Preprocessing and Feature Extraction for Asynchronous Multi-Class Noninvasive Brain Computer Interface based on EEG signal

by

Hamzah S. AlZu'bi

A Thesis Presented to the

Masdar Institute of Science and Technology

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in

Computing and Information Science

©2011 Masdar Institute of Science and Technology

All rights reserved

AUTHOR'S DECLARATION

I understand that copyright in my thesis is transferred to Masdar Institute of Science and Technology.

ACCEPTANCE DECLARATION

This thesis has been accepted and approved by Masdar Institute of Science and Technology on June 23, 2011.

EXAMINATION COMMITTEE MEMBERS

Wei Lee Woon, *Advisor, Masdar Institute of Science and Technology*

Jacob Crandall, *Masdar Institute of Science and Technology*

Zeyar Aung, *Masdar Institute of Science and Technology*

Abstract

A Brain Computer Interface (BCI) is a system which allows direct communication between the brain and a computer. It can be used to allow paralyzed as well as healthy individuals to interact with and control the surrounding environment or to communicate simply by the conscious modulation of thought patterns. Although Noninvasive Electroencephalogram-based (EEG-based) BCI is showing a lot of promise, it is faced with a number of difficult challenges, especially from the perspective of signal processing, because the signals being observed by EEG are extremely weak and typically contain very high levels of noise.

The asynchronous multiclass BCI problem is particular importance because it closely matches realistic operating conditions (as opposed to synchronous problems). However, it is a challenging problem and requires the development of appropriate machine learning and signal processing tools.

This thesis develops an asynchronous multiclass noninvasive EEG-based BCI system which is based on a novel feature extraction method for asynchronous BCI. A set of prepossessing, feature extraction and classification methods were implemented to help support and validate the proposed system and to benchmark it against the latest developed systems in BCI literature.

The proposed system is tested on two separate datasets. The first is a well-known and publicly available dataset, while the second dataset was collected locally using a retail EEG kit.

The developed system showed robust and accurate classification results for both datasets.

This research was supported by the Government of Abu Dhabi to help fulfill the vision of the late President Sheikh Zayed Bin Sultan Al Nayhan for sustainable development and empowerment of the UAE and humankind.

Acknowledgments

First of all, my thanks go to ALLAH who gives me strength and opportunity to complete this work successfully.

I would like to express my gratefulness to all people who helped me achieve this work.

My deep thanks and appreciations go to my friendly advisor, Dr. Wei Lee Woon, who supervised and guided this work; his advices and suggestions have been invaluable. Thanks to my RSC committee members, Dr. Jacob Crandall and Dr. Zeyar Aung, for their valuable comments.

I would like to thank my friends and colleagues in Masdar Institute: Abdullah Sawas and Asma Al Rashedi for the interesting and constructive discussions; Abdel-Halim Hiassat, Husam Suleiman and John Sitler for their participation in our BCI experiment; Sue Tan for providing a powerful workstation.

My deepest gratitude goes to my lovely family: My mother for her pleasant words and endless prayers, great brother Najeh for his great support, wonderful brother

Khaled for his enduring care, beloved brother Tayel for the pleasant moments, my brilliant brother Nayel for all endless support and great inspiration, my lovely sister Khulood for her care and prayers and for my awesome sister Taghreed for her fantastic talks.

To all I would like to say, **Thank you very much.**

Hamzah S. AlZu'bi

Masdar City

June 2011

Contents

1 Introduction **1**

1.1 Understanding the Brain 1

1.2 Brain Computer Interface 3

1.2.1 Introduction 3

1.2.2 Signal Processing and Machine Learning 4

1.3 Challenges and Objectives 6

1.4 Thesis Layout 7

2 Background **8**

2.1 Brain Science 9

2.1.1 Human Brain Anatomy 9

2.1.2 Neuroimaging 11

2.1.3 Review of Brain Computer Interface research 14

2.1.4 Publicly available BCI data: the BCI competition 16
2.2 Signal Processing and Machine Learning Algorithms 19
2.2.1 Preprocessing techniques 19
2.2.1.1 Temporal Filtering 20
2.2.1.2 Common Spatial Patterns (CSP) 23
2.2.2 Feature Extraction Techniques 25
2.2.2.1 Power Spectral Density (PSD) 25
2.2.2.2 Log Variance 27
2.2.2.3 Wavelet Packet Decomposition (WPD) 28
2.2.2.4 Mel Frequency Cepstral Coefficient (MFCC) . . 30
2.2.3 Classifiers 31
2.2.3.1 Random Forests (RF) 32
2.2.3.2 Linear and Quadratic Discriminant Analysis (LDA and QDA) 33
2.2.3.3 Artificial Neural Networks (ANN) 35
2.2.3.4 Support Vector Machines (SVM) 38
2.3 Summary 44

3 Implementation and Methodology 46
3.1 Data Description 46
3.2 Preprocessing 49
3.2.1 Temporal Filtering 49
3.2.2 Common Spatial Patterns (CSP) 50
3.3 Feature Extraction 55
3.3.1 Power Density Estimation (PSD) 55

3.3.2 The Log Variance 56
3.3.3 Wavelet Packet Decomposition (WPD) 58
3.3.4 A Novel Feature Extraction Technique 59
3.4 Summary 62

4 Results and Discussion **63**
4.1 The Experimental System 64
4.2 Evaluation criteria 65
4.3 BBCI competition IV Dataset I 67
4.3.1 Evaluation Process 67
4.3.2 Results 68
4.3.2.1 The Window Size Variation 70
4.3.2.2 Classifiers Variation 71
4.3.2.3 Result Validation 73
4.3.2.4 Computational requirements 75
4.4 Tests using internally collected data 76
4.4.1 Experimental setup 77
4.4.2 Results 79
4.5 Summary 82

5 Conclusion and Future Work **83**
5.1 Conclusion 83
5.2 Future Work 85

A The Emotiv System **87**

B Multidimensional two-sample Kolmogorov-Smirnov test **91**

C Principal Component Analysis **95**

D Abbreviations **96**

List of Tables

2.1 Frequency ranges of the brain rhythms. 14
2.2 A summary of the BBCI competitions and the used techniques by the competitors based on http://www.bbci.de/activities. 17
2.3 Cepstrum Terminology. 32
2.4 Famous Kernal Functions. 44

3.1 The number of selected CSP filters for each pair of tasks. 55

4.1 Averaged confusion matrix for every subject. 4-fold cross-validation on both sessions is used here. 70
4.2 Average accuracy comparison between the implementations given in [2, 37, 36] and our implementation. 4-Fold cross-validation on the first session of BBCI competition IV dataset I was used in both implementations. 76
4.3 Basic information about the three subjects. 77

B.1 P-Value of the K-S Test of each dimensional order. 94

List of Figures

1.1 Schematic of a typical Brain Computer Interface system. 5

2.1 Neuron Structure [from Sanei and Chambers [70]]. 9

2.2 Functional areas of human brain [from Encyclopædia Britannica [16]]. 11

2.3 Motor and Sensory cortex areas [from Rhawn Joseph [43]]. 12

2.4 Electrophysiological activities used in BCI and BBCI competition. 16

2.5 IIR digital filter block diagram. 21

2.6 FIR digital filter block diagram. 21

2.7 FIR in frequency domain. 22

2.8 Some famous wavelet families. 29

2.9 Discrete Wavelet Transform. 30

2.10 MFCC block diagram. 31

2.11 Mel-Scale Function. 32

2.12 Random Forest Classifier. 33
2.13 General Artificial Neural Network Structure. 36
2.14 Linear SVM. 40
2.15 Mapped data using $\kappa\left(x_i, x_j\right) = x_i^2 + x_i^2$. 44

3.1 2D top view of the scalp with the electrodes locations map. 48
3.2 Tasks Timing in Calibration and Evaluation sessions. 49
3.3 The spectra of the EEG signal from C3 channel for two imagery tasks (left/foot) with $r^2 - value$ before IIR filter (subject a). 51
3.4 The spectra of the EEG signal from C3 channel for two imagery tasks (left/foot) with $r^2 - value$ after IIR filter (subject a). 52
3.5 Selected Common Spatial Filters and Patterns for $Left\ hand$, $Right\ hand$ and $Rest$ tasks for subject g. 53
3.6 Spectra for the spatially filtered signal with Laplace and CSP spatial filters and the $r^2 - value$ between two tasks. 54
3.7 Average Log power spectrum for EEG signal of Left hand task trials and Average log band power distribution over scalp surface. 57
3.8 Wavelet packets in different levels using Daubechies order 4. . . . 58
3.9 Wavelet Features extracted from a selected EEG channel using 3 levels WPD. The red and green windows represent averaging windows. Light and dark colors represent high and low WPD value respectively, as the color bar suggests. 59

4.1 System block diagram. 65

4.2 4-Fold cross-validation accuracy for all subjects using LDA classifier, both sessions were used. The 10^5 variance for the 4-fold is presented also. 69
4.3 4-Fold cross-validation MSE for all subjects using LDA classifier, both sessions were used. 71
4.4 Mean 4-Fold cross-validation accuracy with a varying window sizes. 72
4.5 Mean 4-Fold cross-validation mean Square error (MSE) with a variety window sizes. 73
4.6 Average cross-validation accuracy using four classifiers and MFCC feature extraction. Only first session is used. 74
4.7 Average cross-validation mean square error (MSE) using four classifiers and MFCC feature extraction. Only the first session is used. 75
4.8 Average computational time (in seconds) needed by BCI system using 4-fold cross-validation and both data sessions of all subjects which include for subjects $a, b, f and g$ around 52 min (314517 sample), 54 min (325837 sample), 52 min (317478 sample) and 53 min (322394 samples) , respectively. 77
4.9 The three subjects are performing experiment. 78
4.10 The four screens from the first session of the experiment. The white screens with red rectangle request from subjects to perform left and right hand movement, the black and cross screens request from subjects to relax. 79

4.11 Topographies maps for subject-A. Upper-left map shows average power distribution of all trials, upper-right map shows average power distribution of all left trials, lower-left map shows average power distribution of all rest trials and lower-right shows average power distribution of all right trials. 80
4.12 Accuracy of all feature extraction methods from the collected data of all volunteers. 4-fold cross-validation was performed on the data from both sessions. 81
4.13 MSE of all feature extraction methods from the first session of collected data of all volunteers. 4-fold cross-validation was used. . 82

A.1 EPOC electrodes map. 88
A.2 Subject-J is wearing EPOC headset and trying to perform some mental tasks. 89
A.3 a snapshot of the control panel window. 90
A.4 a snapshot of the TestBench window. 90

B.1 CDF of two random variables of different distributions. 92

CHAPTER 1

Introduction

1.1 Understanding the Brain

Understanding the functioning of the brain is one of the biggest mysteries in life and amongst the greatest challenges to science [83]. Physically, the human brain is three pounds (1.5Kg) of grey matter and contains around 10 billion neurons. However, this outwardly unimpressive organ is responsible for human awareness, emotions and sensations [27, 66]. Neuroscientists are trying to discover how these neurons work together and communicate via dense interconnections, and how these eventually result in the phenomenon of consciousness. By carefully studying brain activation patterns, functional maps of the brain have been developed in which the surface of the cortex has been partitioned into a set of functionally

distinct regions. In general, however, neuroscience is a hugely complicated area of research which is highly interdisciplinary in nature. There is little doubt that a complete understanding of the workings of the brain can only be achieved via the coordinated action of researchers from the fields of clinical neurophysiology, psychology, mathematics, physics and computer science. As it currently stands, a number of partial theories have been proposed, but the inner details of brain function are still very poorly understood.

Another important aspect of neuroscience is brain imaging. Scientists monitor brain activity to identify and characterize the different areas of the brain. Identifying the roles played by each of these areas will lead to more accurate diagnosis of mental disorders [27], facilitate the construction of more insightful brain simulations [69], permit the efficacy of medical interventions to be studied more effectively [58] and enable different modalities of brain machine communications. Thus, brain imaging is a tool which can fundamentally change the way in which neuroscientists study the intricacies of the brain [35]. Because each brain imaging technique measures a different aspect of neuronal function, each is better suited to a different set of applications, e.g., the detection of pathologies, functional imaging and brain computer interfaces. The usefulness of a scanning technique to a particular application hence depends on the purchase and operating costs of that technique, spatial and temporal resolutions and also on other factors such as portability and robustness [61].

1.2 Brain Computer Interface

1.2.1 Introduction

In the first international meeting for Brain Computer Interface technology (1999), the following definition for Brain Computer Interfaces was proposed: *"A brain-computer interface is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles"* [84].

Using only thoughts to control the surrounded environment will provide totally paralyzed people with a powerful tool which will not rely on peripheral nerves or on motions. Most of the movement aids currently in use by paralyzed or motion-impaired individuals depend on minute movements and as such are unable to help patients who are completely paralyzed or "locked in". A well known example is ALS, also known as Lou-Gehrig's disease [26, 53], where the patient has full brain functionality but there is a progressive neurodegenerative disease caused by degeneration of motor neurons. Around 20,000 to 30,000 people have ALS and 5,000 people are diagnosed with ALS every year just in the USA [59].

The analysis of brain signals using noninvasive techniques is a very challenging problem and its solution will depend on techniques drawn from the fields of machine learning and signal processing, as well as from the field of neuroscience. It is also an enabling technology for BCI, where the achievable information transfer rates for BCI directly rely on the ability to extract recognizable patterns from recorded brain signals. This thesis seeks to address this challenge by focusing on the development of improved and novel BCI signal processing algorithms.

1.2.2 Signal Processing and Machine Learning

The successful deployment of BCI will depend on the availability of effective signal processing algorithms to enhance and extract relevant aspects of the acquired brain signal and of appropriate machine learning techniques to interpret the brain rhythms and other features embedded in these signals and to associate them with specific brain activities or intent. Generally, a BCI system contains four processing stages as shown in figure 1.1:

1. Signal Preprocessing: In this phase, the aim is to enhance the quality of the recorded brain signal and to prepare it for subsequent processing stages. Examples of techniques used in this stage are:

    - Temporal Filtering: This signal processing technique is used to extract the frequency bands that are associated with the target mental tasks [31].
    - Common Spatial Patterns: A supervised method maximizes the separation between classes by increasing the variance of a certain class and subsequently, minimizing the variance associated with the other class (or classes) [33].
    - Independent Component Analysis: A statistical method that aims to remove interference between multivariate signals to have new statistically independent components [21, 81].

2. Feature Extraction: This is a critical component of any BCI algorithm. The aim of this stage is to identify and generate a set of representative features which target specific aspects of brain activity. In the BCI literature, a wide

variety of features have been proposed. Some of the features which are commonly used in the context of EEG analysis are:

- Signal Variance, which is widely used due to its simplicity and compatibility with the above-mentioned CSP technique. [33].
- Power Spectral Density (PSD) features, where the energy content in various frequency bins are estimated [28, 18, 52].
- Wavelet Packet Decomposition generates a feature set which is composed of the coefficients of the wavelet packet decomposition [78] .

3. Classification: Classifiers seek to identify patterns from the data based on statistical models built using the training data. Linear Discriminant Analysis [29], Support Vector Machine [13] and Artificial Neural Network [7, 4] are the classifiers which are most commonly used for BCI [84].

4. Post processing and Translation: This stage depends on the specific problem being addressed and the goal of the application [20].

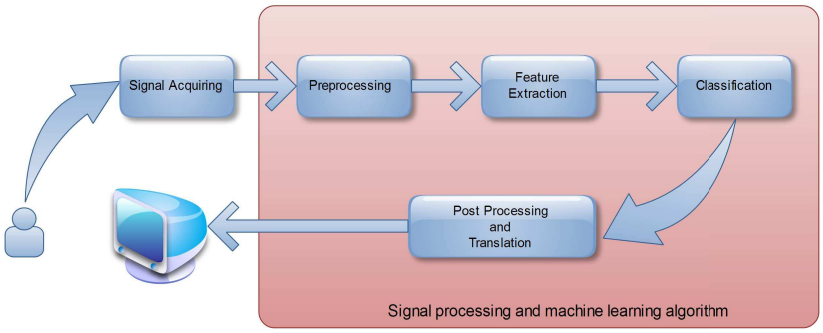


Figure 1.1: Schematic of a typical Brain Computer Interface system.

1.3 Challenges and Objectives

Raw EEG recordings have very low SNRs, and are susceptible to contamination from a variety of noise sources; examples include electrical activity associated with eye blinks and other movements, heartbeats, mains hum and thermal noise. A further issue is the poor spatial resolution; this problem is an inherent limitation of the EEG recording process, which measures potentials occurring at the scalp which would have had to diffuse through multiple layers of intervening skin, muscle and bone. The main challenge for the BCI system is to extract useful information about the targeted mental event from noisy measurements of the underlying brain activity. Given the challenges mentioned above, this can only be achieved if great care is taken to select and apply appropriate noise reduction and pattern processing techniques. The second challenge for BCI is to interpret the measured EEG signal and to determine relationships between this signal and underlying mental activity and events. Neuroscientists have developed many theories to explain brain function and to interpret the meaning of brain signals. Another challenge for BCI is the high dimensionality of the extracted feature vectors.

The research described in this thesis was focused on the signal processing and machine learning side of BCI. More specifically our goal will be to implement and benchmark different feature extraction techniques, to combine them with different classification methods, and finally to propose a new EEG feature extraction method. Towards this end, the following specific objectives have been identified:

1. **Implement a selection of commonly used BCI feature extraction methods.**

2. **Propose a novel technique for BCI feature extraction aimed at address-**

ing some of the shortcomings of the conventional approaches.

3. **Test the complete system using a variety of different classifiers to ensure broad applicability.**

4. **Validate the developed methods using two separate datasets**. The first dataset is publicly available, which should allow us to benchmark our results against previous studies; the second set is an internally collected dataset, which was recorded using a basic, low-cost EEG device.

1.4 Thesis Layout

The rest of the thesis will be organized as follows. Chapter 2 provides an introduction of brain function and signals, the fundamentals of BCI and also the relevant machine learning and signal processing techniques. Chapter 3 describes the implementation of the methods in detail. The procedures for data collection are also explained. Chapter 4 presents and analyzes the results. In particular, the proposed methods are evaluated in light of these results. Finally, chapter 5 concludes the thesis by discussing the main findings and contributions in terms of the original aims and objectives of the project.

CHAPTER 2

Background

In this chapter, key theoretical concepts of brain science, signal processing and machine learning are presented. The first section discusses the physiology of the human brain as well as the functions of its main structures. An overview of widely used neuroimaging techniques is also provided, and describes how these techniques have been used in the development of Brain Computer Interfaces (BCI). The second part of this chapter reviews the latest BCI studies; the BBCI competition, which is the source of much of our data, is also introduced. Finally, a review of the most commonly used pre-processing, feature extraction and classification methods used in BCI is also provided.

2.1 Brain Science

2.1.1 Human Brain Anatomy

The neuron is the basic processing unit in the central nervous system (CNS). Figure 2.1 shows the basic structure of the neuron.

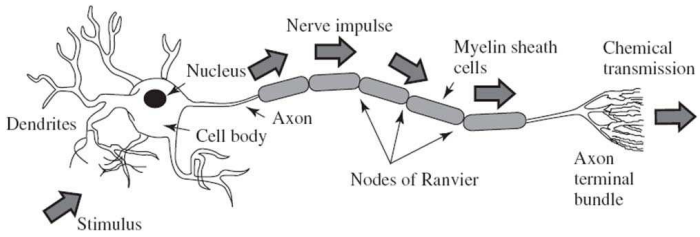


Figure 2.1: Neuron Structure [from Sanei and Chambers [70]].

The three main parts of the neuron are the cell body, axon and dendrites. Neurons are differentiated from other cell types by the presence of axons and dendrites, which are specialized structures for carrying information and which permit the formation of a network of neurons, i.e., the neural network. Information is transmitted via signals which travel along the axons and dendrons; these signals are carried by the propagation of changes in the electrical potential of the axons' membranes, also known as the "action potential" (AP). To generate this potential, special channels are used to pump the positive ions (K^+, Na^+ and Ca^{+2}) and negative ions (Cl^-) between the membranes' interior and exterior. For the human brain this AP can range between -60mV and 10mV [27].

An active neuron produces electrical potential within its dendrites which generates an ionic current that flows in and out of the neuron [55]. This current

generates a magnetic field around the neuron; while the field thus generated is too minute to be measured directly, the aggregate field generated by the synchronized firing of thousands of neurons is strong enough to be measured by Magnetoencephalography (MEG). Similarly, the APs resulting from the action of thousands of neurons can be measured at the surface of the scalp using a technique known as Electroencephalography (EEG).

The human brain contains approximately 10^{10} neurons. In addition, these neurons are massively interconnected via a dense network of synapses to form a large neural network. An adult human brain would have around 5×10^{14} synapses; this number continually increases with age, while conversely the number of neurons will decrease.

Neuroscientists use the term "higher cortical functions" to refer to all conscious mental activities, e.g., thinking, remembering, volunteer movements, etc. Although the detailed workings of the brain are still poorly understood, there are models which spatially map the different cortical functions to different regions of the cerebral cortex [22]. In general, while it is difficult to say if such models are 100% accurate, it is, at least, reasonably clear that different functions and mental states are associated with different regions of the brain, and those complex mental activities can be spatially distributed across the cortex. Figure 2.2 shows an example of a functional map of the brain. The cortex is often thought of as consisting of three main parts: motor areas responsible of volunteer movements, sensory areas interpret the collected information by the sensory organs and association areas are involved in the conscious analysis of this information[75, 73]. The motor and somatic (touch-sensitive) regions of the cortex are frequently referred to as the sensorimotor area. A map of this area of the cortex is shown in figure 2.3.

