Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Parisa Ghane

Entitled
SILENT SPEECH RECOGNITION IN EEG-BASED BRAIN COMPUTER INTERFACE

For the degree of Master of Science in Electrical and Computer Engineering

Is approved by the final examining committee:

Lingxi Li
Chair

Andres Tovar

Lauren Christopher

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Lingxi Li

Approved by: Brian King 12/2/2015
Head of the Departmental Graduate Program Date

SILENT SPEECH RECOGNITION IN EEG-BASED BRAIN COMPUTER INTERFACE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Parisa Ghane

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

December 2015

Purdue University

Indianapolis, Indiana

To my parents whose words of encouragement and push for tenacity ring in my ears and my siblings who have never left my side.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest appreciation to my advisor, professor Lingxi Li, and co-advisor, professor Andres Tovar, for their guidance, encouragement, support, and most of all patience throughout this work. Thank you for providing machines, tools, and study spots to make this research an appealing experience for me. I am also grateful to professor Lauren Christopher for accepting being my Committee member and giving me guidance to do better in this work.

In addition, I wish to thank many faculty and staff members at IUPUI for their helps during my master studies and contributions to my thesis work. Special thanks to professor Brian King for all supports, advices, discussions, inspirations, initiations, motivations, encouragements, and many more that I can not express in words. I would also like to thank Mrs. Sherrie Tucker for her friendly instructions and guides in the department of electrical and computer engineering. In addition, I would like to acknowledge assistance of professor Paul Salama for helping me understanding some new concepts.

Last but not least, I would like to take this opportunity to thank all the people who provided resources, and datasets to make this work easier for me. Thank you all IUPUI library staff and in particular engineering and technology librarian, Mrs. May Jafari, for spending years to collect this strong and rich library database, which has given me access to a wide range of books, journals, and papers throughout this research. Furthermore, I would like to thank Dr. Luis Carlos Sarmiento Vela for conducting the electroencephalographic experiments at the National University of Colombia and letting me use his raw datasets.

TABLE OF CONTENTS

Page

Page

APPENDICES

LIST OF TABLES

LIST OF FIGURES

Figure Page

Appendix Figure

ABSTRACT

Ghane, Parisa. M.S.E.C.E., Purdue University, December 2015. Silent Speech Recognition In EEG-based Brain Computer Interface. Major Professor: Lingxi Li.

A Brain Computer Interface (BCI) is a hardware and software system that establishes direct communication between human brain and the environment. In a BCI system, brain messages pass through wires and external computers instead of the normal pathway of nerves and muscles. General workflow in all BCIs is to measure brain activities, process and then convert them into an output readable for a computer.

The measurement of electrical activities in different parts of the brain is called electroencephalography (EEG). There are lots of sensor technologies with different number of electrodes to record brain activities along the scalp. Each of these electrodes captures a weighted sum of activities of all neurons in the area around that electrode.

In order to establish a BCI system, it is needed to set a bunch of electrodes on scalp, and a tool to send the signals to a computer for training a system that can find the important information, extract them from the raw signal, and use them to recognize the user's intention. After all, a control signal should be generated based on the application.

This thesis describes the step by step training and testing a BCI system that can be used for a person who has lost speaking skills through an accident or surgery, but still has healthy brain tissues. The goal is to establish an algorithm, which recognizes different vowels from EEG signals. It considers a bandpass filter to remove signals' noise and artifacts, periodogram for feature extraction, and Support Vector Machine (SVM) for classification.

1. INTRODUCTION

Communication between human brain and external world is an interesting way for paralyzed people to conduct their daily activities much easier. This excites scholars to work on a new non-muscular path to send commands from ones brain to external tools. There are lots of works on designing a faultless BCI system presented in last few decades. Although there are many big achievement in this field, it looks like applications of BCI for speechless people has not been into consideration that much. Designing a BCI system for speech recognition from EEG signals would be a big step toward the expansion of BCI applications.

1.1 History and background

Working on BCIs started from 1924 by German neurologist ,Hans Berger [1]. He started studying brain circulation, psychophysiology and brain temperature at university. He first started with inserting two silver wires under the patients scalp, at the front and back of the head. Later his research ended with invention of electroencephalography (EEG). He could record the first human brain electrical activity in 1924 and published his paper in 1929 [1]. After a few years, EEG got very popular among researchers in United States, England, and France [2]. Berger was also the first person to introduce different brain waves such as alpha waves, which is also called Berger's waves. His analysis of EEG wave diagrams with brain diseases opened a new window for the research of human brain activities. Nowadays, After several decades of research and laboratory experiments on EEG, we are still far from having EEG recording and applications useful for daily tasks.

[1] http://www.brainvision.co.uk/blog/2014/04/the-brief-history-of-brain-computer-interfaces (last accessed: 12/02/2015)

Defense Advanced Research Projects Agency of USA initiated program to explore brain communications using EEG in 1970. The term Brain Computer Interface (BCI) was made up by Professor Jacques J. Vidal (from University of California, Los Angeles) in his article in 1976 [3–5].

Several years after BCI introduction, in 1998, the first invasive measurement was done to produce high quality brain signals. In 1999, BCI helped a quadriplegic for limited hand movements. Training monkeys to control a computer cursor in 2002 was next big step in BCIs. First BCI game was release to the public in 2003, and first control of a robotic arm by a monkey brain was in 2005 [4, 6].

BCIs has had a significant growth due to the applications of use of neural electrical activities in control of machines. Understanding neural mechanism of communication between human and machine has become a research issue that has been considered more in last few decades. The measurement of brain electrical activities, Electroencephalography (EEG), plays an important role in non-invasive BCI systems. Depending on the function of a BCI system, different electroencephalographical (EEG) signals have to be used [7, 8]. These signals include slow motor imagery potentials [9–11], P300 potentials [12, 13], and visual evoked potentials (VEP) [14, 15].

1.2 Related works

[4, 16] give general reviews of BCI systems. A review of EEG measurements is done by M. Teplan in [17]. Extracting information and features for a BCI system is the main concentration of Waldert and his colleagues in [18]. Also, McFarland with his coauthors and M. J. Safari with his colleagues provide overviews of feature extraction and translation methods [19, 20]. Furthermore, Lotte, et al. concentrate on classifications in [21].

Among successful researches done in BCI fields, we can mention Neural Signals inc. founded by Philip Kennedy in 1988. The first intracortical BCI, which tested on monkeys, was built by him and his colleagues [22]. Cats where the subjects for Yang

Dan's team at university of California at Berkelely research. The cats were shown some pictures and their neural signals were recorded and then were used to reproduce the images in cat's sight [23]. Other examples of successful projects in BCI is the one done by was done by Miguel Nicolelis, a professor of Duke university in North Carolina. He and his team recorded and then decoded mental activities first in rats and then in monkeys to regenerate their movements. The first result of their project was an open loop BCI for remotely control a robot. Then they expanded this system to a closed-loop BCI with feedback meaning that monkeys could see the robot moving and so that the BCI system received feedback [24]. Some other BCI developments are the ones that John Donoghue and Andrew Schwartz have done at Brown University and the University of Pittsburgh respectively. They were the first people who built BCI systems using recording of only a few neuron activities. Donoghue and his team had monkeys to follow the movement of a mark on a computer screen in one project, and control a prosthetic arm [25]. The monkeys brain were examined through the experiments by Schwartz and his research team [26].

Since 2010, there has been an annual BCI research award program hosted by one of the institutes known for BCI research. The host is responsible to judge and award an outstanding and innovative research in the field of Brain-Computer Interfaces. The list bellow includes a record of outstanding works that won the competition from 2010 to 2015 [2].

2010 "Motor imagery-based Brain-Computer Interface robotic rehabilitation for stroke" by Cuntai Guan, Kai Keng Ang, Karen Sui Geok Chua and Beng Ti Ang, from *Agency for Science, Technology and Research, Singapore.*

2011 "What are the neuro-physiological causes of performance variations in brain-computer interfacing?" by Moritz Grosse-Wentrup and Bernhard Schlkopf, from *Max Planck Institute for Intelligent Systems, Germany.*

[2] https://en.wikipedia.org/wiki/Annual_BCI_Research_Award (last accessed: 12/02/2015)

2012 "Improving Efficacy of Ipsilesional Brain-Computer Interface Training in Neurorehabilitation of Chronic Stroke" by Surjo R. Soekadar and Niels Birbaumer, from *Applied Neurotechnology Lab, University Hospital Tbingen and Institute of Medical Psychology and Behavioral Neurobiology, Eberhard Karls University, Tbingen, Germany.*

2013 "A learning-based approach to artificial sensory feedback: intracortical microstimulation replaces and augments vision" by M. C. Dadarlata,b, J. E. ODohertya, P. N. Sabesa,b from *Department of Physiology, Center for Integrative Neuroscience, San Francisco, CA, US, bUC Berkeley-UCSF Bioengineering Graduate Program, University of California, San Francisco, CA, US.*

2014 "Airborne Ultrasonic Tactile Display BCI" by Katsuhiko Hamada, Hiromu Mori, Hiroyuki Shinoda, Tomasz M. Rutkowski, from *The University of Tokyo, JP, Life Science Center of TARA, University of Tsukuba, JP, RIKEN Brain Science Institute, JP.*

2015 The meeting will be held in Chicago, Illinois, USA by the Department of Biosciences and Informatics - Faculty of Science and Technology, Keio University, Japan.

1.3 Applications

Different EEG recordings are used in various areas and applications of BCIs. In the past, the main application of BCI was in medical purposes to help paralyzed people. Nowadays, However, it is being used for people with normal health conditions as well. Gaming and entertainment are examples of applications that motivate designers to make faster, cheaper, and user-friendlier systems. Other applications are in operator monitoring and lie detection [27]. One of the biggest advantages of EEG, over other brain activity measurement methods, is high speed of signal recording. However, it does not have a good spatial resolution and has to be combined with some other methods like MRI when high spatial resolutions are required. The most common applications of measurement and study of EEG data in humans and animals are as following [17]:

- Monitor alertness, coma and brain death.
- Locate areas of damage after an accident, head injury, stroke, tumour, etc.
- Test afferent (the central nervous system) pathways.
- Monitor cognitive engagement.
- Produce biofeedback situations, alpha, etc.
- Control anaesthesia depth.
- Investigate epilepsy and locate seizure origin.
- Test epilepsy drug effects.
- Monitor human and animal brain development.
- Test drugs for convulsive effects.
- Investigate sleep disorder and physiology.

Also, most common applications of BCI systems are in the list bellow:

- Communication with external world.
- Control of robots and objects.
- Game and entertainment.
- Operator monitoring.
- Lie detection.
- Help paralyzed people mentally and physically.

1.4 Problem

BCIs design and implementation have always been a hard problem. Some reasons are difficulties of signal recording and knowledge requirements from multiple fields. The signal recording is difficult since 1) EEG recording always suffer from different types of artifacts, 2) EEG is a very noisy signal (low signal to noise ratio), 3) all sensors record almost the same signals (they are mathematically hard to distinguish from each other), 4) EEG signal depends on several unknown parameters (person specific, task specific, other variables), 5) in capturing EEG signals we have to consider many factors such as non-brain signals, head motions, muscle movements, and some other unexpected stimulus, and 6) large connections of neurons are involved in many different activities. In other words, many neurons fire at the same time, so they all affect the activity measurements. Two reasons are internally generated events and stimulation of cascade of related processes by an external event, for example a flash of light triggers a bunch of changes into the brain.

Furthermore, a wide range of knowledge is required such as methods and models specific to the nature of brain, statistics, linear algebra, signal processing, pattern recognition, and machine learning. However, on the other hand, there are many other problems that are similar to BCI design like speech recognition, pattern recognition, image processing, control systems, and robotics.

One of the problems that has been forgotten in last decades is design of a BCI system that can help speechless people to talk to the environment. This requires a system to recognize imagination of words in mind. Identification of different words needs recognition of different vowels [28]. It needs to record the brain electrical activities from the language area to have signals from speech imagination. Since different functions are assigned to various parts of brain, there is no need to measure activities all around the brain. There will be voltage changes in the posterior-superior temporal lobe, or Wernickes area, and in the posterior inferior frontal gyrus, or Brocas area, when a person mentally pronounces a word [29] [30] (Figure 1.1)[3]. So that, there should be signal recording from left hemisphere. In this project we worked with the signals coming from imagination of one of five English vowels /a/, /e/, /i/, /o/, and /u/. In some literature this process is called silent speech, in which a subject mentally speaks without generating acoustic signals [31].

