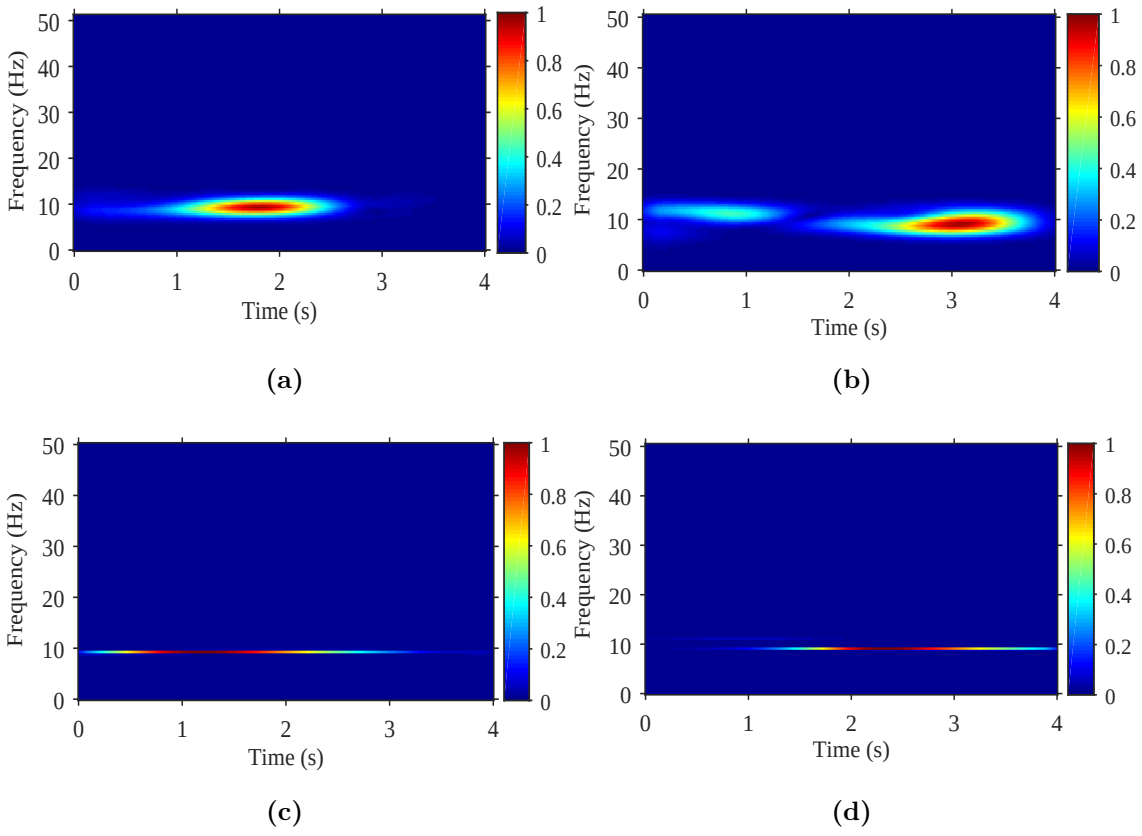
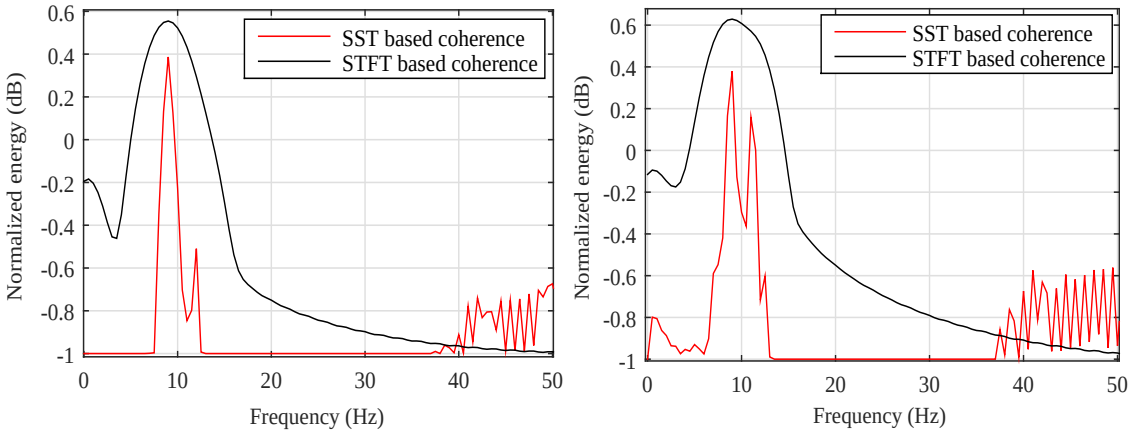


Figure 7.7: TF coherence between channels T7 and T8 based on (a) STFT and (b) SST of left hand movement data, (c) STFT and (d) SST of foot movement data.



(a) (b)

Figure 7.8: Marginal frequency coherences of STFT and SST based coherence between channels T7and T8 (a) left hand movement (b) foot movement.

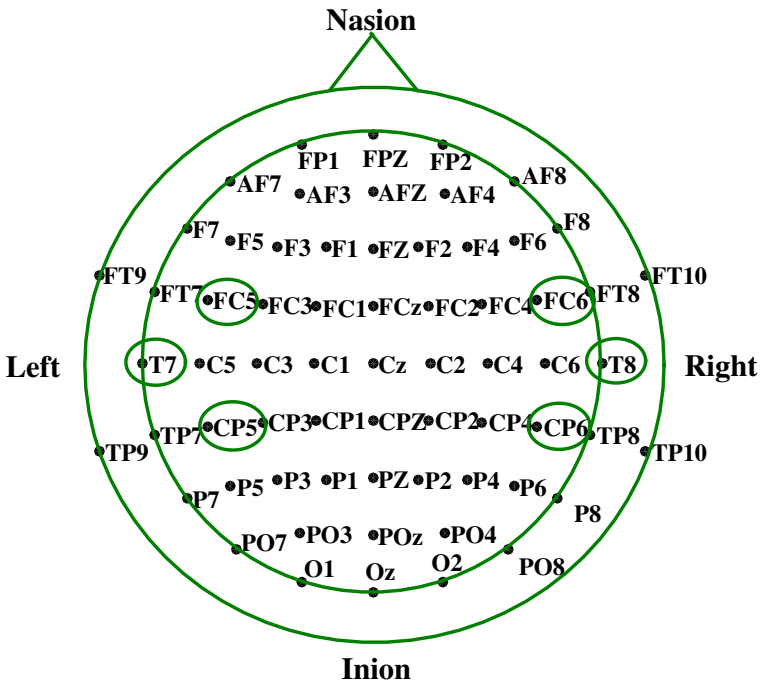


Figure 7.9: The electrodes map of 10/20 EEG system standardized by the American EEG society. The circled electrodes T7, FC5 and CP5 from left hemisphere and T8, FC6 and CP6 from right hemisphere are selected for the dataset used in this experiment.

illustrated in Figure 7.9. The channel pairs FC5FC6, FC5T8, FC5CP6, T7FC6, T7T8, T7CP6, CP5FC6, CP5T8 and CP5CP6 are used to measure time-frequency coherence. Both STFT and SST based time-frequency coherences are measured for all the selected channel pairs. The time-frequency coherences are weighted according to equation (7.1).

$$|C_{x,y}(t,f)|^2_{weighted} = |C_{x,y}(t,f)|^2 \bullet \tilde{C}_{x,y}(f) \tag{7.1}$$

where the symbol $\bullet$ represents binary singleton multiplication operator and the marginal frequency coherence, is used as weight vector. Marginal time coherence is calculated from the weighted time-frequency coherence. The marginal time coherence is defined as

$$\tilde{C}_{x,y}(t) = \arg_f max(|C_{x,y}(t,f)|^2_{weighted}); t = 1,2,3,...,T. \tag{7.2}$$

In this experiment, marginal time coherence is calculated by averaging the marginal time coherence over 100 trials. Figure 7.10 shows the normalized values over time for different channel pairs of left hand and foot movement data. The solid lines represent left hand and the dashes lines represent the coherence for foot movement data. The left panel of Figure 7.10 shows SST based whereas the right panel shows STFT based marginal time coherences. The SST based marginal time coherence can discriminate the left hand and foot movement motor imagery data.

To explore the performance of SST in time-frequency representation, the experiments are performed using EEG signals. The TFR by using STFT and SST of motor imagery of left hand and foot movement data are shown in Figure 7.7. The energy corresponding to the marginal frequency coherence is illustrated in Figure 7.8 for both left hand and foot movement data.

The marginal frequency coherence based on STFT represents poor localization of frequency components, whereas, SST based method illustrates sharp localization of each component within very narrow band of frequencies. In Figure 7.8, the frequencies of 9Hz and 11Hz are well separated with SST based marginal frequency coherence but it is unable to separate those frequencies in STFT based

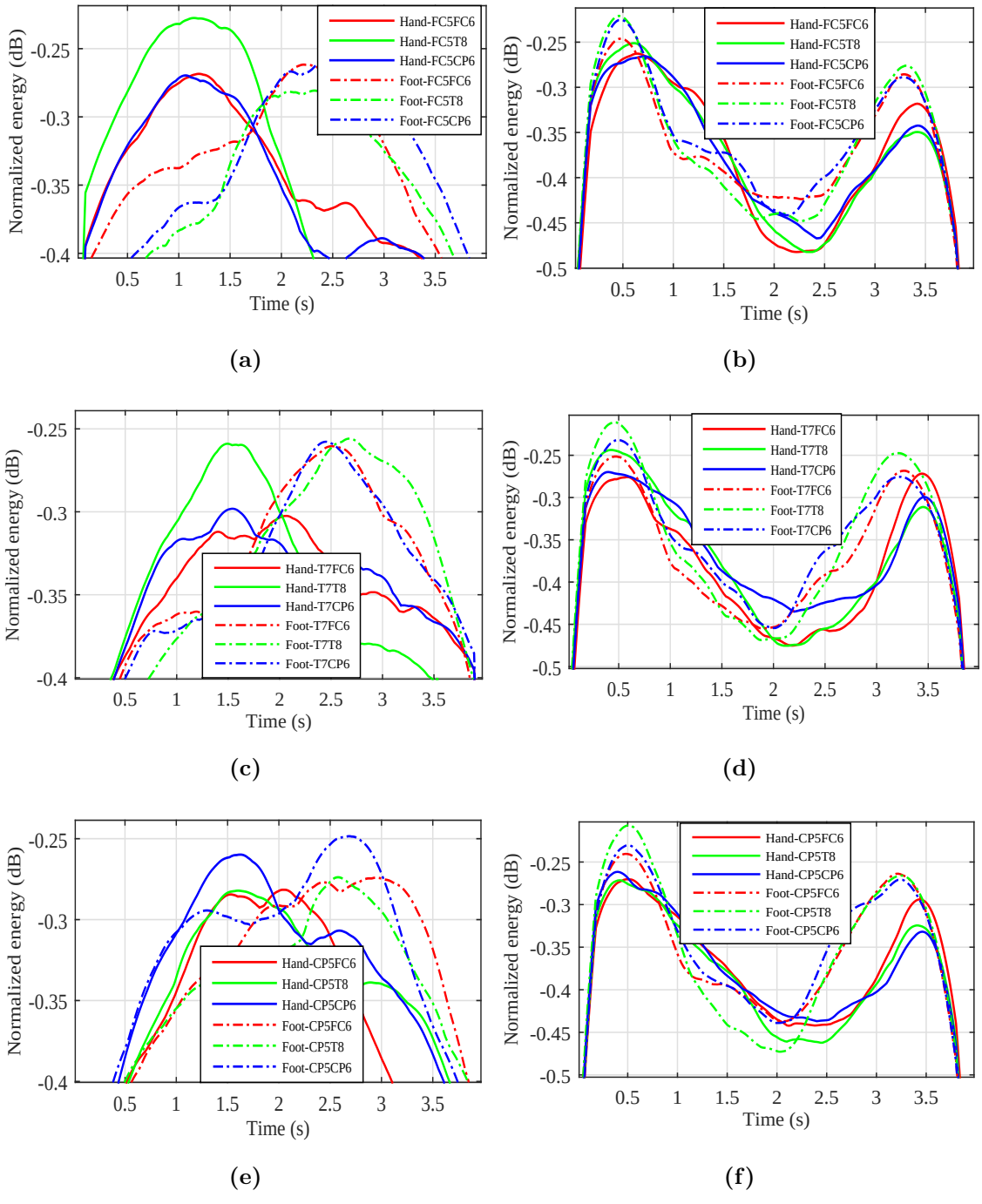


Figure 7.10: Marginal time coherence between different channel pairs of left hand and foot movement data for SST based (left panel) and STFT based (right panel) methods.

approach. Hence, the SST based time-frequency coherence is superior to that of using STFT. The underlying reason is that, the energy in STFT spreads over a wide range of frequency due to the use of window function with overlapping which introduces cross-spectral energy. The SST based marginal time coherence followed by marginal frequency coherence can be used in BCI. In the left panel of Figure 7.10, the coherence value is maximum in time interval 1-2s for left hand movement data, whereas, the maximum coherence value is found in time interval 2-3s for foot movement data with all channel pairs. Hence, the explicit discrimination between left hand and foot movement data is found in the SST based marginal time coherence. On the other hand, no such type of discrimination is found in the STFT based marginal time coherence (in the right panel). The key reason is that the SST has a flexible time-frequency window whereas the STFT has a fixed time-frequency window, making it inaccurate to analyse signals having wide bandwidths that change rapidly with time. In addition, the STFT requires stationarity of the signal during a finite time interval but EEG signal shows non-stationary properties.