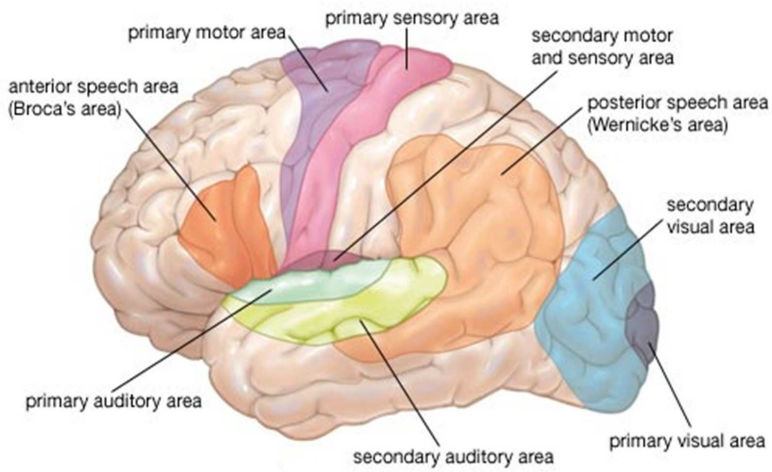


Figure 2.2: Functional areas of human brain [from Encyclopædia Britannica [16]].

2.1.2 Neuroimaging

Physiological activities in the human body, including those occurring in the brain, can be directly measured via electrophysiological signals such as those caused by the aforementioned action potentials. These include electrocardiography (ECG, heart), electroencephalography (EEG, brain), electromyography (EMG, brain and muscular system), magnetoencephalography (MEG, brain), electrogastrography (EGG, stomach) and electrooptigraphy (EOG, eye dipole field). Neuroscientists use a variety of sensing methods to measure brain signals. Some of methods, which are commonly used, are EEG (invasive and noninvasive), magnetoencephalography (MEG), positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and functional Near InfraRed (fNIR) [71].

The three techniques which are used to measure brain activity (as opposed to brain structure) are MEG, fMRI and EEG. In the context of BCI research, each of these methods has its own unique advantages and disadvantages. A brief descrip-

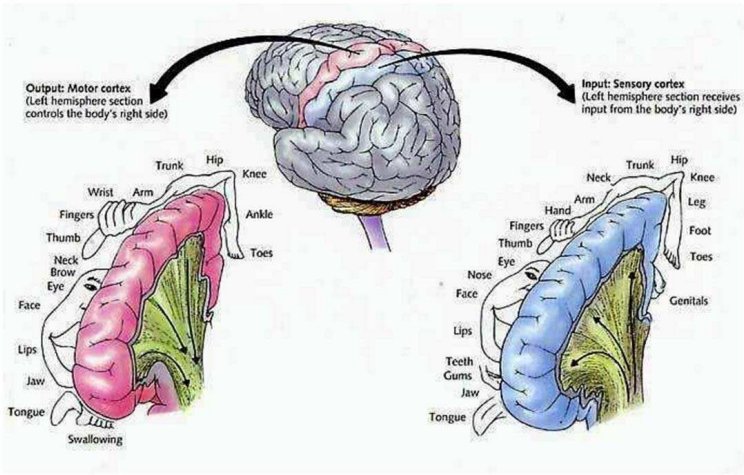


Figure 2.3: Motor and Sensory cortex areas [from Rhawn Joseph [43]].

tion for each method is provided below:

- MEG maps brain activities by recording magnetic fields produced by the electrical activities in the brain. MEG requires expensive and intensive low noise amplifier called superconducting quantum interference device (SQUID), moreover the measurements are sensitive to ferromagnetism therefore MEG equipments should be isolated inside Magnetically Shielded Room (MSR) where MSR will isolate SQUID from all external magnetic field even Earth's magnetic field which is billion time stronger than the raw MEG. MEG is known for having very high temporal and spatial resolution and and can be useful for studying activities which take less than 10 milliseconds [54, 25]. Unfortunately, in terms of BCI, MEG has two very serious problems. Firstly, it is extremely expensive, with MEG devices often costing hundreds of thousands of dollars or more. Secondly, MEG devices are very big and are not suitable for ambulatory applications such as BCI.

- fMRI uses nuclear magnetization of the hydro atoms in the fluids, mainly the blood, to align a powerful magnetic field. Because fMRI depends on the fluids movements inside the body tissues it will be more useful for slow events around several hundred milliseconds [55]. Because of this and other reasons, fMRI is rarely used for BCI.

- EEG signals are obtained by recording fluctuations in the local electric potentials on the surface of the scalp, where it is assumed that these fluctuations originate from the underlying human brain activity. Although EEG contains more noise, EEG signal has low SNR, than MEG and fMRI, EEG is the most used techniques in BCI which represents more than 80% of BCI published work [5] where EEG has very low setup cost and is also very portable. [79]. The EEG rhythm contains a lot of interesting information. For example, each frequency band of the EEG signal is associated with certain brain activities. Neuroscientists have associated each of these frequency bands with a specific set of mental activities or states. Table 2.1 lists the commonly recognized frequency bands, and the respective associated mental activities.

The above mentioned neuroimaging methods have been intensively investigated for BCI research. However, this thesis will focus exclusively on EEG, which is the neuroimaging techniques most commonly used for BCI. It will hence be the focus of subsequent sections.

Name	Frequency Range (Hz)	Mental Activity
alpha (α)	8-13	Relaxation and concentration
theta (θ)	4-7.5	Unconscious, inspiration and deep meditation
beta (β)	14-26	Thinking, attention and solving concrete problems
delta (δ)	0.5-4	Deep sleep
gamma (γ)	30-45	Finger movement, toes and tongue

Table 2.1: Frequency ranges of the brain rhythms.

2.1.3 Review of Brain Computer Interface research

Existing BCI research is based on one of six electrophysiological activities of the brain as shown in figure 2.4. A brief description of each of these activities is provided below:

1. Sensorimotor Activities

   - Brain Rhythms (μ, β, γ) : μ (8-12 Hz) and β (13-30 Hz) rhythms appeared when motor activity process is performed by the subject [41]. Any voluntary movement causes a circumscribed desynchronization in both μ and β rhythms. This circumscribed desynchronization is called Event-Related Desynchronization (ERD). ERD usually starts about 2 seconds prior the movements. Another relevant phenomena in BCI is Event-Related Synchronization (ERS) where β rhythm becomes at the maximum level after about 600 millisecond after the movement [8].

   - Movement Related Potential (MRP): low frequencies start about 1 to

1.5 seconds prior the movement. As closer to the movement as they become contralaterally preponderant [86].

2. Slow cortical potential (SCP): slow potential shifts with frequencies below 2 Hz and can be exist for several seconds [39].

3. P300: is Event-Related Potential (ERP) which has a positive potential peak at about 300 milliseconds after receiving the stimulus. P300 is the widely used with oddball paradigm framework [23].

4. Visual evoked potential (VEP): a low potential is generated as a response to a visual stimulus where its properties depends mainly on the stimulus source, e.g., if the stimulus is presented continually at 5-6 Hz rate then a continuous oscillatory potential effects the visual pathways.

5. Response to mental tasks: activities such as counting, multiplication, imagination, mental object rotation, etc., has certain EEG pattern. Which can be used to recognize mental tasks [47].

6. Activity of neural cells (ANC): the firing rate of neurons depends on the relation between the movement direction and the preferred direction. The firing rate increases when the real movement matches the preferred direction and decreases if not. Such phenomena is used in BMI [24].

While each of the BCI modalities listed above has each own unique properties and advantages, this thesis will focus only on sensorimotor based BCI algorithms (item 1). As such, all subsequent discussions will deal only with this form of BCI.

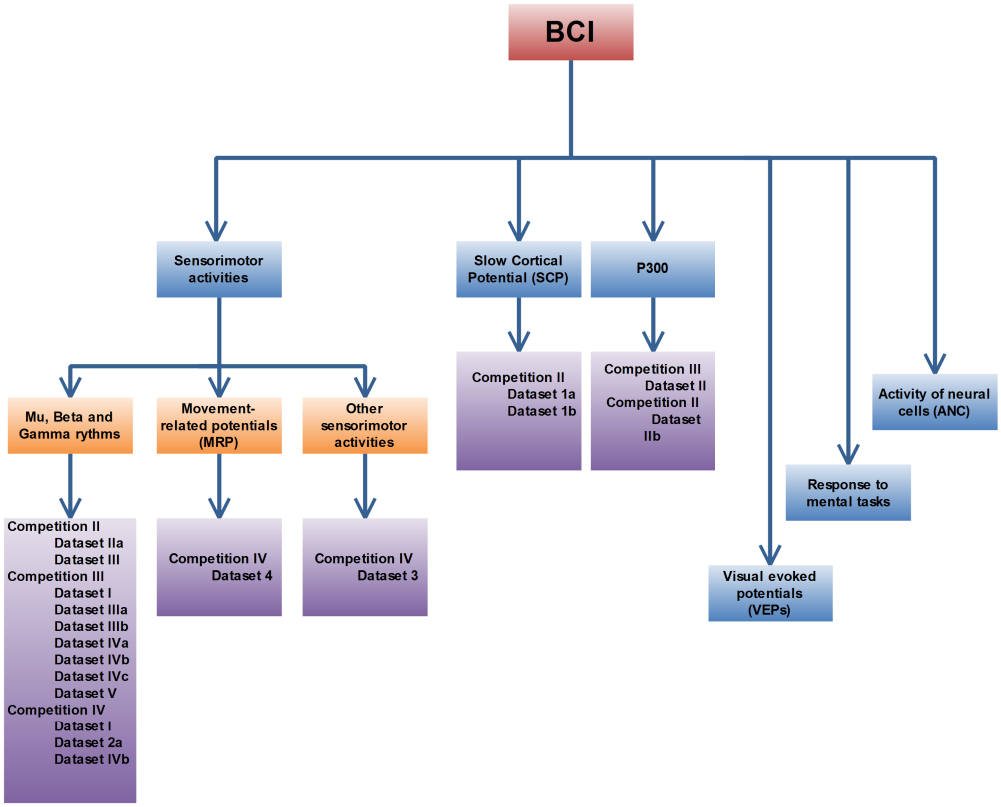


Figure 2.4: Electrophysiological activities used in BCI and BBCI competition.

2.1.4 Publicly available BCI data: the BCI competition

Berlin Brain Computer Interface [1] is one of the leading research groups in Brain Computer Interface in the world. Berlin Brain Computer Interface focuses on understanding the brain waves using modern machine learning techniques. The Berlin Brain-Computer-Interface is a result of a collaboration of several German research organizations [11]. Some of the main organizations are:

- The Institute Computer Architecture and Software Technology of the Fraunhofer Society

- The research group Intelligent Data Analysis (IDA)
- The Neurophysics Research Group
- The department of Neurology at the Campus Benjamin Franklin of the Charité University Medicine
- The Technical University Berlin

Berlin Brain Computer Interface held brain computer interface competitions to motivate brain computer interface groups around the world to find solution for some problems facing brain computer interface. Four competitions have been released to date, Table 2.2 shows a summary for the all BBCI problems and submitted solutions. Figure 2.4 shows the electrophysiological activities used in each competition.

BCI Competition II (2003)						
Dataset	Channels	CSP	AAR	Wavelets	Others	Winner
Dataset III	3	0	4	1	4	Wavelets
BCI Competition III (2005)						
Dataset	Channels	CSP	AAR	Wavelets	Others	Winner
Dataset I	64	3	7	2	17	CSP
Dataset IIIa	60	2	0	0	1	CSP
Dataset IIIb	2	0	2	1	4	Other
Dataset IVa	118	7	2	2	6	CSP,AAR
Dataset IVc	118	2	0	1	4	CSP
BCI CompetitionIV (2008)						
Dataset	Channels	CSP	AAR	Wavelets	Others	Winner
Dataset I	64	9	1	1	14	CSP
Dataset IIa	22	5	0	0	0	CSP
Dataset IIb	3	3	0	1	2	CSP

Table 2.2: A summary of the BBCI competitions and the used techniques by the competitors based on http://www.bbci.de/activities.

This thesis uses the dataset available in competition IV, latest competition, the first problem set. The goal of the first problem of competition IV was to develop asynchronous multi-class EEG based BCI system. The possible applications of the proposed system for such problem are on computer supported workplaces, control a cursor via thoughts and more importantly a useful tool for paralyzed or paraplegic people. The Berlin Brain Computer Interface project aims to motivate researchers to create efficient algorithms to learn brain waves patterns. Thus such efficient and stable algorithm needs to be developed to classify brain waves.

Most of the BCI problems focus on fixed trial window sizes, i.e., subject has to perform mental task within a fixed time frame. Those methods are called synchronous BCI. A more natural human machine interface is called asynchronous (self-paced) BCI, where the user has the control about the tasks and the timing, i.e., s/he controls the type, the start time and the end time of task [80]. Another important problem that is targeted by this dataset is the ability to detect no control status for the user and embed the quality for this detection in the evaluation data by certain weight. Berlin Brain Computer Interface provided a full description for experiment setup. Totally, four healthy subjects sat for this experiment and their brain waves were recorded. In each session each individual subject performed different motor imagery mental tasks without any feedback. Before starting the experiment, each subject selects two motor imagery mental tasks out of three different motor imagery mental tasks, left hand, right hand and foot. The experiment has two main sessions: Calibration session to collect the training set of EEG and Evaluation data to collect testing set of EEG to provide an evaluation for the used algorithm. Calibration Session: For the first two runs in this session, an arrow is pointing to different directions (left, right or down) to guide the subject to per-

form certain imagery mental task related the arrow visual cue. Each visual cue was shown from 4 seconds to let the subject perform the imagery mental task. This visual cue followed by 2 seconds of black screen. After the black interleave screen, a 2 seconds fixation cross will be shown in the center in the computer screen. The recorded EEG signal and the visual cues time were provided. Evaluation Session: after the calibration session ends, the evaluation session takes place with four runs. Every subject was asked to perform a certain motor imagery task by soft acoustic stimuli, e.g., English words left, right and foot. Each motor imagery task length varies for each try bounded by 1.5 seconds minimum and 8 seconds maximum. By the end of each imagery mental task a word *stop* indicated the start for intermit periods which varies randomly between 1.5 to 8 seconds.

2.2 Signal Processing and Machine Learning Algorithms

BCI algorithms consist of different stages as shown in figure 1.1. In this section we will review some of the algorithms used during each stage.

2.2.1 Preprocessing techniques

To prepare the EEG signal for feature extraction and classification, different processes need to be done, e.g., Infinite impulse response filter, Independent component analysis and Common spatial pattern filter. The most common preprocessing methods are as follows:

2.2.1.1 Temporal Filtering

As suggested in the neuroscience literature, different forms of mental activity are associated with the occurrence of specific brain rhythms, which may subsequently be observable via EEG recordings [64]. Mental tasks like motor sensory cause a certain pattern for Mu and Beta rhythms in the EEG signal. For EEG-based BCI studies that detect particular motor sensory activities, one common pre-processing step is to apply band-pass filtering techniques to isolate frequency components in the mu and beta ranges [57]. This is useful for eliminating very low and high frequency artifacts. Another benefit of using temporal filtering is the removal of the DC level which means the filtered EEG signal will be centered at the origin, e.g., zero mean. Infinite Impulse Response and Finite Impulse Response are the two main kinds of temporal filters. In the following we will discuss both of them briefly.

- Infinite Impulse Response filter: An IIR filter is a filter that has infinite impulse response where it contains feedback as shown in figure 2.6. Equation 2.1 shows the output $y[n]$ in term of the input and both feedforward and feedbackward coefficients.

$$y[n] = \frac{1}{a_0}\left(\sum_{i=0}^{P} b_i x[n-i] - \sum_{j=1}^{Q} a_j y[n-j]\right) \tag{2.1}$$

  The transfer function for an IIR filter is given by z-transform.

$$H(z) = \frac{\sum_{i=0}^{P} b_i z^{-i}}{\sum_{j=0}^{Q} a_j z^{-1}} \tag{2.2}$$

- Finite Impulse Response Filter: An FIR filter is a filter which has a finite

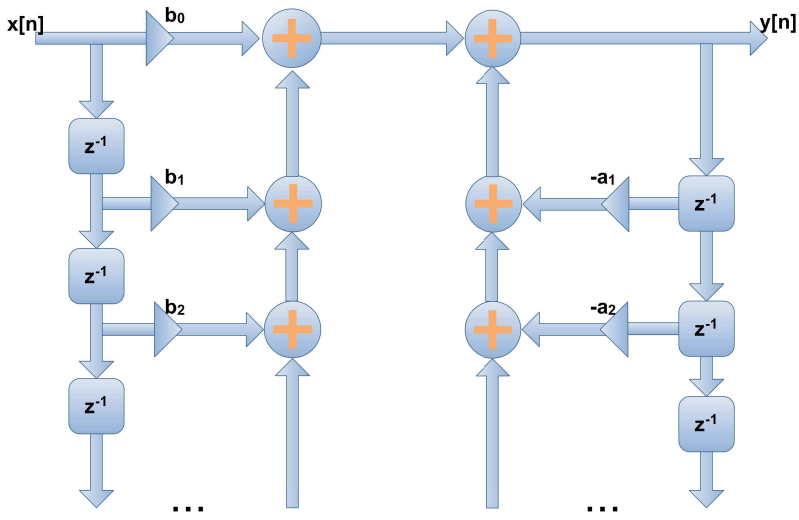


Figure 2.5: IIR digital filter block diagram.

duration as an impulse response because FIR doesn't contain any feedback as shown in figure 2.6 below.

The impulse input fires the output $y[n]$, equation 2.4, for $[N+1]$ samples after that $y[n]$ goes to zero.

$$y[n] = b_0 x[n] + b_1 x[n-1] + \ldots + b_N x[n-N] \tag{2.3}$$

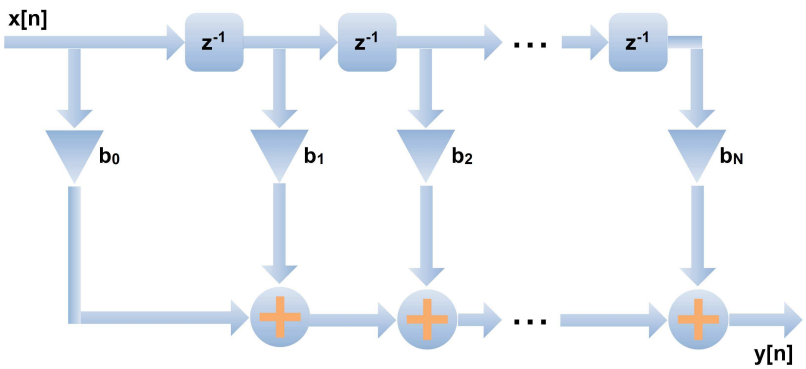


Figure 2.6: FIR digital filter block diagram.

$$y[n] = \sum_{i=0}^{N} b_i x[n-i] \tag{2.4}$$

where, b_i is a filter coefficients and N is the filter order.

The transfer function $H(z)$ for FIR filter is given by the z-transform of the impulse response.

$$H(z) = \sum_{i=-\infty}^{\infty} b_i z^{-i} = \sum_{i=0}^{N} b_i z^{-i} \tag{2.5}$$

The main types of IIR and FIR filter are based on the cutoff frequency. Those types are Low-pass filter, High-pass filter, Band-pass filter, Band-stop filter and By-pass filter. Figure 2.7 shows the frequency response for all types of filter types.

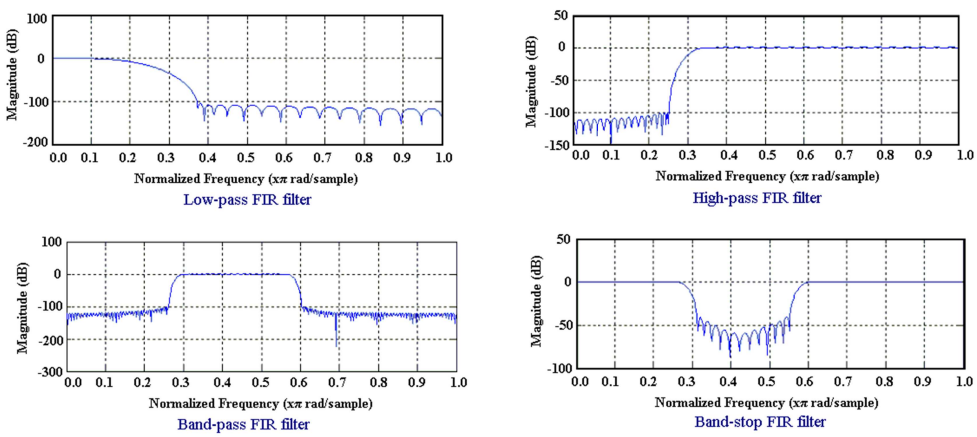


Figure 2.7: FIR in frequency domain.