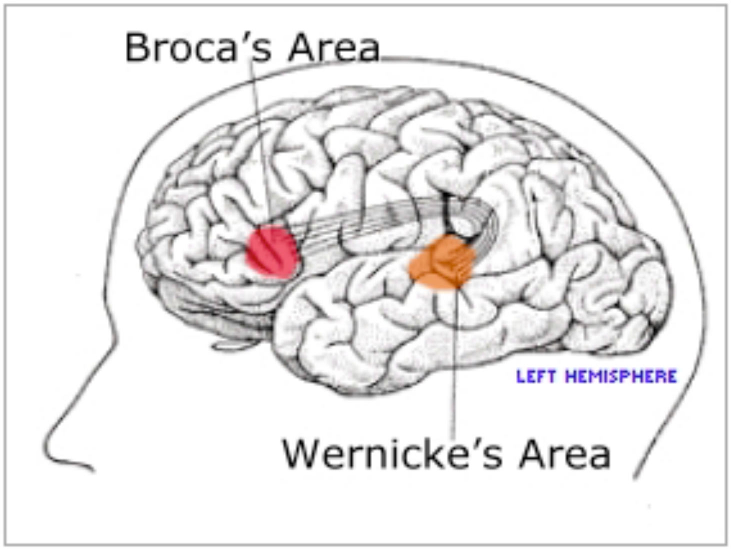

Fig. 1.1: Brain auditory and language areas.

The EEG signals for this work have been collected from 20 individuals at the National University of Colombias Clinical Electrophysiology Laboratory, with the same noise and brightness levels for all individuals. Everyone on the experiment was in good health condition with their eyes closed wearing a neuroheadset with 21 electrodes plus one ground and one reference electrode. All the electrodes were places in the left hemisphere, and labeled from E1 to E21. Figure 1.2 [4] show the position of electrodes in the EEG data recording experiment. The ground electrode was located on the lobe of the left ear, and the reference electrode was located within the EEG

[3]http://mrmikesibpsychology.weebly.com/physiology-and-behaviour.html (last accessed: 12/02/2015)

[4]The experiment was done by Dr. Luis Carlos Sarmiento Vela and his colleagues at the Clinical Electrophysiology Laboratory, National University of Colombia, Bogot, Colombia.

neuroheadset in the medial part of the forehead. In order to locate the neuroheadset in each subject, two reference points were used. The first point was the nasion and the second point was the left preauricular ear. To place each electrode on the scalp, the surface was first cleaned with an abrasive and later a gel conductor was applied. The lights were on and off periodically. The subjects have been asked to think about a specific vowel form /a/, /e/, /i/, /o/, and /u/ while the light is on and try to enter the relaxing state when the lights are turned off. The EEG signals were recorded by NicoletOne (Natus, San Carlos, California), amplified by Nicolet v32 amplifier, and processed using the software Nicolet VEEG (Natus, San Carlos, California). The sampling frequency was set to 500 Hz. The recorded signals are imported into Matlab in the form of two-dimensional arrays.

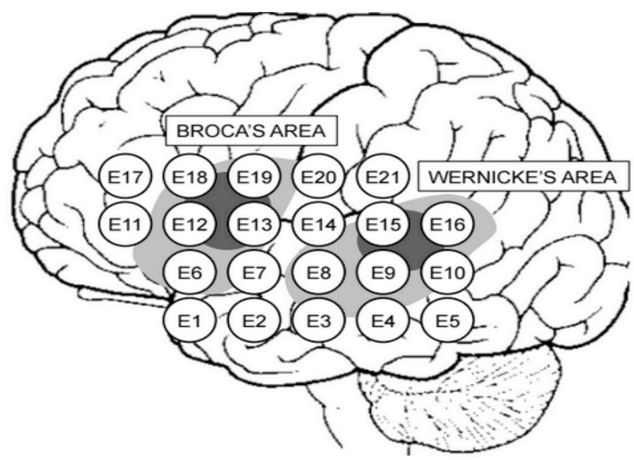

Fig. 1.2: Position of the 21 electrodes on the brains left hemisphere.

Nicole is specifically designed for EEG purposes by neurology team in Natus Medical Incorporated Company provides a high quality diagnostic information. It can also be used to investigate LTM (Long Term Monitoring), Sleep, EP (Evoked Potential), and EMG (Electromyography). Due to NicoletOne's high quality, ease of use, low cost, flexibility, and several available add-on packages, it is widely being used in BCI research labs.

1.5 Goal

The goal is to answer this question: is it possible to build a non-invasive brain computer interface system for vowel recognition? Reaching this goal, we can move forward to the next steps of silent speech recognition. The next step can be either identification of vowels and consonants or manipulation of the algorithm for using a portable headsets with different number of electrodes. Recognition of vowels and consonants will make the system more applicable in a wide range of speech recognition, like recognition of words and sentences. On the other hand, using portable EEG recording tools will give users this opportunity to use the system in their daily activities no matter in which place they actually are. This thesis focuses on the first step, which is recognition of English vowels using a non-portable EEG measurement system.

1.6 Contributions

Considering the problem and goal, we tried to implement an algorithm to take the raw signals as input and return a class label as an output. To reach this purpose, we needed to first preprocess the data and extract some of their important features, and then train a classifier, which can group all the data into 5 classes.

In order to work with the input signals easier and more accurately, we broke them into smaller pieces in such a way that no information gets lost. Then each piece was passed trough a bandpass filter to get rid of the high frequency noise and low frequency artifacts. After normalizing the filtered signals, we used periodogram to find the power spectral density as data features and meanwhile to reduce the size of the data. Sending these features to our classifier training system resulted in 4 binary classifiers based on support vector machines. Finally we used the whole system to predict that to which of 5 groups a new raw signal belongs.

It should be noted that this thesis focuses on 3 major parts: raw signals pre-processing, feature extraction, and classification (classifier training and new signals prediction). Details of each part are explained in this file. The EEG signal acquisition part of our system is done by Dr. Luis Carlos Sarmiento Vela and his colleagues at the National University of Colombia. The rest is literature review and some clarifications on methods that are common in BCI fields and the ones that we used.

1.7 Thesis outline

This thesis is divided into 5 chapters. Here is a list of chapters description:

Chapter 2 defines some of the key concepts that are being used in EEG based BCI systems. It explains about the function of different parts of brain and how to measure these activities using the state of the art technologies. Furthermore, creation of EEG signals and different types of brain waves is described. In this chapter, different classes of a BCI system and technologies are introduced.

Chapter 3 talks about the common structure of BCI systems. It elaborates each part of the system and some of the most used methods for each part. In this chapter, you can find the general workflow for a step by step BCI system design.

Chapter 4 shows the work flow of this thesis including experiments and algorithms. This chapter elaborates the process of training and classifying 5 groups of data using Periodogram, Decision Tree, and Support Vector Machines.

Chapter 5 summarizes the whole project and gives some recommendations for future works.

2. BRAIN FUNCTIONALITY, EEG, AND BCI

The measurement of electrical activities in different parts of the brain is called electroencephalography or EEG. There are lots of sensor technologies with different number of electrodes to record brain activities along the scalp. Each of electrodes captures the weighted sum of each neurons activity from the areas in the brain around that electrode, so more electrodes give more accuracy. This chapter explains about how electrical signals are generated in brain and how electrodes can capture these signals. It also talks about common EEG and BCI technologies.

2.1 Brain and neurons

Active nerves in brain generate electrical current, which in turn produces magnetic field and changes voltages in scalp. These can be measured and recorded as Electromyography (EMG) and Electroencephalography (EEG) respectively [32]. Electroencephalography (EEG, discovered by Hans Berger, helps us to look at brain activities without any health hazard, and even in a lower price rather than other measurement methods. The identification of electrographic patterns requires recognition of electrical sources and fields [33].

The brain consists of almost 100 billion nerve cells or neurons. If we consider a neuron as a switch that has on and off states, we can say that it is off while resting and on during sending electrical signal. This signal is sent through a wire called axon in which the membrane carries ions with electrical charges such as sodium (Na^{+}), potassium (K^{+}), chloride (Cl^{-}), and calcium (Ca^{2+}). Each of axons of these billion neurons generate a very small electrical charge, which helps the neurons to transmits information through electrical and chemical signals. Neurons can connect to each other to form neural networks.

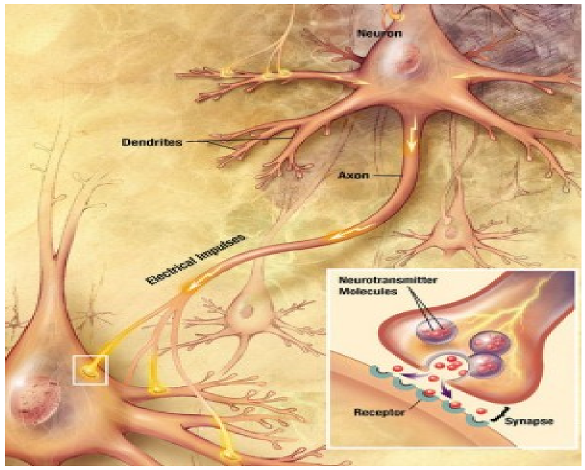

Fig. 2.1: A signal propagating down an axon to the cell body and dendrites of the next cell.

A neuron typically has a cell body (soma), dendrites, and an axon. Dendrites arise from the cell body and travel from micrometer to meters in different species, having several branches. The cell body of a neuron often has multiple dendrites, but never more than one axon. However, an axon may branch hundreds of times before it terminates. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another. Figure 2.1 [1], explains a signal propagation of a neuron and its transmission to the next neuron.

As explained, a neuron generates electro-chemical signals when transmitting information. This neurons' activities are measured in microvolts (μV), and frequency spectrums. The amplitude of the EEG is about 100 μV when measured on the scalp, and about 1-2 mV when measured on the surface of the brain. The bandwidth of this signal is from under $1Hz$ to about $50Hz$, as demonstrated in Figure 2.2 [34]. The combination of electrical activity of the brain is commonly called a brainwave pattern, because of its cyclic wave-like nature. Brainwaves are produced by synchronized electrical pulses from masses of neurons communicating with each other. Brainwaves are divided into smaller bandwidth intervals or different brainwave types. Depending

[1]https://en.wikipedia.org/wiki/Neuron (last accessed: 12/02/2015)

on what a person is doing, brainwaves will change and move from one type to another. Following is different brain waves with their specifications like frequencies, amplitudes, mental states, and such that.

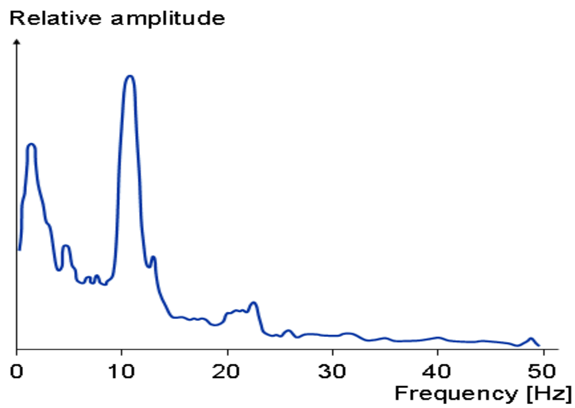

Fig. 2.2: Frequency spectrum of normal EEG in a random trial from a random subject.

2.2 Brain waves

Brain waves are detected using sensors placed on the scalp. They are divided into bandwidths to describe their functions (as defined bellow), and are a metric of relaxation and consciousness. For example, delta waves are slow, but gamma waves are fast, sensitive, and complex (Figure 2.3 [2]). Our brainwaves change according to what we are doing and feeling. When slower brainwaves are dominant we feel tired, slow, lazy, or dreamy. The higher frequencies are dominant when we feel wired, or hyper-alert. Brainwave speed is measured in Hertz (cycles per second) and they are dived into bands of slow, moderate, and fast waves.

[2] http://www.zenlama.com/the-difinitive-guide-to-increasing-you-mind-power (last accessed: 12/02/2015)

2.2.1 Delta waves (0.5 to 4 Hz)

Delta waves has lowest frequency. They are the slowest but high-amplitude brainwaves. They happen in deep dreamless sleep mostly when person is unconscious. healing, regenerating and resetting of the body happens only in this state. They have also been rarely found in some continues attention tasks [35].

2.2.2 Theta waves(4 to 8 Hz)

Theta waves occur in light sleep and extreme relaxation. This state is known as a gateway to deep learning and memory. In theta, person has a very limited sense from external world but a very strong focus on a specific thing.These signals are also dominant in deep meditation. In this state, person is in a dream, or in a very special state of relaxation or intuition beyond normal conscious awareness. It is very receptive mental state that has proven useful for hypnotherapy, as well as self-hypnosis using recorded affirmations and suggestions [36, 37].

2.2.3 Alpha waves (8 to 12 Hz)

Alpha waves is the state of awake but relaxed resting and also not processing much information. They are dominant during quietly flowing thoughts, and in some meditative states. The brain is naturally in this state when an individual gets up in the morning and just before sleep. When one closes eyes, brain automatically starts producing more alpha waves. Studying EEG activities of experienced meditators reveal strong increases in alpha activity. Alpha activity has also been connected to the ability to recall memories, lessened discomfort and pain, and reductions in stress and anxiety. For more information about alpha waves see [38–42].

2.2.4 Beta waves (12 to 35 Hz)

Beta waves are present in our normal waking state of consciousness. This state is also known as wild awake. Brain goes into this state when a person is completely conscious and active. Active calm, focusing, stress, anxiety, judgement, decision making, problem solving, and all conscious activities fall into this type of waves. Some mental or emotional disorders such as depression and ADD can be caused from lack of beta activities in one's brain. Stimulating beta activity can improve emotional stability, energy levels, attentiveness and concentration. For more information on beta waves see [43, 44].