7.2 Spatial Filters and Features Selection

This paper presents a novel method for the selection of spatial filters and features in electroencephalography (EEG) based motor imagery classification. The analyzing of EEG data is divided into training and test sets. The training set is used to select appropriate spatial filters with dominant features. To accomplish such features, the EEG of training set is segmented again into two subsets termed as training subset and test subset. The features of both subsets are extracted using common spatial pattern (CSP). Then features of training subset are ranked using mutual information based approach. Besides, the features of test subset are also ranked according to the order of the training subset features. The initial classification performance using training and test subsets are obtained by using linear discriminant analysis (LDA). Then a grid search method is employed to select the

effective number of spatial filter pairs as well as the discriminative features. Thus obtained spatial filter and features are used in actual classification accuracy of the test set of EEG. The experimental results show that the proposed approach yields comparatively superior classification performance compared to prevailing methods. The performance of the proposed method is evaluated by conducting classification experiments on MI movement data. We applied the proposed method on two undermentioned datasets, where the datasets are widely used publicly available dataset of Dataset IIIa and IVa from BCI competition III [204].

Dataset IIIa, BCI Competition III (BCIC III-IIIa): The EEG data was recorded with a 64-channel EEG amplifier from Neuroscan, using the left mastoid for reference and the right mastoid as ground. The EEG was sampled with 250Hz, it was filtered between 1 and 50Hz with Notchfilter on. The subject sat in a relaxing chair with armrests. The task was to perform imagery left hand, right hand, foot or tongue movements according to a cue. The order of cues was random. The experiment consists of several runs (at least 6) with 40 trials each; after the trial begin, the first 2s were quite, at t=2s an acoustic stimulus indicated the beginning of the trial, and a cross + is displayed; then from t=3s an arrow to the left, right, up or down was displayed for 1s; at the same time the subject was asked to imagine a left hand, right hand, tongue or foot movement, respectively, until the cross disappeared at t=7s. Each of the 4 cues was displayed 10 times within each run in a randomized order. In this experiment, only EEG signals corresponding to left and right hand MI is used. EEG signals are recorded using 60 electrodes with a sample rate of 250 Hz. A training set and a test set is available for each subject. Both sets contain 45 trials per class for subject *k3b*, and 30 trials per class for subject *k6b* and *l1b*.

Dataset IVa, BCI Competition III (BCIC III-IVa): The dataset contains data from four initial sessions without feedback and recorded from five healthy subjects (labelled *aa*, *al*, *av*, *aw*, *ay*). Visual cues indicated for 3.5s which of the following three motor imageries the subject performed: left hand (L), right hand (R), right foot (F). In this experiment, only EEG signals corresponding to right

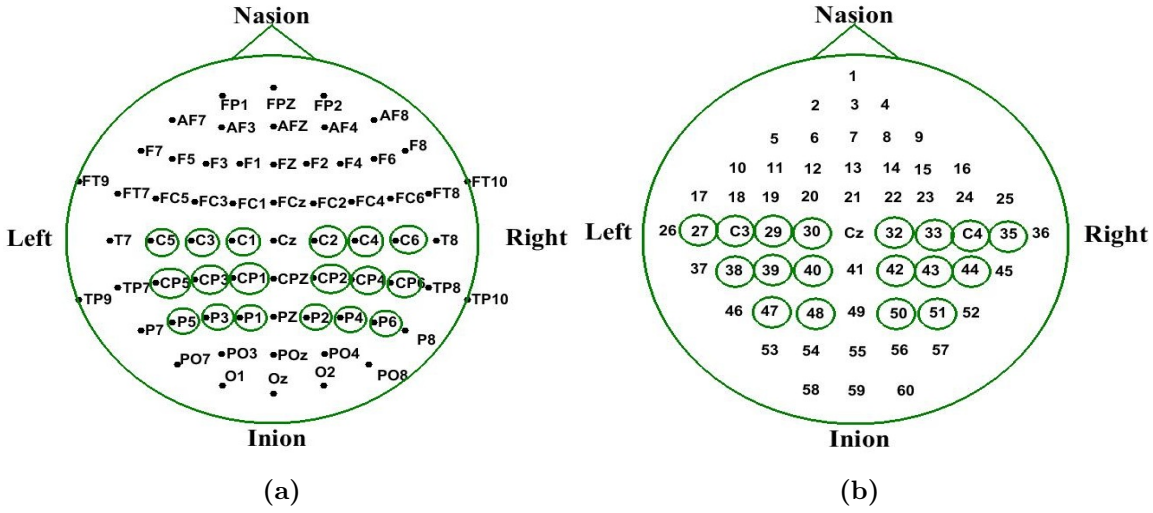
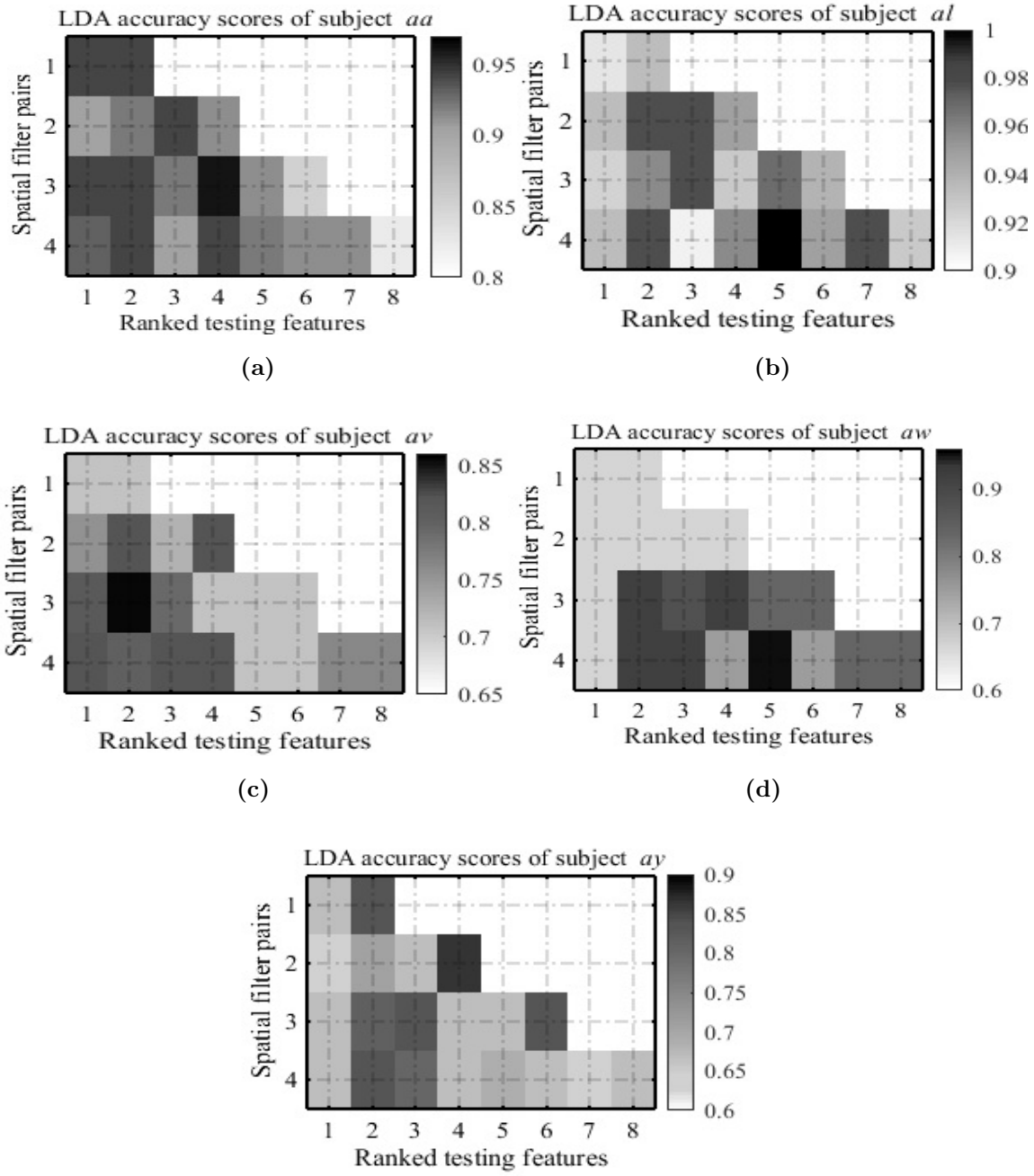


Figure 7.11: Selected electrodes (circled) for (a) BCI competition III dataset IVa and (b) BCI competition III dataset IIIa.

hand and right foot MI are used. The presentation of target cues were intermitted by periods of random length, 1.75 to 2.25s, in which the subject could relax. There were two types of visual stimulation: (1) where targets were indicated by letters appearing behind a fixation cross (which might nevertheless induce little target-correlated eye movements), and (2) where a randomly moving object indicated targets (inducing target-uncorrelated eye movements). A total of 118 electrodes are used for recording EEGs with a sample rate of 100 Hz. A training set and test set with different size for each subject is available. The data for each subject comprises 280 trials, among which 168, 224, 84, 56, and 28 composed the training set for subject *aa*, *al*, *av*, *aw*, and *ay* respectively, the remaining of the trials composing their test set.

Preprocessing: In this experiment, a class is assigned to each trial, i.e., the discrete classification of each class is considered. The time of 0-0.5s and 3.5-4.0s are imagination preparation stage and post imagination stage respectively. During this imagination, features do not contain significant information [223]. For each dataset and trial, the data from time segment located from 0.5s to 2.5s after the visual cue instructing the subject to perform MI is considered. The brain rhythmic



(a) **(b)** **(c)** **(d)** **(e)**

Figure 7.12: GS-CSP based LDA accuracy scores for all the five subjects of dataset BCIC III-IVa.

components alpha contains frequency band 8-12Hz and beta contains frequency band 12-30Hz. A fourth-order Butterworth band pass IIR filter with zero phases is used to filter the alpha and beta rhythmic components (8-30Hz) from each trial.

Channel Selection: The motor imagery response of brain is more active in the motor cortex [175]. In this experiment, out of the 118 EEG channels of BCIC III-IVa and 60 EEG channels of BCIC III-IIIa, from both datasets 18 electrodes around the sensory motor cortex are manually selected for classification as mentioned in [241]. Selected electrodes for the two datasets are shown in Figure 7.11.

Spatial Filter Pair and Features Selection: We employed the proposed state-of-the-art GS-CSP to select the most significant spatial filter pair and discriminative features in different subjects of aforementioned datasets. The training subset comprises 80 percent trials of the EEG training set and the test subset includes remaining trials. To extract features from both training subset and test subset data, the CSP algorithm with CSP filter pairs m (as it varies from 1 to 4 in most of the existing work) is used in this experiment. For every combination of spatial filter pair and number of features (FP, Fs), a grid of accuracy scores is produced. For different subject of BCIC III-IVa and BCIC III-IIIa, the GS-CSP based LDA accuracy scores are shown in Figure 7.12 and in Figure 7.13 respectively. These figures represent the calculated LDA accuracy scores for every value of spatial filter pair with their corresponding discriminant hyperplane coefficients and ranked test subset features. Based on the maximum score we select the best combination which is subsequently used in classification. In this experiment, the combinations (3, 4), (4, 5), (3, 2), (4, 5) and (2, 4) are selected for the subject *aa*, *al*, *av*, *aw* and *ay* respectively as shown in Figure 7.12. In Figure 7.13, for the subject *k3b*, *k6b* and *l1b* the selected combinations are (4, 8), (3, 5) and (3, 4) respectively in the occasion of left hand versus right hand binary classification. The combination with the smallest number of filter pair is selected if more than one combination produces maximum score since the combination required less training time. In this experiment, the appropriate combinations that are selected for

Table 7.1: Selected combination of spatial filter pair and number of features (FP, Fs) for various motor imagery task and subject of BCI competition III dataset IIIa.

Motor Imagery Tasks	Subjects		
	k3b	*k6b*	*l1b*
Left hand-Right hand	(4,8)	(3,5)	(3,4)
Left hand-Foot	(3,6)	(2,4)	(1,2)
Left hand-Tongue	(4,7)	(4,8)	(3,4)
Right hand-Foot	(1,2)	(3,4)	(4,6)
Right hand-Tongue	(4,5)	(4,8)	(2,3)
Foot-Tongue	(3,6)	(2,4)	(3,4)

various motor imagery tasks and subjects are tabulated in Table 7.1. The selected spatial filter pair and features are then used for the classification of test set EEG of respective subjects.