2.2.1.2 Common Spatial Patterns (CSP)

Common Spatial Patterns is a supervised method for building spatial filters, which will result in a set of "virtual channels" that are optimized for discriminating between classes. CSP was introduced to EEG signal processing by [45] to detect abnormal EEG signals. Then CSP was subsequently used for EEG classification by [57]. CSP is based on simultaneous diagonalization of the covariance matrices to construct the special filters. CSP uses the covariance of the EEG signal because all the information of the class distribution exists only in the covariance ; also, the inputs to the CSP algorithm has to be of zero mean (using temporal filtering will guarantee that this condition is adhered to). CSP algorithm starts by finding the mean normalized spatial covariance matrices $\left\{\bar{C}_1, \bar{C}_2\right\}$ of the EEG signal for each class.

$$\bar{C}_i = \frac{\sum_{j=1}^{N_i} \frac{E_{ij}E_{ij}^{'}}{trace\left(E_{ij}E_{ij}^{'}\right)}}{N_i} \tag{2.6}$$

$$\begin{aligned} where, \quad & E_{ij} \; is \; the \; j^{th} \; EEG \; trial \; of \; class \; i \\ & N_i \; is \; number \; of \; EEG \; trials \; for \; class \; i \\ & E_{ij}^{'} \; donates \; the \; transpose \; of \; E_{ij} \end{aligned}$$

Whitening the composite spatial covariance matrices (C_c) will generate a linear matrix (P) that transforms the composite spatial covariance matrix into a white vector (S_c) which has uncorrelated random variables and their variances equal unity.

$$C_c = \bar{C}_1 + \bar{C}_2 \tag{2.7}$$

$$C_c = U_c \lambda_c U_c^{'} \tag{2.8}$$

$$P = \sqrt{\frac{1}{\lambda_c}} U_c^{'} \tag{2.9}$$

$$S_c = P C_c P^{'} \tag{2.10}$$

Instead of using P to whiten C_c, it can be used to whiten C_1 and C_2 as follows:

$$S_1 = P \bar{C_1} P^{'} \tag{2.11}$$

$$S_2 = P \bar{C_2} P^{'} \tag{2.12}$$

Since S_1 and S_2 share the same eigenvectors, the sum of their diagonal matrix of eigenvalues will be the identity matrix.

$$S_1 = \beta \lambda_1 \beta^{'} \tag{2.13}$$

$$S_2 = \beta \lambda_2 \beta^{'} \tag{2.14}$$

Because both classes share the same eigenvector β and according to equation 2.13, the eigenvectors which are associated with the largest eigenvalues for class 1 (λ_1) have the smallest eigenvalues for class 2 (λ_2) and vice versa. Eigenvectors become useful for classification because of this property, especially eigenvectors which are associated with largest and least eigenvalues. The projection matrix W is the projection of the whitened EEG trials on selected eigenvectors. The columns of W^{-1} are time-invariant vectors which are called common spatial patterns.

$$W = \left(\beta^{'} P\right)^{'} \tag{2.15}$$

CSP can deal with multiclass problems where it uses the one-versus-rest technique. To find the spatial filter for a certain class, the trials are divided into target and non-target trials.

2.2.2 Feature Extraction Techniques

Feature extraction is a special form of dimensionality reduction. During the feature extraction process, representative characteristics of the data (simple examples are the mean and variance) are calculated and stored as "feature vectors" which may then be passed to classifiers and other processing stages. This allows us to deal only with the properties of the data that are relevant to the task at hand, and helps to facilitate the classification of this data into classes of interest. We now discuss several feature extraction methods, some of which are widely used for EEG signals.

2.2.2.1 Power Spectral Density (PSD)

PSD is used in 41% of sensorimotor based BCI studies, making it one of the most commonly used feature extraction techniques [8]. Imagined motor tasks are considered as event-related desynchronization (ERD) of neurons signals. PSD finds the power density for filtered EEG recordings. Typically, a bandpass FIR filter with passband of 8 to 30 Hz is used to pre-process the raw data prior to the use of PSD extraction algorithms. PSD finds signal power contained in these frequency ranges. This is because ERD is known to occur mainly in this range of frequencies [6, 12, 15, 19, 46, 48]. Mathematically, PSD is defined as a Fourier

Transform of the autocorrelation sequence of the time series.

$$EEG(f) = \int_{t_1}^{t_2} eeg(t)\, e^{-2\pi i f t} dt \tag{2.16}$$

$$PSD(f) = 2\frac{\|EEG(f)\|^2}{(t_2 - t_1)} \tag{2.17}$$

$EEG(f)$ is the complex spectrum of $eeg(t)$ in the range t_1 to t_2. Equation 2.16 and 2.17 are using Fourier transform for autocorrelation signal then it will be hard to find the exact values for EEG signal since it is a stochastic signal. To find estimate PSD value for EEG signal, two main approaches have been used nonparametric and model-based methods.

1. Nonparametric methods: Equation 2.16 and 2.17 are the basis of nonparametric methods of power spectral density estimation. Because of the random nature of the EEG signal, many methods have been developed to reduce the variance of the obtained spectrum. In the following some of nonparametric power spectral density estimation methods:

    - The mean and variance of periodograms: This method finds the mean value of the periodogram through the expected value and calculating the variance too.
    - Bartlett Method: This method averages several periodograms from different segments of the EEG signal in order to reduce the variance in the estimated periodogram.
    - Welch Method: This method divides the EEG signal into overlapped windowed segments and then averages the periodograms for all seg-

ments. Welch method uses windowed segments to reduce the effect of the discontinuous of the segments.

2. Model-Based methods: In model-based or parametric methods, a model of the EEG signal is used to extrapolate the autocorrelation function out of available data samples range (out of the segments' samples). Model-based methods have a better resolution than nonparametric methods because the nonparametric methods assume the correlation is zero for unavailable samples. In the following a list of some model-based methods:

   - Maximum-Entropy Spectral Estimation: This method estimates the autocorrelation sequence according to the most random (highest entropy) signal inside the defined segment.
   - Autoregressive Power Spectrum Estimation: In this method a linear prediction model is used to model the signal and to estimate the power spectrum density.

2.2.2.2 Log Variance

Calculating the variance for the EEG signal is a useful feature after applying common special filter (CSP) where the goal of common special filter is to minimize the variance of certain class and maximize the variance of the other class [57]. By extracting the variance for the EEG signal it will be a clear feature for discriminating between classes.

$$f_p = \log\left(\frac{Var_p}{\sum_{i=1}^{P} Var_i}\right) \tag{2.18}$$

After calculating the variance for the spatially filtered EEG signal components, the logarithm of the normalized variance is used to find the feature vector. The computational requirements of the log variance method are extremely modest when compared with the other methods discussed in this section.

2.2.2.3 Wavelet Packet Decomposition (WPD)

The wavelet transform is a spectral estimation method where any function can be expressed by an infinite series of wavelets. The main benefit for using wavelet is the time-frequency localization. About 13% of the published work in BCI for sensorimotor activities use wavelets as feature extraction [8, 9, 30, 67]. The idea behind wavelet is to express the EEG signal by a linear combination of functions generated by scaling and shifting pre-define signal called mother wavelet. Wavelet gives the flexibility to extract multi-resolution features from the signal. Unlike the Fourier Transform which used only sine waves as bases functions, Wavelets can be constructed using any "mother wavelet" function which satisfies the following conditions:

$$\int_{-\infty}^{\infty} \psi(u)\, du = 0 \tag{2.19}$$

$$\int_{-\infty}^{\infty} \psi(u)^2 du = 1 \tag{2.20}$$

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(u)\, e^{-2i\pi f u} du \tag{2.21}$$

where $\int_{-\infty}^{\infty} \frac{|\Psi(f)|^2}{|f|} df < \infty$.

Equation 2.19 specifies that the mother function ψ has to oscillate around the

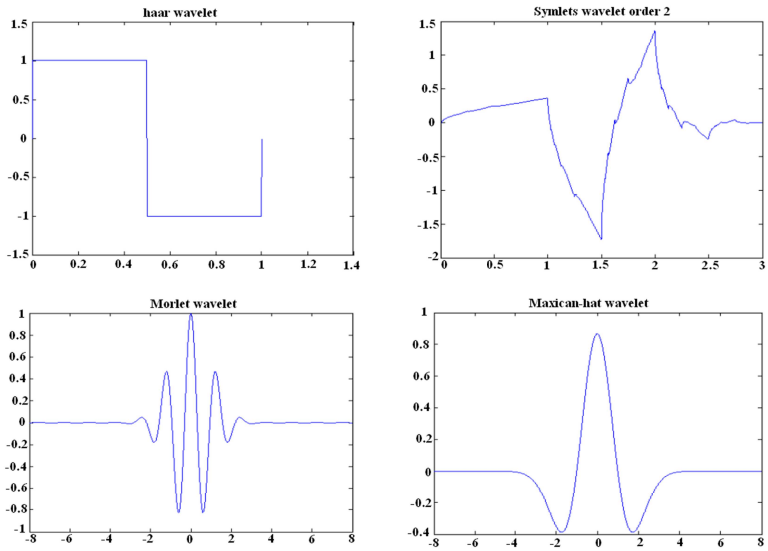


Figure 2.8: Some famous wavelet families.

zero equally, so the positive side cancels the negative side. While Equation 2.20 forces the mother function ψ to have oscillation around the zero. Equation 2.21 was designed to ensure that wavelet transform is invertible. Figure 2.8 shows some of the commonly encountered wavelet functions, e.g., The Haar function is one of the simplest mother wavelet functions to use. Selecting certain wavelet mother function depends on the nature of the signal, since the characteristics of the mother wavelets should match those of the underlying signal components as much as possible. Wavelet has two main classes' continuous and discrete. Continuous wavelet transform is considered computationally expensive comparing to the discrete wavelet transform. The most commonly used approach for implementing the discrete transform is a bank of digital filters levels. Each level consists of low and high pass filters to create detail and approximation coefficients for the input from the previous level. Figure 2.9 shows the structure of the discrete wavelet transform, where H[t] and L[t] are high-pass filter and low-pass filter re-

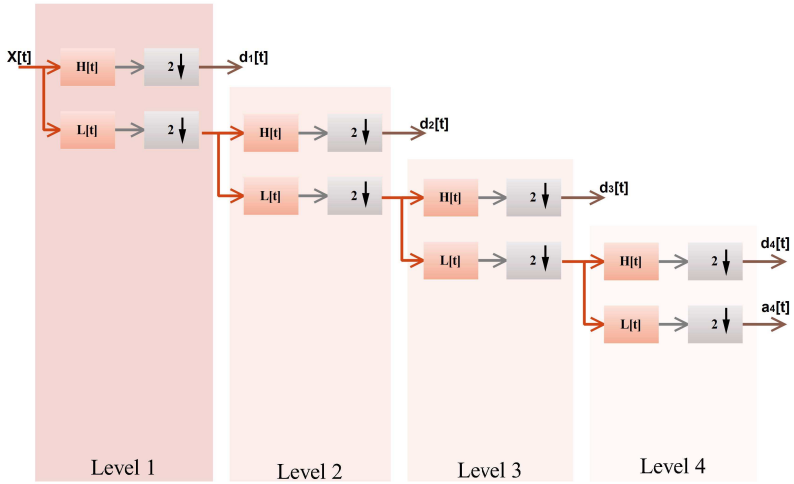


Figure 2.9: Discrete Wavelet Transform.

spectively. Each level of discrete wavelet transform level splits the data into two orthonormal vectors the detailed and approximation coefficients. This method, in particular, was used to generate a feature vector of wavelet packet decomposition coefficients of the EEG signal for this thesis.

2.2.2.4 Mel Frequency Cepstral Coefficient (MFCC)

The MFCC is widely used for speech recognition that has been used to detect echoes in a sound signal in 1963 [14]. However, while it has since become the primary feature extraction method for speech recognition, it can also be applied in other domains where noisy signals have been encountered, such as analysis of seismic data, control and vibration analysis of engine noise [3, 34]. Figure 2.10 shows the processes to extract the MFCC features for EEG signal.

Initially the EEG signal was divided into equally size windows. Each vector of MFCC features will represent one window. The Fourier transform of each window was calculated to extract the frequency components present in that window, this

known as "short time Fourier transform (STFT)". Next, the magnitude of this spectrum is warped by the Mel scale as shown in figure 2.11. This converts the frequency to Mels which is a perceptual scale of pitches judged by listeners to be equal in distance from one another as proposed by Stevens, Volkman and Newman [76] in 1937. The conversion between the linear scale of frequencies and the Mel scale is given in equation 2.22.

$$m = 2595 \log_{10} \left(\frac{f}{700} + 1 \right) \tag{2.22}$$

The logarithm for the Mel-scaled spectrum is taken. Finally the Fourier transform is taken for the log Mel-scaled spectrum. Working with cepstral introduces confusions since most terms get mixed up with the spectrum terms. To solve this terminology confusion, some terms was introduced as in table 2.3. For example, the frequency for the spectrum of the EEG signal is called Quefrency.

2.2.3 Classifiers

Classifiers build a model to discriminate between different classes based on a training data. This thesis investigates the performance of different classifiers. A brief discussion about some famous classifier is in the following :

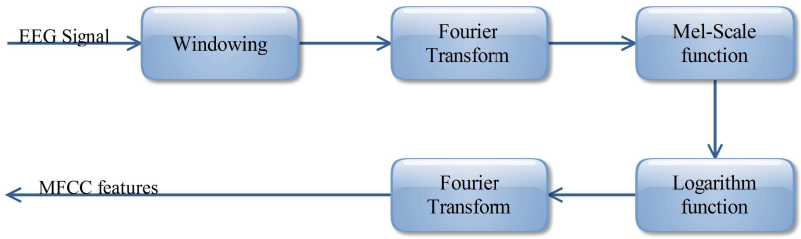


Figure 2.10: MFCC block diagram.

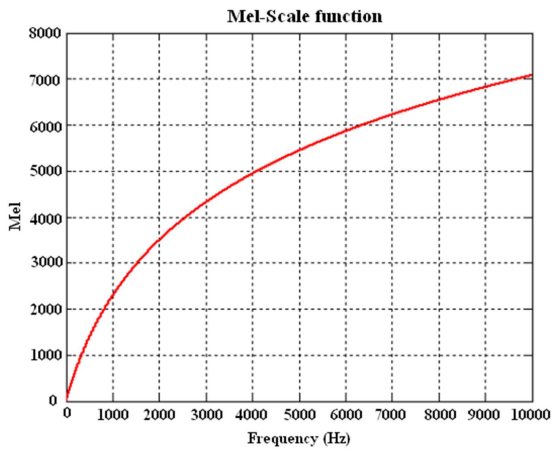


Figure 2.11: Mel-Scale Function.

Spectrum Terminology	Cepstrum Terminology
Frequency	Quefrency
Spectrum	Cepstrum
Phase	Saphe
Amplitude	Gamnitude
Filtering	Liftering
Harmonic	Rahmonic
Period	Repiod

Table 2.3: Cepstrum Terminology.

2.2.3.1 Random Forests (RF)

Random Forest is an ensemble classifier which contains $N - number$ of trees. Each tree is grown on randomly selected subset of the training data using bootstrap sampling. The final predication formed by a majority voting of the all independently trained trees [10, 38]. Figure 2.12 shows random forest consists of N tree. Usually, no pruning step performed on the trees.

The major benefit of using random forest is to lower the variance of the output. Another advantage of using random forest is avoiding overfitting where the boot-

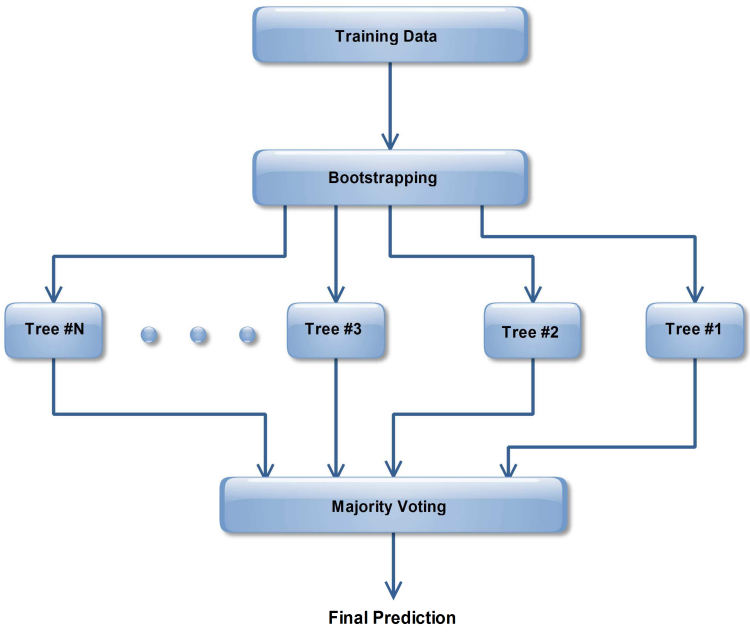


Figure 2.12: Random Forest Classifier.

strap process and growing N-tree independently will avoid the overfitting problem.

2.2.3.2 Linear and Quadratic Discriminant Analysis (LDA and QDA)

LDA and QDA can be used for both classification or dimension reduction [15, 32, 44, 42]. As a classifier LDA finds a linear hyper plane that separates between each two classes. LDA assumes the distribution for each class is multivariate Gaussian distribution as in equation 2.23.

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)} \tag{2.23}$$

Where k is the class label, μ_k is the mean value of class k, Σ_k is the covariance of class k. The posterior probability of optimal classification is $P(C|X)$

$$P(C=k|X=x) = \frac{f_k(x)\pi_k}{\sum_{i=1}^{K} f_i(x)\pi_i} \tag{2.24}$$

π_i is the prior probability of class k, where $\sum_{k=1}^{K}\pi_k = 1$. To find a discrimination hyper plane, we need to compare each two classes by looking to the log ratio of the two posterior probabilities as in equation 2.25. Equation 2.25 is based on the assuming that all classes share the same covariance matrix $\Sigma_k = \Sigma \ \ \forall k$.

$$\begin{aligned} \log\frac{P(C=k|X=x)}{P(C=\iota|X=x)} &= \log\frac{f_k(x)}{f_\iota(x)} + \log\frac{\pi_k}{\pi_\iota} \\ &= \log\frac{\pi_k}{\pi_\iota} - \frac{1}{2}(\mu_k-\mu_\iota)^T\Sigma^{-1}(\mu_k-\mu_\iota) + x^T\Sigma^{-1}((\mu_k-\mu_\iota)) \end{aligned} \tag{2.25}$$

The linear discriminant functions $\delta_k(x)$ as shown in equation 2.26 are the decision rule where $G(x) = \arg\max_k \delta_k(x)$.

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k \tag{2.26}$$

We can estimate the parameters of the Gaussian distributions using the training data sets. Where:

$$\begin{aligned} \hat{\pi}_k &= \frac{N_k}{N} \\ \hat{\mu}_k &= \sum_{i\in X_k}\frac{x_i}{N_k} \\ \hat{\Sigma} &= \sum_{k=1}^{K}\sum i\in X_k\frac{(x_i-\hat{\mu}_k)(x_i-\hat{\mu}_k)^T}{N-K} \end{aligned}$$

Quadratic discriminant analysis uses a quadratic hyper-plane instead of linear hyper plane. Quadratic discriminant analysis relaxes the assumption of equally class's covariance. As a result of this relaxation we get the discrimination function $\delta_k(x)$ as in equation 2.27.

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log\pi_k \tag{2.27}$$

The decision boundary between two different classes is expressed by a quadratic equation of x $\delta_k(x) = \delta_l(x)$. The same linear discriminant analysis parameters estimation will be used for quadratic discriminant analysis.

2.2.3.3 Artificial Neural Networks (ANN)

An ANN is a computational model that is inspired by biology. The most commonly used ANN architecture is the feed-forward Neural Network, which consists mainly of three layers which are input layer, hidden layers and output layers. About 25% used classifiers in BCI are based on Neural Network [8, 60, 50, 40]. Figure 2.13, shows the structure of feed forward ANN. ANN can be seen as a series of activation function for where each layer contains certain type of activation function.

$$\alpha_j = \sum_{i=1}^{D} \omega_{ji}^{(1)} x_i + \omega_{j0}^{(1)} \tag{2.28}$$

$$z_j = h(\alpha_j) \tag{2.29}$$

$$\alpha_k = \sum_{i=1}^{M} \omega_{ki}^{(2)} z_i + \omega_{k0}^{(2)} \tag{2.30}$$

Where α_i the activation value of neuron i (which is the weighted sum of all

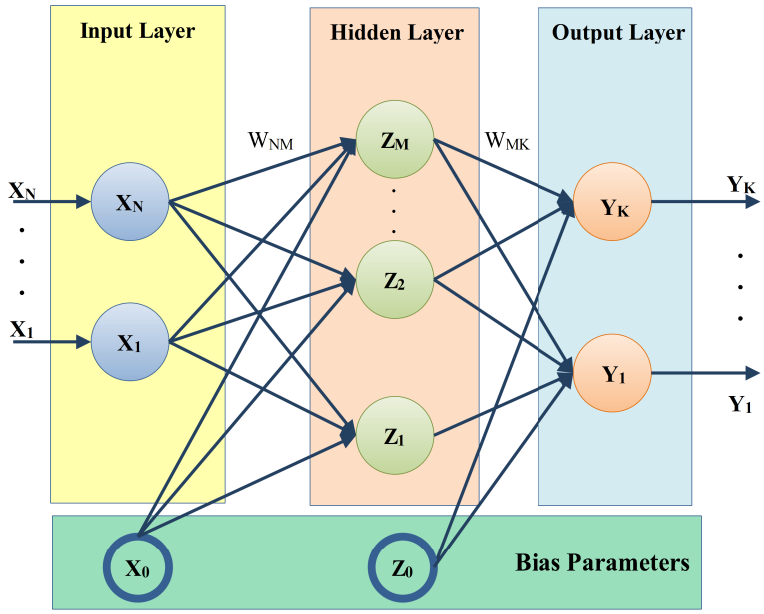


Figure 2.13: General Artificial Neural Network Structure.

the output from the previous layer). $h(.)$ is the activation function and $\omega_{ji}^{(n)}$ is the weight parameter between neuron j in layer n and neuron i in layer $n+1$. $\omega_{j0}^{(n)}$ are the bias parameters.