2.2.5 Gamma waves (35 Hz and up)

Gamma waves are the fastest type of brain waves (high frequency) and relate to simultaneous processing of information from different brain areas. In this type, information are passed rapidly and happens when brain is highly active. Gamma waves have been shown to disappear during deep sleep induced by anesthesia, but return with the transition back to a wakeful state. For more information see [45, 46].

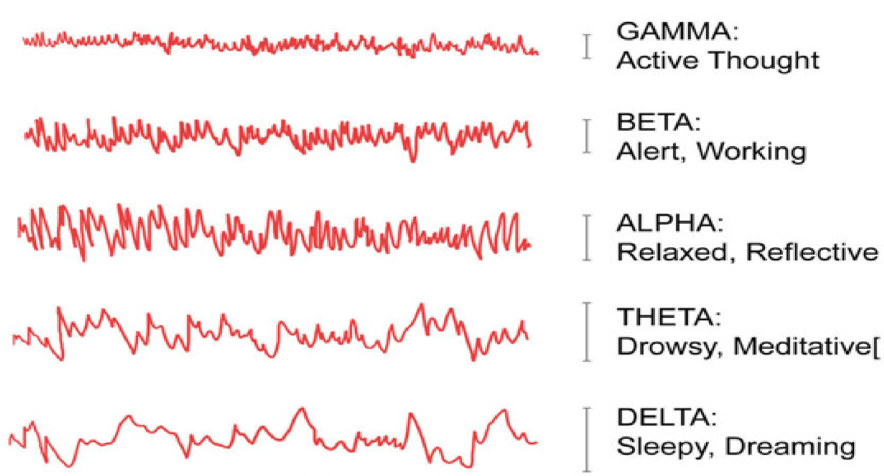

Fig. 2.3: Normal adult brain waves.

The signals and messages that are generated by BCI users' brain pass through external wires and computers instead of the normal pathway of nerves and muscles. BCI systems make it easy to interact with the people who can not or do not like to use their muscles for any reason [4]. BCIs can be placed into two categories of dependent and independent. In order to have a dependent BCI, we need to have some knowledge about the activities that is being carried out. For example in the experiment of a screen with flashing letter, the user can choose a letter by directly looking at it. In contrast, an independent BCI is a case that does not need any of muscle activities or brains normal output pathway, like the screen with flashing letter in which the person can choose a letter not by gazing at it, but by thinking about it.

2.3 BCI classes

Generally, everything that is being controlled by a computer can be controlled by a BCI system. Depending on the required speed and accuracy, applications, patient states, and available equipment different types of BCI has been offered. There are basically three types of BCI systems: invasive, Partially invasive, and noninvasive. Figure 2.4 and table 2.1 [3] briefly show these three types of BCIs.

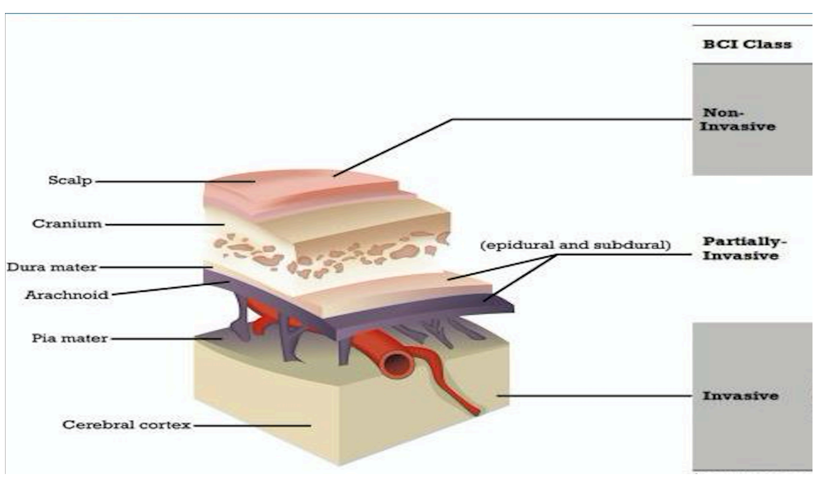

Fig. 2.4: BCI classes in different layers covering the brain [47].

[3]Source: Neurosurg Focus @ 2010 American Association of Neurological Surgeons

2.3.1 Invasive BCI systems

Invasive BCI systems are those that need to implant electrodes on or near the surface of the brain, directly into the grey matter, during neurosurgery. Signals are in the highest resolution and quality but after a while the signals become weaker, or even non-existent, as the body reacts to a foreign object in the brain. On the other hand it requires electrodes to be implanted through a surgery with health hazards One of the under research goals for invasive BCIs are to help people with sight damage and those who are seriously paralyzed [48].

2.3.2 Partially invasive BCI systems

Partially invasive BCI systems are another type of BCIs which records signals from electrodes placed inside of the skull, but outside the grey matter. An electrode grid is being implanted by a surgical process. They produce better resolution signals than non-invasive BCIs, because there is no bone tissue to deflects and deforms signals and have a lower risk of forming scar-tissue in the brain than fully invasive BCIs. Partially invasive BCI shows potential for real world application for people with motor disabilities [49, 50].

2.3.3 Non-invasive BCI systems

Non-invasive BCI systems are those which do not need any surgery to penetrate or break any part of scalp. Non-invasive systems are easy to wear without any discomfort. However, non-invasive implants produce poor signal resolution because the the electromagnetic field generated by neurons are dispersed and blurred while passing through the skull. Although the waves can still be detected it is more difficult to determine the area of the brain that created them or the actions of individual neurons. FMRI is one of the well-known technologies that measures brain activities by detecting changes in blood volume. this method has two main advantages: no surgery hazard and high space resolution Most non-invasive systems use electrodes placed on the scalp. Non-invasive measurements are commonly used in research and most of medical applications. In this research we consider this type of BCIs.

EMG and regular EEG are non-Invasive measurements, which means there is no need for penetrates or breaks in scalp. On the other hand, there are some brain activities recording that are known as invasive measurements. An example is Electrocorticography (ECoG) or intracranial EEG (iEEG) in which electrodes should be directly implanted on the cortex surface. They shows higher resolution compared to EMG and EEG but requires electrodes to be implanted through a surgery with health hazards. Medical clinics prefer to use non-invasive methods in neuroimaging. FMRI is one of the well-known technologies that measures brain activities by detecting changes in blood volume. this method has two main advantages: no surgery hazard and high space resolution

2.4 EEG capturing tools in non-invasive BCIs

Electrophysiological experiments consider the electrical features activities of biological cells and tissues. In this field we put measurement tools in some important zones and study changes in recorded voltages. However, hemodynamic activities are those ones which are involved with the process of body adjustment while delivering glucose and oxygen to some tissues. Delivered glucose and oxygen increase activity of neurons in those body tissues. Some clinical methods like Functional Magnetic Resonance Imaging (FMRI) can measure this change in neuronal brain activities. Table 2.2 gives a summary of neuroimaging methods [16].

The electrodes, whether invasive or non-invasive, capture neuron activities. These activities are sent to a computer, which has to use a software to translate the brain signals into computer commands. Different BCI systems use different types of EEG headsets with different types of EEG electrodes. There are basically two types of electrodes: wet and dry.

Table 2.1: Advantages and limitations in different BCI classes.

BCI Class	Predominant Platform	Advantages	Limitations
Non-invasive	EEG	• Inexpensive. • No surgical implantation. • Less training requirements for using.	• Inference from non cortical. stimuli • Low quality.
Partially Invasive	ECoG	• Higher resolution than non-invasive. • No cortical damage. • Good signal-to-noise ratio. • Low clinical risk. • Long-term stability.	• Neurosurgical implantation required. • Not studied in motor-impaired patients.
Invasive	Cortical multi-electrode array	• High performance and rate of data output. • Direct recording of cortical neuron potentials.	• Vascular and neuronal damage. • Lack of signal stability over time. • Neurosurgical implantation required.

Table 2.2: Neuroimaging methods.

Neuroimaging method	Activity measured	Direct/Indirect measurement	Spatial resolution	Risk	Portablility
EEG	Electrical	Direct	$\sim$ 10 mm	Non-invasive	Portable
MEG	Magnetic	Direct	$\sim$ 5 mm	Non-invasive	Non-Portable
ECoG	Electrical	Direct	$\sim$ 1 mm	Invasive	Portable
fMRI	Metabolic	Indirect	$\sim$ 1 mm	Non-invasive	Portable

BCI has been started with wet electrodes, and conductive gel. Preparation of wet electrodes are more difficult than the dry ones, which is an obstacle to use EEG-BCI system day to day for patients with impaired mobility. Scrubbing of the scalp significantly increase the signal resolutions, but it would be an unpleasant experience for almost all subjects. Also, having wet electrodes on the skin for several frequent sessions heighten skin-sensitivity. Besides, over hours of use, it requires regular maintenance as the conductive gel dries and degrades signal quality. So that there has been several efforts to develop dry electrodes. However, studies show that additional work is needed before dry electrodes become an alternative to standard wet electrodes for the recording of EEG signals in clinical and other applications with long-term exposures [51].

Fig. 2.5: Dry electrode with pin.

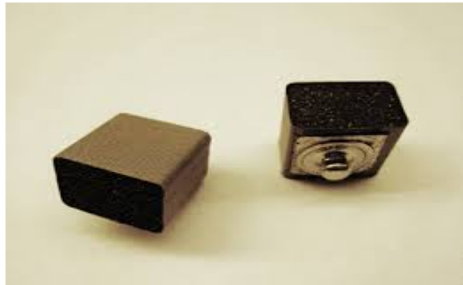

Fig. 2.6: Dry foam electrode fabricated by electrically conductive polymer.

Nowadays new dry EEG electrodes progresses give rise to a wide range of applications other than clinical applications. Some advantages of dry electrodes, such as gel-free operation, make them easy to frequently use. Robust dry EEG electrodes are one of the key issues to practically develop BCI technologies. Dry electrodes are subject to several challenges since they do not use the electrolytic gel to penetrate hair and contact the skin. They have to be designed in a way to directly touch the scalp. Also, the location of the electrode should be accurate to reduce artifact and noises as much as possible. Figures 2.5 [4] and 2.6 [5] show two types of common dry electrodes. The electrodes can be installed on a headset or helmet. They can be whether with a conductive gel, or dry electrodes which does not use the gel. There are two major types of tools that are being used to capture EEG signals in non-invasive BCIs: EEG caps with wet electrodes and EEG headsets with dry electrodes [17].

[4] http://www.gtec.at/Products/Electrodes-and-Sensors/g.Electrodes-Specs-Features (last accessed: 12/02/2015)
[5] http://mindo.com.tw/en/goods.php (last accessed: 12/02/2015)

3. BCI STRUCTURE

BCI systems can be used in a wide variety of applications such as communication and control, operator monitoring, lie detection, gaming and entertainment, health, and help to paralyzed people. However, before reaching any of these goals several steps have to be passed. Similar to any other systems, a BCI has inputs, outputs, several components in between to translate input to outputs, and in some BCIs a feedback circuit to make the system stable all the time. More precisely, first step to design a BCI system is to provide a biocompatible EEG signals recording system, second is to develop an algorithm which can decode the messages in the EEG data with a good accuracy, and last is to create commands to send to the target system, which can be a computer, or a moving object, or a emotion detection system, or such that [4].

3.1 Signal acquisition

BCI systems use information from brain activities to identify the user's intention. As shown in figure 3.1 [1], different functions are assigned to various parts of brain, so depending on the activity and intention, different groups of neurons will be activated. The brain is a very busy organ and is the control center for body. It runs all organs such as heart and lungs. All of senses, sight, smell, hearing, touch, and taste depend on brain functionality. For example, tasting food with the sensors on tongue is only possible if the signals from taste buds are sent to the brain. Once in the brain the signals are decoded. The sweet flavor of an orange is only sweet if the brain tells so. This highlights the role of recording brain activities as well as converting them to electrical signals.

[1] http://brainpictures.org/Brain-Functions-Pictures.php (last accessed: 12/02/2015)

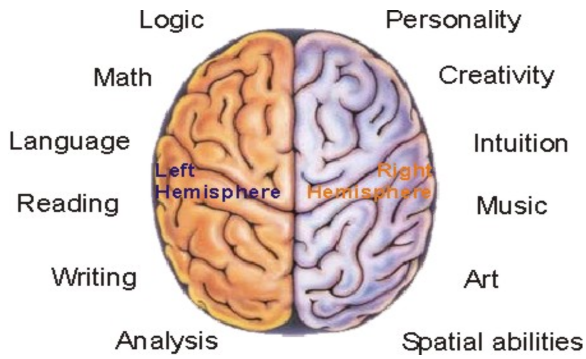

Fig. 3.1: Left and right hemispheres.

EEG recording systems measure difference of potentials between two electrodes, which have been placed on an active neuron, and another neuron in resting state. As explained before, EEG signals depend on several unknown parameters (person specific, task specific, other variables) and always suffer from different types of artifacts. The reason of this variability can be some biological or experimental facts like: 1) relevant functional map differs across individuals, 2) sensor locations differ across recording sessions, 3) brain dynamics are not the same at all time scales (each week something will be change in ones brain), and so on.