In this research, feature extraction and classification performance of the proposed method are compared to CSP and some modified version of the CSP. Regularized CSP (RCSP) is one of the modified version of the CSP, aiming at computing optimal spatial pattern by familiarizing a regularization term to the CSP formula [213, 149]. The RCSP method attempts to use a priori knowledge in the spatial filter optimization by imposing variant constraints in the CSP's formulation [72]. Some examples of RCSP method are **(i) Composite CSP (CCSP) algorithm:** it works on the basis of useful information transfer from subject to subject by regularizing the covariance matrices and using other subjects data [113], **(ii) Regularized CSP with Generic Learning (GLRCSP) approach:** the aim of this approach is to regularize the covariance matrix estimation, using data from other subjects [151]. It uses regularization parameters to shrink the covariance matrix towards both the identity and a generic covariance matrix, **(iii) Spatially Regularized CSP (SRCSP):** the main target of this approach is to obtain spatially smooth filters for which neighboring electrodes have similar weights [149]. Here, the a priori knowledge that neighboring neurons tend to have as a similar function is used, **(iv) CSP with Tikhonov Regularization (TRCSP):** the

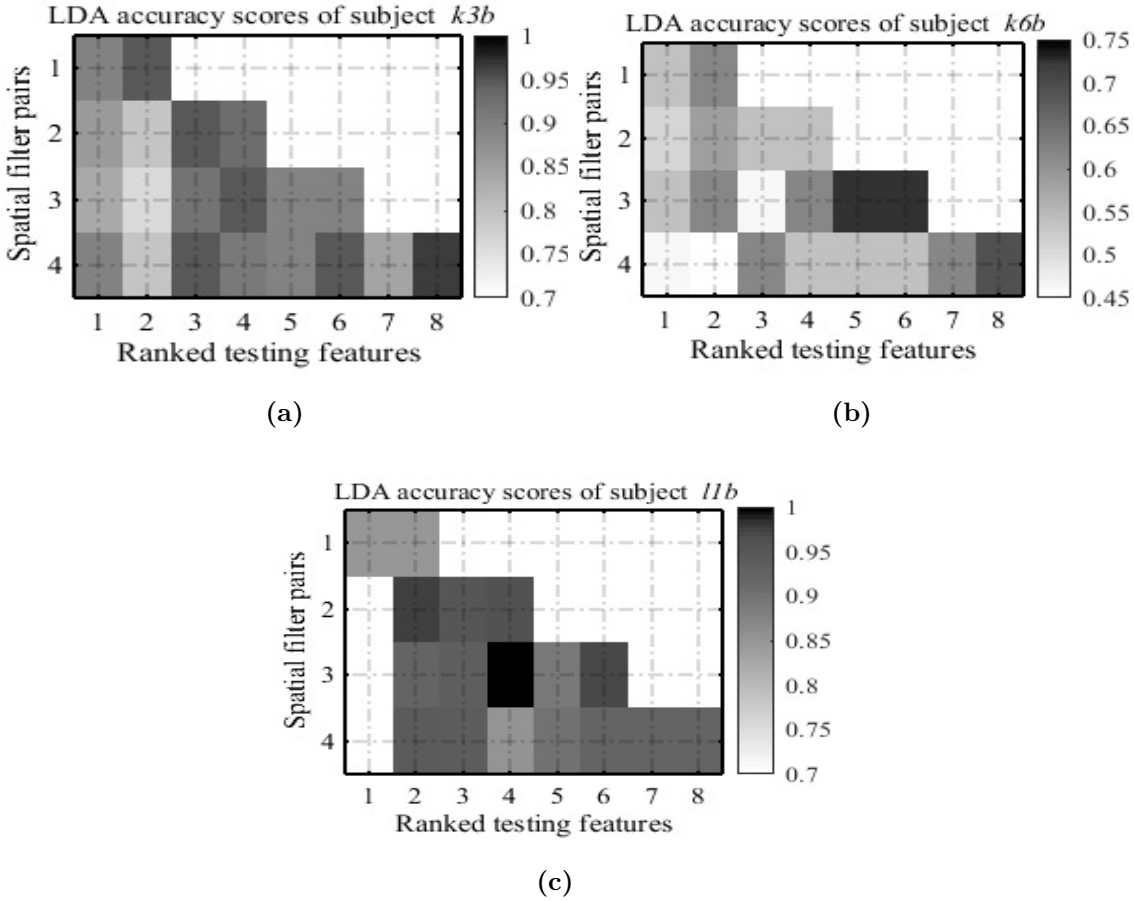


Figure 7.13: GS-CSP based LDA accuracy scores for all the three subjects of dataset BCIC III-IIIa.

TRCSP algorithm [149] functions based on the regularization of the CSP. The regularization involves penalizing results with large weights. Since the CSP formulation considers the variances of two classes, CSP is very sensitive to outlier that sometimes limits the effectiveness of the approach.

Figure 7.14 shows an example about separation of classes by using CSP, CCSP, SRCSP, GLRCSP, TRCSP and GS-CSP. To scatter plot the features, spatial filter pair 3 (m=3) and the highest and the lowest ranked features are used. In the case of GS-CSP the selected (3, 2) combination, FP=3 and Fs=2 is considered. As shown in Figure 7.14, with the proposed approach the extracted features are properly organized compared to other methods.

Classification: For each subject, the classification performance is evaluated, using the selected spatial filter pair and a number of ranked CSP features. The

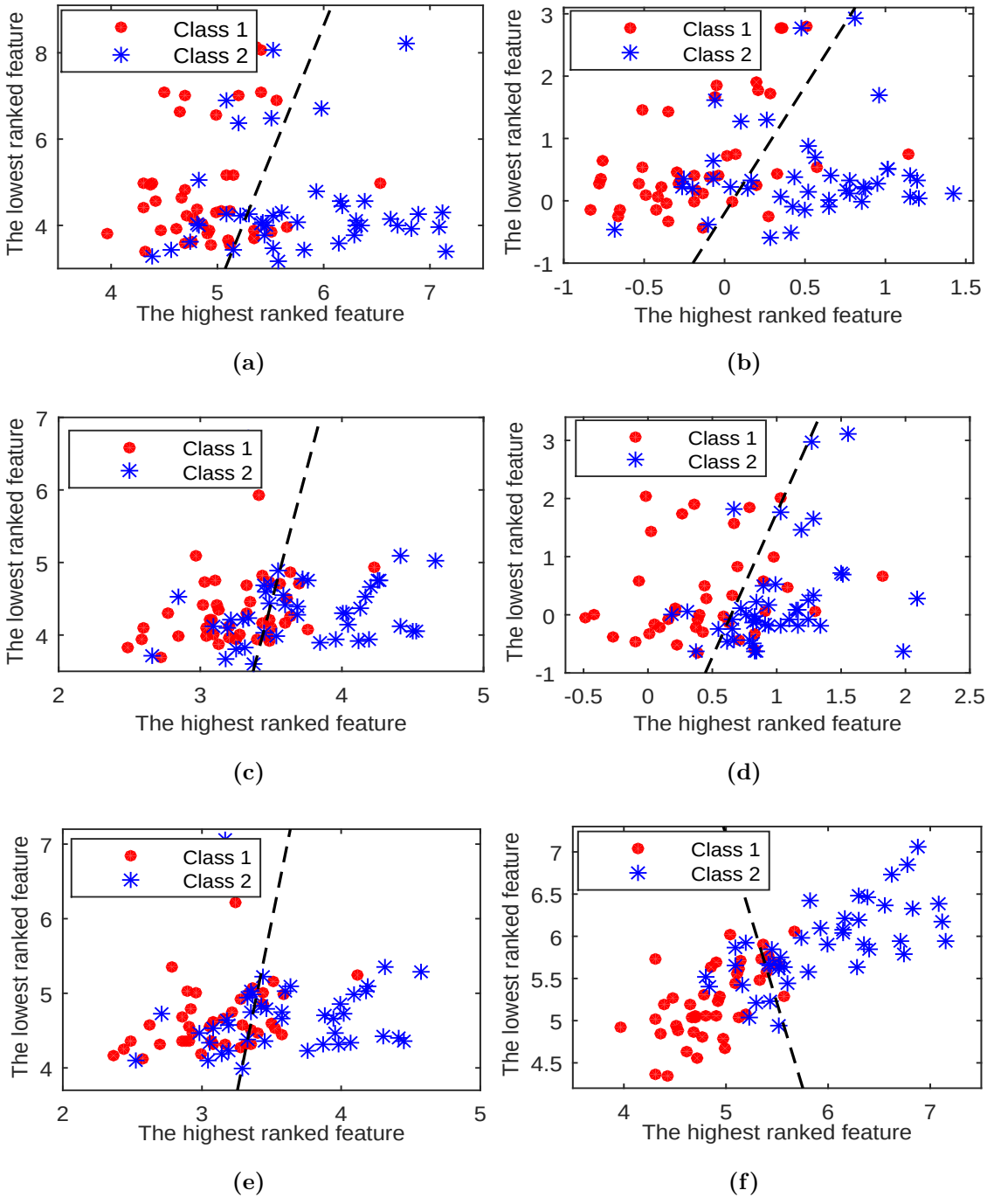


Figure 7.14: Features extracted with different types of spatial filtering method using spatial filter pair m=3 and the GS-CSP using FP=3, Fs=2 from the training set of subject *av* of dataset BCIC III-IVa: (a) CSP, (b) CCSP, (c) SRCSP, (d) GLRCSP, (e) TRCSP and (f) GS-CSP. The class borders are indicated by dashed lines.

Table 7.2: Classification accuracies (mean and standard deviation (SD) in %) of the BCI competition III dataset IVa (Right hand vs Right foot) using the proposed GS-CSP and reported CSP variants. The best result for each subject is displayed in bold characters.

Methods	Spatial filter pairs	BCIC III-IVa					
		Subjects					Mean±SD
		aa	*al*	*av*	*aw*	*ay*	
CSP	1	81.89	93.68	70.75	84.71	66.71	79.55±10.88
	2	82.75	94.29	69.89	87.18	**93.46**	85.51±9.93
	3	85.50	96.18	70.04	87.75	90.46	85.99±9.77
	4	85.95	95.86	70.07	87.43	90.50	85.96±9.66
CCSP	3	81.29	95.71	72.39	87.75	89.17	85.26±8.88
SRCSP	3	87.25	94.32	66.71	87.32	89.60	85.04±10.64
GLRCSP	3	80.43	95.46	69.92	72.82	87.85	81.29±10.55
TRCSP	3	87.00	95.25	65.53	87.75	92.78	85.66±11.77
GS-CSP	1,2,3,4	**87.96**	**97.78**	**72.54**	**87.95**	**93.46**	**87.94±9.55**

selected spatial filter pair is used during features extraction. Subsequently, only the selected number of mutual information based ranked features are considered to test classification accuracy. The binary classification (right-hand versus right-foot) performances of datasets BCIC III-IVa is presented in Table 7.2. The performance of dataset BCIC III-IIIa is presented in Table 7.3 and Table 7.4. Table 7.3 demonstrates the result of left-hand versus right-hand, left-hand versus foot and left-hand versus tongue motor imagery binary classification and the binary classification of right-hand versus foot, right-hand versus tongue, and foot versus tongue are demonstrated in Table 7.4. Compared to the other methods, the maximum mean accuracy of 87.94% is obtained for the classification between right hand versus right foot MI task by the proposed method as shown in Table 7.2. Moreover, the proposed method yields the maximum mean accuracies of 91.85%, 92.93%, 95.71%, 94.01%, 95.07% and 89.89% for the binary classification of left hand versus right hand, left hand versus foot, left hand versus tongue, right hand versus foot, right hand versus tongue and toot versus tongue MI tasks respectively.

Table 7.3: Classification accuracies (mean and standard deviation (SD) in %) of the BCI competition III dataset IIIa (Left hand vs Right hand, Left hand vs Foot and Left hand vs Tongue) using the proposed GS-CSP and reported CSP variants. The best result for each subject is displayed in bold characters.

Methods	Spatial filter pairs	BCIC III-IVa (Left hand-Right hand)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	93.58	73.88	79.21	82.22±10.19
	2	92.97	76.21	94.29	87.82±10.08
	3	93.67	76.24	97.02	88.98±11.16
	4	**95.30**	80.66	96.03	90.66±8.67
CCSP	3	94.80	80.24	92.97	89.34±**7.93**
SRCSP	3	94.11	78.92	97.34	90.12±9.84
GLRCSP	3	94.56	65.16	94.13	84.62±16.85
TRCSP	3	93.86	80.34	95.41	89.87±8.29
GS-CSP	1,2,3,4	**95.30**	**81.69**	**98.55**	**91.85**±8.94

Methods	Spatial filter pairs	BCIC III-IVa (Left hand-Foot)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	94.38	80.39	**97.37**	90.71±9.06
	2	95.84	**84.78**	94.88	91.83±**6.13**
	3	**96.65**	80.16	94.72	90.51±9.02
	4	94.36	80.96	94.42	89.91±7.75
CCSP	3	**96.65**	78.65	94.72	90.01±9.88
SRCSP	3	**96.65**	82.42	93.09	90.72±7.41
GLRCSP	3	94.65	76.98	95.83	89.15±10.56
TRCSP	3	96.65	82.43	92.90	90.66±7.37
GS-CSP	1,2,3,4	**96.65**	**84.78**	**97.37**	**92.93**±7.07

Methods	Spatial filter pairs	BCIC III-IVa (Left hand-Tongue)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	97.79	87.17	95.13	93.36±5.53
	2	98.17	87.43	95.51	93.70±5.59
	3	96.46	88.50	95.25	93.40±4.28
	4	96.83	**90.35**	94.64	93.94±3.29
CCSP	3	96.12	82.43	95.25	91.26±7.66
SRCSP	3	98.46	88.53	95.25	94.08±5.07
GLRCSP	3	98.02	88.37	95.25	93.88±4.97
TRCSP	3	98.46	89.22	95.25	94.31±4.69
GS-CSP	1,2,3,4	**98.90**	**90.35**	**97.88**	**95.71**±4.66

Table 7.4: Classification accuracies (mean and standard deviation (SD) in %) of the BCI competition III dataset IIIa (Right hand vs Foot, Right hand vs Tongue and Foot vs Tongue) using the proposed GS-CSP and reported CSP variants. The best result for each subject is displayed in bold characters.