By minimizing the square error function for the output we will train the neural network on the optimal parameters.

$$E(\omega) = \frac{1}{2}\sum_{n=1}^{N} \|y(x_n, \omega) - t_n\|^2 \tag{2.31}$$

Where x_n is the nth training feature vector and t_n is the true label for x_n. In simplify the classification problem we consider t as string of binary flags where each flag indicate the label of the training example $t_k \in \{0,1\}$, where $k = 1,\ldots,K$. K is the total number of classes. For such multi-label classification problem, a K

output neural network is used.

$$p(t|x,\omega) = \prod_{k=1}^{K} y_k(x,\omega)^{t_k}[1-y_k(x,\omega)]^{1-t_k} \tag{2.32}$$

By taking the negative logarithm of the likelihood we form the error function.

$$E(\omega) = -\sum_{i=1}^{N}\sum_{k=1}^{K}[t_{ik}\log y_{ik}(x_i,\omega) + (1-t_{ik})\log(1-y_{ik}(x_i,\omega))] \tag{2.33}$$

The target vector t is a binary vector which sum to exactly one, because t-vector represents 1-of-K scheme.

$$\begin{aligned} \sum_{i=1}^{K} t_i &= 1 \\ & \quad where \;\; t_i \in \{0,1\}\,\forall i \end{aligned} \tag{2.34}$$

Then the neural network output vector can be represented as $y_i(x,\omega) = p(t_i = 1|x)$. The error function will be as following based on this representation.

$$E(\omega) = -\sum_{i=1}^{N}\sum_{k=1}^{K} t_{ik}\log y_{ik}(x_i,\omega) \tag{2.35}$$

We will consider the output function as a softmax activation function as in equation 2.36.

$$y_i(x,\omega) = \frac{e^{\alpha_i(x,\omega)}}{\sum_j e^{\alpha_j(x,\omega)}} \tag{2.36}$$

Softmax function satisfies the boundary conditions for the binary output where

$0 \leq y_i \leq 1$, and the sum for the outputs are 1, $\sum_i y_i = 1$. Sequential gradient descent, equation 2.37, is one of the methods to find the optimal parameters (ω) that minimize the error function.

$$\omega^{(\tau+1)} = \omega^{(\tau)} - \eta \nabla E_n \left(\omega^{(\tau)}\right) \tag{2.37}$$

Where, E_n is the error function for single training point x_n . To evaluate the gradient of the error function $(\nabla E_n(\omega))$, Error backpropagation is used. Where, error value is propagated from layer to layer.

2.2.3.4 Support Vector Machines (SVM)

An SVM classifies data by constructing an N-dimensional hyperplane which optimally separates the classes. The basic version of the SVM was introduced by Vladimir Vapnik [82] which classifies data by constructing deterministic binary linear classifier. Soft margin SVM was developed to deal with the trade-off between maximizing the margins and minimizing the misclassification error. Nonlinear decision hyperplane was introduced to deal with nonlinear separable classification problems by mapping the features to higher dimensional space or dealing with kernel tricks. Classification methods based on SVM formed about 11% of BCI publications [8, 85, 56, 29]. This section will show the formulation of the linear SVM, soft margin SVM and nonlinear SVM.

- Binary Linear SVM: Binary Linear SVM aims to find a $(N-1)$ linear hyperplane that separate the dataset into two classes. An infinite number of such hyperplanes can be found, so to choose between them we will maximize the margins between the two classes and the hyperplane. Such a hy-

perplane is called a maximum-margin hyperplane. The maximum margin hyperplane H_s is described by Equation 2.38.

$$\theta \cdot x - \theta_0 = 0 \tag{2.38}$$

The classification is determined by the location of the point according to the H_s. Equation 2.39 shows the classification function.

$$\begin{aligned} c_i &= sign\,(\theta \cdot x_i - \theta_0) \\ & \quad where \;\; c \in \{-1, 1\}\,,\; x \in \Re^D \end{aligned} \tag{2.39}$$

To maximize the margin between H_s and the nearest point (support vectors), we introduce H_1 and H_2 hyperplanes where they are parallel to H_s.

$$x_i \cdot \theta - \theta_0 = +d \quad for \;\; H_1 \tag{2.40}$$

$$x_i \cdot \theta - \theta_0 = -d \quad for \;\; H_2 \tag{2.41}$$

Note H_s is in the middle of gap which separates H_1 and H_2, as is shown in figure 2.14.

To achieve the optimal discriminating hyperplane we need to maximize the distance D based on equation 2.42.

$$D = \frac{d + \theta_0}{\|\theta\|} - \frac{-d + \theta_0}{\|\theta\|} = \frac{2d}{\|\theta\|} \tag{2.42}$$

To simplify the optimization problem, we divide equation 2.40 and 2.41 by

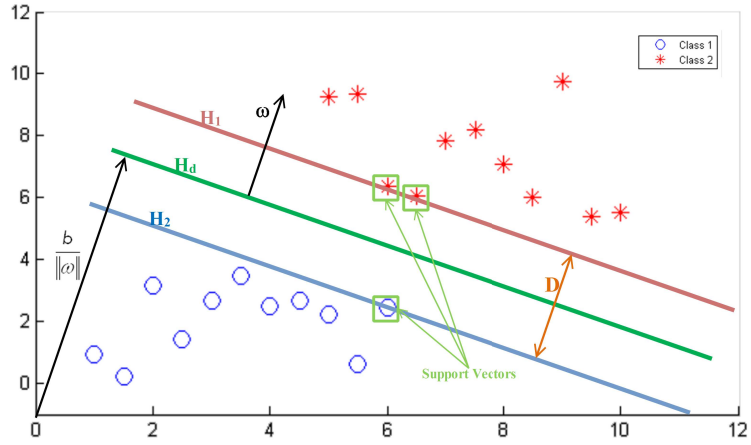


Figure 2.14: Linear SVM.

d and update θ and θ_0 to be $\theta \leftarrow \frac{\theta}{d}$ and $\theta_0 \leftarrow \frac{\theta_0}{d}$. The objective function will minimize $\|\theta\|$. We minimize $\frac{1}{2}\|\theta\|^2$ instead of $\|\theta\|$ to simplify the optimization problem. In the end, minimizing either of these two quantities are equivalent since they are monotonically related.

$$\min \frac{1}{2}\|\theta\|^2 \tag{2.43}$$

$$\begin{aligned} \textit{where}, \qquad & c_i\left(x_i \cdot \theta - \theta_0\right) \geq 1 \\ & c_i \in \{-1, 1\}, \quad \textit{class label} \end{aligned}$$

Using Lagrange multipliers α to take enforce these constrains in the minimization problem.

$$L_P = \frac{1}{2}\|\theta\|^2 - \sum_{i=0}^{n-1} \alpha_i \left(c_i\left(x_i \cdot \theta - \theta_0\right) - 1\right) \tag{2.44}$$

Equation 2.44 can be solved by converting it from the primal to the dual

form. This process can be done by finding $\frac{\partial L_p}{\partial \theta}$ and $\frac{\partial L_p}{\partial \theta_0}$ and assign it to zero to find the minimum according to θ and θ_0 as in equation 2.45 and 2.46.

$$\begin{aligned} \frac{\partial L_p}{\partial \theta} = \theta - \sum_{i=0}^{n-1} \alpha_i c_i x_i &= 0 \\ \theta = \sum_{i=0}^{n-1} \alpha_i c_i x_i &= 0 \end{aligned} \tag{2.45}$$

$$\begin{aligned} \frac{\partial L_p}{\partial \theta_0} = 0 - \sum_{i=0}^{n-1} \alpha_i c_i &= 0 \\ \sum_{i=0}^{n-1} \alpha_i c_i &= 0 \end{aligned} \tag{2.46}$$

By substituting equation 2.45 and 2.46 into equation 2.44, we get the dual form for the equation 2.44.

$$\begin{aligned} L_D &= -\frac{1}{2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \alpha_i \alpha_j c_i c_j x_i x_j + \sum_{i=0}^{n-1} \alpha_i \\ & subject\ \ to: \quad \alpha_i \geq 0 \ \ \forall i \end{aligned} \tag{2.47}$$

L_D depends on α only, and it is a convex quadratic optimization problem which can be solved using QP-solver. By finding α we will be able to find θ according to the training dataset using equation 2.45. To find all the parameters for the decision hyperplane, we still need to find θ_0. We can find θ_0 by using the support vectors as in equation 2.48.

$$c_i(\theta \cdot x_i - \theta_0) = 1 \tag{2.48}$$

$$where \qquad (x_i, c_i) \in \ Support\ Vectors$$

By substituting equation 2.45 in equation 2.48 and multiplying both sides by c_i , we find that:

$$\begin{aligned} \theta_0 &= -c_i + \sum_{j \in S} \alpha_j c_j x_j x_i \\ where \qquad & (x_i, c_i) \in \ Support\ Vectors \end{aligned} \tag{2.49}$$

Note that $c_i^2 = 1$. We can obtain θ_0 using only one support vector (x_i, c_i) but it is preferable to take the average for different θ_0 values using all the support vectors as in equation 2.50.

$$\theta_0 = \frac{1}{N_{SupportVectors}} \sum_{(x_i,c_i) \in S_{SV}} \left(-c_i + \sum_{(x_j,c_j) \in S_{SV}} \alpha_j c_j x_j x_i \right) \tag{2.50}$$

Using the training dataset we can find an estimate for both θ and θ_0 and then we will have a classification function as in equation 2.39.

- Soft margin SVM: for a practical dataset, there is often some overlap between classes, such that it is impossible to separate them perfectly. Margin error was introduced to SVM to deal with this overlap which allow the misclassification error for a penalty cost in the objective function. Using a nonnegative slack variable ξ to measure the degree of constraint violation. Our new objective function, which is given in equation 2.51, thus includes a term to represent the degree to which the constraints were violated, and

which hence needs to be minimized.

$$\min \frac{1}{2}\|\theta\|^2 + C\sum_{i=0}^{n} \xi_i \tag{2.51}$$

And these constraints will also be relaxed, as shown in equation 2.52 below.

$$\begin{aligned} c_i\left(\theta \cdot x_i + \theta_0\right) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \;\; \forall i \in \{1, \ldots, N\} \end{aligned} \tag{2.52}$$

Lagrange multipliers can then be used to solve the relaxed problem as in the linear SVM. The cost parameter C in equation 2.51 is used to manage the tradeoff between the misclassification errors and maximizing the margin (minimizing θ).

- Non-linear SVM: to deal with nonlinearly separable dataset, SVMs can use non-linear discrimination hyperplane. SVM maps the original features to higher dimensional space to form a linearly separable dataset. Figure 2.15 shows data that is not linearly-separable in a two dimensional space (left), but which becomes linearly separable when mapped to the three dimensional space (right). By applying the linear SVM in the higher dimensional space, SVM will be able to produce a nonlinear discriminating hyperplane capable of correctly classifying the data.

  Instead of mapping the feature space into higher dimensional space, we use the idea of linear kernel functions, where $\left(x_i, x_j\right) = x_i \cdot x_j$. We can introduce any kernel function and use SVM equation 2.47 to find the discrimination hyperplane. Nonlinear kernel functions can be used. Table 2.4 shows some

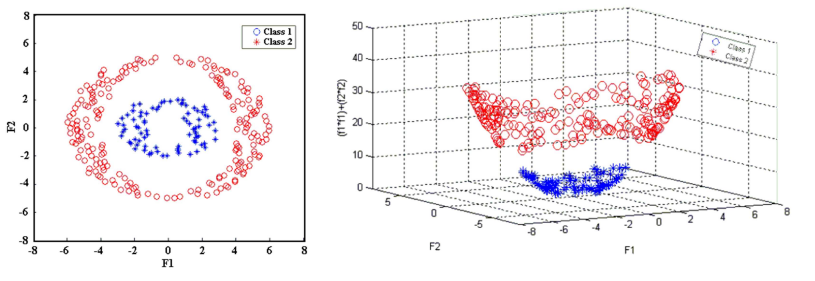


Figure 2.15: Mapped data using $\kappa\left(x_i, x_j\right) = x_i^2 + x_i^2$.

nonlinear kernel functions.

Kernal Family	$\kappa\left(x_i, x_j\right)$
Polynomial	$\left(x_i \cdot x_j\right)^d$
Sigmoidal	$\tanh\left(\alpha x_i \cdot x_j - \beta\right)$
Radial Basis	$e^{-\gamma\left\|x_i \cdot x_j\right\|^2}$
Gaussian Radial Basis	$e^{-\frac{\left\|x_i \cdot x_j\right\|^2}{2\sigma^2}}$

Table 2.4: Famous Kernal Functions.

The SVM classifier, in this thesis, was based on LIBSVM implementation from [17].

2.3 Summary

Among the different neuroimaging techniques that are used for BCI research, EEG is the most practical and commonly used, and will be the focus of this study. BCI exploits different neuromechanisms where each one is useful to distinguish certain mental tasks. It was decided that sensorimotor-based BCI will be the modality used in this work. This chapter also provided a description of the data that will be used to validate the proposed methods. Two datasets are used: one is obtained

from the publicly available BBCI Competition IV database, while the other dataset was recorded internally at the Institute. The signal processing and machine learning algorithms that will be used, or which were considered for use, have also been discussed in some detail. The following chapter will cover. The next chapter will cover the implementation and optimization details for preprocessing and feature extraction methods.

CHAPTER 3

Implementation and Methodology

This chapter describes our implementation of the proposed system, as well as the procedures used for the selection of appropriate parameters. The use of IIR and CSP as pre-processing stages are described. Finally, a discussion of the signal feature extraction and classification used in this project is provided.

3.1 Data Description

The research described here utilized data from two main sources: data provided by the BBCI Competition[1], and data that was recorded at Masdar Institute using the Emotiv EPOC EEG device [2].

[1] http://www.bbci.de/competition/iv/desc_1.html
[2] Data is available from hamzah_sul@yahoo.com

This thesis will use data provided by the BBCI Competition IV Dataset I, which is available for public use. As described in section 2.1.4 , data was collected from four healthy subjects (a, b, f *and* g) while they were performing asynchronous BCI tasks. This dataset was selected because it includes a variety of challenges and aims to develop natural and flexible BCI systems. BBCI Competition IV Dataset I uses the BrainAmp MR plus amplifier and a 59-Ag/AgCl electrode cap to record the EEG signal from subjects. The 59 electrodes were distributed as shown in figure 3.1, most of the electrodes were distributed over the sensorimotor areas. Each subject was asked to select two mental tasks to perform out of three different tasks: left hand movements, right hand movements, or foot movements (side chosen by subject and can be both feet). Then each subject was required to undergo two recording sessions: a calibration and an evaluation session. These are as follows:

- Calibration session: Subjects were asked to perform a certain mental task based on the visual cues on the computer screen. Each cue appeared for 4 seconds, after that a blank screen was shown for 2 seconds followed by a fixation cross for another 2 seconds.

- Evaluation session: Subjects were asked to perform the mental tasks indicated by soft acoustic cues, which ranged from 1.5 to 8 second in length. Each stimulus was followed by intermitting command (beep), which indicates the start of a rest period which is for a variable lengths ranging from 1.5 to 8 seconds.

Complete data was provided as follows:

- EEG recordings for training and evaluation sessions. The EEG recordings

are provided as $n \times t$ matrices where n is the number of channels, and t is the number of time steps.

- Cue timing and types. As shown in figure 3.2. An array of start times of the cues was provided and another array includes the labels.

- Electrodes placements as in figure 3.1. The x and y coordinations were provided with the electrodes name.

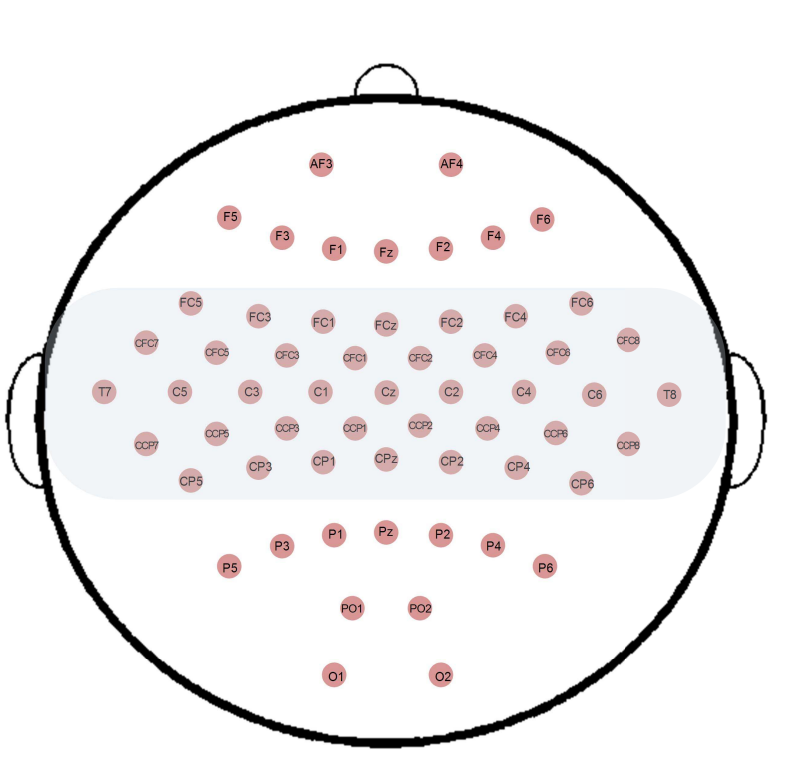


Figure 3.1: 2D top view of the scalp with the electrodes locations map.

The Dataset I of the BBCI Competition IV also includes three artificially generated datasets. The purpose of this artificially generated datasets is to have artificial EEG signals which have realistic properties close to real EEG recordings.

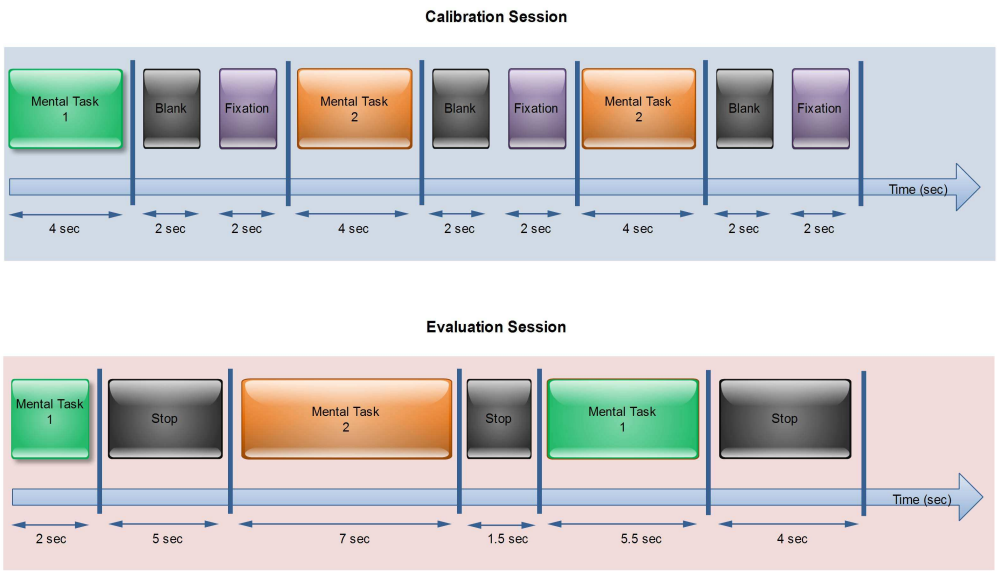


Figure 3.2: Tasks Timing in Calibration and Evaluation sessions.

Artificial EEG signals can be used for test and analysis if they have realistic properties. The locally collected dataset using Emotiv was based on the same experiment that is designed by BBCI research group. Emotiv includes only 14 EEG electrodes; more information can be found about Emotiv in appendix A.

3.2 Preprocessing

3.2.1 Temporal Filtering

To filter out frequencies that do not fall within the μ and β ranges, bandpass filtering was used. This was implemented using a 10th order Butterworth filter using the filter design toolbox from Matlab [51]. The choice of this range of frequencies is based on the fact that sensorimotor activities are associated with [8-30Hz] frequency band but it was found that this frequency band is subject dependent, i.e.,

the frequency band is different from subject to another [77]. To find out the most useful frequency band for each subject, we used bi-serial correlation coefficient $r^2 - value$, Equation 3.1, of the difference between two average power of each pair of tasks. $r^2 - value$ evaluates the discriminability between two classes based on their distribution.

$$r^2 = \left[\frac{\sqrt{N_{T1} \cdot N_{T2}}}{N_{T1} + N_{T2}} \cdot \frac{mean\,(EEG_{T1}) - mean\,(EEG_{T2})}{std\,(EEG_{T1} \cup EEG_{T2})} \right]^2 \tag{3.1}$$

The frequency band with $r^2 - value$ greater than a certain threshold was used for Butterworth IIR filter. Figure 3.3 shows the averaged spectra of two mental tasks (left hand vs. foot) of subject a from the BBCI competition. As figure 3.3 shows the $r^2 - value$ for each frequency and the light green area shows the selected frequency band based on the $r^2 - value$, where a range of frequencies were selected around the peak of r^2 value. Notice that this spectrum was taken before applying any spatial filtering to study the effect of IIR filter. The figure shows the spectra for $C4$ channel which was selected based on the average $r^2 - value$, $C4$ has the maximum $r^2 - value$ among all other channels for this case. The automatic selection of channels confirms the standard positions of used sensorimotor areas. Figure 3.4 shows the average spectra of filtered EEG signal of the $C4$ channel and the selected band power to be filtered.