In order to deal with some of EEG measurement variabilities, standard systems are being used to locate sensors on scalp. These methods are developed to ensure standardized reproducibility so that a subject's studies could be compared over time and subjects could be compared to each other. One of popular international methods is ten-twenty system, which is based on proportional measurements between easily identified skull landmarks and provides adequate coverage of all parts of the head. Figure 3.2 shows internationally standardized ten-twenty system electrode setup.

According to the ten-twenty system there are 21 electrodes placed on the scalp, as shown in Figure 3.2. To locate these electrodes the following steps has been considered: there are two reference points which are called Nasion and Inion. Nasion is the delve at the top of the nose, level with the eyes; and Inion is the bony lump at the base of the skull on the midline at the back of the head. From these points, the skull

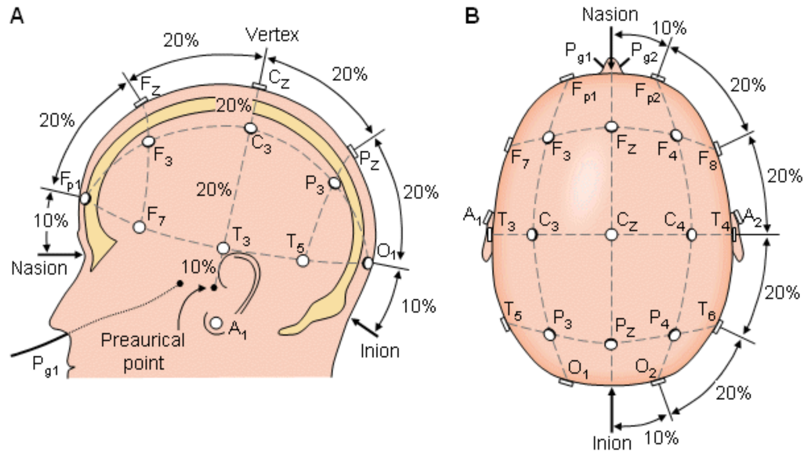

Fig. 3.2: 10-20 standard system [34].

perimeters are measured in the transverse and median planes. Electrode locations are determined by dividing these perimeters into 10% and 20% intervals. Ten-twenty system has been widely used in the EEG cap designs since it was introduced.

The international ten-twenty system of electrode placement, originally proposed in 1958 [52]. It is widely being used as a recommended standard method for recording scalp EEG. Recently, the American EEG Society has made some modifications to the original alphanumeric. According to this new modified system, the original T3, T4, T5 and T6 are now referred to as T7, T8, P7 and P8 respectively. This modification allows standardized extension of electrode placement in the sub-temporal region (e.g., F9, T9, P9, F10, T10, P10) and designates named electrode positions in the intermediate coronal lines between the standard coronal lines (e.g., AF7, AF3, FT9, FT7, FC5, FC3, FC1, TP9, TP7, CP5, CP3, CP1, PO7, PO3 and so on)(Figure 3.3). Letters specifying the spatial locations are in table 3.1[2].

[2]https://en.wikipedia.org/wiki/10-20_system_(EEG) (last accessed: 12/02/2015)

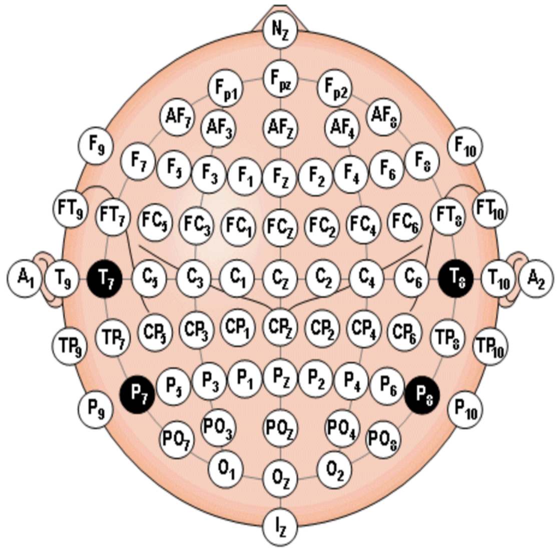

Fig. 3.3: 10-20 system modified by American EEG society [34].

Table 3.1: Spatial letters.

Letter	Location
C	Central
P	Parietal
F	Frontal
T	Temporal
O	Occipital

3.2 Signal preprocessing

Once the EEG signal is captured, we are good to move toward processing, pattern recognition, classifier training, and classification of the data. In every EEG recording experiments, possible non-neural signals, such as noises and artifacts, are inevitably added up with the actual brain signals. Removing these non-neural part of the data is the first step to make signals ready for later analysis. In EEG signal processing, first, we need to use filter to remove some unwanted components of the signal. Filtering is a method of signal processing, which deletes some frequencies and pass the rest to remove some background noises and artifacts of the signal. There are several types of filters for EEG signals processing. Some common filters used in EEG signal processing are:

- Constant filters, which have constant dynamic for all sampling times.
- Spatial filters like Independent Component Analysis (ICA) [53, 54].
- Temporal filters like moving average filters [55, 56].
- Frequency selective filters like high pass, low pass, band pass, band stop filter [28, 57].

As explained before, EEG signals are in the frequency domain below $60Hz$. A high pass filter can remove any components of the signal which are in frequency higher than $60Hz$. These components mostly come from background noise [58].

3.3 Feature extraction

Classification of patterns based on sampled waveforms results in poor performance. Hence, extraction of characteristic features from the data can increase the classification performance. Recorded EEG signals consist of a large number of simultaneous fired neurons. In order to select a suitable classifier, it is required to find any or all of the the sources, properties, and features of the data. Four most common groups of features are time-domain features (TDF), frequency-domain features (FDF), wavelet features (WF), and cepstral features (CF) [59].

The recorded EEG data can be quantified as voltage versus time, in the time-domain analysis, and as power versus frequency, in frequency domain analysis. Both forms of analysis can be used for EEG-based communication [60]. In time-domain, changes in the form or magnitude of voltage can function as a command. They are referred to as an evoked potential or evoked response. For example, in flashing letters experiments, the evoked potentials shows that if the person wants to pick that letter or not. [60,61] In the frequency-domain, the commands are the changes in amplitude of the signal in a specific frequency band. They are referred to as a rhythm. The major works done on cursor control on a computer screen is an example of EEG processing in frequency domain [62–64]. Bellow is a discussion of different common methods of interest. Table 3.2 from [65] compares the performance of these feature extraction methods and their general advantages and disadvantages.

3.3.1 Fast Fourier Transform (FFT) & Power Spectral Density (PSD)

Fast Fourier Transform (FFT) in signal processing is the most widespread method that analyzes data using mathematical tool. Generally, Fourier transform finds the spectrum of any type of signals using the equation 3.1. However, it is not being used for non-deterministic variables. Instead people use Power Spectral Density (PSD) for stochastic processes, in which the distribution of signal's power is studied over different frequencies. PSD can be interpreted as the Fourier transform of the auto-correlation function and is calculated as equation 3.2.

$$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt \tag{3.1}$$

where $x(t)$ is a deterministic signal in time domain and $X(\omega)$ is the represents the Fourier transform of the signal $x(t)$.

$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau)e^{-j\omega\tau}d\tau \tag{3.2}$$

where $R_{xx}(.)$ is autocorrelation function and $S_{xx}(\omega)$ shows PSD.

There are many features to extract from data for machine learning purposes and specifically, in the field of signal processing, power spectral density is widely used as a feature of the signals of under study because it is both easily measurable and observable [66,67]. It shows that how much power in each frequency a signal has and in this way describes the specific features from different data.

3.3.2 Wavelet Transform (WT)

In Wavelet Transform (WT) the process includes finding a set of basis function and decompose the signals onto them. These basis functions are called wavelets. The prototype wavelets are called mother wavelets. The basic functions are made up from extended, contracted, and/or shifted versions of mother wavelets [68]. As an example of a basic function, figure 3.4 shows a Gaussian wavelet of order three. Equation 3.3 shows the general wavelet transform formula, which also represents the Continues Wavelet Transform (CWT) [69].

$$X_w(a,b) = \int_{-\infty}^{+\infty} x(t)\psi^*_{a,b}(t)dt \tag{3.3}$$

where $x(t)$ is the unprocessed signal, which can be EEG signals, a stands for dilation (scaling parameter), and b represents translation parameter. $*$ is the complex conjugate symbol, function $\psi_{a,b}(t)$ is calculated from $\psi(t)$, and $\psi(t)$ is the wavelet that we has been chosen as the mother wavelet.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{3.4}$$

The wavelet transform or wavelet analysis can be regarded as a successful solution for shortcomings of the Fourier transform. In fact Fourier transform is a special case of wavelet transform where $\psi^*_{a,b} = e^{-j\omega t}$. The main difference is that Fourier transform decomposes the signal into sines and cosines, but the wavelet transform uses functions that are localized in both the real and Fourier space [69].

Wavelet transform is widely being used to reduce the signals parameters without much changes in the original signal [70]. It has also been applied on the EEG data as a spectral estimation strategy for the feature extraction purpose. The main focus in wavelet problems is to re-express any function as an infinite series of wavelets [71–73]. Wavelet transforms can be continuous wavelet transform (CWT) or Discrete Wavelet Transforms (DWT) [74, 75]. The DWT is a modified version of CWT when it can only be scaled and translated in discrete steps.

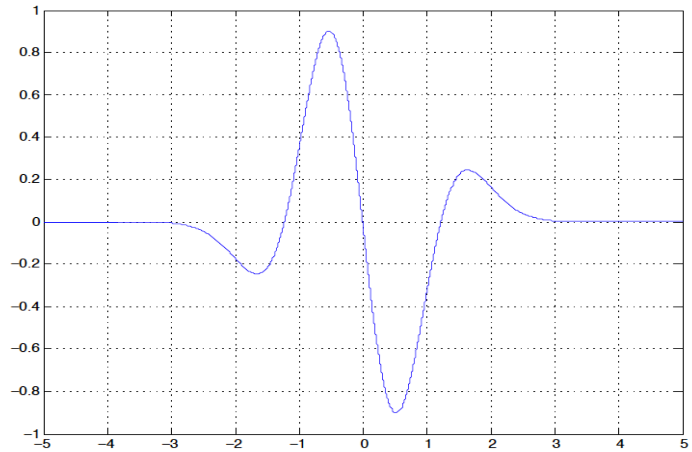

Fig. 3.4: Gaussian wavelet of order three, an example for continues wavelets.

3.3.3 Eigenvectors

This method is used to find frequency and power of a signal which is artifact dominated measured. There are a few eigenvector methods such as Principal Component Analysis (PCA), MUSIC, and Pisarenkos method [76–78]. Eigenvector methods are mostly used to reduce the dimensionality of the data. In many cases, feature extraction can be only a dimensionality reduction technique, where a subset of new features will be extracted as the new dimensions. This subset shows the most involved dimensions and keeps as much information in the data as possible [79]. In many cases, working with a feature space with high dimensions increase the computational complexity and the probability of having classification error [80]. Hence, dimensionality of the feature space is reduced before sending features to classification algorithm.

Principal Component Analysis (PCA)

One of important aspects of BCI systems design is extracting and selecting the main features of data. However, a bad feature selection can result in a very poor classification. On the other hand, classification of features only based on sampled waveforms would end in computationally complex and inaccurate classification performance [59]. PCA provides features that are robust to small amounts of noise since it keeps maximum variance. PCA is widely being used in many applications and forms of analysis such as neuroscience and computer graphics. It simplifies complex high dimension confusing data by extracting uncorrelated and relevant information of the data. In other words, it provides a roadmap for revealing the hidden dynamic and important features as well as reducing the complexity and dimensionality.

Principal component analysis has been introduced to compute the most meaningful basis of a noisy, garbled data set [81]. Using this new basis can filter out much of noise and artifacts. For each of experimental trials, an experimenter records a set of data consisting of multiple measurements like voltage or position. The dimension of the data set is the number of measurement types.

Generally, each data sample is a vector in space of dimension m, where is the number of measurement types. This m-dimensional vector space is spanned by a possible orthonormal basis. PCA tries to find a linear combination of the original basis that best re-expresses the original data set [82]. This linear combinations of unit length basis vectors produce all measurement vectors in the space.

Generally, PCA maps huge correlated data to small-uncorrelated data by using orthogonal linear transformation. Presenting PCA is quite simple after obtaining the form of a data matrix. There are two solutions for PCA: the eigenvectors of the covariance matrix, and the singular value decomposition. The second one is more general. PCA is closely related to singular value decomposition (SVD).

Pisarenkos method

Pisarenko's method, is a method of frequency estimation, which is used to evaluate power spectral density (PSD). It considers a signal $x(n)$ as a sum of p complex exponentials in the presence of white noise [83]. Pisarenko's technique estimates the frequencies from the eigenvector corresponding to the minimum eigenvalue of the autocorrelation matrix. The polynomial $A(f)$ which contains zeros on the unit circle is used to estimate the PSD [77]. This method is sometimes limited in its usefulness because it is sensitive to noise and the number of complex exponentials must be known. For more about Pisarenko's method see [77, 78, 84].