Methods	Spatial filter pairs	BCIC III-IVa (Right hand-Foot)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	**96.01**	75.89	93.91	88.60±11.06
	2	95.01	80.29	94.06	89.78±8.24
	3	94.01	87.86	93.78	91.88±3.48
	4	94.01	87.29	96.59	92.63±4.80
CCSP	3	94.01	87.86	93.72	91.86±3.47
SRCSP	3	94.01	78.38	94.10	88.83±9.05
GLRCSP	3	96.62	75.10	95.88	89.20±12.22
TRCSP	3	93.12	87.86	94.10	91.69±**3.36**
GS-CSP	1,2,3,4	**96.01**	**89.34**	**96.67**	**94.01**±4.05

Methods	Spatial filter pairs	BCIC III-IVa (Right hand-Tongue)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	96.31	83.43	94.53	91.42±6.98
	2	96.20	86.62	94.97	92.60±5.21
	3	96.20	85.31	94.81	92.10±5.93
	4	96.73	**89.39**	94.70	93.60±3.79
CCSP	3	98.20	89.30	91.72	93.07±4.60
SRCSP	3	97.20	88.66	94.06	93.31±4.32
GLRCSP	3	96.20	84.04	91.88	90.71±6.16
TRCSP	3	96.20	89.20	93.99	93.13±**3.58**
GS-CSP	1,2,3,4	**98.90**	**89.39**	**96.93**	**95.07**±5.01

Methods	Spatial filter pairs	BCIC III-IVa (Foot-Tongue)			
		Subjects			Mean±SD
		k3b	*k6b*	*l1b*	
CSP	1	86.16	90.31	73.57	83.35± **8.72**
	2	86.55	**96.94**	75.17	86.22±10.89
	3	**93.79**	94.32	75.79	87.63±10.29
	4	91.47	94.23	73.25	86.32±11.40
CCSP	3	89.79	93.32	73.82	85.64±10.39
SRCSP	3	90.64	96.32	74.18	87.05±11.50
GLRCSP	3	89.79	93.32	75.77	86.29±9.28
TRCSP	3	90.75	96.32	75.41	87.49±10.83
GS-CSP	1,2,3,4	**93.79**	**96.94**	**78.94**	**89.89**±9.61

Although there are some variabilities in classification performance over subjects, the Tables show that the proposed method clearly outperforms the other reported methods. Furthermore, on the basis of execution time, the proposed method outperforms the reported methods as it gives less execution time. To estimate the performance a validation technique is used in this experiment. In all our experiments 10-fold cross validation is used. In this cross validation technique, the dataset is randomly divided into 10 equal subsets where one of the subsets is used for the test while rests 9 are used to the training. The cross-validation is repeated 10 times, and then the results of 10 times are averaged to yield a single classification rate. The execution time is estimated by summing up the expended time during spatial filtering (training) and classification (testing). To calculate the execution time the 10-fold cross validation with the technical software MATLAB R2014b is used. During the performance estimation the system environment was Intel(R) Core(TM) i5-4310U @ 2.00GHz processor, 4.00 GB RAM and 64-bit Windows 7 Ultimate operating system. Figure 7.15 shows the average execution time for various motor imagery tasks of the reported methods including the proposed one. It is found that the average execution time of the proposed method is less than one second. Due to more training time of other methods, they need more execution time compared to the proposed method.

7.3 Auditory Stimuli Based BCI

In this research a new data containing imagined and vocalized phonemic and single-word prompts is analyzed. The EEG data is preprocessed and binary classification of phonological categories are performed with the proposed GS-CSP. These data may be used to develop the brain computer interface.

The KARA ONE Database: Phonological categories in imagined and articulated speech: To collect the data four female and eight male participants (mean age = 27.4, σ = 5, range = 14) were employed from the University of

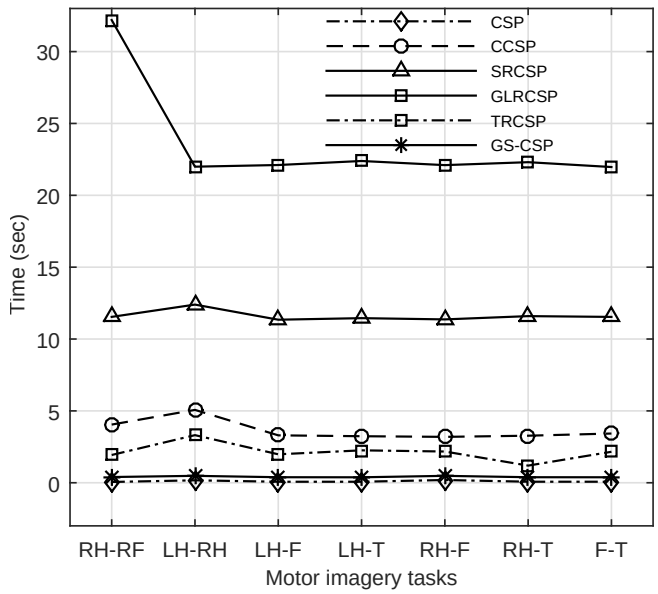


Figure 7.15: Average execution time of different subject for various motor imagery tasks; RH-RF: Right Hand versus Right Foot, LH-RH: Left Hand versus Right Hand, LH-F: Left Hand versus Foot, LH-T: Left Hand versus Tongue, RH-F: Right Hand versus Foot, RH-T: Right Hand versus Tongue, and F-T: Foot versus Tongue.

Toronto campus. All the twelve participants were right-handed, had at least some post-secondary education, had no visual, hearing, or motor impairments, and had no history of neurological conditions or drug abuse. Additionally, ten of them identified North American English as their first language and the remaining two spoke North American English at a fluent level, having learned the language at a mean age of 6.

The data collection was carried out an office environment at the Toronto Rehabilitation Institute. Subject was seated in a chair before a computer monitor. To record the participants speech a Microsoft Kinect (v.1.8) camera was placed next to the screen. A research assistant placed an appropriately-sized EEG cap on the participants head and injected a small amount of gel to improve electrical conductance. A 64-channel Neuroscan Quick-cap was used, where the electrode placement follows the 10-20 system. Four electrodes were placed above and below the left eye and to the lateral side of each eye to control for artifacts arising from eyemovement. The EEG data were recorded using the SynAmps RT amplifier and

sampled at 1 kHz. Impedance levels were usually maintained below 10 k. The subject was instructed to look at the computer monitor and to move as little as possible. Over the course of 30 to 40 minutes, individual prompts appeared on the screen one-at-a-time. Seven phonemic/syllabic prompts (*/iy/*, */uw/*, */piy/*, */tiy/*, */diy/*, */m/*, */n/* and four words derived from Kents list of phonetically-similar pairs (i.e., */pat/*, */pot/*, */knew/*, and */gnaw/*) were used [118].

Each trial consisted of 4 successive states:

1. The subject was instructed to relax and clear their mind of any thoughts during a 5-second rest state.

2. During the stimulus state the prompt text would appear on the screen and its associated auditory utterance was played over the computer speakers. This was followed by a 2-second period in which the subject moved their articulators into position to begin pronouncing the prompt.

3. A 5-second imagined speech state, in which the subject imagined speaking the prompt without moving.

4. A speaking state, in which the subject spoke the prompt aloud. The Kinect sensor recorded both the audio and facial features during this stage.

When the subject has finished speaking, one of the investigators would proceed to the next trial. Each prompt was presented 12 times for a total of 132 trials. The phonemic/syllabic prompts were first presented followed by the 4 Kent words, and the trials were randomly permuted within each of those two sections.The subject was given the chance to rest after every 40 trials. Data from 4 of the 12 participants were discarded due to unattached ground wires and two participants falling asleep during recording. Also ethical approval was obtained from both the University of Toronto and the University Health Network, of which Toronto Rehab is a member.

Preprocessing: The recorded EEG was pre-processed with EEGLAB [58], with the removal of ocular artifacts using blind source separation [86]. The data were band-pass filtered between 1 Hz and 50 Hz, and the mean values were subtracted from each channel. Moreover a small Laplacian filter was applied to the data, using the neighbourhood of adjacent channels. The EEG data were seg-

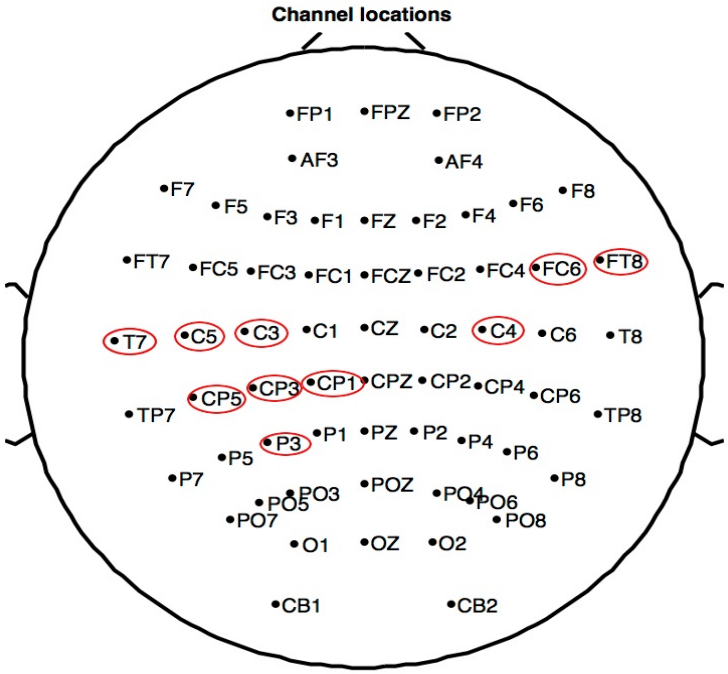


Figure 7.16: The electrodes map of 10/20 EEG system standardized by the American EEG society. The circled electrodes T7, FT8, CP3, C5, FC6, P3, CP5, C3, C4, CP1 are selected for the dataset used in this research.

mented into different trials, and each trial was further segmented into the 4 states described above.

Channel Selection: It is investigated that the channels of central and temporal locations T7, FT8, CP3, C5, FC6, P3, CP5, C3, C4, CP1, generally around the auditory cortex of brain are more dominated during the planning of speech articulation [259]. All of the 62 EEG channels over all imagined speech segments in the dataset are shown in Figure 7.16. In this research, the data of the 10 selected channels are used.

Classification: In this experiment, the classification performance is evaluated using the proposed GS-CSP based approach. Here, 10 of 62 channels and 5 of 12 subjects (MM05, MM08, MM09, MM16 and MM19) are used. We have used 6 phonemic/syllabic prompts /*iy*/, /*uw*/, /*piy*/, /*tiy*/, /*m*/ and /*n*/ in this experiment. The binary classification operation between vowel-only versus consonant(C/V) is performed. At first, binary classifications between single vowel /*iy*/ and single consonant /*piy*/, between single vowel /*iy*/ and single consonant

Table 7.5: Classification performance on the KARA ONE database (phonological categories in imagined and articulated speech). Classification accuracies (in %) obtained for each subject for the proposed GS-CSP method.

Subject	Prompts					
	/iy/ vs. */piy/*	*/iy/* vs. */m/*	*/uw/* vs. */tiy/*	*/uw/* vs. */n/*	*/iy//uw/* vs. */piy//n/*	*/iy//uw/* vs. */tiy//m/*
MM05	50	100	100	100	66.67	75.00
MM08	100	75	80	100	88.89	83.33
MM09	75	100	75	75	75.86	66.67
MM16	100	75	100	100	77.78	77.78
MM19	100	80	75	80	66.67	88.89
Mean	**85**	**86**	**86**	**91**	**75.17**	**78.33**

Table 7.6: Classification performance on the KARA ONE database (phonological categories in imagined and articulated speech). Mean accuracies (in %) of the five subjects MM05, MM08, MM09, MM16 and MM19 for the different methods

Methods	Prompts					
	/iy/ vs. */piy/*	*/iy/* vs. */m/*	*/uw/* vs. */tiy/*	*/uw/* vs. */n/*	*/iy//uw/* vs. */piy//n/*	*/iy//uw/* vs. */tiy//m/*
CSP	83.25	82.58	80.54	90.32	74.98	77.33
CCSP	81.45	80.34	79.33	83.88	73.24	72.69
SRCSP	82.30	78.50	78.50	82.70	74.50	73.77
GLRCSP	79.40	76.70	78.50	83.70	73.24	73.77
TRCSP	83.50	84.00	81.90	90.32	73.58	77.33
GS-CSP	**85**	**86**	**86**	**91**	**75.17**	**78.33**

/*m*/, between single vowel /*uw*/ and single consonant /*tiy*/ and between single vowel /*uw*/ and single consonant /*n*/ are performed. Moreover, the classification performances between two vowel /*iy*//*uw*/ and two consonant /*piy*//*n*/ and finally between two vowel /*iy*//*uw*/ and two consonant /*tiy*//*m*/ are verified. The classification performance for the dataset is shown in Table 7.5. In this study, the minimum mean accuracy of 85% is obtained between single vowel /*iy*/ and single constant /*piy*/ whereas the maximum mean accuracy of 91% is obtained between single vowel /*uw*/ and single consonant /*n*/. Besides, the maximum mean accuracy of 78.33% is obtained between two vowel /*iy*//*uw*/ and two consonant /*tiy*//*m*/.