3.2.2 Common Spatial Patterns (CSP)

To enhance the discriminability between classes, the method of Common Spatial Patterns (CSP) was used. Because our problem includes three classes, one-versus-rest CSP was used. The idea behind one-versus-rest CSP is to divide the problem

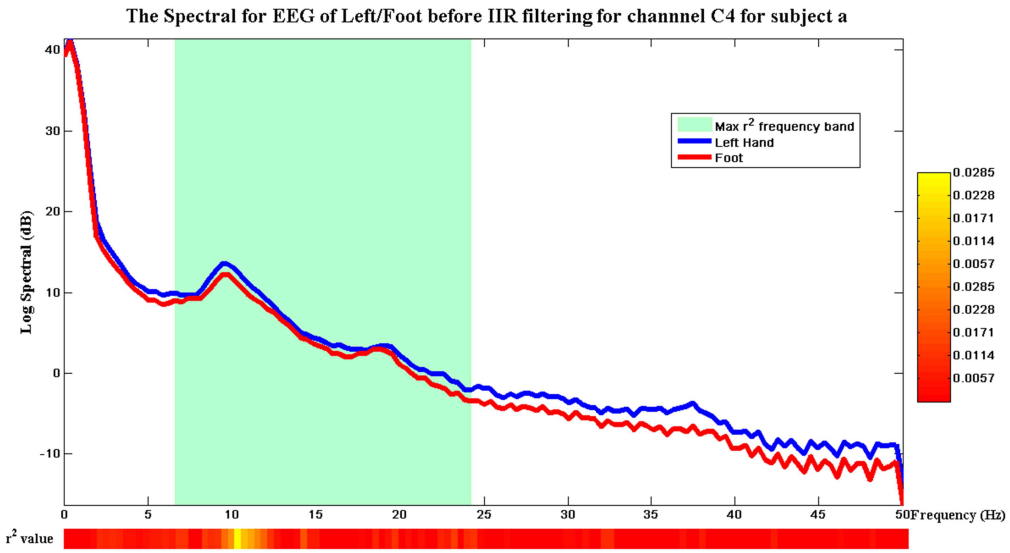


Figure 3.3: The spectra of the EEG signal from C3 channel for two imagery tasks (left/foot) with $r^2 - value$ before IIR filter (subject a).

into three subproblems where the EEG trials for a certain class are differentiated from the trials for the other two classes. This process is subsequently repeated for each of the three classes, resulting in three separate CSP decomposition matrices. The rows of the decomposition matrices are the spatial filters and the columns of its inverse are the common spatial patterns. Figure 3.5 shows 2D top view maps of the scalp, where the colors indicate the values of the respective filters at the sites of the 59 electrodes. The maps on the top row of the figure correspond to the stationary spatial patterns, and the maps on the bottom row correspond to filters generated using CSP, for the trial which has the minimum variance for each class $left$, $right\ and\ foot$. The patterns cover larger area from the Scalp and they are smoother than the filters, as shown in figure 3.5. CSP filters will be useful to weight the channels based on their relevance to the mental task. Notice the spatial patterns for the left hand mental task filter out the regions out of the right

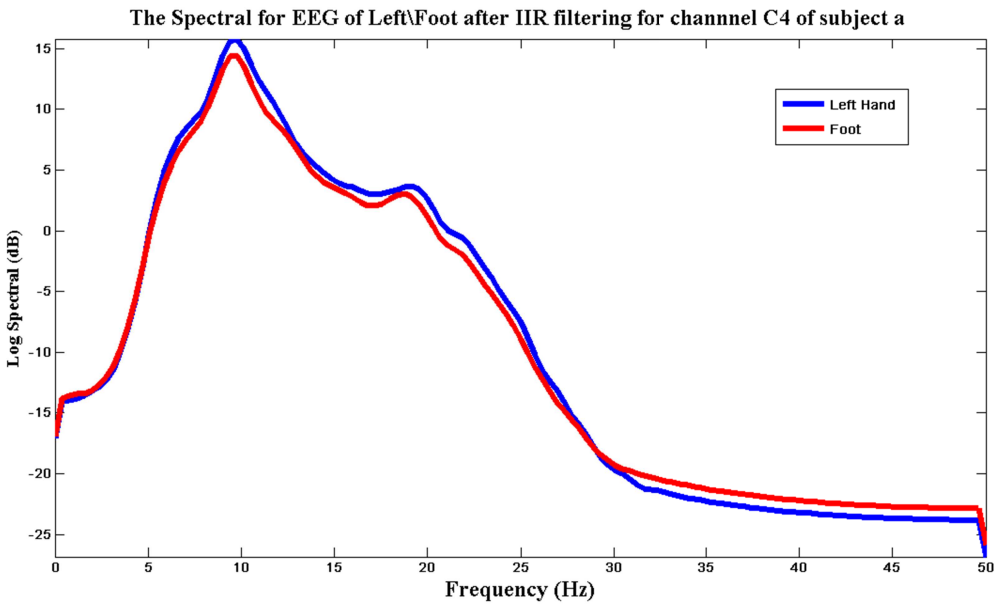


Figure 3.4: The spectra of the EEG signal from C3 channel for two imagery tasks (left/foot) with $r^2 - value$ after IIR filter (subject a).

hemisphere sensorimotor as suggested by neurophysiologists. The same for the spatial patterns of the right hand mental task which filter out the regions out of the left hemisphere sensorimotor and the idle task which is distributed over the middle of the scalp.

CSP was benchmarked to Laplace spatial filter and with pure signal without using any spatial filter. Laplace filter is a high pass spatial filter where each channel is averaged with its neighbors, i.e., 5 nearest neighbors was selected. Figure 3.6 compares the log power of the two tasks after application of the selected CSP filter and after the using the Laplace spatial filter. It can be seen that CSP outperforms the Laplace filter in improving the discriminability of two mental tasks, i.e., the $r^2 - value$ is much bigger after using CSP (0.09 to 0.45) than using Laplace filter (0.006 to 0.028).

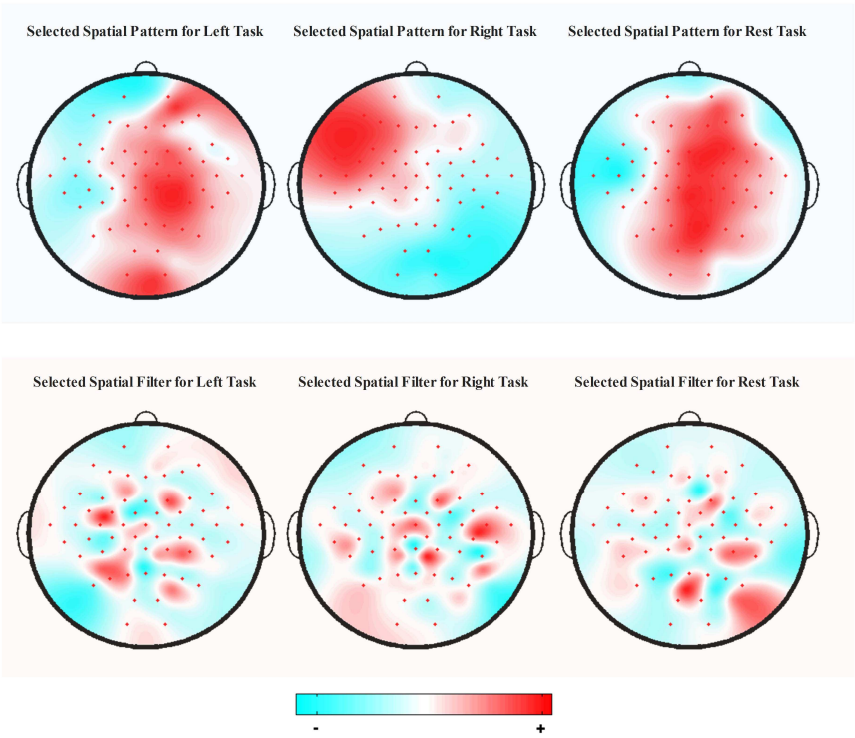


Figure 3.5: Selected Common Spatial Filters and Patterns for *Left hand*, *Right hand* and *Rest* tasks for subject g.

For each task we generate 59 CSP filter, though only some of them are prominent filters and can be used to gain higher discrimination between classes. To select the best filters we used $r^2 - value$ to evaluate each filter and decide to select or ignore it. We apply the following algorithm to select the best CSP filters to be used.

1. Compute the CSP Filters for each mental task.

2. Calculate the projection of the the decomposition matrices on the EEG signal.

3. Find the power spectrum for each row of each projection matrix.

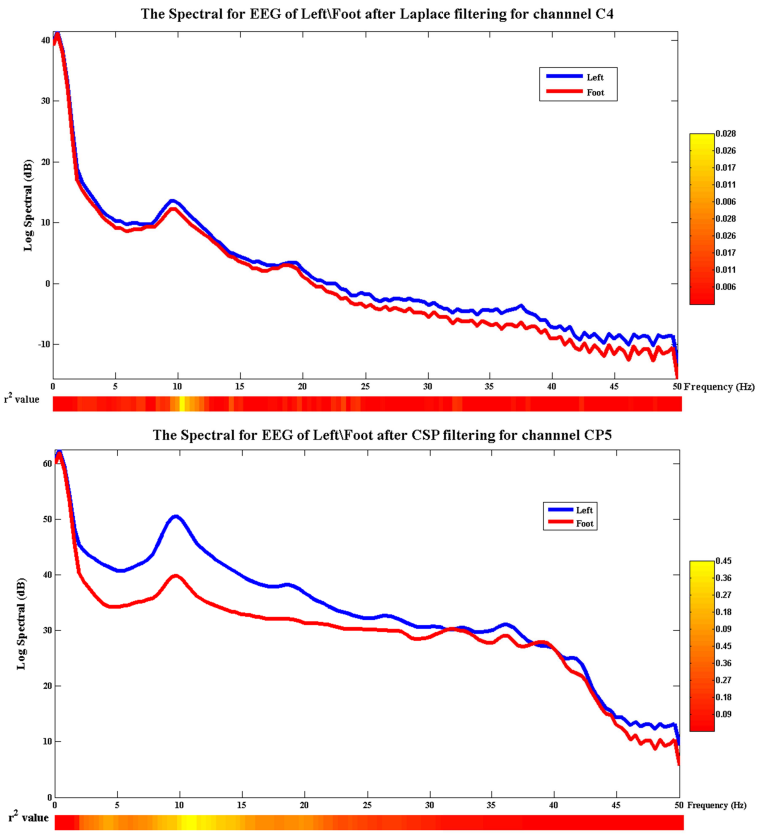


Figure 3.6: Spectra for the spatially filtered signal with Laplace and CSP spatial filters and the $r^2 - value$ between two tasks.

4. Calculate $r^2 - value$ between each two spectrum matrices resulted from the previous step.

5. Find power component for each channel for each matrix with the maximum $r^2 - value$.

6. Divide the vector of maximum $r^2 - values$ into two halves and find the index of the maximum for each half.

7. Select the first and last C filter from the CSP decomposition matrix for each mental task where C is the largest between the maximum index of the two

halves.

After applying the above algorithm we got the following CSP configuration for each subject.

Subject	**Tasks**	***C*-value**
a	Left Hand vs. Foot	2
	Left Hand vs. Rest	3
	Foot vs. Relax	24
b	Left Hand vs. Right Hand	1
	Left Hand vs. Relax	3
	Right Hand vs. Relax	1
f	Left Hand vs. Foot	2
	Left Hand vs. Relax	3
	Foot vs. Relax	3
g	Left Hand vs. Right Hand	3
	Left Hand vs. Relax	6
	Right Hand vs. Relax	6

Table 3.1: The number of selected CSP filters for each pair of tasks.

3.3 Feature Extraction

3.3.1 Power Density Estimation (PSD)

PSD is the most commonly used feature extraction method for sensorimotor mental tasks because μ, β, *and* γ rhythms have been strongly linked to motor function, both real or imaginary, e.g., left hand, right hand or foot movements. To extract PSD feature vectors from raw EEG recordings, the signal was first band-pass filtered then 'virtual channels' are extracted using CSP filters. Each of these virtual channels was then divided into non-overlapping windows of a hundred samples each. For each of these windows the power spectral density was then es-

timated using the Welch method, which in turn was set to use a 1 second hamming window with 30% overlap. The estimated power spectra were then divided into 2-Hz frequency bands where the power in each band was the average of the power in the corresponding PSD coefficients. Finally, these values are combined to form the PSD feature vector corresponding to the respective EEG window. Each time segment will have 11 features per selected CSP channel. Figure 3.7 shows topographic maps which depict the average power density in each frequency band. This figure was computed for the average power trials of all left hand movement mental tasks after the application of IIR and Laplace spatial filters. Figure 3.7 was based on Laplace spatial filter to show average band power density estimation over scalp because CSP filters linearly transforms the spatial channels into new set of channels; hence, we will not be able to use CSP channels to show power density over the scalp.

Some bands in the feature vector have better spatial information to describe the mental task than other bands in the feature vector, e.g., band number 5, 16-18 Hz, shows a spot on the sensorimotor area of Left hand. With CSP spatial filtering, the features will be more discriminative.

3.3.2 The Log Variance

Log variance feature extraction method exploits the variance of EEG trials to extract useful features for discrimination between classes. Log variance method is useful with CSP filtered signal where CSP increases the discriminability of labels, i.e., CSP is maximizing the variance of one class while minimizing the variance of other classes. Log variance method requires low computational power to ex-

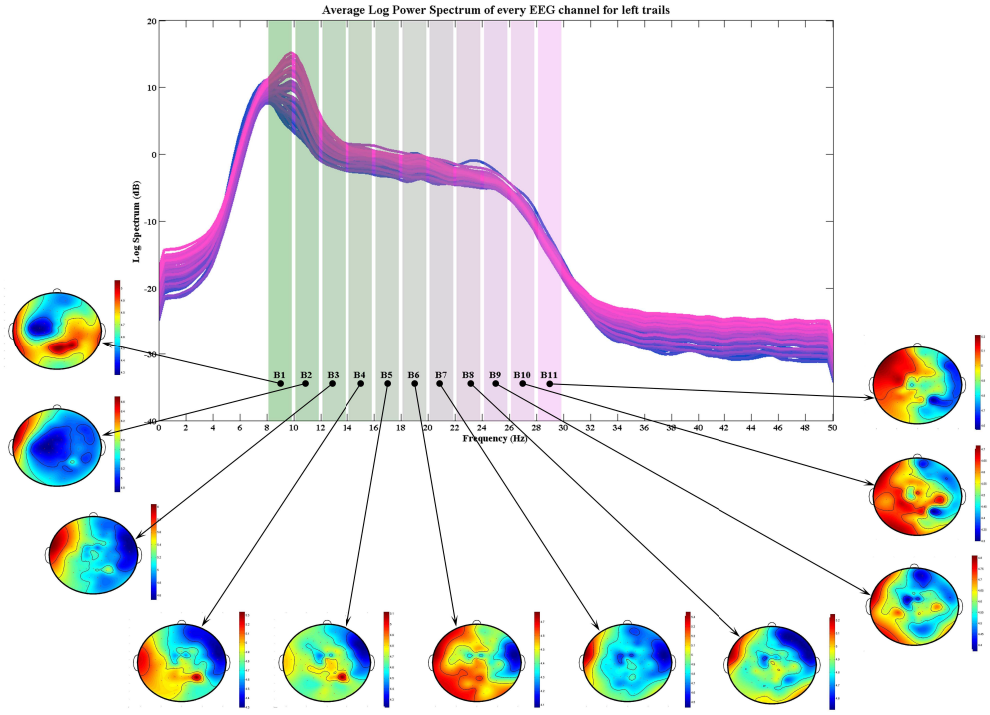


Figure 3.7: Average Log power spectrum for EEG signal of Left hand task trials and Average log band power distribution over scalp surface.

tract features; furthermore, the obtained feature vector is in minimum size because this method extracts one dimensional feature vector per channel. To extract log variance features, EEG signal was divided into non-overlapped equally size windows, 100 sample window size was used, then the variance of each window was computed and marginalized over all channel's variances. Finally marginalized variance vector was log-transformed. Equation 2.18 explains this method. The feature vector, obtained from training data, was used to train different classifiers. Then the classifier was used to obtain the final classification of the evaluation dataset.

3.3.3 Wavelet Packet Decomposition (WPD)

Wavelet Packet Decomposition (WPD) separates the EEG signal into two orthonormal subspaces A and B at each decomposition level. Applying $l - level$ WPD on an EEG signal, consisting of n samples, produces 2^l packets, where each packet has n samples. Figure 3.8 shows all subspaces in each level of WPD where the Daubechies order 4 mother wavelet is used.

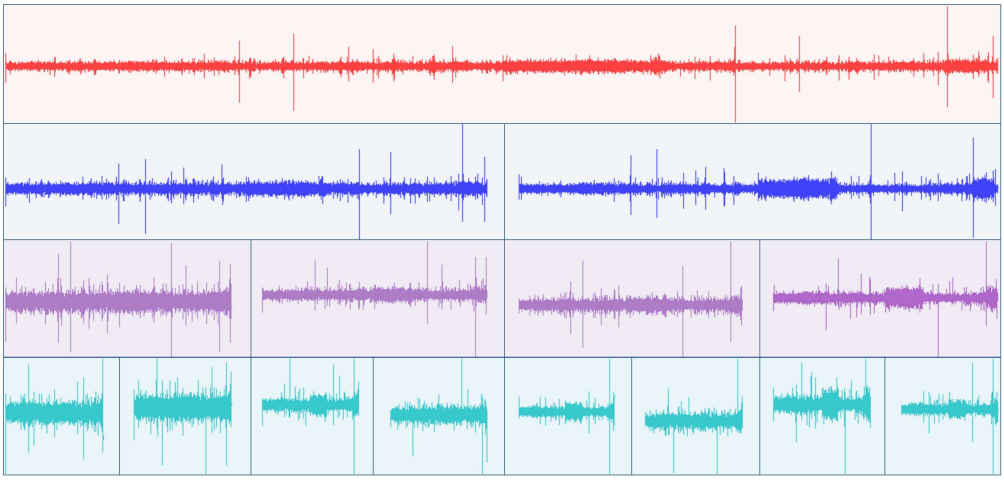


Figure 3.8: Wavelet packets in different levels using Daubechies order 4.

We used 3 levels to generate 8 packets per channel. To construct the feature vector, non-overlapping windows were used to average the wavelet packets which had the same length as the original EEG signal. Using more than 3 levels will increase the number of features exponentially and complicate the classification process. Figure 3.9 shows a set of features extracted from 8 packets by taking the mean of all samples in non-overlapping window.

The window size was chosen to be less than the minimum time taken to perform any task, we choose 1 second because the minimum time to perform a task in

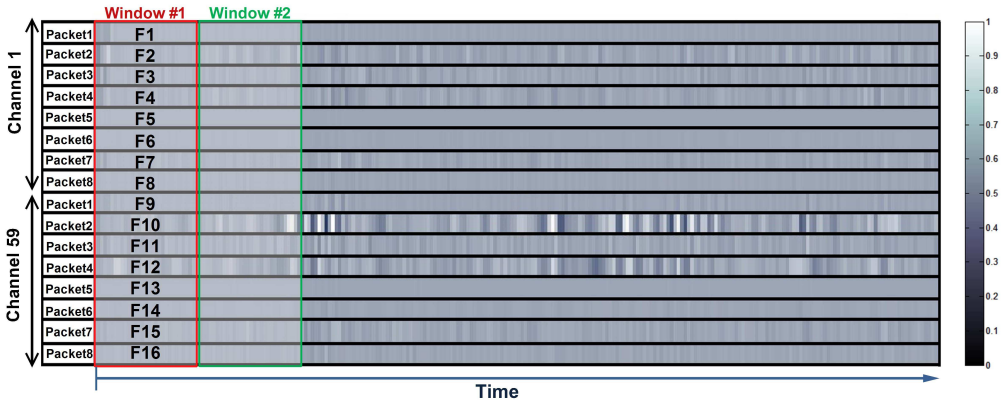


Figure 3.9: Wavelet Features extracted from a selected EEG channel using 3 levels WPD. The red and green windows represent averaging windows. Light and dark colors represent high and low WPD value respectively, as the color bar suggests.

the test data was 1.5 second. Choosing smaller window will increase the number of samples to be classified and it will introduce noise representation for the signal.

3.3.4 A Novel Feature Extraction Technique

Feature extraction is a critical component in any BCI algorithm. The choice of an appropriate feature extraction technique could mean the difference between extracting the relevant information embedded in the signal of merely amplifying the noise content. Many feature extraction methods have been used in BCI research. (PSD) is the most widely used technique and has been used in numerous studies. However, as mentioned before, it is far from perfect and is only capable of producing very modest accuracy levels. In searching for a novel feature extraction technique, we started by identifying a set of theoretical requirements which would have to be met by the proposed method. These were as follows:

1. Regardless of the specific technique used, it is important that the resultant

feature vector be insensitive to noise, as EEG signals are notoriously susceptible to contamination by a variety of environmental noise sources.

2. Frequency based: most widely studied brain signals are sensorimotor rhythms, which are inherently linear (frequency based). To capture these signals, a method based on the power spectra would be required.

3. Time-shift invariant: the feature vector should be time independent, because mental tasks are time independent and are not limited by the time window in the evaluation set.

4. Ability to cope with transient signals: as the phenomena being studied in BCI are inherently transient since commands such as 'left' and 'right' occur arbitrarily and are not necessarily required to be stationary. As such, the proposed feature extraction method must be able to effectively detect transient occurrences in the EEG time series.

Devising an entirely new feature extraction technique is beyond the scope of the thesis, but what we tried to do was to identify existing techniques which might satisfy these requirements. We found a similar area to BCI, which is voice recognition, that deals with noisy signals, e.g., recorded voice. Although EEG signals are much more complex and noisy than voice signals, they share some similarities, such as:

- Both signals have interference from different sources. For the voice signal, multi-speakers environment, e.g., the cocktail party problem is similar to multiple observed active brain regions. Both signals suffer from this interference and usually ICA is used to separate signal sources.

- Both signals carry underlying information within a certain band of frequencies where the information extraction process focuses on this frequency band. For example, sensorimotor activities fall within certain frequency bands, which is the case with the human voice.

- Both recorded signals have low SNR values. This is because of the very noisy environment in which both signals are typically recorded.

Guided by this intuition, we then searched for techniques which have been successfully applied in voice recognition, and, hence, were likely to meet all or most of the above requirements. We discovered that one of the most widely used and efficient methods in voice recognition appeared to fit the bill [62, 87]. This method is called Mel frequency cepstral coefficients (MFCC) which is based on cepstral analysis of the signal.