Multiple Signal Classification (MUSIC)

The MUSIC method is the one that looks for the frequency content of a signal. It uses autocorrelation matrix and eigenspace methods to form power spectral density as 3.5 [84].

$$P_{MUSIC}(f) = \frac{1}{\frac{1}{k}\sum_{i=0}^{K-1} |A_i(f)|^2} \tag{3.5}$$

where K is the dimension of noise subspace and $A_i(f)$ is the desired polynomial that corresponds to all the eigenvectors of the noise subspace.

$$A(f) = \sum_{k=0}^{m} a_k e^{-j2\pi fk} \tag{3.6}$$

where a_k are coefficients of $A(f)$, and m is the order of $A(f)$.

3.3.4 Autoregressive Method (AR)

Autoregressive (AR) is another method which is used to estimate the power spectrum density (PSD) of the signal using a parametric approach. Since AR methods use parametric approaches, they give better frequency resolution and rarely face problem of spectral leakage [65]. Two examples of AR methods are Yule-Walker method and Burgs method. For more about AR methods see [67].

Table 3.2: Comparison between performance of different EEG feature extraction methods.

Method name, domain, and suitability	Advantages	Disadvantages
Fast Fourier transform in frequency domain, suitable for narrow-band stationary signals.	• Good for stationary signal processing. • Appropriate for narrow-band signal, such as sine wave. • Has an enhanced speed over virtually all other available methods in real-time applications.	• Does not have good spectral estimation and cannot be employed for analysis of short EEG signals. • Cannot reveal the localized spikes and complexes that are typical among epileptic seizures in EEG signals. • Suffers from large noise sensitivity, and it does not have shorter duration data record.

continued on next page

Table 3.2: *continued*

Method name, domain, and suitability	Advantages	Disadvantages
Wavelet transform in both time and frequency domain, suitable for transient and stationary signal.	• Has a varying window size, being broad at low frequencies and narrow at high frequencies. • Well suited for analysis of sudden and transient signal changes.	Needs selecting a proper mother wavelet.
Eigenvector in both time and frequency domain, suitable for signals buried with noise.	Provides suitable resolution to evaluate the sinusoid from the data.	Lowest eigenvalue may generate false zeros when Pisarenkos method is employed.

continued on next page

Table 3.2: *continued*

Method name, domain, and suitability	Advantages	Disadvantages
AR in frequency domain, suitable for signal with sharp spectral features.	• Limits the loss of spectral problems and yields improved frequency resolution. • Gives good frequency resolution. • Spectral analysis based on AR model is particularly advantageous when short data segments are analyzed, since the frequency resolution of an analytically derived AR spectrum is infinite and does not depend on the length of analyzed data.	• The model order in AR spectral estimation is difficult to select. • Gives poor spectral estimation once the estimated model is not appropriate, and models orders are incorrectly selected. • Is readily susceptible to heavy biases and even large variability.

3.4 Features classification

Generally, a good BCI system needs a good pattern identification and translation part, which depends on classification algorithms used in the system [19]. A classification algorithm is designed to train a system that can predict the class of an input data based on its features [85]. In other words, it finds out that a new observation belongs to which sub-population or category. There are several factors that should be considered in choosing the classifier. An example of problems in classifications is the curse of dimensionality [21].

The curse of dimensionality is a concern when the training set is small but the dimensionality of the feature vector is high. The system needs enough data to describe the different categories and find the proper class for a newcomer signal. Depending on feature vector dimensionality, the required data will be increased exponentially [86]. For a good classification performance, it is suggested to have training samples more than at least five times rather than the features dimensionality [87].

3.4.1 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical pattern recognition and machine learning method for classification problems. It is also used for dimensionality reduction before later classifications. LDA looks for a linear combination of variables, which can best describe data, and in this point of view it is similar to PCA [88, 89]. Two major advantages of LDA over PCA for some types (not all) of data are explained in the following.

PCA is based on the covariance matrix. Covariance in sensitive to individual large values, so if someone takes a single attribute of the data and multiplies it by a large number, the PCA will be easily messed up. It wrongly shows that large attribute as the dominant component of the whole data. The other problem with PCA is that it considers the space as a linear one, so that in one dimensional it finds a straight line and in two dimensional it finds a flat sheet.

Most times we use PCA to reduce dimensionality before classification problems. It sometimes helps but sometimes not. It can sometimes hurt the data preprocessing. A reason is that PCA only looks at the datapoint coordinates but does not consider the classes labels. As a result the dimension that it picks might be very bad for later classifications. Take, for example, the data points in figure 3.5 [3]. There are two classes of red and green. PCA takes the dimension with the greatest variance (dimension along the blue axis shown in the figure 3.5), and projects all the datapoint on this axis. This makes the two classes completely mixed into each others. There is no way to separate the data in new dimensions. However, if different dimension is picked up (like the red axis), then the projected data will be easily seperable. Hence, we need to find another way to choose a dimension as close as possible to this red axis. This is where LDA got introduced. LDA is a version of PCA that reduces the dimensionality in such a way that is most useful for the classifier [90].

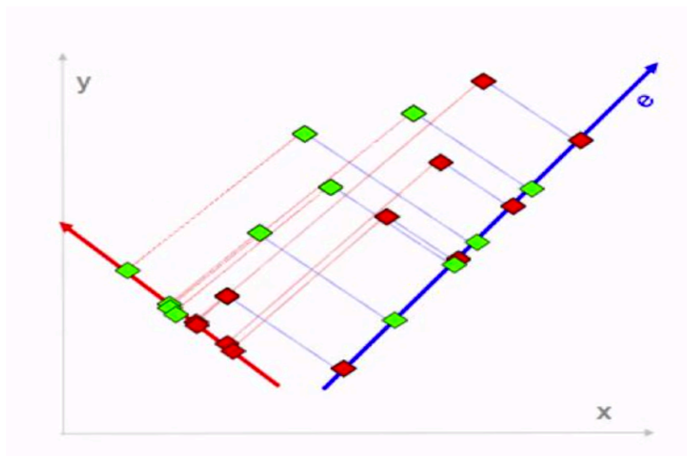

Fig. 3.5: LDA versus PCA in finding a new axis for separation of red and green data. Both methods project the data on their new axis. LDA finds the red axes and PCA finds the blue one.

LDA is very similar to PCA, but it takes advantage of class labels when picking a new dimension. This new dimension gives maximum separation between means of projected classes, while having minimum variance within each projected class. LDA algorithm takes a few assumptions such as the data is Gaussian and there is a simple boundary between the data points. As result LDA does not always guarantee a

[3] linear discriminant analysis, Introductory Applied Machine Learning (IAML) course by Victor Lavrenko at the University of Edinburgh

better projection for the classifier. It usually fails when separating information is not in the mean of datapoint in each class, but is in the variance the data. So, generally, sometimes LDA works better (like in figure 3.6 [4]), sometimes PCA gives better result (like in figure 3.7 [5]). They should be tried out in experiments.

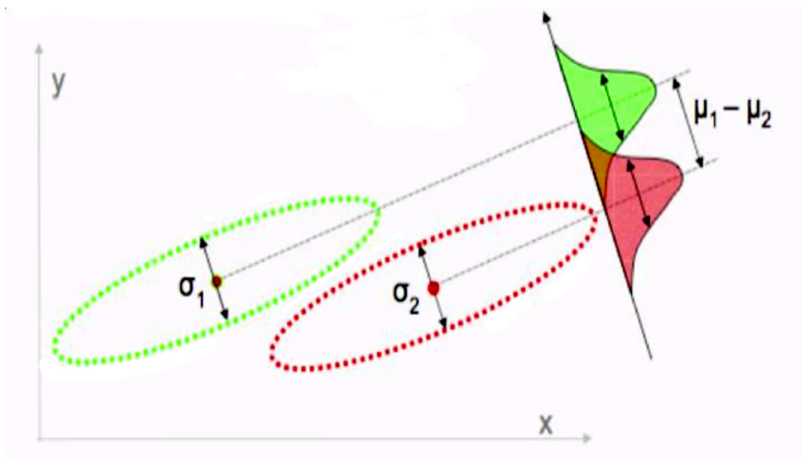

Fig. 3.6: LDA works better than PCA in this case, because the mean of two datasets are easily distinguishable.

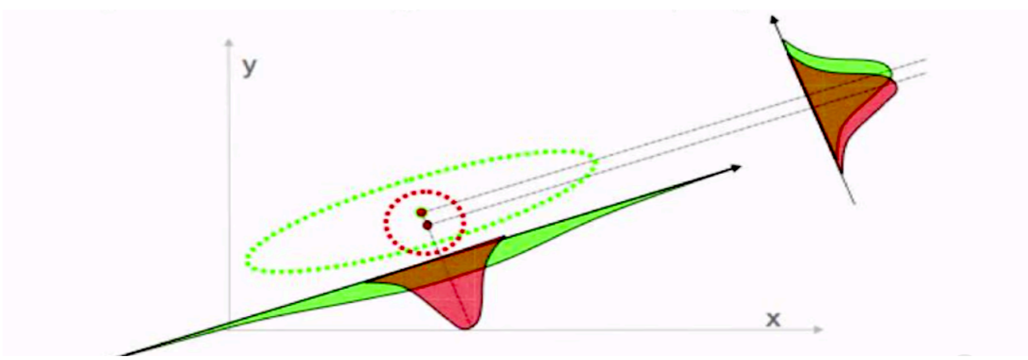

Fig. 3.7: PCA performs better than LDA in this case, because the means of two variables are very close to each other but their variance can distinguish them.

[4]linear discriminant analysis, Introductory Applied Machine Learning (IAML) course by Victor Lavrenko at the University of Edinburgh

[5]linear discriminant analysis, Introductory Applied Machine Learning (IAML) course by Victor Lavrenko at the University of Edinburgh

3.4.2 Support Vector Machine (SVM)

In 1964, Vapnik and Chervonenkis introduced a new algorithm which constructs an optimal separating hyperplane when the training data is separable. They proposed the method with the simplest possible case, which is to have linear machines on separable data. At that time, they called it the generalized portrait method [91]. In 1995, Vapnik and Cortez generalized the method to a non-separable set of training data [92]. The concept of SVM is given in the following.

Considering that the problem is to divide positives from negatives in the figure 3.8, we are trying to find the line (plane and hyperplane in higher dimensions) that can separate the data into two groups. There can be many lines (planes or hyperplanes) to separate the training data into groups, but the question is which one is the best. According to [93], the best choice is the one that leaves the maximum margin from both classes. This is also called the widest approach because we are looking for the widest street that separates the positives from the negatives. As result, the SVM method tries to put the line in such a way that the street is as wide as possible.

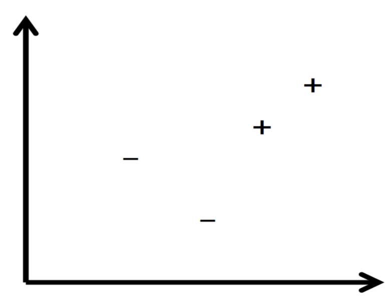

Fig. 3.8: Two classes of positives and negatives.

Vapnik tried to first find the decision rules that make that decision boundary. Imagine that we have a vector $\bar{w}$ of any length perpendicular to the median line of the street and also an unknown vector $\bar{u}$. The question is to figure out the vector $\bar{u}$ is on which side of the street. To find the answer, we would need to find the distances

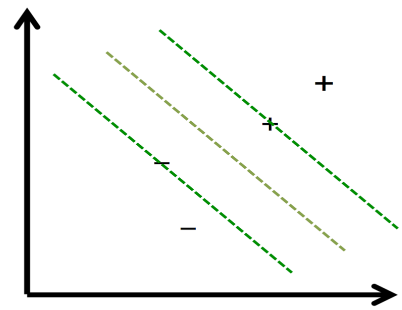

Fig. 3.9: Widest possible street between closest elements of two groups.

of $\bar{u}$ with any lines of the street. So that we project the vector $\bar{u}$ onto a vector that is perpendicular to the street, like $\bar{w}$. The starting point is to is to see if the inner product of $\bar{w}$ and $\bar{u}$ is greater than a constant c or not.

$$\bar{w} \cdot \bar{u} \geq c$$

For our positive-negative example, if the equation above is true, then the vector $\bar{u}$ shows a positive sample. The dot product lets us apply the directional growth of one vector to another, and can give us the projection of $\bar{u}$ on the $\bar{w}$. If we take $b = -c$, without loss of generality we can write our first decision rule for positive samples as:

$$\bar{w} \cdot \bar{u} + b \geq 0 \tag{3.7}$$

The equation 3.7 is not specific, because there can be many b and $\bar{w}$. In order to fix a particular b and $\bar{w}$ there should be more constraints. Let's take the inner product of $\bar{w}$ with a positive sample x_+ that is on the positive side of the street, we will have the equation:

$$\bar{w} \cdot \bar{x}_+ + b \geq 1$$

Likewise if take the inner product of $\bar{w}$ with a negative sample x_- that is on the negative side of the street, we can we will have the equation:

$$\bar{w} \cdot \bar{x}_- + b \leq -1$$

Since dealing with two equations is not mathematically convenient, we can introduce a new variable y_i such that,

$$y_i = -1 \qquad \textit{for negative samples}$$

$$y_i = +1 \qquad \textit{for negative samples}$$

where the value of y show that in which group our sample is. Multiplying y_i with the decision rules, we will have both equations equal to a single equation defined as:

$$y_i(\bar{w} \cdot \bar{x}_i + b) \geq 1 \quad or$$

$$y_i(\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0$$

It is equal to zero $(= 0)$ for x_i on the borders of the street. So the second decision rule can be:

$$y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0 \qquad \textit{for } x_i \textit{ on the borders of the street.} \tag{3.8}$$

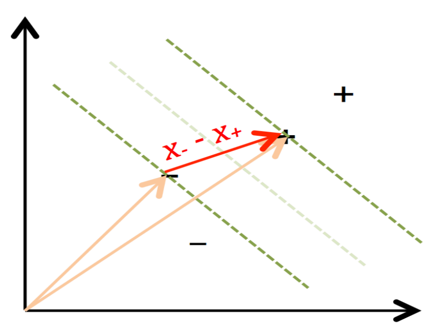

Fig. 3.10: Vector of deference of positive support vector and negative support vector. The dot product of this vector and a normal unit vector gives the width of the street.