In the proposed method, the selected combinations of spatial filter pair and number of features (FP, Fs) are (4, 2), (3, 4), (2, 3), (4, 2) and (3, 6) for the subject MM05, MM08, MM09, MM16 and MM19 respectively. The filter pair $m = 3$ is considered for the other methods. The mean accuracies (in %) for the GS-CSP method are compared to the reported methods in Table 7.6. Table 7.6 shows that the mean accuracies for the C/V are obtained from 85% to 91% for the proposed method whereas the accuracies are yielded between 80% and 84% for the other methods. Therefore, the proposed method can be promising to implement in audio stimui based BCI.

Chapter 8

Conclusions and Future Works

8.1 Contributions

To analyze the time-frequency (TF) coherence between a pair of signals, a novel method is presented in this research. The densities of the cross and auto spectrum are measured for the specified signals in time frequency domain. Then the spectral densities are smoothed using non-identical smoothing operators. The TF coherence is estimated with the smooth spectral densities for synthetic signals with STFT and SST based time-frequency representation. The proposed SST based coherence estimation method is applied to real EEG signals of different motor imagery. The performance of the both techniques are compared and observed that SST based method is more efficient than STFT for localization of frequency components with higher resolution in coherence domain. Then marginal time coherences are calculated from both SST and STFT based coherences. It is noticeably found that the SST based marginal time coherences exhibits very clear discrimination between left hand and foot movement data whereas the STFT based marginal time coherences are unable to do this.

A novel method to classify EEG of imagined movement is presented in this research. The EEG is filtered into the most dominant frequency bands alpha and beta (8-30Hz). The training and test sets of the EEG are separated and GS-CSP is applied to only the training set in order to select spatial filter pair and discriminative number of CSP features. In doing so, the training set is further divided into two sets; training subset and test subset. The discriminative CSP

features are then extracted from both parts. Mutual information concerning the training subset CSP features and their corresponding given class are computed. Later on features are put in order based on the mutual information. Subsequently, the CSP features of the test subset are ranked in accordance with the order of the ranked training subset CSP features. The discriminant hyperplane coefficients are computed for all of the ranked training CSP features. These coefficients are used to calculate the LDA accuracy scores for the respective testing CSP features. The number of spatial filter pair doubles the CSP features and influences classification accuracy. In most of the existing BCI study a certain pair of spatial filter is selected manually, however, that does not assure the best accuracy result. Spatial filter pair selection differs from study to study mostly from 1 to 4. The highest result can be achieved by appropriate selection of spatial filter pair and number of CSP features. In this study, all of the spatial filter pairs form 1 to 4 and their corresponding CSP features are considered. The LDA accuracy scores are calculated for all of the combinations of spatial filter pairs and CSP features, (FP, Fs) where FP= 1,...,4 and Fs= 1,...,2m. Based on the grid search the top accuracy score is obtained and hence corresponding combination is selected for classification. The accuracy scores are subject to the combination e.g., the combination of (3, 4), (4, 5), (3, 2), (4, 5) and (2, 4) that yields top scores for the subject *aa*, *al*, *av*, *aw* and *ay* respectively as shown in Figure 7.12. In the case of left hand versus right hand motor imagery task, the best score for the subject *k3b*, and *l1b* are produced for the combination of (4, 8) and (3, 4) respectively as shown in Figure 7.13. Two combinations e.g., (3, 5) and (3, 6) generate the highest score for the subject *k6b* but only (3, 5) is selected for classification because of less features. The combination with less filter pair should be selected if more than one combination with different filter pair produces the highest score since the pair gives a reduced amount of training time. Finally, the selected combination is used to compute the classification accuracy of the test set EEG. It is concluded that on average even for individual subject the maximum classification accuracy is achieved by the proposed method. Moreover, in the case of execution time

the proposed method outperforms some other methods as shown in Figure 7.15. The execution time of the methods CCSP, SRCSP, GLRCSP and TRCSP is more than one second because of more training time. The algorithms uses data from other subjects and regularizes the estimation of the covariance matrix. Since the methods consider the information of other subject during training period, the algorithms consume much more training time than standard CSP. The proposed feature selection approach GS-CSP searches the best feature combination using the standard CSP. The approach needs little more time compared to CSP for the searching but much lower than the other methods as shown in Figure 7.15. Therefore, experimental studies on two public EEG datasets (BCI competition III dataset IVa and BCI competition III dataset IIIa) showed that the proposed GS-CSP method yielded higher overall classification accuracy and lower execution time in contrast to some other stated methods.

The proposed method is applied to classify audio stimuli based EEG. For the experiment the KARA ONE database (Computational Linguistics Lab, University of Toronto, Canada) is used in this research. On the basis of classification results, it is concluded that the proposed approach will be useful for audio stimuli based BCI research. The Audio stimuli EEG classification can take part for the auditory related brain development for children and diagnosis of human brain.

8.2 Future Works

The work described in this thesis has left some unanswered questions deserving further investigation. This section describes the future goals of the research. Though some promising results are achieved in the work, there are many opportunities to improve the performance of the system. A number of techniques could be examined to enhance the performance of the classification method with the purpose of enabling fast and accurate decision. It is needed to improve the presented BCI to be a practical reality and potentially be a part of the future generations of physically disable people. The approach described in this research is designed for

motor-imagery based BCIs and deals with two class problems using linear classifier. In future, the approach can be extended to include a neural network which is able to produce nonlinear decision boundary and effective for multiclass problem.

It is necessary to select the effective number of channels to improve the classification performance. Different frequency bands can be used to enhance the accuracy. Furthermore, it would be interesting to design the BCI system using very short period of time for the intended activity that maximizes the chance of performing the actual task for the specific subject.

In this research, effective dominant features are selected with mutual information approach. Future work could deal with the exploration of approaches such as eigenvalue centrality [210] and Laplacian score [99]. The development of better computational model to implement a practical auditory BCI system for patients with physical disabilities can be another direction of future work. Furthermore, the development of audio stimuli based BCI technology that would be promising for a substitute of standard computer input device for both healthy and disabled computer users is also left for future work.

Bibliography

[1] Alireza Ahrabian, David Looney, Ljubiša Stanković, and Danilo P Mandic. Synchrosqueezing-based time-frequency analysis of multivariate data. *Signal Processing*, 106:331–341, 2015.

[2] Brendan Allison, Thorsten Luth, Diana Valbuena, Amir Teymourian, Ivan Volosyak, and Axel Graser. Bci demographics: How many (and what kinds of) people can use an ssvep bci? *IEEE transactions on neural systems and rehabilitation engineering*, 18(2):107–116, 2010.

[3] Amir Amedi, Rafael Malach, and Alvaro Pascual-Leone. Negative bold differentiates visual imagery and perception. *Neuron*, 48(5):859–872, 2005.

[4] Charles W Anderson and Zlatko Sijercic. Classification of eeg signals from four subjects during five mental tasks. In *Solving engineering problems with neural networks: proceedings of the conference on engineering applications in neural networks (EANN96)*, pages 407–414. Turkey, 1996.

[5] Keith Andrews, Lesley Murphy, Ros Munday, and Clare Littlewood. Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *Bmj*, 313(7048):13–16, 1996.

[6] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 2390–2397. IEEE, 2008.

[7] François Auger and Patrick Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE*

Transactions on signal processing, 43(5):1068–1089, 1995.

[8] François Auger, Patrick Flandrin, Yu-Ting Lin, Stephen McLaughlin, Sylvain Meignen, Thomas Oberlin, and Hau-Tieng Wu. Time-frequency reassignment and synchrosqueezing: An overview. *IEEE Signal Processing Magazine*, 30(6):32–41, 2013.

[9] Armando B Barreto, Annette M Taberner, and Luis M Vicente. Classification of spatio-temporal eeg readiness potentials towards the development of a brain-computer interface. In *Southeastcon'96. Bringing Together Education, Science and Technology., Proceedings of the IEEE*, pages 99–102. IEEE, 1996.

[10] Guilherme A Barreto, Rewbenio A Frota, and Fátima NS de Medeiros. On the classification of mental tasks: a performance comparison of neural and statistical approaches. In *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop*, pages 529–538. IEEE, 2004.

[11] Gerhard Bauer, Franz Gerstenbrand, and Erik Rumpl. Varieties of the locked-in syndrome. *Journal of neurology*, 221(2):77–91, 1979.

[12] Andrei Belitski, Jason Farquhar, and Peter Desain. P300 audio-visual speller. *Journal of Neural Engineering*, 8(2):025022, 2011.

[13] Hans Berger. On the electroencephalogram of man. third report. *Electroencephalography and clinical neurophysiology*, pages Suppl–28, 1969.

[14] Niels Birbaumer and Leonardo G Cohen. Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of physiology*, 579(3):621–636, 2007.

[15] Niels Birbaumer, Nimr Ghanayim, Thilo Hinterberger, Iver Iversen, Boris Kotchoubey, Andrea Kübler, Juri Perelmouter, Edward Taub, and Herta Flor. A spelling device for the paralysed. *Nature*, 398(6725):297, 1999.

[16] Christopher M Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[17] DH Blackwood and WJ Muir. Cognitive brain potentials and their applica-

tion. *The British Journal of Psychiatry*, 1990.

[18] Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Classifying single trial eeg: Towards brain computer interfacing. In *Advances in neural information processing systems*, pages 157–164, 2002.

[19] Benjamin Blankertz, K-R Muller, Gabriel Curio, Theresa M Vaughan, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlogl, Christa Neuper, Gert Pfurtscheller, Thilo Hinterberger, et al. The bci competition 2003: progress and perspectives in detection and discrimination of eeg single trials. *IEEE transactions on biomedical engineering*, 51(6):1044–1051, 2004.

[20] Benjamin Blankertz, K-R Muller, Dean J Krusienski, Gerwin Schalk, Jonathan R Wolpaw, Alois Schlogl, Gert Pfurtscheller, Jd R Millan, Michael Schroder, and Niels Birbaumer. The bci competition iii: Validating alternative approaches to actual bci problems. *IEEE transactions on neural systems and rehabilitation engineering*, 14(2):153–159, 2006.

[21] Benjamin Blankertz, Michael Tangermann, Carmen Vidaurre, Siamac Fazli, Claudia Sannelli, Stefan Haufe, Cecilia Maeder, Lenny E Ramsey, Irene Sturm, Gabriel Curio, et al. The berlin brain–computer interface: non-medical uses of bci technology. *Frontiers in neuroscience*, 4:198, 2010.

[22] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and K-R Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2008.

[23] Jaimie F Borisoff, Steven G Mason, Ali Bashashati, and Gary E Birch. Brain-computer interface design for asynchronous control applications: improvements to the lf-asd asynchronous brain switch. *IEEE Transactions on Biomedical Engineering*, 51(6):985–992, 2004.

[24] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[25] Mary AB Brazier. A history of the electrical activity of the brain: The first half-century. 1961.

[26] John-Stuart Brittain, David M Halliday, Bernard A Conway, and Jens Bo Nielsen. Single-trial multiwavelet coherence in application to neurophysiological time series. *IEEE Transactions on Biomedical Engineering*, 54(5):854–862, 2007.

[27] Andrea Brovelli, Mingzhou Ding, Anders Ledberg, Yonghong Chen, Richard Nakamura, and Steven L Bressler. Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by granger causality. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26):9849–9854, 2004.

[28] Roberto Brunelli. *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[29] Peter Brunner, S Joshi, Samuel Briskin, Jonathan R Wolpaw, Horst Bischof, and Gerwin Schalk. Does the p300speller depend on eye gaze? *Journal of neural engineering*, 7(5):056013, 2010.

[30] Marie-Aurélie Bruno, Steve Majerus, Mélanie Boly, Audrey Vanhaudenhuyse, Caroline Schnakers, Olivia Gosseries, Pierre Boveroux, Murielle Kirsch, Athena Demertzi, Claire Bernard, et al. Functional neuroanatomy underlying the clinical subcategorization of minimally conscious state patients. *Journal of neurology*, 259(6):1087–1098, 2012.

[31] Ethan Buch, Cornelia Weber, Leonardo G Cohen, Christoph Braun, Michael A Dimyan, Tyler Ard, Jurgen Mellinger, Andrea Caria, Surjo Soekadar, Alissa Fourkas, et al. Think to move: a neuromagnetic brain-computer interface (bci) system for chronic stroke. *Stroke*, 39(3):910–917, 2008.

[32] Scott C Bunce, Meltem Izzetoglu, Kurtulus Izzetoglu, Banu Onaral, and Kambiz Pourrezaei. Functional near-infrared spectroscopy. *IEEE engineering in medicine and biology magazine*, 25(4):54–62, 2006.

[33] Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[34] Gemma A Calvert. Crossmodal processing in the human brain: insights

from functional neuroimaging studies. *Cerebral cortex*, 11(12):1110–1123, 2001.

[35] René Carmona, Wen-Liang Hwang, and Bruno Torresani. *Practical Time-Frequency Analysis: Gabor and wavelet transforms, with an implementation in S*, volume 9. Academic Press, 1998.

[36] Gail A Carpenter, Stephen Grossberg, Natalya Markuzon, John H Reynolds, and David B Rosen. Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on neural networks*, 3(5):698–713, 1992.

[37] G Clifford Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75(2):236–255, 1987.