MFCC is implemented in three stages. Firstly, the power spectrum is computed. Next the power spectral is transformed using equation 2.22. Finally, the cepstrum is computed for the Mel scaled spectrum. The process of computing the power spectrum is similar to the PSD process where non-overlapping windows are taken and then the power spectrum was computed using the Welch method with hamming window of size of 1 second and overlapping by 30%. We used the whole resulting spectrum as input to the Mel scale transformer, equation 2.22. The process of computing the power spectrum was repeated. Finally the resulting quefrency was divided into 8 bins and the average cepstrum was taken as feature vector.

3.4 Summary

Preprocessing filtering is one of the essential stages to enhance the distinction between classes and to remove artifacts. In this thesis two basic preprocessing methods, temporal filtering and Spatial filtering, were used. The preprocessing parameters, CSP channels and the cutoff frequencies, were optimized based on $r^2 - value$. In this chapter, we discussed our implementation of four different feature extraction techniques (Log variance, PSD, WPD and MFCC) and how to use them to extract the most representative feature set. The next chapter will discuss and compare the classification results of all feature extraction and classification methods.

CHAPTER 4

Results and Discussion

This chapter presents and discusses the results of the experiments conducted to evaluate the proposed methods. A comparison between MFCC and a variety of other feature extraction techniques was performed, and each technique was also used with a range of different classifiers to eliminate classifier specific issues. The comparison between the methods of feature extraction was based on a number of different evaluation criteria: accuracy, mean square error and confusion matrix. Two datasets were used to validate the results: the first is dataset one of BBCI competition IV, and the second dataset was collected internally using the Emotiv-EPOC portable EEG device.

4.1 The Experimental System

Although a variety of feature extraction and classification methods were investigated, the same experimental procedure was used in all the experiments. The diagram in figure 4.1 summarizes the entire process. The test and training data was collected using cross-validation. The proposed BCI system composed of the following set of steps:

- Both training and test EEG recordings were bandpass filtered, using an IIR filter, to remove irrelevant frequencies.
- CSP filters were constructed (only the training data is used here to avoid "information leakage"). Then it was used to spatially filter training and test EEG recordings.
- A set of features was extracted. The feature extraction techniques used were Log Variance, PSD, Wavelets and MFCC.
- Principle Component Analysis (PCA) was used to reduce the dimensionality of the feature vectors. Appendix C explains the use of PCA in dimension reduction.
- The resulting data was then divided into training and test sets, which were then processed using the classification techniques mentioned before.
- Finally, a variety of scoring metrics were used to measure the performance of the system.

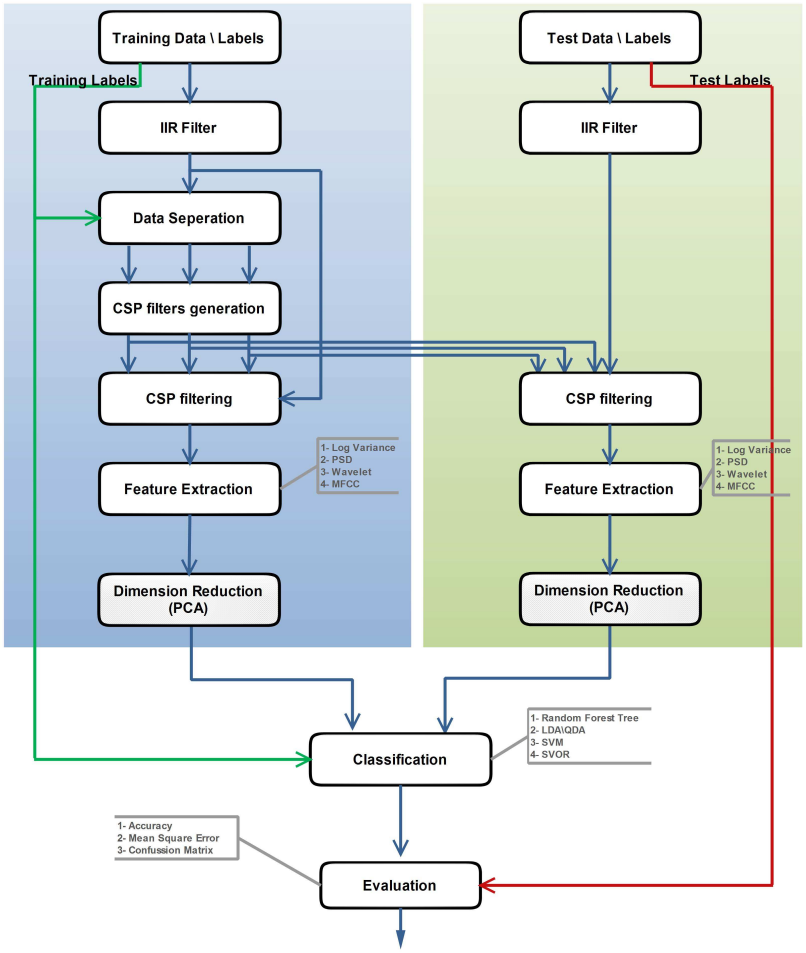


Figure 4.1: System block diagram.

4.2 Evaluation criteria

The performance of the BCI system was evaluated using three separate, but commonly used [72], criteria: mean squared error, accuracy and the confusion matrix. These will now be briefly discussed:

- Mean Squared Error (MSE): This evaluation method is particularly useful when attempting multiclass problems. As such, MSE is suitable for the present study as it involves multiple classes, and the costs of the various

misclassification errors are non-uniform, e.g., if the specified task was to imagine a movement in the left hand, and the classifier detects an imagined movement in the right hand, then it is a more serious error than if it was classified as the resting state. Equation 4.1 shows how this is calculated.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \tag{4.1}$$

Where N is number of samples in test data, Y_i and $\hat{Y}_i$ are the true and predicted label for sample i respectively.

- Accuracy (ACC): this is the most common evaluation criteria in BCI research. used in BCI research, and is defined as follows:

$$ACC = \frac{1}{N}\sum_{i=1}^{M} n_i \tag{4.2}$$

Where n_i is the number of correctly classified samples within class i, M is the number of classes.

- Confusion Matrix: describes the relationship between the true and predicted labels. The diagonal elements in the confusion matrix represent the correctly labeled samples and the off-diagonal elements shows the wrongly classified labels. Confusion matrices give a clearer picture of the error distribution. However, unlike MSE and accuracy, it does not provide a single score with which the goodness of the procedure tested can be evaluated.

4.3 BBCI competition IV Dataset I

This section will discuss and compare the performance of the different feature extraction methods when applied to the BBCI competition IV Dataset I. As described in section 3.1, the first dataset of the BBCI competition IV is a multiclass and asynchronous problem.

4.3.1 Evaluation Process

Dataset I from the BBCI Competition IV was used, which is a dataset for multiclass asynchronous BCI, to assess and compare the different feature extraction methods. The dataset contains two sessions; note that, as was mentioned in section 3.1, these two sessions were conducted according to different experimental protocols. We expected to have two different data distributions between the two sessions. To confirm that this is indeed the case, multidimensional two-sample Kolmogorov-Smirnov test was performed on the two sessions, this technique has been used to check the IID assumption [68]. The test was modified from the two dimensional Kolmogorov-Smirnov test in [49]. The multidimensional two-sample Kolmogorov-Smirnov test was performed on the two five-dimension PSD feature vectors, which were extracted from training and evaluation sessions. The result confirms that the datasets belong to different distributions. Table B.1 shows the outcome of the Kolmogorov-Smirnov test, which includes the 2^5 possible ordering of dimensions and their p-value. As it is shown all p-values are less than 0.05, this value is commonly used in literature, allowing the null hypothesis to be rejected. Appendix B explains the multidimensional two-sample Kolmogorov-Smirnov test. To deal with this problem, 4-fold cross-validation was used over a

combination set of the two sessions after removing all the samples for transient periods as stated in the competition [1] , i.e., the data from both sessions are consolidated and 4-fold cross-validation is used to allow testing on it. This will allow the classifier to recognize both distributions. We benchmark the results with [37], [2] and [36] which used the same evaluation method.

4.3.2 Results

All feature extraction methods, $Log-Var$, PSD, $Wavelet$, $MFCC$, were applied to all the human subjects, a, b, f and g, from BBCI competition IV. A fixed window size of 1 second was used to allow fair comparisons to be made between the feature extraction methods. Figure 4.2 shows the mean 4-fold cross-validation classification accuracy for each feature extraction method and using the LDA classifier. All data from both sessions were used to compute the accuracy in this figure. Figure 4.2 shows:

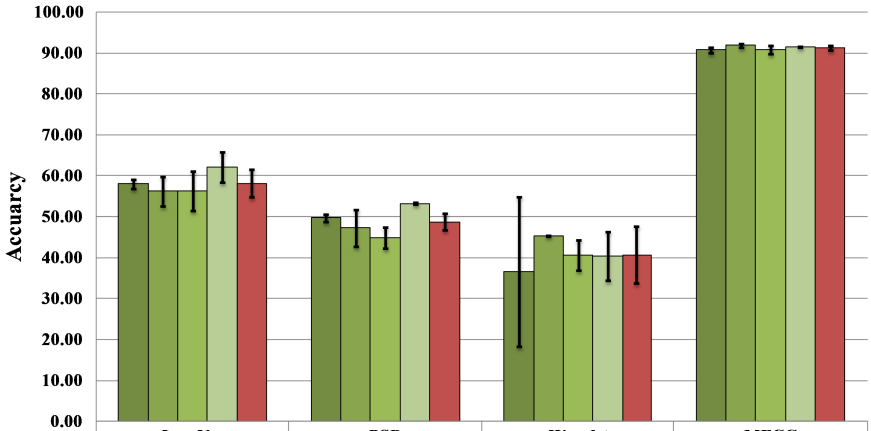
1. MFCC outperforms other feature extraction methods. It scored in average over all subjects of about 91% accuracy.

2. Other feature extraction methods, Log-Var, PSD and Wavelet, achieve 58%, 48% and 40% accuracy, respectively.

3. The variance in classification accuracy over the four folds is very small (on the order of 10^{-5}) which shows the consistency of the algorithms.

These results seemed to indicate that MFCC was able to extract information that was very relevant to the classification of the EEG signals. To investigate fur-

[1] http://www.bbci.de/competition/iv/desc_1.html



	Log-Var	PSD	Wavelet	MFCC
Subj-a	58.01	49.71	36.48	90.70
Subj-b	56.19	47.25	45.32	91.81
Subj-f	56.28	44.89	40.50	90.79
Subj-g	62.17	53.23	40.36	91.48
Average	58.16	48.77	40.67	91.20

Figure 4.2: 4-Fold cross-validation accuracy for all subjects using LDA classifier, both sessions were used. The 10^5 variance for the 4-fold is presented also.

ther, the other evaluation methods discussed previously were then applied. Table 4.1 shows the confusion matrix for MFCC method using LDA classifier.

From the confusion matrix, it can be seen that most of the misclassified samples fall in the idle mental task. For example, a misclassified right hand imaginary movement is far more likely to be classified into the idle class than into the left hand movement class. This helps to improve the reliability of the method, since in a BCI setting the idle class would translate to "No Operation (NOP)", probably the safest option if the intent of the user is unclear. For example, for a wheelchair application it is more acceptable to mistranslate the $left$ command as $idle$, then as $right$, for obvious reasons.

This observation can be captured by using MSE which gives higher penalty for critical misclassification than a misclassification from idle task. Figure 4.3 shows

Subj-a					Subj-b				
		Predicted Labels					Predicted Labels		
		Left	Idle	Foot			Left	Idle	Right
True Labels	Left	**17672**	2829	9	True Labels	Left	**20421**	2408	9
	Idle	809	**35737**	805		Idle	896	**34391**	977
	Foot	3	2855	**17911**		Right	43	2335	**19980**

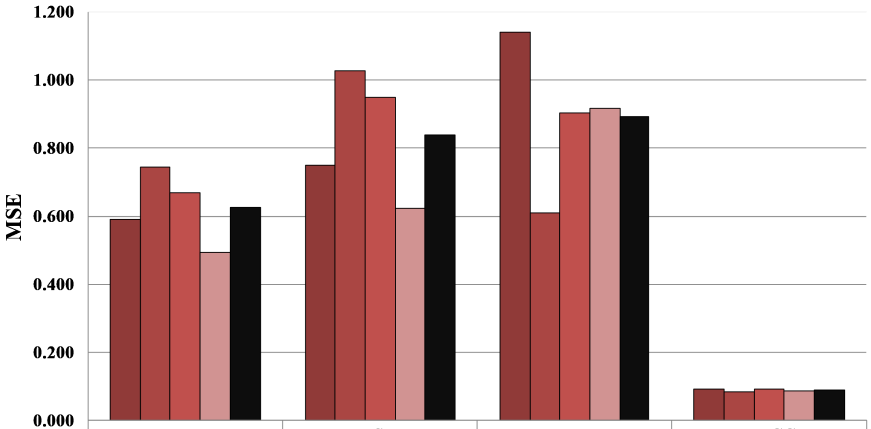
Subj-f					Subj-g				
		Predicted Labels					Predicted Labels		
		Left	Idle	Foot			Left	Idle	Right
True Labels	Left	**17597**	2386	0	True Labels	Left	**18522**	2217	17
	Idle	1107	**37055**	656		Idle	943	**37443**	827
	Foot	15	3149	**17405**		Right	9	2852	**17770**

Table 4.1: Averaged confusion matrix for every subject. 4-fold cross-validation on both sessions is used here.

the average MSE over 4-fold cross-validation MSE for all subjects for the four feature extraction methods. All MSE measurements in figure 4.3 were collected using LDA classifier and a window size of 1 second.

4.3.2.1 The Window Size Variation

One of the most important parameters in our system is the size of the nonoverlapping window which is used to compute the feature vector within that window. To investigate its effect on the performance of the system, a range of different window sizes were used and the classification accuracy was recorded in each instance. Figure 4.4 and 4.5 show the accuracy and MSE, respectively, changes with the window size, LDA was used as a classifier and we took the average accuracy over the four folds and all subjects. Only the training data was used.



	Log-Var	PSD	Wavelet	MFCC
Subj-a	0.592	0.751	1.140	0.093
Subj-b	0.746	1.029	0.611	0.084
Subj-f	0.670	0.948	0.904	0.093
Subj-g	0.494	0.623	0.917	0.086
Average	0.625	0.838	0.893	0.089

Figure 4.3: 4-Fold cross-validation MSE for all subjects using LDA classifier, both sessions were used.

Increasing the window size decreases the accuracy for MFCC, PSD and Log-Var. This is because the non-parametric power estimation methods such as Welch method, which is used to estimate the power in PSD and MFCC feature extraction methods, suffers from spectral leakage effects. This spectral leakage is relative to the number of samples in the window, i.e., with a small number of samples, the leakage effect will be larger. Though spectral leakage limits the quality of the power estimation, it removes weak signals from the spectrum and then increase the classification accuracy [65].

4.3.2.2 Classifiers Variation

Next, the experiments with the MFCC features were repeated using four different classifiers to rule out any classifier or implementation specific effects. Figure

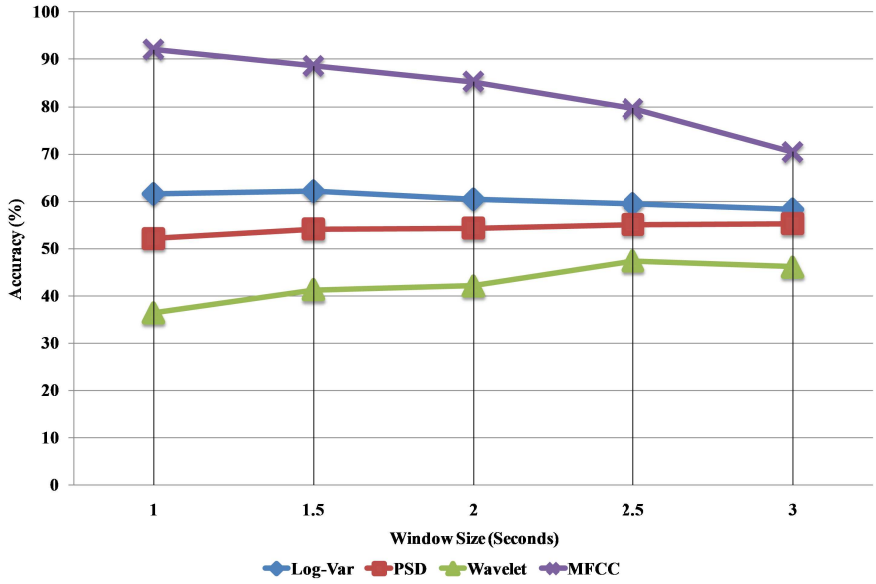


Figure 4.4: Mean 4-Fold cross-validation accuracy with a varying window sizes.

4.6 and 4.7 show the average cross-validation classification accuracy and mean square error (MSE), respectively. In both cases, MFCC features were used and only data from the first session was used. The accuracy was consistent for random forest ensemble, support vector machine and linear discrimination analysis classifiers. Surprisingly, the ANN had the worst performance, producing accuracy values which were significantly lower than those obtained using the other classifiers. The scaled conjugate gradient backpropagation was used to train the neural network which is much faster than Levenberg-Marquardt -LM- training function. Although LM showed better accuracy, around 44.4% comparing with 28% using scaled conjugate gradient, it also took a long time to converge (around 5 hours [2] were required to train the neural network four times, one for each fold from the cross-validation, for one subject using one feature extraction method). Be-

[2] Average CPU utilization is 90% of 2 processors (2.8GHz), 6 cores each and 24GB RAM

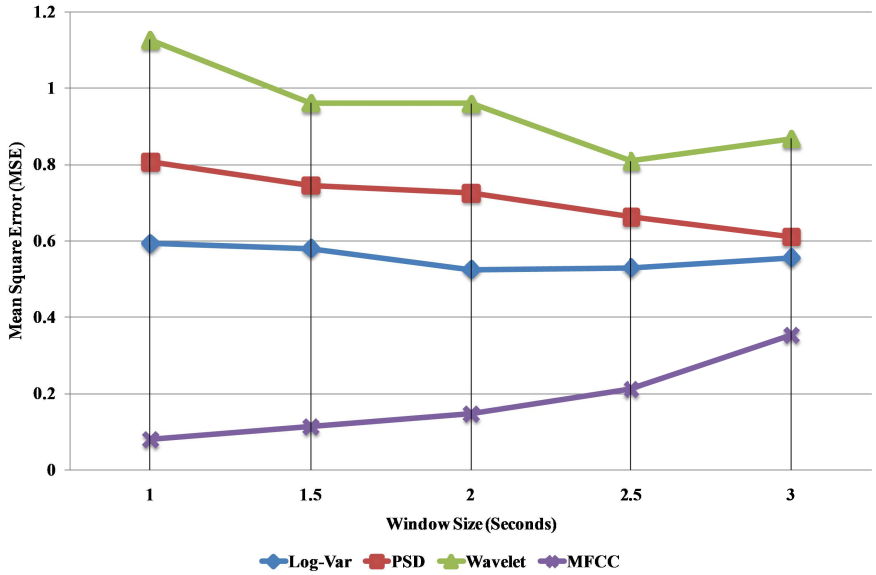


Figure 4.5: Mean 4-Fold cross-validation mean Square error (MSE) with a variety window sizes.

cause of the intensive computational time needed by LM training algorithm, LM was not used in this thesis. However, with the exception of the ANN, there were no significant changes in the performance when different classifiers were used, which provided some confidence that the results were independent of the exact classification algorithm used.

4.3.2.3 Result Validation

Results from [2, 37, 36] were used to benchmark our implementation. In [2, 37, 36] band power feature extraction and LDA for classification were used. A 4-fold cross-validation was performed on the first session from the BBCI competition IV dataset I to measure the accuracy. Table 4.2 shows a comparison between the results from [2, 37, 36] and our implementation of band power and LDA. The

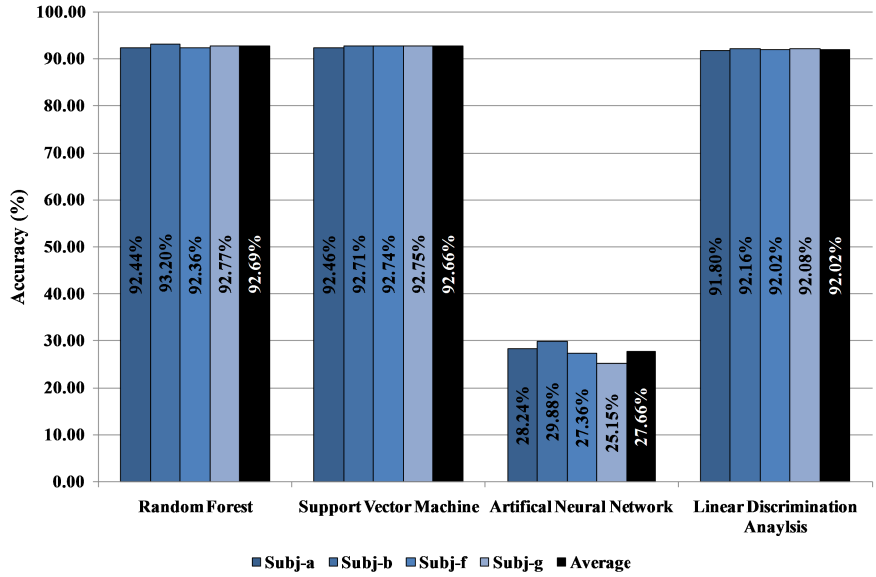


Figure 4.6: Average cross-validation accuracy using four classifiers and MFCC feature extraction. Only first session is used.

results are alike with small differences because the optimization algorithm which was used by [2, 37, 36]. We believe the minor differences, less than 7% accuracy difference, in the results was because of:

- [2, 37, 36] used swarm optimization to find the best possible system parameters. In our case, the far less comprehensive $r^2 - based$ procedure described in section 3.2.1 was used. This was a far simpler approach but at the same time it still yielded results which were very impressive, and is less likely to suffer from over-fitting.