At this point we can have the vector for the negative and the vector for the positive samples on the border. The dot product of the vector $(\bar{x}_+ - \bar{x}_-)$ and a normal unit vector, gives us the width of the street in figure 3.10. We have assumed that $\bar{w}$ is a normal, so $\frac{\bar{w}}{||w||}$ is a normal unit vector.

$$width = (\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{w}}{||w||} \tag{3.9}$$

From equation 3.8 for x_+ (positive samples) $y_i = 1$ and for x_- (negative samples) $y_i = -1$ so that,

$$\begin{cases} x_+.\bar{w} = 1 - b \\ x_-.\bar{w} = -1 - b \end{cases}$$

substituting these values for in the equation 3.9:

$$width = (\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{w}}{||w||} = \frac{1 - b - (-1 - b)}{||w||} = \frac{2}{||w||} \tag{3.10}$$

Considering the equation 3.10, in order to maximize the width of the street, $\frac{2}{||w||}$ should be maximized. We can instead minimize $||w||$, which means we can solve the following optimization problem:

$$\begin{array}{lll} minimize & \frac{1}{2}(||w||)^2 & \\ subject\ to & y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0 & for\ i = 1, 2, 3, \cdots, N \end{array} \tag{3.11}$$

The method of Lagrange multiplier is used to find the minimum of the function in equation 3.11. Generally the Lagrange multiplier method (introduced by Joseph Luis Lagrange) is a strategy that is used in optimization problems to find the local minimum or maximum of a function by putting the main function and the constraints in a single equation [94, 95]. As result, the Lagrange multiplier method gives us a new expression (L) that can be maximize or minimize without thinking about the constraints any more. L is the function minus the summation of constraints while each constraint is equal to 0 and has a multiplier.

$$L = \frac{1}{2}||w||^2 - \sum_i \alpha_i \left[y_i(\bar{w} \cdot \bar{x}_i + b) - 1\right] \tag{3.12}$$

We are minimizing L in respect to variables $\bar{w}$ and b. Because the differentiating in respect to a vector is same as the differentiation in respects to a scalar, the extremum of L is easy by finding the derivatives in respect to all variables and set them to 0.

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum_i \alpha_i y_i \bar{x}_i = 0 \Rightarrow \bar{w} = \sum_i \alpha_i y_i \bar{x}_i \tag{3.13}$$

So, the vectors $\bar{w}$ is a linear sum of $x_i s$.

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \Rightarrow \sum_i \alpha_i y_i = 0 \tag{3.14}$$

Substituting $\bar{w}$ in equation 3.12:

$$L = \frac{1}{2}\left(\sum_i \alpha_i y_i \bar{x}_i\right)\left(\sum_j \alpha_j y_j \bar{x}_j\right) - \left(\sum_i \alpha_i y_i \bar{x}_i\right)\left(\sum_j \alpha_j y_j \bar{x}_j\right) - \sum \alpha_i y_i b + \sum_i \alpha_i$$

simplifying the equation:

$$L = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j \tag{3.15}$$

Equation 3.15 shows that the optimization depends only on the dot product of samples. Rewriting the decision rule, it turns out that the decision rule also depends only on the dot product of those samples with the vector $\bar{u}$.

$$if \qquad \sum_i \alpha_i y_i \bar{x}_i \cdot \bar{u} + b \geq 0 \qquad \textit{Then the sample is a positive sample} \tag{3.16}$$

In conclusion, the classification problem will turn to a simple optimization problem, and all we need to do is to solve the equation 3.15 while considering constraint 3.16.

4. METHODOLOGY AND EXPERIMENTS

This chapter describes the methods that has been used throughout the thesis work and why the have been chosen among the ones mentioned in chapter 3. It also explains the process of EEG data gathering. Considering the goal of this research, we sketched a schematic of a general BCI without feedback as shown in figure 4.1.

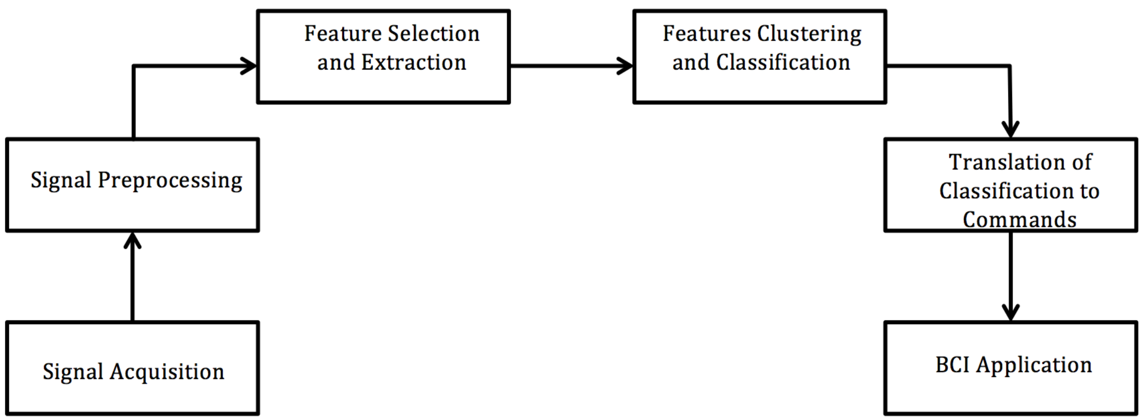

Fig. 4.1: Schematic of a common BCI.

4.1 EEG signal recording

In this thesis we were allowed to use the data recorded at the National University of Colombias Clinical Electrophysiology Laboratory. In their experiment, the brain EEG signals have been collected from 20 individuals (17 men and 3 women) with the same noise and brightness levels for all individuals. The individuals were in good health condition, while their eyes are closed. They wore a neuro-headset with 21 electrodes, plus one ground and one reference electrode. All the electrodes were places on the left hemisphere Werinckes and Brocas area, and labeled from E1 to E21 (Figure 4.2 [1]). The ground electrode was located on the lobe of the left ear, and the reference electrode was located within the EEG neuro-headset in the medial part of the forehead.

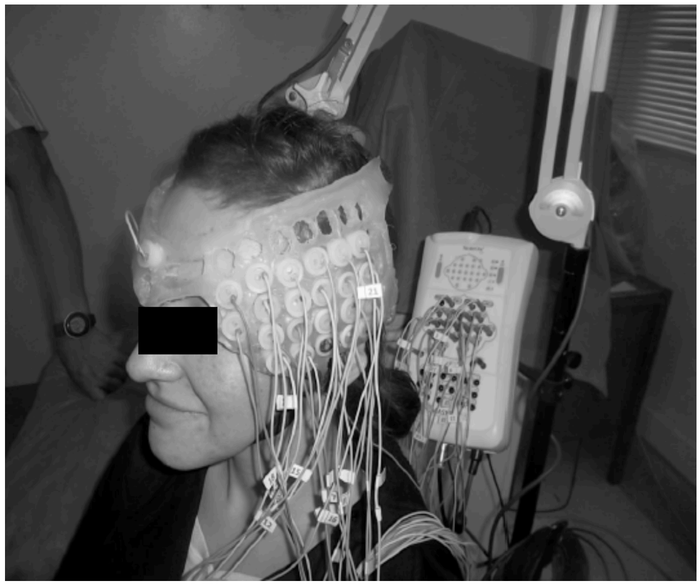

Fig. 4.2: Subject wearing the silicon EEG neuro-headset with 21 electrodes on the left hemisphere plus one frontal reference electrode.

To place each electrode on the scalp, the surface was first cleaned with an abrasive solution and later a gel conductor was applied. The lights were turned on and off periodically. The subjects have been asked to think about a specific vowel form /a/, /e/, /i/, /o/, and /u/ while the light is on and try to enter the relaxation state when the lights are turned off.

The EEG signals were amplified by an amplifier NicoletOne V32 (Natus, San Carlos, California) and processed using the software Nicolet VEEG (Natus, San Carlos, California). The sampling frequency was set to $500Hz$. The recorded signals were imported into Matlab in the form of two-dimensional arrays, where the columns represent the number of electrodes and rows show the number of the data samples. In

[1]Experiment conducted at National University of Colombias Clinical Electrophysiology Laboratory

summary, the signals that are being used in this project have been captured by an equipment with 21 channels with sampling frequency of $500Hz$. The signals are taken in periods of 6 seconds, 3 seconds of relaxation and 3 seconds of thinking. This is repeated for 11 times and the experiment is started with the relaxation mode.

4.2 Signal pre-processing

As mentioned in chapter 3, there are a number of choices for EEG signals filtering. All of those methods individually or in combination has been widely used in BCI designs. In this project, a frequency selective filter is used to remove the non-brain components of the data.

4.2.1 Neural and non-neural components of EEG data

Although EEG is invented and designed to capture electrical activities arising form brain neurons, it, unfortunately, records some non-neural activities from non-neural sites as well. These additional recorded activities are called "**Artifacts**". A necessary step in designing a BCI system is to remove artifacts from the actual EEG signals. As described before, brain EEG activities are in specific frequency bands. Therefore, a filtering can be regarded as one of the artifact rejection techniques. Figure 4.3 briefly shows the frequency band of actual brain EEG signals and major portion of artifacts (high frequency noise and low frequency muscle movements).

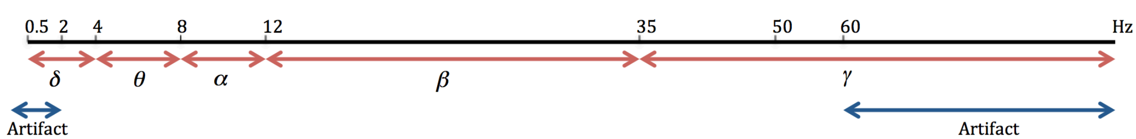

Fig. 4.3: Frequency bandwidth of components of recorded EEG.

Considering the bellow analysis of artifacts the fact that the important EEG data for a awake healthy normal person is between 4 to 35 Hz, a Low Pass Filter (LPF) and a High Pass Filter (HPF) were tested on the subjects' EEG data. These two filters were replaces with a Band Pass Filter (BPF) with same specifications but higher order. Major EEG artifacts are:

- External devices artifacts, like the a constant sound from the mechanical part of tools or electrical power supply of devices in the recording room. These types of artifacts are called "**Noise**", which has the frequency of $60Hz$ in North America and $50Hz$ in most other parts of the world.($frequency \geq 50Hz$).

- Electrode artifacts, which happens when an electrode moves in the middle of recording. This rarely happens and is almost impossible when electrodes are attached to a cap. Even if this happens, it will be less than twice per second ($frequency \leq 2Hz$).

- Muscle artifacts, which are resulted from subject's movement. An essential action before starting the experiment is to explain the sensitivity of EEG recording to the subject, so that the subject tries to get relaxed and prevent moving during the experiment. Since the EEG recording in each trial is sone in a very short amount of time (at most a few minutes), it is rarely happens that the subject moves. Same as electrode artifacts, even if a muscle artifact happens, it will be less than twice per second ($frequency \leq 2Hz$).

- Ocular artifacts, like eye blink or right and left eye movements. In the process of data acquisition for this thesis, all the subjects are given enough time to relax and close their eyes, so that we would not worry about the ocular artifacts. However, they still might happen even with close eyes. To make sure that we have considered the possibility of presence of ocular artifacts, lets imagine that the subject produces this type of artifact at most twice over second ($frequency \leq 2Hz$).

- Cardiac artifacts, which are the consequence of heart beats and is not ignorable in any situation. For a normal relaxed person, the the heart rate is between 60 to 100 beat per minute. All the subjects in the experiment of this thesis were healthy and relaxed. Let's say that they heart rate is at most 120 beat per minute, which means 2 beats per second. So that, the cardiac artifact has happened in a frequency of less that $2Hz$ ($frequency \leq 2Hz$).

4.3 Feature extraction

There are several features that can be extracted from a set of data. In different machine learning problems, different features are extracted depending on the type of the data, goal, and applications. Normally, thousands of trials are required in order to train a reliable system. Similarly, design of a BCI requires a large number of EEG recording for each of experiments on desired states. Desired states are those that will be a command later, like the vowel imaginations in this project.