[38] Gavin C Cawley, Nicola L Talbot, and Mark Girolami. Sparse multinomial logistic regression via bayesian l1 regularisation. In *Advances in neural information processing systems*, pages 209–216, 2007.

[39] Hubert Cecotti, Ivan Volosyak, and Axel Gräser. Reliable visual stimuli on lcd screens for ssvep based bci. In *Signal Processing Conference, 2010 18th European*, pages 919–923. IEEE, 2010.

[40] Moonjeong Chang, Nozomu Nishikawa, Zbigniew R Struzik, Koichi Mori, Shoji Makino, D Mandic, and Tomasz M Rutkowski. Comparison of p300 responses in auditory, visual and audiovisual spatial speller bci paradigms. *arXiv preprint arXiv:1301.6360*, 2013.

[41] JM Charcot. De la sclérose latérale amyotrophique. symptomatologie. *Lecons sur les Maladies du Systeme Nerveux Faites a la Salpetriere*, 2:227–242, 1880.

[42] Silvia Chiappa and Samy Bengio. Hmm and iohmm modeling of eeg rhythms for asynchronous bci systems. Technical report, IDIAP, 2003.

[43] F Cincotti, A Scipione, A Timperi, D Mattia, AG Marciani, J Millan, S Salinari, L Bianchi, and F Bablioni. Comparison of different feature classifiers for brain computer interfaces. In *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, pages 645–647. IEEE, 2003.

[44] Luca Citi, Riccardo Poli, Caterina Cinel, and Francisco Sepulveda. P300-based bci mouse with genetically-optimized analogue control. *IEEE transactions on neural systems and rehabilitation engineering*, 16(1):51–61, 2008.

[45] Ed AK Cohen and Andrew T Walden. A statistical study of temporally smoothed wavelet coherence. *IEEE Transactions on Signal Processing*, 58(6):2964–2973, 2010.

[46] Marco Congedo, Fabien Lotte, and Anatole Lécuyer. Classification of movement intention by spatially filtered electromagnetic inverse solutions. *Physics in medicine and biology*, 51(8):1971, 2006.

[47] Thomas M Cover and Joy A Thomas. Elements of information theory 2nd edition. 2006.

[48] Thomas M Cover and Jan M Van Campenhout. On the possible orderings in the measurement selection problem. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(9):657–661, 1977.

[49] DH Coyle, Aine Carroll, Jacqueline Stow, Alison McCann, Aneesa Ally, and Jacinta McElligott. Enabling control in the minimally conscious state in a single session with a three channel bci. The 1st international DECODER Workshop, 2012.

[50] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[51] FH Lopes Da Silva. The generation of electric and magnetic signals of the brain by local networks. In *Comprehensive human physiology*, pages 509–531. Springer, 1996.

[52] Janis J Daly and Jonathan R Wolpaw. Brain–computer interfaces in neurological rehabilitation. *The Lancet Neurology*, 7(11):1032–1043, 2008.

[53] Ingrid Daubechies, Jianfeng Lu, and Hau-Tieng Wu. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool. *Applied and computational harmonic analysis*, 30(2):243–261, 2011.

[54] Ingrid Daubechies and Stephane Maes. A nonlinear squeezing of the con-

tinuous wavelet transform based on auditory nerve models. *Wavelets in medicine and biology*, pages 527–546, 1996.

[55] Beatrice De Gelder, Gilles Pourtois, and Lawrence Weiskrantz. Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences*, 99(6):4121–4126, 2002.

[56] Jean Decety and David H Ingvar. Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation. *Acta psychologica*, 73(1):13–34, 1990.

[57] Jose del R Millan, Josep Mouriño, Marco Franzé, Febo Cincotti, Markus Varsta, Jukka Heikkonen, and Fabio Babiloni. A local neural classifier for the recognition of eeg patterns associated to mental tasks. *IEEE transactions on neural networks*, 13(3):678–686, 2002.

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

[59] Jennifer E Doble, Andrew J Haig, Christopher Anderson, and Richard Katz. Impairment, activity, participation, life satisfaction, and survival in persons with locked-in syndrome for over a decade: Follow-up on a previously reported cohort. *The Journal of head trauma rehabilitation*, 18(5):435–444, 2003.

[60] Oliver Doehrmann and Marcus J Naumer. Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain research*, 1242:136–150, 2008.

[61] Emanuel Donchin, Kevin M Spencer, and Ranjith Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE transactions on rehabilitation engineering*, 8(2):174–179, 2000.

[62] Guido Dornhege. *Toward brain-computer interfacing*. MIT press, 2007.

[63] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and Klaus-Robert Müller. Optimizing spatio-temporal filters

for improving brain-computer interfacing. In *Advances in Neural Information Processing Systems*, pages 315–322, 2006.

[64] Richard O Duda and Peter E Hart. Pattern recognition and scene analysis, 1973.

[65] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification second edition john wiley & sons. *New York*, 58, 2001.

[66] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[67] Jennifer G Dy and Carla E Brodley. Feature subset selection and order identification for unsupervised learning. In *ICML*, pages 247–254, 2000.

[68] Jennifer G Dy and Carla E Brodley. Feature selection for unsupervised learning. *Journal of machine learning research*, 5(Aug):845–889, 2004.

[69] Carlos Escolano, Ander Ramos Murguialday, Tamara Matuz, Niels Birbaumer, and Javier Minguez. A telepresence robotic system operated with a p300-based brain-computer interface: initial tests with als patients. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4476–4480. IEEE, 2010.

[70] Owen Falzon, Kenneth P Camilleri, and Joseph Muscat. The analytic common spatial patterns method for eeg-based bci data. *Journal of neural engineering*, 9(4):045009, 2012.

[71] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

[72] Davood Fattahi, Behrooz Nasihatkon, and Reza Boostani. A general framework to estimate spatial and spatio-spectral filters for eeg signal classification. *Neurocomputing*, 119:165–174, 2013.

[73] Siamac Fazli, Márton Danóczy, Florin Popescu, Benjamin Blankertz, and Klaus-Robert Müller. Using rest class and control paradigms for brain computer interfacing. In *Brain-Computer Interfaces*, pages 55–70. Springer, 2010.

[74] Torsten Felzer and B Freisieben. Analyzing eeg signals using the probability estimating guarded neural classifier. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(4):361–371, 2003.

[75] Alexandra Fort, Claude Delpuech, Jacques Pernier, and Marie-Hélène Giard. Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, 12(10):1031–1039, 2002.

[76] Walter J Freeman III. The physiology of perception. *Scientific American*, 264(2), 1991.

[77] Jerome H Friedman. On bias, variance, 0/1loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, 1(1):55–77, 1997.

[78] Gerhard M Friehs, Vasilios A Zerris, Catherine L Ojakangas, Mathew R Fellows, and John P Donoghue. Brain–machine and brain–computer interfaces. *Stroke*, 35(11 suppl 1):2702–2705, 2004.

[79] Pascal Fries. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual review of neuroscience*, 32:209–224, 2009.

[80] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.

[81] Adrian Furdea, Sebastian Halder, DJ Krusienski, Donald Bross, Femke Nijboer, Niels Birbaumer, and Andrea Kübler. An auditory oddball (p300) spelling system for brain-computer interfaces. *Psychophysiology*, 46(3):617–625, 2009.

[82] Ferran Galán, Marnix Nuttin, Eileen Lew, Pierre W Ferrez, Gerolf Vanacker, Johan Philips, and J del R Millán. A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. *Clinical neurophysiology*, 119(9):2159–2169, 2008.

[83] Marissa L Gamble and Steven J Luck. N2ac: An erp component associated with the focusing of attention within an auditory scene. *Psychophysiology*, 48(8):1057–1068, 2011.

[84] Joseph T Giacino, Stephen Ashwal, Nancy Childs, Ronald Cranford, Bryan

Jennett, Douglas I Katz, James P Kelly, Jay H Rosenberg, JOHN Whyte, RD Zafonte, et al. The minimally conscious state definition and diagnostic criteria. *Neurology*, 58(3):349–353, 2002.

[85] Joseph T Giacino, Kathleen Kalmar, and John Whyte. The jfk coma recovery scale-revised: Measurement characteristics and diagnostic utility1. *Archives of physical medicine and rehabilitation*, 85(12):2020–2029, 2004.

[86] Germán Gómez-Herrero, Wim De Clercq, Haroon Anwar, Olga Kara, Karen Egiazarian, Sabine Van Huffel, and Wim Van Paesschen. Automatic removal of ocular artifacts in the eeg without an eog reference channel. In *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*, pages 130–133. IEEE, 2006.

[87] Henry Gray. *Anatomy of the human body*. Lea & Febiger, 1918.

[88] Georgia G Gregoriou, Stephen J Gotts, Huihui Zhou, and Robert Desimone. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *science*, 324(5931):1207–1210, 2009.

[89] Moritz Grosse-Wentrup and Bernhard Schölkopf. A review of performance variations in smr-based brain- computer interfaces (bcis). In *Brain-Computer Interface Research*, pages 39–51. Springer, 2013.

[90] C Guger, G Edlinger, W Harkam, I Niedermayer, and G Pfurtscheller. How many people are able to operate an eeg-based brain-computer interface (bci)? *IEEE transactions on neural systems and rehabilitation engineering*, 11(2):145–147, 2003.

[91] Jing Guo, Shangkai Gao, and Bo Hong. An auditory brain–computer interface using active mental response. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(3):230–235, 2010.

[92] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[93] Stelios Hadjidimitriou, A Zacharakis, Panagiotis Doulgeris, K Panoulas, L Hadjileontiadis, and S Panas. Sensorimotor cortical response during motion reflecting audiovisual stimulation: evidence from fractal eeg analysis.

Medical & biological engineering & computing, 48(6):561–572, 2010.

[94] S Halder, M Rea, R Andreoni, F Nijboer, EM Hammer, SC Kleih, N Birbaumer, and A Kübler. An auditory oddball brain–computer interface for binary choices. *Clinical Neurophysiology*, 121(4):516–523, 2010.

[95] Matti Hämäläinen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalographytheory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413, 1993.

[96] Mohammad Rubaiyat Hasan, Muhammad Ibn Ibrahimy, SMA Motakabber, and Shahjahan Shahid. Classification of multichannel eeg signal by linear discriminant analysis. In *Progress in Systems Engineering*, pages 279–282. Springer, 2015.

[97] Ernst Haselsteiner and Gert Pfurtscheller. Using time-dependent neural networks for eeg classification. *IEEE transactions on rehabilitation engineering*, 8(4):457–463, 2000.

[98] Stefan Haufe, Matthias S Treder, Manfred F Gugler, Max Sagebaum, Gabriel Curio, and Benjamin Blankertz. Eeg potentials predict upcoming emergency brakings during simulated driving. *Journal of neural engineering*, 8(5):056001, 2011.

[99] Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selection. In *Advances in neural information processing systems*, pages 507–514, 2006.

[100] Grit Hein, Oliver Doehrmann, Notger G Müller, Jochen Kaiser, Lars Muckli, and Marcus J Naumer. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *Journal of Neuroscience*, 27(30):7881–7887, 2007.

[101] N Jeremy Hill, Thomas N Lal, Karin Bierig, Niels Birbaumer, and Bernhard Schölkopf. An auditory paradigm for brain-computer interfaces. In *Advances in neural information processing systems*, pages 569–576, 2005.

[102] N Jeremy Hill, Thomas Navin Lal, Karin Bierig, Niels Birbaumer, and

B Scholkopf. Attention modulation of auditory event-related potentials in a brain-computer interface. In *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, pages S3–5. IEEE, 2004.

[103] NJ Hill and Bernhard Schölkopf. An online brain–computer interface based on shifting attention to concurrent streams of auditory stimuli. *Journal of neural engineering*, 9(2):026011, 2012.

[104] Akira Hiraiwa, Katsunori Shimohara, and Yukio Tokunaga. Eeg topography recognition by neural networks. *IEEE Engineering in Medicine and Biology magazine*, 9(3):39–42, 1990.

[105] Leigh R Hochberg, Mijail D Serruya, Gerhard M Friehs, Jon A Mukand, Maryam Saleh, Abraham H Caplan, Almut Branner, David Chen, Richard D Penn, and John P Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164, 2006.

[106] Bo Hong, Fei Guo, Tao Liu, Xiaorong Gao, and Shangkai Gao. N200-speller using motion-onset visual response. *Clinical neurophysiology*, 120(9):1658–1666, 2009.

[107] Tetsuya Hoya, Gen Hori, Hovagim Bakardjian, Tomoaki Nishimura, Taiji Suzuki, Yoichi Miyawaki, Arao Funase, and Jianting Cao. Classification of single trial eeg signals by a combined principal+ independent component analysis and probabilistic neural network approach. In *Proc. ICA2003*, volume 197, 2003.

[108] Han-Jeong Hwang, Kiwoon Kwon, and Chang-Hwang Im. Neurofeedback-based motor imagery training for brain–computer interface (bci). *Journal of neuroscience methods*, 179(1):150–156, 2009.

[109] Anil Jain and Douglas Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE transactions on pattern analysis and machine intelligence*, 19(2):153–158, 1997.

[110] Anil K Jain, Robert P. W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):4–37, 2000.

[111] Marc Jeannerod and Victor Frak. Mental imaging of motor activity in humans. *Current opinion in neurobiology*, 9(6):735–739, 1999.

[112] Bryan Jennett and Fred Plum. Persistent vegetative state after brain damage: a syndrome in search of a name. *The Lancet*, 299(7753):734–737, 1972.