- [2, 37, 36] used Sequential Floating Forward Search (SFFS) for feature selection, whereas in this study PCA was used for dimensionality reduction.

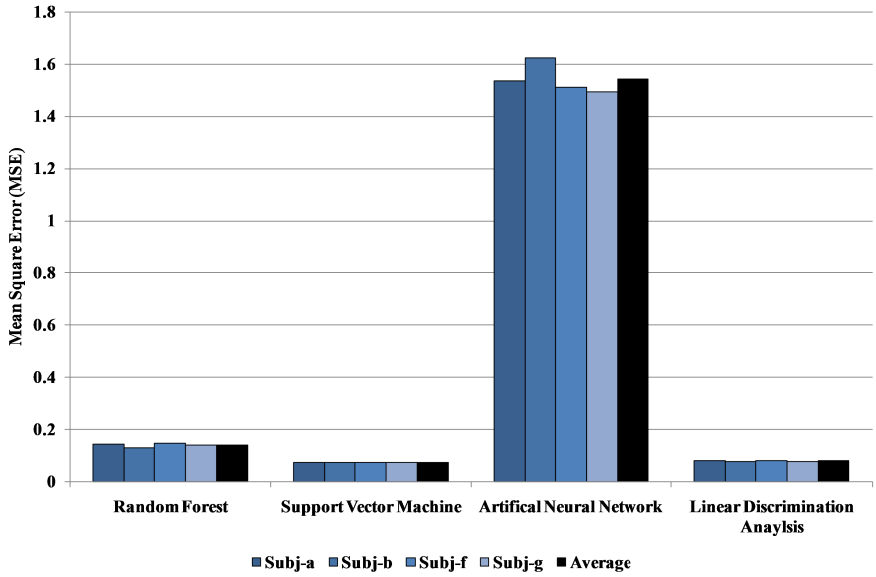


Figure 4.7: Average cross-validation mean square error (MSE) using four classifiers and MFCC feature extraction. Only the first session is used.

4.3.2.4 Computational requirements

All of the experiments discussed here were conducted offline, i.e., on pre-recorded data. As such, the computational cost of MFCC has not been discussed much in this thesis thus far. Furthermore, since MFCC has been used extensively in speech recognition, it was assumed that it would have acceptable performance characteristics for BCI. However, in an actual (online) BCI setting, performance is very important as it would strongly affect the responsiveness of the system. To address this, we compared the performance of an MFCC-based BCI system to that of other, more conventional approaches. The results are shown in figure 4.8, which shows the computational time for a set of systems consisting of cross-validation, data pre-processing, feature extraction and processing and classification stages. It can be seen from figure 4.8, that MFCC requires greater computational effort than

Subject	SFFS-PSD-LDA [2, 37, 36]	PCA-PSD-LDA (our Implementation)	The difference
Subj-a	58.61%	53.06%	5.55%
Subj-b	52.79%	48.66%	4.13%
Subj-f	58.45%	52.60%	5.85%
Subj-g	61.26%	54.86%	6.40%
Average	57.78%	52.30%	5.48%

Table 4.2: Average accuracy comparison between the implementations given in [2, 37, 36] and our implementation. 4-Fold cross-validation on the first session of BBCI competition IV dataset I was used in both implementations.

other feature extraction methods. MFCC needs twice the time that is required by PSD; this is because each MFCC feature vector requires the Welch method to be used twice, while with PSD it is only used once. Among all classifiers LDA scored the lowest computational time, because LDA is a simple classifier comparing to RF, SVM and ANN.

4.4 Tests using internally collected data

This section will discuss the feasibility of using "retail grade" EEG devices for asynchronous BCI tasks. For this purpose, the Emotiv EPOC EEG device was used (more about Emotiv can be found in appendix A). It was attempted to follow an experimental protocol which was as similar to that used in the BBCI computation IV data set 1 as possible. Three healthy subjects (table 4.3) volunteered to participate in the experiment, figure 4.9 shows the three subjects the sessions were underway.

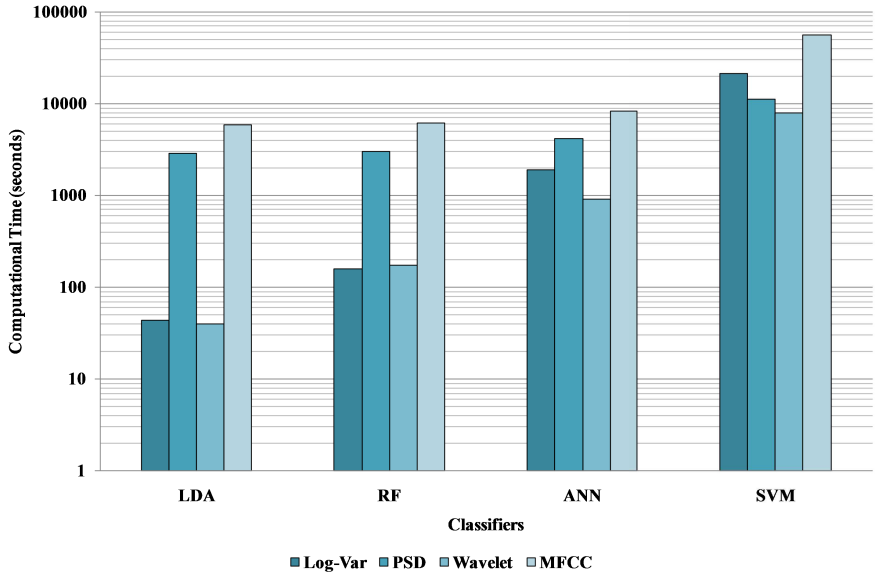


Figure 4.8: Average computational time (in seconds) needed by BCI system using 4-fold cross-validation and both data sessions of all subjects which include for subjects $a, b, f and g$ around 52 min (314517 sample), 54 min (325837 sample), 52 min (317478 sample) and 53 min (322394 samples) , respectively.

4.4.1 Experimental setup

The experiment was conducted in two sessions. In the first session, visual cues were used while acoustic cues were presented during the second session. The experiment was done as follows:

1. The Emotiv-EPOC was fixed on the head of the subject until it get clear EEG signal and the subject feels comfortable wearing the device. Because

Subject	Gender	Age	Handedness
Subj-A	Male	24	Left-Handedness
Subj-H	Male	26	Right-Handedness
Subj-J	Male	28	Right-Handedness

Table 4.3: Basic information about the three subjects.

Figure 4.9: The three subjects are performing experiment.

of the usability of this device, this process took less than 5 minutes per subject.

2. A short demonstration, 3 mental tasks for each session, was performed to ensure that the subject had a clear understanding of the required tasks.

3. The first session of the experiment started and the subject was asked to perform 200 mental task for 4 seconds each. Each task was also followed by a rest period, which was for the same length of time. The 200 mental tasks were randomly chosen between left and right hand movements. Figure 4.10 shows the visual cues which was used in the first session.

4. The second session was performed after a short break, length of the break

depends on the subject. The second session depends on acoustic cues, where the subject was asked to perform 100 mental tasks with a rest after each task. Each mental task last for 1.5 to 8 seconds where its length was selected randomly. The tasks were selected randomly from left and right hand movements.

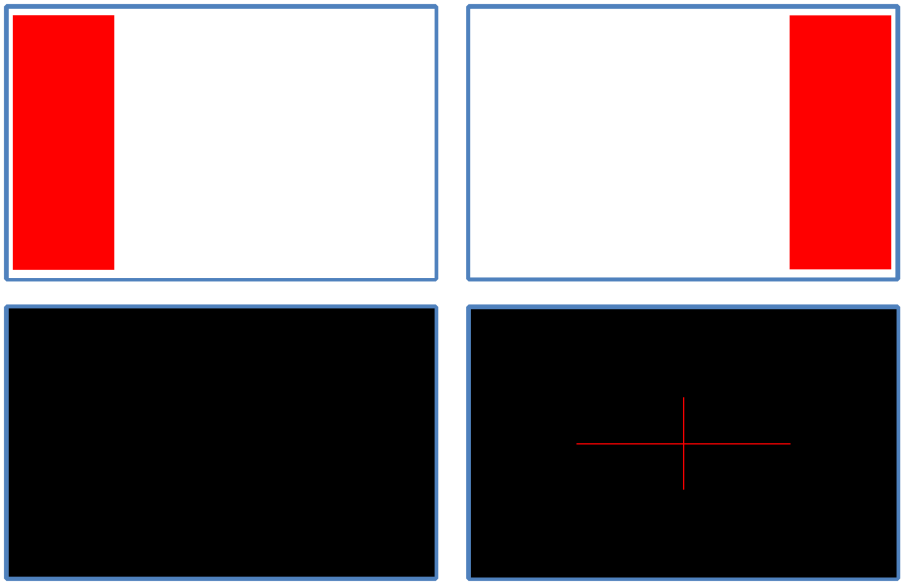


Figure 4.10: The four screens from the first session of the experiment. The white screens with red rectangle request from subjects to perform left and right hand movement, the black and cross screens request from subjects to relax.

4.4.2 Results

The system was tested with the three healthy subjects to test the robustness of the system. All feature extraction methods were applied on the experimental data to allow them to be compared. The average power distribution for each mental task over the skull was shown in figure 4.11. These topographic maps show clearly

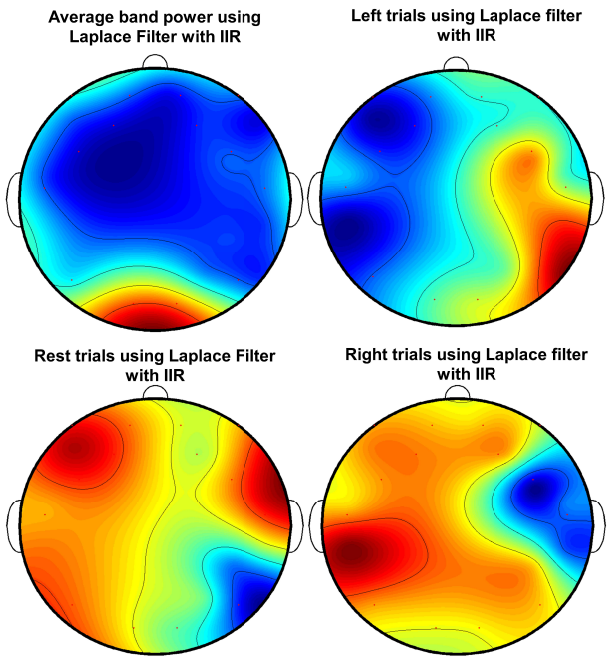


Figure 4.11: Topographies maps for subject-A. Upper-left map shows average power distribution of all trials, upper-right map shows average power distribution of all left trials, lower-left map shows average power distribution of all rest trials and lower-right shows average power distribution of all right trials.

the concentration of the power on the related cortex center, e.g., for the right hand imagery movement, most of the power concentrates over the left side cortex area.

Again, the MFCC features resulted in very high accuracy, around 90%, scores when tested on this data. Figure 4.12 shows a comparison between the four feature extraction methods according to the accuracy measurements, LDA was used for classification. Also, note that while the performance of the conventional feature-sets experienced significant degradation when applied to this data (as a result of the presumed lower quality of the data) the performance of the MFCC features dropped only slightly or, in some cases, even showed an improvement. This trend, which is especially clear when using the MSE metric, seems to confirm our initial

intuition that the MFCC features would provide improved noise resistance.

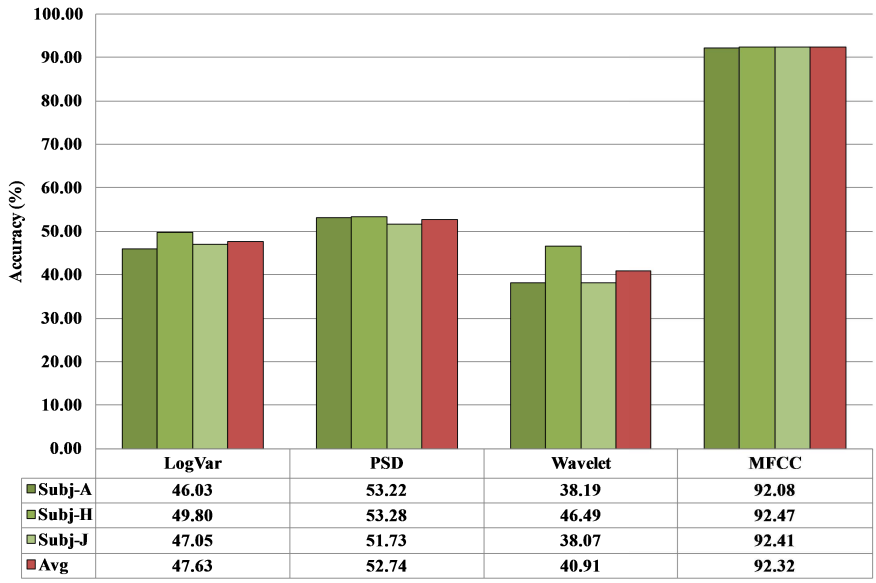
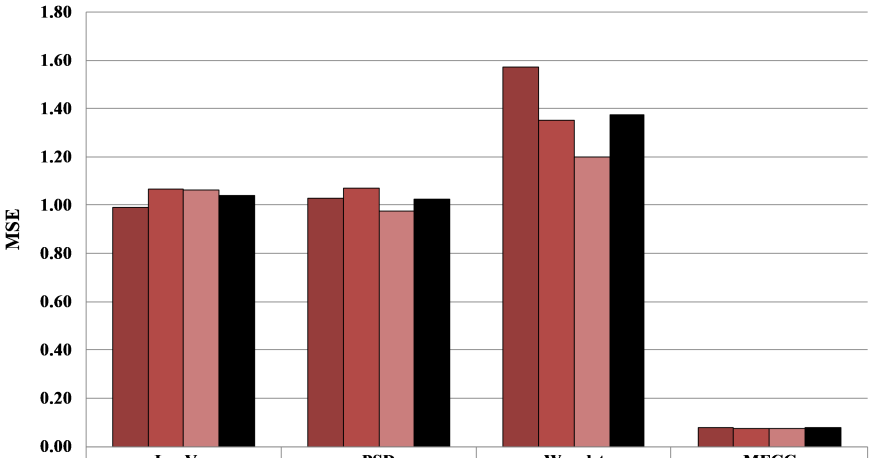


Figure 4.12: Accuracy of all feature extraction methods from the collected data of all volunteers. 4-fold cross-validation was performed on the data from both sessions.

MSE was measured to confirm the results. Figure 4.13 shows the average MSE for all subjects for all feature extraction methods using LDA as classifier. Figure 4.13 confirm the reported accuracy in figure 4.12, where MFCC scored the lowest MSE, around 0.08, among all feature extraction methods.

Results proved that Emotiv can be used to achieve high classification accuracy for asynchronous BCI problems. Moreover, Emotiv is considered as one of the best inexpensive devices in term of usability according to [74].



	LogVar	PSD	Wavelet	MFCC
Subj-A	0.99	1.03	1.57	0.08
Subj-H	1.07	1.07	1.35	0.08
Subj-J	1.06	0.98	1.20	0.08
Average	1.04	1.02	1.37	0.08

Figure 4.13: MSE of all feature extraction methods from the first session of collected data of all volunteers. 4-fold cross-validation was used.

4.5 Summary

The first part of this chapter compared the MFCC feature extraction technique to other commonly used feature extraction methods. MFCC outperform other feature extraction. Also, a variety of classification methods were used and they achieved consistent accuracies, except ANN. The second part of this chapter demonstrated the possibility of using inexpensive EEG devices for BCI problem. The experiments conducted in the first part (with the BBCI competition dataset) were repeated on the internally collected data, and as far as we could tell, the results were very similar. This was particularly impressive given the low cost and specifications of the Emotiv EPOC compared to the BBCI data, which were collected using medical-grade equipment (BrainAmp MR plus amplifiers and a Ag/AgCl electrode cap) and by trained professionals.

CHAPTER 5

Conclusion and Future Work

This chapter discusses the main findings and contributions of this research, and presents our recommendations for future work.

5.1 Conclusion

This work aims to achieve the set of objectives which were specified at the beginning of this thesis. In the following, a review of the objectives and achievements of this thesis:

- **Implement a selection of commonly used BCI feature extraction methods.** Three such techniques were tested: Log Variance, PSD and the Wavelet Transform.

- **Propose a novel technique for BCI feature extraction aimed at addressing some of the shortcomings of the conventional approaches.** MFCC, a feature extraction technique commonly used in speech recognition, was tested and was able to produce very high classification accuracies. Note that, based on our review of the literature, this appears to be the first time that MFCC has been used for Asynchronous BCI.

- **Test the complete system using a variety of different classifiers to ensure broad applicability.** SVM, RF and LDA were used to classify the MFCC feature vectors, and they showed consistent performance with a variation of the computational time, where LDA required the least computational time. ANN was tested too, but it showed low classification accuracy and was also computationally more expensive. This was justified by the training algorithm of the ANN which was used to train ANN. The training function has huge impact on the accuracy and the training time too.

- **Validate the developed methods using two separate datasets.** The first set is publicly available data, which should allow us to benchmark our results against previous studies; the second set is an internally collected dataset, which was recorded using a basic, low-cost EEG device. The publicly available dataset, BBCI Competition IV dataset I, was selected to test the developed system. Also, Emotiv EPOC EEG recording device was used to internally collect our dataset. Both datasets allowed us to study the robustness of the developed method and the feasibility of using low-cost devices for asynchronous multiclass noninvasive BCI.

5.2 Future Work

The developed system in this thesis raises a lot of questions for more investigations. This section will present some of these question and suggestions:

- Online BCI algorithms: online BCI algorithms allow for real time applications. The developed system in this thesis was based on offline algorithm, where the training and testing data were processed offline. Adapting the developed system to online BCI system will bring up some challenges such as the real time classification process and the imbalanced labels problem. Such problems were studied before in BCI and machine learning field and there are some proposed solutions but those solutions need to be investigated and adjusted to be suitable for online asynchronous multiclass BCI problems.

- Parameter optimization: the system developed in this thesis includes effective parameters, such as the cutoff frequencies of the bandpass IIR filter, the number of the selected CSP channels and the number of quefrency bins of MFCC. Such parameters need to be optimized for each subject to find the optimal value of these parameters. In this thesis, simple optimization was performed to select the value of each parameter. More advanced optimization methods can be used to select the optimal value of the system's parameters, which could further increase the classification accuracy.

- Feature selection and reduction: most of the feature extraction methods used produced a large number of features, which in turn required more computational power to run efficiently. In this work, PCA was used for feature reduction methods but more advanced feature reduction can be used, which

would hopefully reduce information loss. Feature selection, such as mutual information, can be used to select the most representative feature combination.

APPENDIX A

The Emotiv System

The Emotiv system is relatively a new personal interface for brain computer interface. The Emotiv system consists of the Emotiv EPOC which is neuro-signal acquisition and a processing wireless neuroheadset, and software suit which process and analyzes the data. The Emotiv EPOC headset contains 14 EEG electrodes, figure A.1 shows the electrodes locations map. Also it contains a gyroscope which can detect head orientation. Emotiv EPOC communicates with specific software using wireless communication channel at 2.4GHz band, a USB dongle and an embedded 12 hours lithium battery are provided within the kit. Figure A.2 shows a subject wearing the headset.

Emotiv offers six different software kits with the headset each kit is suitable for different cluster of users. Each kit contains different set of software. The

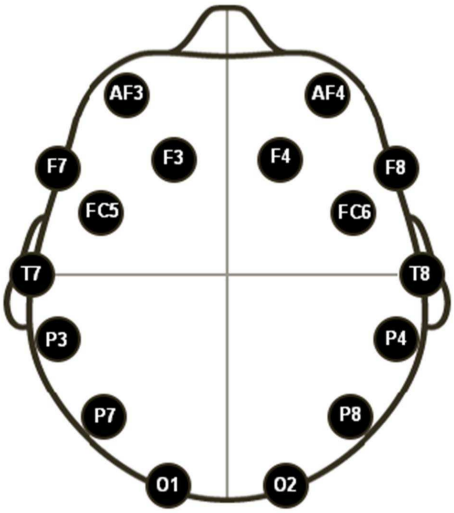


Figure A.1: EPOC electrodes map.

most suitable one of researchers is the research kit which includes Control Panel, EmoComposer, EmoKey, TestBench and Raw EEG data API. A short description of the most important software follows:

- Control Panel: is a graphical user interface with embedded functionalities like preprocessing, classifying input, providing feedback to the user, expression and emotion detection and mouse emulator. Another main function for control panel is to communicate and operate the EPOC headset and detect the signal strength. We used this software to communicate with EPOC headset. Figure A.3 shows a snapshot of the control panel software.

- TestBench: raw signals can be recorded using this software and professional process can be done using this software, e.g, FFT and power spectral density estimation. We used this software to record the EEG signal and send it to our model in Matlab format. Figure A.4 shows a snapshot of TestBench

Figure A.2: Subject-J is wearing EPOC headset and trying to perform some mental tasks.

software.

More information can be found in the Emotiv website www.emotiv.com.

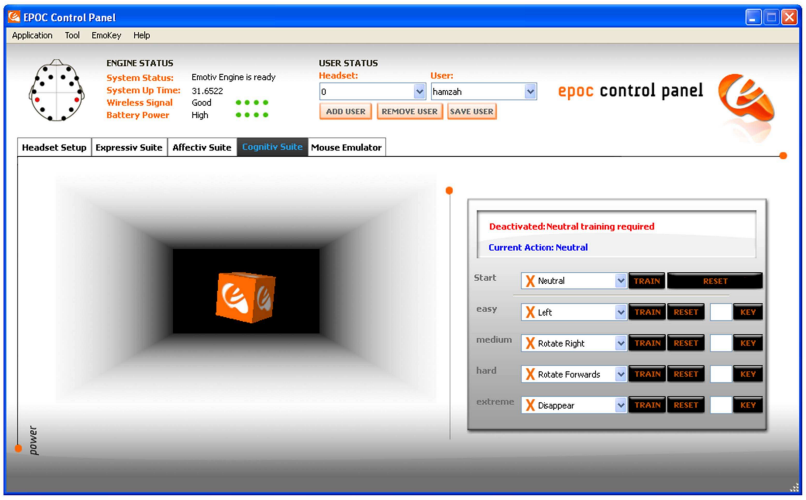


Figure A.3: a snapshot of the control panel window.