Among all feature extraction methods mentioned in chapter 3, PSD of the signals has been extracted as the features. Although Fourier transform and power spectral density are more appropriate for narrow band signals, they have the lowest running time. This becomes significant while processing a large amount of data and specially for real time application. Since EEG signals are non-deterministic, PSD should be used to see how the power of a signal is distributed over different frequencies. Power spectral densities are very often widely used as a features and can be calculated as the Fourier transform of the estimated autocorrelation of signals as demonstrated in equation 3.2.

4.4 Classification

The main focus of this thesis is to design a BCI using supervised machine learning techniques for a multi-class classification problem. Some examples of classification algorithms are linear regression, random forest, neural network, naive Bayes, support vector machine, etc. Our concern is to achieve a good classifier with a high performance in terms of time, complexity, and accuracy. Many of classification approaches are designed and offered for a binary class classification, in which there are only two groups of data. For a multi-class classification problem, we would need to use a combination of classifiers trained by binary classification methods. Among all, Support Vector Machine (SVM) technique has been chosen for the classification part of this project.

SVM is a state of the art method for supervised classification in machine learning problems. It is optimal in terms of both running time and complexity. Most classification methods use all datapoints in their process of groups separation. According to SVM, it is not required to consider all datapoints, but only a number of them. As described in section 3, an SVM only uses support vectors to divide the data into two groups and does not care bout the rest of the data. This approach substantially increases the speed of training a BCI system and classifying a new set of data. Besides, this method reduces the size of data in its algorithm without decreasing the accuracy. However, the described original SVM algorithm looks powerful when the data is easily separable. Figure 3.8 in section 3.4.2 shows such a separable data.

In the situation when the data is not linearly separable (like figure 4.4), kernel functions are being used. According to this approach the space can be changed to another space in which the data can be separated (like figure 4.5). Therefore, we need to find a transformation $\phi(\bar{x})$ that takes the data from the cuurent space, into a space where their separation is more convenient.

Since the SVM optimization problem depends only on the dot product, it is sufficient to find the dot product of the transformation of one vector and the transformation of another vector to find the minimum. We do not need to know the transformation. All we need to know is a Kernel function (K) that is the dot product of vectors in the new space $\phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$. Both equations 3.15 and 3.16 will be functions of $\phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$ instead of $\bar{x}_i \cdot \bar{x}_j$.

$$K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i) \cdot \phi(\bar{x}_j) \tag{4.1}$$

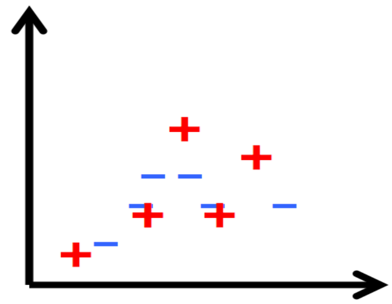

Fig. 4.4: Positives and negatives are not easily separable in this space.

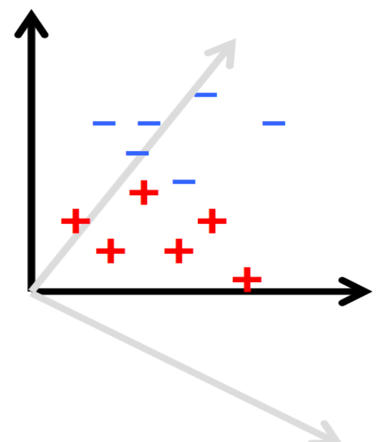

Fig. 4.5: Changing the space will result in easily separable datapoints.

At the end, the constraints in the new optimization problem will be independent of the transformation of samples $\phi(\bar{x_i})$s:

$$\begin{aligned} &minimize && L = \sum_i \alpha_i - \tfrac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\bar{x_i}, \bar{x_j}) \\ &constraints: && \sum_i \alpha_i y_i = 0 \qquad for\ i = 1, 2, 3, \cdots, N \end{aligned} \tag{4.2}$$

Different kernel functions will result in different performances and classification accuracies. Choosing the kernel function depends on the type of the classification problem. In this thesis a Gaussian Radial Basis Function (RBF) is used as the kernel. This choice was because the RBF Kernel showed the highest performance in test and trials.

The RBF kernel resulted in a higher accuracy compared to some other popular Kernel functions , while the running time of classifier training and testing for all Kernels were almost same. These popular kernels are linear kernel, quadratic kernel, polynomial kernel with order 3, and multilayer perceptron kernel. The RBF function used in this thesis is obtained from equation 4.3.

$$k(\bar{x}_i, \bar{x}_j) = e^{\frac{||\bar{x}_i - \bar{x}_j||^2}{2\sigma^2}} \tag{4.3}$$

where $\bar{x}_i$ and $\bar{x}_j$ are two dimensional vectors, $k(\bar{x}_i, \bar{x}_j)$ is the kernel of $\bar{x}_i$ and $\bar{x}_j$ and σ is the scaling factor, which we assume to be equal to 1. To make a better classification algorithm, we can apply optimization methods on parameters of the chosen kernel function.

5. RESULTS

Current goal is to work on a BCI that can help speechless people to communicate with the environment. In some literature this process is called silent speech, in which a subject mentally speaks without generating acoustic signals [31]. This requires recognition of the imagination of words. An important part in identifying different words from each other is to distinguish different vowels [28]. This needs to take the signals from the language area. Since different functions are assigned to various parts of brain, when a person mentally pronounces a word, there will be changes in the Wernickes area and Brocas area [50] on left brain hemisphere. In this project we considered five English vowels /a/, /e/, /i/, /o/, and /u/.

5.1 Data pre-processing

5.1.1 EEG segmentation

As mentioned in chapter 4, section 4.1, the data of this thesis has been captured from 20 subjects through a non-invasive process using a EEG recording system with 21 electrodes placed on the left hemisphere. All subjects were in good health condition and were given time to close their eyes and relax.

Each trial includes 3 seconds of relaxation and 3 seconds of thinking, so total of 6 seconds per trial. Every recording session contains 11 of trials for an individual, so that the time of each single recording is 66 seconds. The sampling frequency was set to $500Hz$, so that each signal has been generated in 33000 samples. Figure 5.2 shows a plot of a set of 21 amplified EEG signals taken in one experiment from one subject.

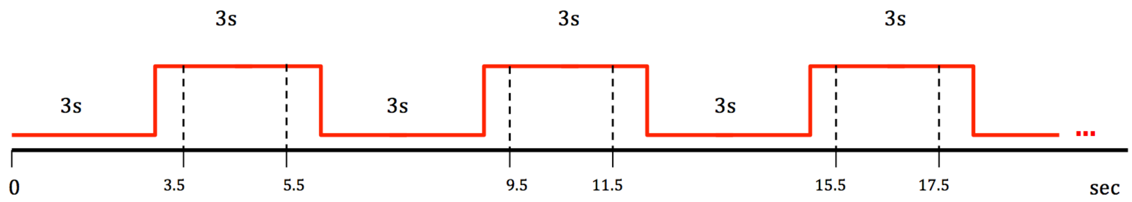

Fig. 5.1: EEG recording time frames.

$$\begin{cases} \text{sampling frequency} = 500Hz = 500\ \frac{datapont}{second} \\ \text{total time of the experiment} = (3+3) \times 11 = 66\ seconds \\ \text{total datapoint in the signal} = 500 \times 66 = 33000 \end{cases}$$

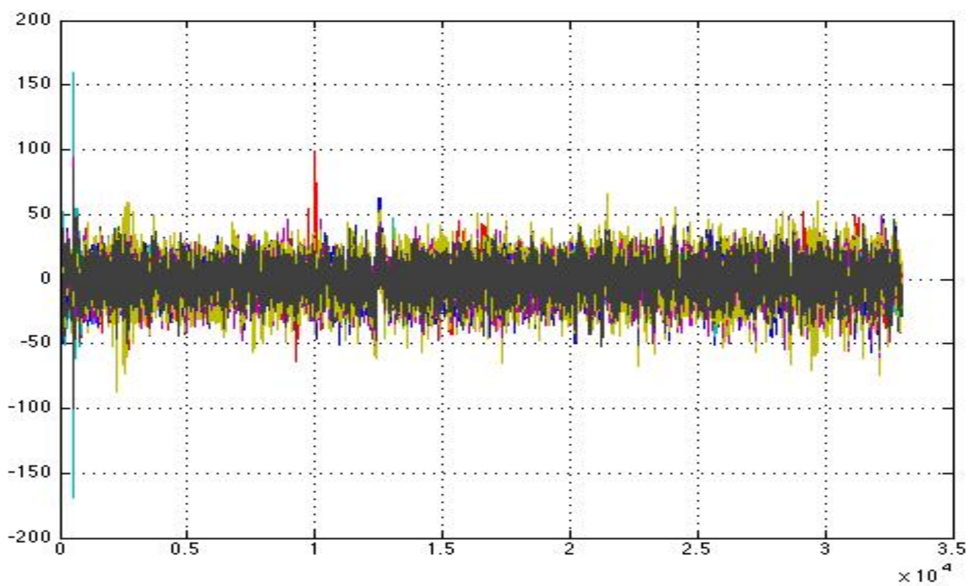

Fig. 5.2: Raw signals from 21 channels taken from one subject including 11 consequtive trials.

Figure 5.1 shows the recording timing. As demonstrated in this figure we would ignore the first and last 0.5sec of each thinking session. In other words, if we take the first period (first 6 seconds). We have $33000 \div 11 = 3000\frac{samples}{period}$. So we have 1500 samples in each 3 seconds. we have to consider the 3 seconds that the person is thinking about something, meaning from $t = 3$ to $t = 6$. To remove the transition state effects, we consider from $t = 3.5$ to $t = 5.5$ seconds, which is the samples 1750 to 2750 as shown in figure 5.3.

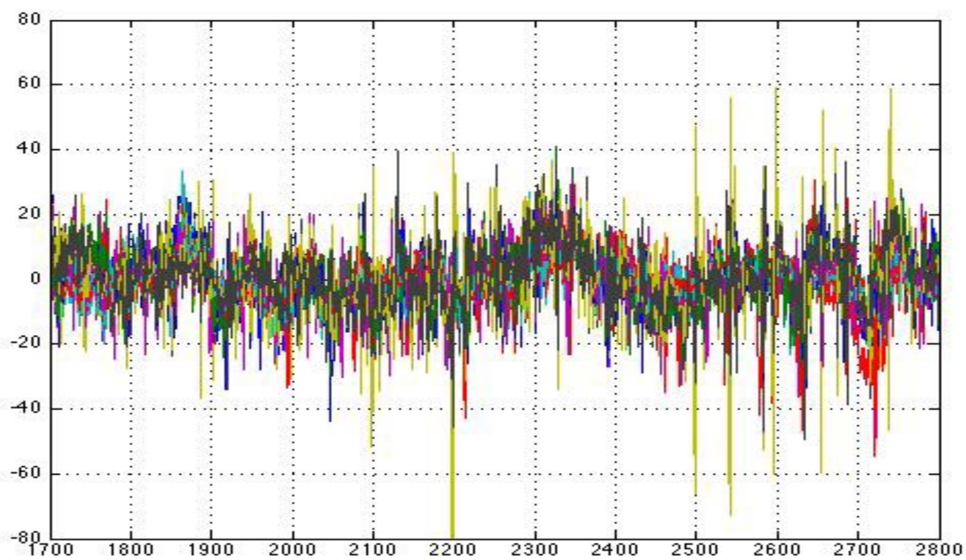

Fig. 5.3: Raw signals for 21 channels in one trial.

5.1.2 Filtering

There are many low frequency artifacts and high frequency noise that have to be removed from the signal before extracting features. For this purpose, first, two elliptic filters were designed. We let Matlab to find the minimum order which match our specifications. One of the filters is an elliptic LPF with the cutoff frequency of

50Hz ($\frac{1}{2}(F_{stop} + F_{stop})$), stopband attenuation of 80dB (A_{stop}), and passband gain of 1dB (A_{pass}). The other is an elliptic High Pass Filter (HPF) with the cutoff frequency of 2 Hz ($\frac{1}{2}(F_{pass} + F_{stop})$), stopband attenuation of 80dB (A_{stop}), and passband gain of 2dB (A_{pass}). Later, these two filters were replaced with a Band Pass Filter (BPF) with the same specifications. Figures 5.4 and 5.5 show the specifications and designed BPF filter respectively.