[113] Hyohyeong Kang, Yunjun Nam, and Seungjin Choi. Composite common spatial pattern for subject-to-subject transfer. *IEEE Signal Processing Letters*, 16(8):683–686, 2009.

[114] Shin'ichiro Kanoh, Ko-ichiro Miyamoto, and Tatsuo Yoshinobu. A brain-computer interface (bci) system based on auditory stream segregation. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 642–645. IEEE, 2008.

[115] Mehmed Kantardzic. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.

[116] Zachary A Keirn and Jorge I Aunon. Man-machine communications through brain-wave processing. *IEEE Engineering in Medicine and biology magazine*, 9(1):55–57, 1990.

[117] Zachary A Keirn and Jorge I Aunon. A new mode of communication between man and his surroundings. *IEEE transactions on biomedical engineering*, 37(12):1209–1214, 1990.

[118] Ray D Kent, Gary Weismer, Jane F Kent, and John C Rosenbek. Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54(4):482–499, 1989.

[119] B Abou Khalil and KE Misulis. Atlas of eeg & seizure semiology, 2006.

[120] Kunal Khanna, Ajit Verma, and Bella Richard. the locked-in syndrome: Can it be unlocked? *Journal of Clinical Gerontology and Geriatrics*, 2(4):96–99, 2011.

[121] Do-Won Kim, Han-Jeong Hwang, Jeong-Hwan Lim, Yong-Ho Lee, Ki-Young Jung, and Chang-Hwan Im. Classification of selective attention to auditory stimuli: toward vision-free brain–computer interfacing. *Journal of neuroscience methods*, 197(1):180–185, 2011.

[122] Daniela S Klobassa, TM Vaughan, Peter Brunner, NE Schwartz, JR Wolpaw, Christa Neuper, and EW Sellers. Toward a high-throughput auditory p300-based brain–computer interface. *Clinical neurophysiology*, 120(7):1252–1261, 2009.

[123] Kunihiko Kodera, Roger Gendrin, and Claude Villedary. Analysis of time-varying signals with small bt values. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):64–76, 1978.

[124] Jens Kohlmorgen, Guido Dornhege, Mikio Braun, Benjamin Blankertz, Klaus-Robert Müller, Gabriel Curio, Konrad Hagemann, Andreas Bruns, Michael Schrauf, and Wilhelm Kincses. Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, pages 409–422, 2007.

[125] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[126] Zoltan Joseph Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical eeg. *Electroencephalography and clinical Neurophysiology*, 79(6):440–447, 1991.

[127] GV Kondraske. Neurophysiological measurements. *Biomedical Engineering and Instrumentation*, pages 138–179, 1986.

[128] Aleksander Kostov and Mark Polak. Parallel man-machine training in development of eeg-based cursor control. *IEEE Transactions on Rehabilitation Engineering*, 8(2):203–205, 2000.

[129] Matthias Krauledat, Konrad Grzeska, Max Sagebaum, Benjamin Blankertz, Carmen Vidaurre, Klaus-Robert Müller, and Michael Schröder. Playing pinball with non-invasive bci. In *Advances in neural information processing systems*, pages 1641–1648, 2009.

[130] Roman Krepki, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. The berlin brain-computer interface (bbci)–towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, 33(1):73–90, 2007.

[131] Dean J Krusienski, Eric W Sellers, François Cabestaing, Sabri Bayoudh, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. A comparison of classification techniques for the p300 speller. *Journal of neural engineering*, 3(4):299, 2006.

[132] Dean J Krusienski, Eric W Sellers, Dennis J McFarland, Theresa M Vaughan, and Jonathan R Wolpaw. Toward enhanced p300 speller performance. *Journal of neuroscience methods*, 167(1):15–21, 2008.

[133] Andrea Kübler and Niels Birbaumer. Brain–computer interfaces and communication in paralysis: extinction of goal directed thinking in completely paralysed patients? *Clinical neurophysiology*, 119(11):2658–2666, 2008.

[134] Andrea Kübler, Femke Nijboer, Jürgen Mellinger, Theresa M Vaughan, Hannelore Pawelzik, Gerwin Schalk, Dennis J McFarland, Niels Birbaumer, and Jonathan R Wolpaw. Patients with als can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777, 2005.

[135] Edmund C Lalor, Simon P Kelly, Ciarán Finucane, Robert Burke, Ray Smith, Richard B Reilly, and Gary Mcdarby. Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment. *EURASIP journal on applied signal processing*, 2005:3156–3164, 2005.

[136] Tian Lan, Deniz Erdogmus, Andre Adami, Santosh Mathan, and Misha Pavel. Channel selection and feature projection for cognitive load estimation using ambulatory eeg. *Computational intelligence and neuroscience*, 2007:8–8, 2007.

[137] Mikhail A Lebedev and Miguel AL Nicolelis. Brain–machine interfaces: past, present and future. *TRENDS in Neurosciences*, 29(9):536–546, 2006.

[138] Steven Lemm, Christin Schafer, and Gabriel Curio. Bci competition 2003-data set iii: probabilistic modeling of sensorimotor/spl mu/rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.

[139] Eric C Leuthardt, Kai J Miller, Gerwin Schalk, Rajesh PN Rao, and Jeffrey G Ojemann. Electrocorticography-based brain computer interface-the

seattle experience. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):194–198, 2006.

[140] Chuan Li and Ming Liang. A generalized synchrosqueezing transform for enhancing signal time–frequency representation. *Signal Processing*, 92(9):2264–2274, 2012.

[141] Jundong Li and Huan Liu. Challenges of feature selection for big data analytics. *IEEE Intelligent Systems*, 32(2):9–15, 2017.

[142] Yuanqing Li, Jiahui Pan, Fei Wang, and Zhuliang Yu. A hybrid bci system combining p300 and ssvep and its application to wheelchair control. *IEEE Transactions on Biomedical Engineering*, 60(11):3156–3166, 2013.

[143] Hualou Liang, Steven L Bressler, Mingzhou Ding, Robert Desimone, and Pascal Fries. Temporal dynamics of attention-modulated neuronal synchronization in macaque v4. *Neurocomputing*, 52:481–487, 2003.

[144] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE Transactions on Biomedical Engineering*, 54(6):1172–1176, 2007.

[145] Huan Liu and Lei Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on knowledge and data engineering*, 17(4):491–502, 2005.

[146] Paul C Liu. Wavelet spectrum analysis and ocean wind waves. *Wavelets in geophysics*, 4:151–166, 1994.

[147] M-A Lopez, Hector Pomares, Francisco Pelayo, Jose Urquiza, and Javier Perez. Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in brain–computer interfaces. *Neurocomputing*, 72(16):3617–3623, 2009.

[148] Fabien Lotte. *Study of electroencephalographic signal processing and classification techniques towards the use of brain-computer interfaces in virtual reality applications*. PhD thesis, INSA de Rennes, 2008.

[149] Fabien Lotte and Cuntai Guan. Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms. *IEEE Transactions*

on biomedical Engineering, 58(2):355–362, 2011.

[150] Eric Lowet, Mark J Roberts, Pietro Bonizzi, Joël Karel, and Peter De Weerd. Quantifying neural oscillatory synchronization: a comparison between spectral coherence and phase-locking value approaches. *PloS one*, 11(1):e0146443, 2016.

[151] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Regularized common spatial patterns with generic learning for eeg signal classification. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 6599–6602. IEEE, 2009.

[152] Dorothée Lulé, Quentin Noirhomme, Sonja C Kleih, Camille Chatelle, Sebastian Halder, Athena Demertzi, Marie-Aurélie Bruno, Olivia Gosseries, Audrey Vanhaudenhuyse, Caroline Schnakers, et al. Probing command following in patients with disorders of consciousness using a brain–computer interface. *Clinical Neurophysiology*, 124(1):101–106, 2013.

[153] Dorothée Lulé, Claudia Zickler, S Häcker, Marie-Aurelie Bruno, Athena Demertzi, Frederic Pellas, Steven Laureys, and A Kübler. Life can be worth living in locked-in syndrome. *Progress in brain research*, 177:339–351, 2009.

[154] Kaushik Majumdar. Constraining minimum-norm inverse by phase synchronization and signal power of the scalp eeg channels. *IEEE Transactions on Biomedical Engineering*, 56(4):1228–1235, 2009.

[155] Joseph N Mak and Jonathan R Wolpaw. Clinical applications of brain-computer interfaces: current state and future prospects. *IEEE reviews in biomedical engineering*, 2:187–199, 2009.

[156] Stéphane Mallat. *A wavelet tour of signal processing*. Academic press, 1999.

[157] Duvinage Matthieu, Castermans Thierry, P Mathieu, et al. A p300-based quantitative comparison between the emotiv epoc headset and a medical eeg device. *International Journal of Biomedical Engineering*, 12(56):201, 2013.

[158] Karl A McCreadie, Damien H Coyle, and Girijesh Prasad. Sensorimotor learning with stereo auditory feedback for a brain–computer interface. *Med-*

ical & biological engineering & computing, 51(3):285–293, 2013.

[159] Dennis J McFarland and Jonathan R Wolpaw. Brain-computer interfaces for communication and control. *Communications of the ACM*, 54(5):60–66, 2011.

[160] Saeid Mehrkanoon, Michael Breakspear, Andreas Daffertshofer, and Tjeerd W Boonstra. Non-identical smoothing operators for estimating time-frequency interdependence in electrophysiological recordings. *EURASIP Journal on Advances in Signal Processing*, 2013(1):73, 2013.

[161] Matthew Middendorf, Grant McMillan, Gloria Calhoun, and Keith S Jones. Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE transactions on rehabilitation engineering*, 8(2):211–214, 2000.

[162] Jd R Millan, Frederic Renkens, Josep Mourino, and Wulfram Gerstner. Non-invasive brain-actuated control of a mobile robot by human eeg. *IEEE Transactions on biomedical Engineering*, 51(6):1026–1033, 2004.

[163] Tom M Mitchell. Machine learning, ser. *Computer Science Series. Singapure: McGraw-Hill Companies, Inc*, 1997.

[164] Pabitra Mitra, CA Murthy, and Sankar K. Pal. Unsupervised feature selection using feature similarity. *IEEE transactions on pattern analysis and machine intelligence*, 24(3):301–312, 2002.

[165] Sophie Molholm, Walter Ritter, Micah M Murray, Daniel C Javitt, Charles E Schroeder, and John J Foxe. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive brain research*, 14(1):115–128, 2002.

[166] Eva Möller, Bärbel Schack, Matthias Arnold, and Herbert Witte. Instantaneous multivariate eeg coherence analysis by means of adaptive high-dimensional autoregressive models. *Journal of neuroscience methods*, 105(2):143–158, 2001.

[167] Andrew W Moore. Support vector machines. *Tutorial. School of Computer Science of the Carnegie Mellon University. Available at http://www. cs. cmu. edu/˜ awm/tutorials.[Accessed August 16, 2009]*, 2001.

[168] Melody M Moore. Real-world applications for brain-computer interface technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):162–165, 2003.

[169] Klaus-Robert Müller, Matthias Krauledat, Guido Dornhege, Gabriel Curio, and Benjamin Blankertz. Machine learning techniques for brain-computer interfaces. 2004.

[170] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial eeg-analysis: from brain–computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167(1):82–90, 2008.

[171] Johannes Müller-Gerking, Gert Pfurtscheller, and Henrik Flyvbjerg. Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clinical neurophysiology*, 110(5):787–798, 1999.

[172] Gernot R Muller-Putz and Gert Pfurtscheller. Control of an electrical prosthesis with an ssvep-based bci. *IEEE Transactions on Biomedical Engineering*, 55(1):361–364, 2008.

[173] GR Muller-Putz, Reinhold Scherer, Christa Neuper, and Gert Pfurtscheller. Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE transactions on neural systems and rehabilitation engineering*, 14(1):30–37, 2006.

[174] Hideyuki Nakashima, Hamid Aghajan, and Juan Carlos Augusto. *Handbook of ambient intelligence and smart environments*. Springer Science & Business Media, 2009.

[175] Christa Neuper, Gernot R Müller-Putz, Reinhold Scherer, and Gert Pfurtscheller. Motor imagery and eeg-based control of spelling devices and neuroprostheses. *Progress in brain research*, 159:393–409, 2006.

[176] Miguel AL Nicolelis. Actions from thoughts. *Nature*, 409(6818):403–407, 2001.

[177] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams &

Wilkins, 2005.

[178] Femke Nijboer, Adrian Furdea, Ingo Gunst, Jürgen Mellinger, Dennis J McFarland, Niels Birbaumer, and Andrea Kübler. An auditory brain–computer interface (bci). *Journal of neuroscience methods*, 167(1):43–50, 2008.

[179] Anton Nijholt, Danny Plass-Oude Bos, and Boris Reuderink. Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment computing*, 1(2):85–94, 2009.

[180] Nozomu Nishikawa, Shoji Makino, and Tomasz M Rutkowski. Spatial auditory bci paradigm based on real and virtual sound image generation. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pages 1–5. IEEE, 2013.

[181] Paul L Nunez and Brian A Cutillo. *Neocortical dynamics and human EEG rhythms*. Oxford University Press, USA, 1995.

[182] Bernhard Obermaier, Christoph Guger, Christa Neuper, and Gert Pfurtscheller. Hidden markov models for online classification of single trial eeg data. *Pattern recognition letters*, 22(12):1299–1309, 2001.