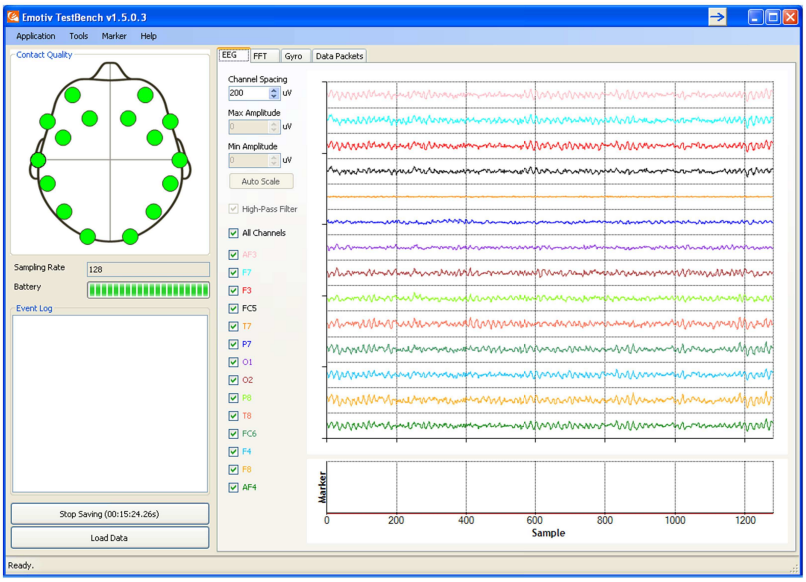


Figure A.4: a snapshot of the TestBench window.

APPENDIX B

Multidimensional two-sample Kolmogorov-Smirnov test

K-S test is a nonparametric test which can be used to compare the distribution of two sets of samples. In this thesis, we used K-S test to find out if a two sets of samples belong to the same distribution or not, i.e., follow the IID assumption or not. Basically, K-S test uses the maximum vertical distance between the two cumulative distribution function (CDF) of the two sets. As figure B.1 illustrates the maximum difference between two CDF, notes distance D which represents the maximum difference between the two CDFs.

Let us explain a little about multi dimensional two samples K-S test. Let us assume $H_0 : F = G$, and $H_1 : F \neq G$ where H and G are the CDF of the two samples. The maximum distance between two CDFs is defined as in equation below.

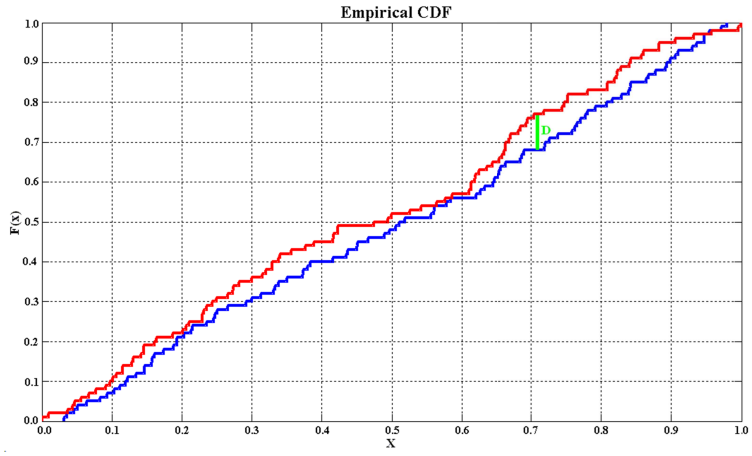


Figure B.1: CDF of two random variables of different distributions.

$$D_{mn} = \sqrt{\left(\frac{mn}{m+n}\right)} sup_x \left|F_m(x) - G_n(x)\right| \tag{B.1}$$

The decision function of accepting the null hypothesis is as in equation below.

$$\delta = \begin{cases} H_0 & : \quad D_n \leq c \\ H_1 & : \quad D_n > c \end{cases} \tag{B.2}$$

The decision is based on the c threshold which can be calculated based on the following equation.

$$Y(c) = 1 - \alpha \tag{B.3}$$

where, α threshold set to be 0.05 as a default value, and c can be obtained by the inverse of the following function,

$$Y(c) = 1 - 2\sum_{i=1}^{\infty}((-1)^{i-1} e^{-2i^2 c} \tag{B.4}$$

We need to consider all possible combinations for multidimensional samples. And choose the maximum D_n resulting from 2^d combinations, d is the number of dimensions.

For mathematical proves of the K-S test, refer to [63].

Combinations	**P-Value**	**Accepted or Rejected**
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 \leq V$	0	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 > V$	0	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 \leq V$	0.0001	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 \leq V$	0	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 > V$	0.0464	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 \leq V$	0	Rejected
$D_1 \leq X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 \leq V$	0.0014	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 > V$	0.0036	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 \leq V$	0.0012	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 > V$	0.0015	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 \leq V$	0	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 > V$	0.0002	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 \leq V$	0	Rejected
$D_1 \leq X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 \leq V$	0	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 \leq V$	0.0001	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 \leq V$	0.0232	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 > V$	0.0093	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 \leq V$	0.0003	Rejected
$D_1 > X \wedge D_2 \leq Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 \leq V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 \leq W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 \leq V$	0.0002	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 \leq Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 \leq V$	0.0001	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 \leq W \wedge D_5 > V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 \leq V$	0	Rejected
$D_1 > X \wedge D_2 > Y \wedge D_3 > Z \wedge D_4 > W \wedge D_5 > V$	0	Rejected

Table B.1: P-Value of the K-S Test of each dimensional order.

APPENDIX C

Principal Component Analysis

Principal Component Analysis (PCA) is the most commonly used dimension reduction technique. PCA is an orthogonal linear transformation which projects the data on a new coordinate system. The PCA algorithm is presented as follows:

- Compute the mean of the data. $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$. N is the number of data points, and x_i is the ith data point.
- Compute the covariance matrix of the data. $\Sigma = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$.
- Find the eigenvalues λ and eigenvectors v of the covariance matrix Σ.
- Select M eigenvectors with the largest eigenvalues to be the principal components. M represents the number of new dimensions.
- Find the projection of data on the selected principal components.

APPENDIX D

Abbreviations

ALS Amyotrophic Lateral Sclerosis

CNS Central Nervous System

AP Action Potential

ERD Event-Related Desynchronization

ERS Event-Related Synchronization

ERP Event-Related Potential

MRP Movement Related Potential

VEP Visual evoked potential

ANC Activity of Neural Cells

ECG Electrocardiography

EEG Electroencephalography

EMG Electromyography

MEG Magnetoencephalography

SQUID Superconducting Quantum Interference Device

PET Positron Emission Tomography

EGG Electrogastrography

EOG Electrooptigraphy

fMRI functional Magnetic Resonance Imaging

fNIR functional Near InfraRed

MSR Magnetically Shielded Room

SNR Signal to Noise Ratio

BCI Brain Computer Interface

BMI Brain Machine Interface

BBCI Berlin Brain Computer Interface

FIR Finite Impulse Response

IIR Infinite Impulse Response

ICA Independent Component Analysis

PCA Principle Component Analysis

CSP Common Spatial Pattern

PSD Power Spectral Density

MFCC Mel Frequency Cepestral Coefficients

ANN Artificial Neural Network

LM Levenberg Marquardt

WPD Wavelet Packet Decompositions

IID Independent and Identically Distributed

CDF Cumulative Distribution Function

SFFS Sequential Floating Forward Search

Bibliography

[1] Berlin Brain-Computer Interface. http://www.bbci.de, April 2011.

[2] N. Al Moubayed, B.A.S. Hasan, J.Q. Gan, A. Petrovski, and J. McCall. Binary-sdmopso and its application in channel selection for brain-computer interfaces. In *Computational Intelligence (UKCI), 2010 UK Workshop on*, pages 1–6. IEEE.

[3] I. Alvarez, G. Cortes, A. de la Torre, C. Benitez, L. García, P. Lesage, R. Arambula, and M. Gonzalez. Improving feature extraction in the automatic classification of seismic events. application to colima and arenal volcanoes. In *Geoscience and Remote Sensing Symposium, 2009 IEEE International, IGARSS 2009*, volume 4, pages IV–526. IEEE.

[4] C.W. Anderson and Z. 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 507–414. Citeseer, 1996.

[5] Joao Carreiras Pedro Osorio Andre Zamith, Antonio Caetano. Brain Computer Interface (BCI) based on Electroencephalography (EEG). December 2006.

[6] F. Babiloni, F. Cincotti, L. Lazzarini, J. Millan, J. Mourino, M. Varsta, J. Heikkonen, L. Bianchi, and MG Marciani. Linear classification of low-resolution EEG patterns produced by imagined hand movements. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):186–188, 2000.

[7] D. Balakrishnan and S. Puthusserypady. Multilayer perceptrons for the classification of brain computer interface data. In *Bioengineering Conference, 2005. Proceedings of the IEEE 31st Annual Northeast*, pages 118–119. IEEE, 2005.

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

[9] A. Bashashati, R.K. Ward, and G.E. Birch. A new design of the asynchronous brain computer interface using the knowledge of the path of features. In *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pages 101–104. IEEE, 2005.

[10] R. Bellazzi, A. Abu-Hanna, and J. Hunter. *Artificial intelligence in medicine: 11th Conference on Artificial Intelligence in Medicine, AIME 2007, Amster-*

dam, The Netherlands, July 7-11, 2007 : proceedings. Lecture Notes in Artificial Intelligence. Springer, 2007.

[11] T.W. Berger, J.K. Chapin, G.A. Gerhardt, and D.J. McFarland. *Brain-computer interfaces: an international assessment of research and development trends*. Springer, 2008.

[12] G. Blanchard and B. Blankertz. BCI competition 2003-data set IIa: spatial patterns of self-controlled brain rhythm modulations. *Biomedical Engineering, IEEE Transactions on*, 51(6):1062–1066, 2004.

[13] B. Blankertz, G. Curio, and K.R. Műller. Classifying single trial EEG: Towards brain computer interfacing. In *Advances in neural information processing systems 14: proceedings of the 2002 conference*, page 157. MIT Press, 2002.

[14] B.P. Bogert, M.J.R. Healy, and J.W. Tukey. The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the Symposium on Time Series Analysis*, pages 209–243, 1963.

[15] R. Boostani and M.H. Moradi. A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier. *Journal of Neural Engineering*, 1:212, 2004.

[16] Encyclopædia Britannica. Brain: functional areas. http://www.britannica.com/EBchecked/media/100577/Functional-areas-of-the-human-brain, April 2011.

[17] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[18] S. Chiappa and S. Bengio. HMM and IOHMM modeling of EEG rhythms for asynchronous BCI systems. In *European Symposium on Artificial Neural Networks, ESANN*, pages 193–204. Citeseer, 2004.

[19] B.H. Cho, S. Kim, D.I. Shin, J.H. Lee, S. Min Lee, I. Young Kim, and S.I. Kim. Neurofeedback training with virtual reality for inattention and impulsiveness. *Cyberpsychology & Behavior*, 7(5):519–526, 2004.

[20] E. Cinar. A study of recent classification algorithms and a novel approach for biosignal data classification. Master's thesis, Rochester Institute of Technology.

[21] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press. Elsevier, 2010.

[22] L.E. Davis, M.K. King, and J.L. Schultz. *Fundamentals of neurologic disease*. Demos Medical Pub., 2005.

[23] E. Donchin, K.M. Spencer, and R. Wijesinghe. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):174–179, 2000.

[24] J.P. Donoghue. Connecting cortex to machines: recent advances in brain interfaces. *nature neuroscience*, 5:1085–1088, 2002.

[25] M. Ernst and J.M. Rumsey. *Functional neuroimaging in child psychiatry*. Cambridge University Press, 2000.

[26] F.F. Ferri. *Ferri's color atlas and text of clinical medicine*. Expert Consult Series. Saunders/Elsevier, 2008.

[27] Society for Neuroscience. *Brain facts: a primer on the brain and nervous system*. Society for Neuroscience, 2008.

[28] F. Galán, P.W. Ferrez, F. Oliva, and J. Guardia. Feature extraction for multi-class BCI using canonical variates analysis. In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pages 1–6. IEEE, 2007.

[29] D. Garrett, D.A. Peterson, C.W. Anderson, and M.H. Thaut. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):141–144, 2003.

[30] E.L. Glassman. A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs. *Biomedical Engineering, IEEE Transactions on*, 52(11):1851–1862, 2005.

[31] B. Graimann, B. Allison, and G. Pfurtscheller. *Brain-Computer Interfaces: Non-Invasive and Invasive Technologies*. The Frontiers Collection. Springer, 2010.

[32] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller. How many people are able to operate an EEG-based brain-computer inter-

face (BCI)? *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):145–147, 2003.

[33] C. Guger, H. Ramoser, and G. Pfurtscheller. Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI). *Rehabilitation Engineering, IEEE Transactions on*, 8(4):447–456, 2000.

[34] M.M. Khatter A. M. Osheiba H. S. El-Sayed, F. M. EL-Khouly and F. E. Abd El-Samie. Automatic speed estimation for the permanent magnet brushless dc motor with a cepstral approach. *Mechanical Systems and Signal Processing, to appear*.

[35] S.J. Hanson and M. Bunzl. *Foundational Issues in Human Brain Mapping*. Bradford Books. MIT Press, 2010.

[36] B.A.S. Hasan and J.Q. Gan. Multi-objective particle swarm optimization for channel selection in brain-computer interfaces. In *The UK Workshop on Computational Intelligence (UKCI2009)*.

[37] Bashar Awwad Shiekh Hasan. *Adaptive Methods Exploiting the Time Structure in EEG for Self-paced Brain-Computer Interfaces*. PhD thesis, University of Essex.

[38] T. Hastie, R. Tibshirani, and J.H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer, 2009.

[39] T. Hinterberger, J.M. Houtkooper, and B. Kotchoubey. Effects of feedback

control on slow cortical potentials and random events. In *Parapsychological Association Convention*, pages 39–50, 2004.

[40] C.I. Hung, P. Lee, Y.T. Wu, L.F. Chen, T.C. Yeh, and J.C. Hsieh. Recognition of motor imagery electroencephalography using independent component analysis and machine classifiers. *Annals of biomedical engineering*, 33(8):1053–1070, 2005.

[41] H. Jasper and W. Penfield. Electrocorticograms in man: effect of voluntary movement upon the electrical activity of the precentral gyrus. *European Archives of Psychiatry and Clinical Neuroscience*, 183(1):163–174, 1949.

[42] W. Jia, X. Zhao, H. Liu, X. Gao, S. Gao, and F. Yang. Classification of single trial EEG during motor imagery based on ERD. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 5–8. IEEE, 2004.

[43] R. Joseph. *Neuropsychiatry, neuropsychology, and clinical neuroscience: emotion, evolution, cognition, language, memory, brain damage, and abnormal behavior*. Williams & Wilkins, 1996.

[44] S.P. Kelly, E.C. Lalor, R.B. Reilly, and J.J. Foxe. Visual spatial attention tracking using high-density SSVEP data for independent brain-computer communication. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 13(2):172–178, 2005.

[45] Z.J. Koles. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79(6):440–447, 1991.

[46] M. Krauledat, G. Dornhege, B. Blankertz, F. Losch, G. Curio, and K.R. Műller. Improving speed and accuracy of brain-computer interfaces using readiness potential features. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 2, pages 4511–4515. IEEE, 2004.

[47] A. Kubler, B. Kotchoubey, J. Kaiser, J.R. Wolpaw, and N. Birbaumer. Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, 127(3):358–375, 2001.

[48] R. Leeb and G. Pfurtscheller. Walking through a virtual city by thought. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 2, pages 4503–4506. IEEE, 2004.

[49] R.H.C. Lopes, I. Reid, and P.R. Hobson. The two-dimensional Kolmogorov-Smirnov test. In *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Nikhef, Amsterdam, the Netherlands, April 23-27, 2007*.

[50] B. Mahmoudi and A. Erfanian. Single-channel EEG-based prosthetic hand grasp control for amputee subjects. In *[Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 3, pages 2406–2407. IEEE, 2002.

[51] Inc MathWorks. *Filter design toolbox for use with MATLAB: user's guide*. MathWorks, 2000.

[52] J. Millan and J. Mouriño. Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):159–161, 2003.

[53] H. Mitsumoto. *Amyotrophic Lateral Sclerosis: A Guide for Patients and Families*. Demos Health Series. Demos Health, 2009.

[54] R.J. Moore and J. Paice. *Biobehavioral Approaches to Pain*. Springer, 2008.

[55] C. Mulert and L. Lemieux. *EEG-fMRI: Physiological Basis, Technique and Applications*. Springer, 2010.

[56] K.R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz. Machine learning techniques for brain-computer interfaces. *Biomed. Tech*, 49(1):11–22, 2004.

[57] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical neurophysiology*, 110(5):787–798, 1999.

[58] D.J. Nutt. *Drugs and the future: brain science, addiction and society*. Academic Press, 2006.

[59] Office of Communications, National Institute of Neurological Disorders Public Liaison, and MD 20892 Stroke, National Institutes of Health Bethesda. Als (amyotrophic lateral sclerosis) fact sheet. http://www.ninds.nih.gov/disorders/amyotrophiclateralsclerosis/detail_ALS.htm#172314842, April 2011.

[60] R. Palaniappan. Brain computer interface design using band powers extracted during mental tasks. In *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pages 321–324. IEEE, 2005.

[61] A.C. Papanicolaou. *Fundamentals of functional brain imaging: a guide to the methods and their applications to psychology and behavioral neuroscience*. Studies on neuropsychology, development, and cognition. Swets & Zeitlinger, 1998.

[62] D. Pearce, H.G. Hirsch, et al. The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *Proc. ICSLP00*, volume 4, pages 29–32. Citeseer, 2000.

[63] W.R. Pestman. *Mathematical statistics: an introduction*. De Gruyter textbook. Walter de Gruyter, 1998.

[64] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and clinical Neurophysiology*, 103(6):642–651, 1997.

[65] J.G. Proakis and D.G. Manolakis. *Digital signal processing*. Pearson Prentice Hall, 2007.

[66] B.E. Publishing and K. Rogers. *The Brain and the Nervous System*. Human Body. Encyclopaedia Britannica, 2010.

[67] L. Qin and B. He. A wavelet-based time–frequency analysis approach for

classification of motor imagery for brain–computer interface applications. *Journal of neural engineering*, 2:65, 2005.

[68] D.S. Rosario. *Algorithm development for hyperspectral anomaly detection*. PhD thesis, University of Maryland, 2008.

[69] M. RUZEK and T. BRANDEJSKY. Homeostasis and artificial neuron. *networks*, 4:5.

[70] S. Sanei and J. Chambers. *EEG signal processing*. John Wiley & Sons, 2007.

[71] G. Schalk and J. Mellinger. *Introduction to Brain-Computer Interfacing Using BCI2000*. Human-Computer Interaction Series. Springer, 2010.

[72] A. Schlogl, J. Kronegg, JE Huggins, and S.G. Mason. Evaluation criteria for BCI research, 2007.

[73] L. Sherwood. *Human physiology: from cells to systems*. Human Physiology. Brooks/Cole, Cengage Learning, 2008.

[74] K. Stamps and Y. Hamam. Towards inexpensive bci control for wheelchair navigation in the enabled environment–a hardware survey. *Brain Informatics*, pages 336–345, 2010.

[75] C. Starr and B. McMillan. *Human Biology*. Brooks/Cole, 2008.

[76] SS Stevens, J. Volkmann, and EB Newman. The Mel scale equates the magnitude of perceived differences in pitch at different frequencies. *J. Acoust. Soc. Am*, 8(3):185–190, 1937.

[77] N. Suppakun and S. Maneewongvatana. Automatic and semi-automatic approaches for Selecting prominent spatial filters of CSP in BCI applications. *Brain Informatics*, pages 203–213, 2009.

[78] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong. EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement*, 41(6):618–625, 2008.

[79] S. Tong and N.V. Thakor. *Quantitative EEG analysis methods and clinical applications*. Engineering in Medicine and Biology. Artech House, 2009.

[80] C.S.L. Tsui, J.Q. Gan, and S.J. Roberts. A self-paced brain–computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training. *Medical and Biological Engineering and Computing*, 47(3):257–265, 2009.

[81] M. Ungureanu, C. Bigan, R. Strungaru, and V. Lazarescu. Independent component analysis applied in biomedical signal processing. *Measurement science review*, 4(2), 2004.

[82] V.N. Vapnik. An overview of statistical learning theory. *Neural Networks, IEEE Transactions on*, 10(5):988–999, 1999.

[83] Akash Vedi. Top ten unsolved problems of science. ax3.in/top-ten-unsolved-problems-of-science.html, April 2011.

[84] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, and T.M. Vaughan. Brain-computer interface technology: a review of the first international

meeting. *Rehabilitation Engineering, IEEE Transactions on*, 8(2):164–173, 2000.

[85] H. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *IEEE transactions on knowledge and data engineering*, pages 1186–1198, 2005.

[86] W. Yu, H. He, and N. Zhang. *Advances in Neural Networks - ISNN 2009: 6th International Symposium on Neural Networks, ISNN 2009, Wuhan, China, May 26-29, 2009 Proceedings*. Number pt. 2 in Lecture Notes in Computer Science. Springer, 2009.

[87] A. Zolnay, R. Schluter, and H. Ney. Acoustic feature combination for robust speech recognition. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, PA*. Citeseer, 2005.