[183] Bernhard Obermaier, Christa Neuper, Christoph Guger, and Gert Pfurtscheller. Information transfer rate in a five-classes brain-computer interface. *IEEE Transactions on neural systems and rehabilitation engineering*, 9(3):283–288, 2001.

[184] J Vernon Odom, Michael Bach, Colin Barber, Mitchell Brigell, Michael F Marmor, Alma Patrizia Tormene, Graham E Holder, et al. Visual evoked potentials standard (2004). *Documenta ophthalmologica*, 108(2):115–123, 2004.

[185] Seiji Ogawa, David W Tank, Ravi Menon, Jutta M Ellermann, Seong G Kim, Helmut Merkle, and Kamil Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 89(13):5951–5955, 1992.

[186] Ramaswamy Palaniappan. Brain computer interface design using band pow-

ers extracted during mental tasks. In *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pages 321–324. IEEE, 2005.

[187] Ramaswamy Palaniappan, Raveendran Paramesran, Shogo Nishida, and Naoki Saiwaki. A new brain-computer interface design using fuzzy artmap. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10(3):140–148, 2002.

[188] Jiahui Pan, Qiuyou Xie, Yanbin He, Fei Wang, Haibo Di, Steven Laureys, Ronghao Yu, and Yuanqing Li. Detecting awareness in patients with disorders of consciousness using a hybrid brain–computer interface. *Journal of neural engineering*, 11(5):056007, 2014.

[189] Xiaomei Pei, Dennis L Barbour, Eric C Leuthardt, and Gerwin Schalk. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of neural engineering*, 8(4):046028, 2011.

[190] Carlos Andrés Pena-Reyes and Moshe Sipper. Evolutionary computation in medicine: an overview. *Artificial Intelligence in Medicine*, 19(1):1–23, 2000.

[191] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[192] William D Penny and Stephen J Roberts. Eeg-based communication via dynamic neural network models. In *Neural Networks, 1999. IJCNN'99. International Joint Conference on*, volume 5, pages 3586–3590. IEEE, 1999.

[193] William D Penny, Stephen J Roberts, Eleanor A Curran, and Maria J Stokes. Eeg-based communication: a pattern recognition approach. *IEEE transactions on Rehabilitation Engineering*, 8(2):214–215, 2000.

[194] Gert Pfurtscheller and FH Lopes Da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

[195] Gert Pfurtscheller, Doris Flotzinger, and Joachim Kalcher. Brain-computer interfacea new communication device for handicapped persons. *Journal of Microcomputer Applications*, 16(3):293–299, 1993.

[196] Gert Pfurtscheller, C Neuper, C Guger, WAHW Harkam, Herbert Ramoser, Alois Schlogl, BAOB Obermaier, and MAPM Pregenzer. Current trends in graz brain-computer interface (bci) research. *IEEE Transactions on Rehabilitation Engineering*, 8(2):216–219, 2000.

[197] Gert Pfurtscheller and Christa Neuper. Motor imagery activates primary sensorimotor area in humans. *Neuroscience letters*, 239(2):65–68, 1997.

[198] Gert Pfurtscheller and Christa Neuper. Dynamics of sensorimotor oscillations in a motor task. In *Brain-Computer Interfaces*, pages 47–64. Springer, 2009.

[199] Terence W Picton. The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4):456–479, 1992.

[200] Anne K Porbadnigk, Jan-N Antons, Benjamin Blankertz, Matthias S Treder, Robert Schleicher, Sebastian Möller, and Gabriel Curio. Using erps for assessing the (sub) conscious perception of noise. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 2690–2693. IEEE, 2010.

[201] Anne K Porbadnigk, Simon Scholler, Benjamin Blankertz, Arnd Ritz, Matthias Born, Robert Scholl, Klaus-Robert Müller, Gabriel Curio, and Matthias S Treder. Revealing the neural response to imperceptible peripheral flicker with machine learning. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 3692–3695. IEEE, 2011.

[202] Pavel Pudil, Jana Novovičová, and Josef Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994.

[203] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[204] Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, 8(4):441–446, 2000.

[205] Rajesh PN Rao. *Brain-computer interfacing: an introduction*. Cambridge University Press, 2013.

[206] KVR Ravi and R Palaniappan. A minimal channel set for individual identification with eeg biometric using genetic algorithm. In *Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on*, volume 2, pages 328–332. IEEE, 2007.

[207] Boris Reuderink and Mannes Poel. Robustness of the common spatial patterns algorithm in the bci-pipeline. 2008.

[208] Siamak Rezaei, Kouhyar Tavakolian, Ali Moti Nasrabadi, and S Kamaledin Setarehdan. Different classification techniques considering brain computer interface applications. *Journal of Neural Engineering*, 3(2):139, 2006.

[209] Marko Robnik-Šikonja and Igor Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53(1-2):23–69, 2003.

[210] Giorgio Roffo and Simone Melzi. Features selection via eigenvector centrality. *Proceedings of New Frontiers in Mining Complex Patterns (NFMCP 2016)(Oct 2016)*, 2016.

[211] Tomasz M Rutkowski, A Cichocki, and Danilo P Mandic. Spatial auditory paradigms for brain computer/machine interfacing. In *International workshop on the principles and applications of spatial hearing*, pages 11–13, 2009.

[212] Emilio Salinas and Terrence J Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature reviews. Neuroscience*, 2(8):539, 2001.

[213] Wojciech Samek, Carmen Vidaurre, Klaus-Robert Müller, and Motoaki Kawanabe. Stationary common spatial patterns for brain–computer interfacing. *Journal of neural engineering*, 9(2):026013, 2012.

[214] Korn Saranyasoontorn, Lance Manuel, Paul S Veers, et al. A comparison of standard coherence models for inflow turbulence with estimates from field

measurements. *Transactions of the ASME-N-Journal of Solar Energy Engineering*, 126(4):1069–1082, 2004.

[215] Alois Schlögl, Felix Lee, Horst Bischof, and Gert Pfurtscheller. Characterization of four-class motor imagery eeg data for the bci-competition 2005. *Journal of neural engineering*, 2(4):L14, 2005.

[216] Caroline Schnakers, Audrey Vanhaudenhuyse, Joseph Giacino, Manfredi Ventura, Melanie Boly, Steve Majerus, Gustave Moonen, and Steven Laureys. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC neurology*, 9(1):35, 2009.

[217] Martijn Schreuder, Benjamin Blankertz, and Michael Tangermann. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PloS one*, 5(4):e9813, 2010.

[218] Ruth Schubert, Michael Tangermann, Stefan Haufe, Claudia Sannelli, Michael Simon, EA Schmidt, WE Kincses, and Gabriel Curio. Parieto-occipital alpha power indexes distraction during simulated car driving. *International journal of psychophysiology*, 69(3):214, 2008.

[219] Andrew B Schwartz, X Tracy Cui, Douglas J Weber, and Daniel W Moran. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron*, 52(1):205–220, 2006.

[220] Eric W Sellers and Emanuel Donchin. A p300-based brain–computer interface: initial tests by als patients. *Clinical neurophysiology*, 117(3):538–548, 2006.

[221] S Solhjoo and MH Moradi. Mental task recognition: A comparison between some of classification methods. In *BIOSIGNAL 2004 International EURASIP Conference*, pages 24–26, 2004.

[222] Soroosh Solhjoo, Ali Motie Nasrabadi, and Mohammad Reza Hashemi Golpayegani. Classification of chaotic signals using hmm classifiers: Eeg-based mental task classification. In *Signal Processing Conference, 2005 13th European*, pages 1–4. IEEE, 2005.

[223] Le Song and Julien Epps. Classifying eeg for brain-computer interface: Learning optimal filters for dynamical system features. *Computational intelligence and neuroscience*, 2007, 2007.

[224] Le Song, Alex Smola, Arthur Gretton, Karsten M Borgwardt, and Justin Bedo. Supervised feature selection via dependence estimation. In *Proceedings of the 24th international conference on Machine learning*, pages 823–830. ACM, 2007.

[225] Heung-Il Suk and Seong-Whan Lee. A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):286–299, 2013.

[226] Kevin Talbot and Martin R Turner. Oculomotor dysfunction in amyotrophic lateral sclerosis. *Arch Neurol*, 68(7):857–861, 2011.

[227] Durk Talsma and Marty G Woldorff. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of cognitive neuroscience*, 17(7):1098–1114, 2005.

[228] Kouhyar Tavakolian and Siamak Rezaei. Classification of mental tasks using gaussian mixture bayesian network classifiers. In *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, pages S3–6. IEEE, 2004.

[229] WA Teder-Sälejärvi, JJ McDonald, F Di Russo, and SA Hillyard. An analysis of audio-visual crossmodal integration by means of event-related potential (erp) recordings. *Cognitive Brain Research*, 14(1):106–114, 2002.

[230] David J Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

[231] Paul Tiesinga, Jean-Marc Fellous, and Terrence J Sejnowski. Regulation of spike timing in visual cortical circuits. *Nature reviews. Neuroscience*, 9(2):97, 2008.

[232] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, 1998.

[233] Christopher Torrence and Peter J Webster. Interdecadal changes in the enso–monsoon system. *Journal of Climate*, 12(8):2679–2690, 1999.

[234] Matthias S Treder and Benjamin Blankertz. (c) overt attention and visual speller design in an erp-based brain-computer interface. *Behavioral and brain functions*, 6(1):28, 2010.

[235] Vladimir Vapnik, Steven E Golowich, and Alex J Smola. Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems*, pages 281–287, 1997.

[236] Bastian Venthur, Benjamin Blankertz, Manfred F Gugler, and Gabriel Curio. Novel applications of bci technology: psychophysiological optimization of working conditions in industry. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 417–421. IEEE, 2010.

[237] François-Benoît Vialatte, Monique Maurice, Justin Dauwels, and Andrzej Cichocki. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in neurobiology*, 90(4):418–438, 2010.

[238] Jacques J Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

[239] Jacques J Vidal. Real-time detection of brain events in eeg. *Proceedings of the IEEE*, 65(5):633–641, 1977.

[240] Haixian Wang, Qin Tang, and Wenming Zheng. L1-norm-based common spatial patterns. *IEEE Transactions on Biomedical Engineering*, 59(3):653–662, 2012.

[241] Suogang Wang and Christopher J James. Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis. *Computational intelligence and neuroscience*, 2007, 2007.

[242] Yijun Wang, Shangkai Gao, and Xiaornog Gao. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5392–5395. IEEE, 2006.

[243] Yijun Wang, Zhiguang Zhang, Yong Li, Xiaorong Gao, Shangkai Gao, and Fusheng Yang. Bci competition 2003-data set iv: an algorithm based on cssd and fda for classifying single-trial eeg. *IEEE Transactions on Biomedical Engineering*, 51(6):1081–1086, 2004.

[244] Jason Weston, André Elisseeff, Bernhard Schölkopf, and Mike Tipping. Use of the zero-norm with linear models and kernel methods. *Journal of machine learning research*, 3(Mar):1439–1461, 2003.

[245] Langford B White and Boualem Boashash. Cross spectral analysis of non-stationary processes. *IEEE Transactions on Information Theory*, 36(4):830–835, 1990.

[246] Jonathan Wolpaw and Elizabeth Winter Wolpaw. *Brain-computer interfaces: principles and practice*. OUP USA, 2012.

[247] Jonathan R Wolpaw. Brain–computer interfaces as new brain output pathways. *The Journal of physiology*, 579(3):613–619, 2007.

[248] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

[249] Jonathan R Wolpaw and Dennis J McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854, 2004.

[250] Hau-Tieng Wu, Patrick Flandrin, and Ingrid Daubechies. One or two frequencies? the synchrosqueezing answers. *Advances in Adaptive Data Analysis*, 3(01n02):29–39, 2011.

[251] Yongxin Xi, Uri Hasson, Peter J Ramadge, and Zhen J Xiang. Boosting with spatial regularization. In *Advances in Neural Information Processing Systems*, pages 2107–2115, 2009.

[252] Yan Xu, Simon Haykin, and Ronald J Racine. Multiple window time-frequency distribution and coherence of eeg using slepian sequences and hermite functions. *IEEE Transactions on Biomedical Engineering*, 46(7):861–

866, 1999.

[253] Zenglin Xu, Irwin King, Michael Rung-Tsong Lyu, and Rong Jin. Discriminative semi-supervised feature selection via manifold regularization. *IEEE Transactions on Neural networks*, 21(7):1033–1047, 2010.

[254] Han Yuan and Bin He. Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61(5):1425–1435, 2014.

[255] Thorsten O Zander and Christian Kothe. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, 8(2):025005, 2011.

[256] Dan Zhang, Huaying Song, Honglai Xu, Wei Wu, Shangkai Gao, and Bo Hong. An n200 speller integrating the spatial profile for the detection of the non-control state. *Journal of neural engineering*, 9(2):026016, 2012.

[257] Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface. *Journal of neuroscience methods*, 255:85–91, 2015.

[258] QiBin Zhao, LiQing Zhang, and Andrzej Cichocki. Eeg-based asynchronous bci control of a car in 3d virtual reality environments. *Chinese Science Bulletin*, 54(1):78–87, 2009.

[259] Shunan Zhao and Frank Rudzicz. Classifying phonological categories in imagined and articulated speech. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 992–996. IEEE, 2015.

[260] Zheng Zhao and Huan Liu. Semi-supervised feature selection via spectral analysis. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 641–646. SIAM, 2007.

[261] Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M Aarts. A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, 2010:1, 2010.