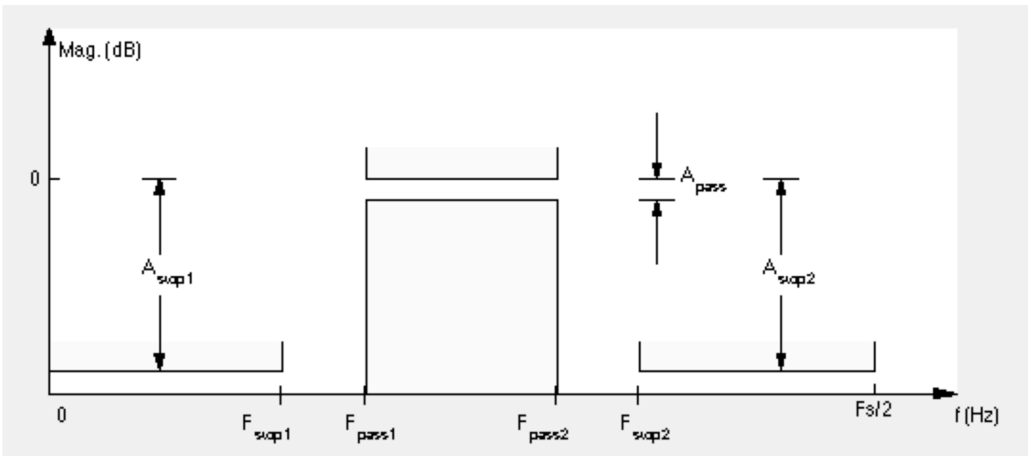

Fig. 5.4: Bandpass filter specifications.

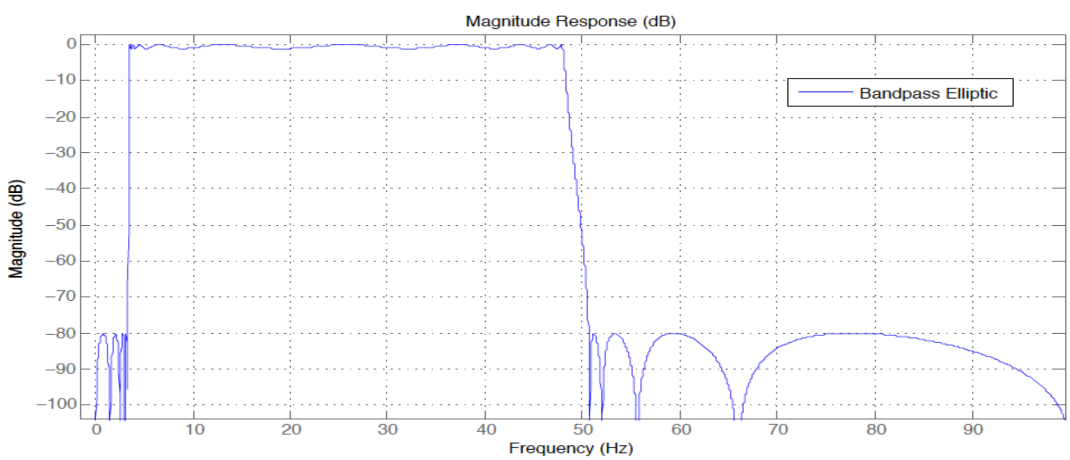

Fig. 5.5: Band pass filter.

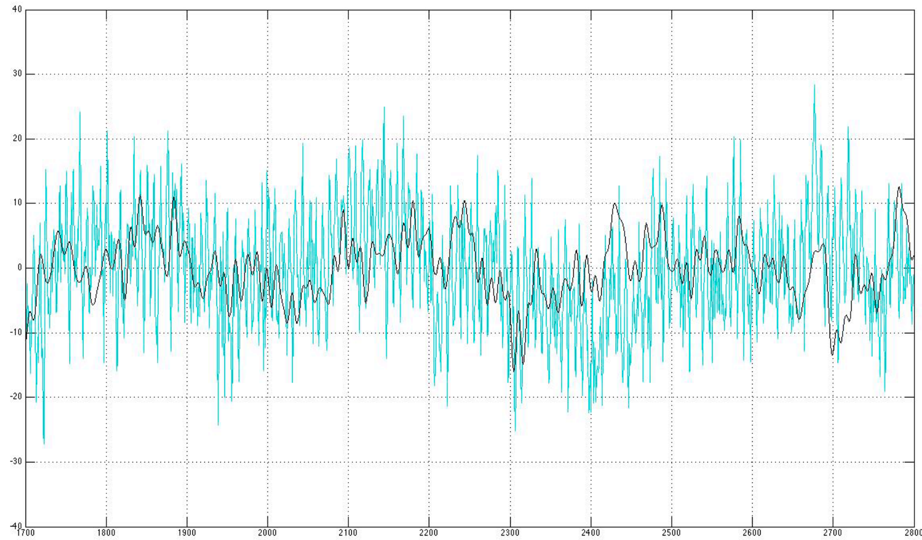

Fig. 5.6: Signal from one electrode before and after BPF(noise and artifact rejection).

To show the workflow only one sample of signal from one electrode from one subject is plotted in each step. Figure 5.6 shows the signal from the first subject, for the first trial of thinking about /a/, before and after filtering.

5.2 Feature extraction

In this project, the frequency spectrum of the signal has been extracted as the feature for later classification. The data used in this project is stochastic, so Power Spectral Density (PSD) was used to find the frequency spectrum of the signals. To reduce the size of the data matrix, we used the periodogram technique while taking the PSD [28].

The periodogram is generally the process of breaking the time series data into small pieces, working with each piece separately, and then combine all pieces again. It can be summarized in 4 steps:

- Breaking up the signal into small pieces.
- Taking the PSD in each piece.
- Taking the average of the PSD values in each piece.
- Considering each piece as a sample and representing the averaged value as the PSD of that piece.

In the work flow of this project, there are several strategies for reducing the size and dimension of the data. Since we have total of 21 electrodes, the frequency spectrum of all the signals in each trial create vectors of 21 values. Considering the number of trials equal to 1100 and number of samples equal to 1000 in each trial, we would have 1100 data vectors with dimensions 1000×21. All the data can be shown in Matlab with a 3 dimensional matrix with dimensions $1100 \times 1000 \times 21$, which will be $1100 \times 513 \times 21$ after periodogram. Still, in order to use matlab functions in an efficient way, we have to reduce the dimension from three to two. For this purpose, we take the average of all datapoint for each electrode for each trial. Figures 5.7 and 5.8 show the average of periodogram values for the first trial from subjects one and two. These two figures show bar plots of the values of PSD in each of our desired 5 classes for each 21 channels are included.

Comparing figures 5.7 and 5.8, there seem to be some similarities between the change in the behavior of the PSD values. There looks like to be a relationship between electrodes in each class, which is different from other classes. This can be used as the signal characteristics that can be as input for later classification.

5.3 Classifier training using known data

Among the method explained in section "Feature classification" in chapter 3, the last method (Support Vector Machines (SVM)) is used in this work. There are two main reasons for this choice: 1- SVM is the state of the art of supervised machine learning problems specially for non-deterministic and stochastic data, 2- SVM has

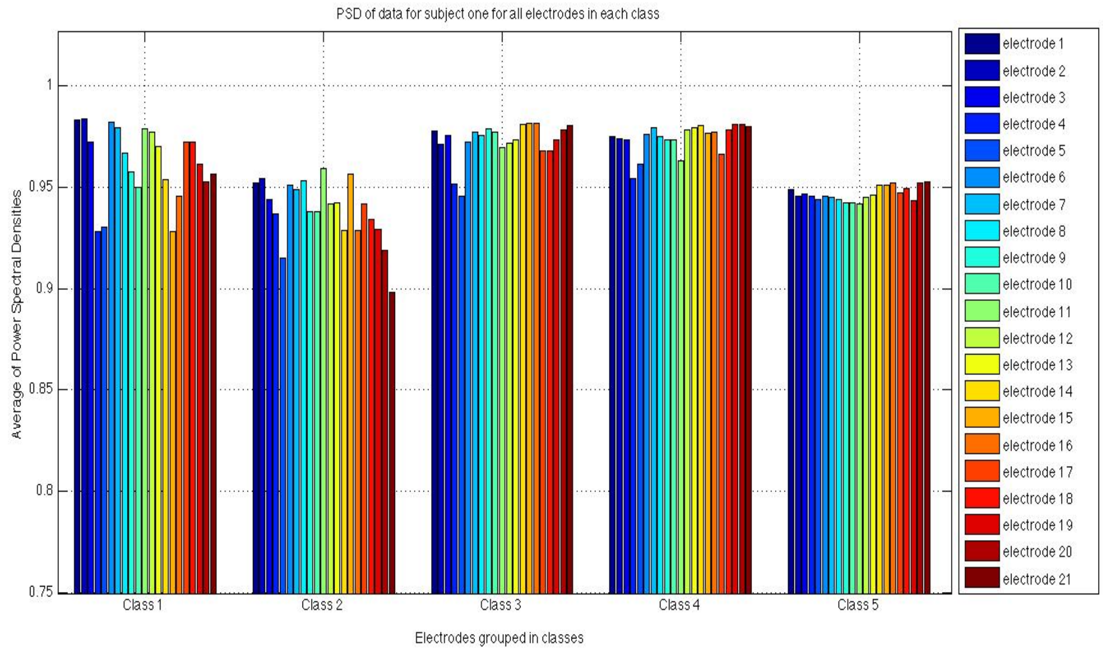

Fig. 5.7: Taking average of periodograms over all datapoints gives a value for each channel. This figure shows this value for 21 channels in each class in first trial with subject one.

had the best performance for classification problems among other methods [96, 97]. However, the SVM is designed for binary classification problems only. Since our data contain 5 classes, we can not use the SVM algorithm directly. Instead, we have to implement an algorithm that allows to take advantages of SVM for multi-class classification problems. Two approaches that have been widely used in multi class classifications are one-against-all and one-against-one methods.

The one-against-all approach uses a binary SVM to distinguish samples of each class from all other classes. Using this approach, an N-class problem will be broken into n two-class problem. In each of these binary problems, the class labels will be set as $y = 1$ for those data that are placed in the specified class , and $y = -1$ for the rest of data. According to this approach, at the end, a data point will be placed in a

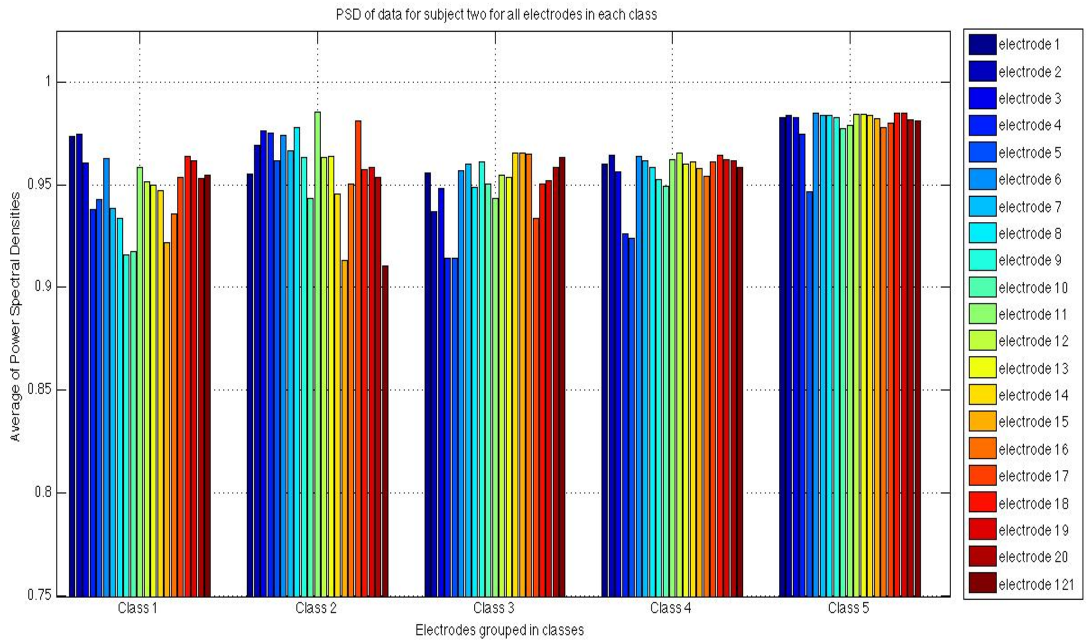

Fig. 5.8: Taking average of periodograms over all datapoints gives a value for each channel. This figure shows this value for 21 channels in each class in first trial with subject two.

specific class only if it is recognized as a class member by the SVM of that class and is rejected by all other SVMs for other classes, meaning that x_i belongs to class i if only for class i we have $y = 1$ and for all other classes $y = -1$ [92]. In this method, one SVM per class is required, which means N SVMs are needed to classify N groups. Beside, some of the data points will remain undecided if they are not accepted by any SVM or by more than one SVMs.

The one-agains-one approach uses a binary SVM to classify each possible pairs of classes. At the end, a datapoint will be placed in the class that is chosen by the maximum number of binary SVMs [98]. In this approach, $N(N-1)/2$ SVMs are required to classify N groups. Beside, it requires a set of computations to chose the correct class.

In 2002, Takahashi and Abe [99] proposed a new approach based on decision tree to solve multi-class SVM problems. They suggest 4 types of decision trees:

Type 1 Each classifier (SVM in our problem) at each node separates one class from the remaining classes using the Euclidean distance.

Type 2 Each classifier at each node separates some class from the remaining classes using the Euclidean distance.

Type 3 Each classifier at each node separates one class from the remaining classes using the Mahalanobis distance.

Type 4 Each classifier at each node separates some class from the remaining classes using the Mahalanobis distance.

In this thesis, we proceed with the following steps to create a type 2 decision tree. We will refer to these steps as the tree algorithm:

Step 1 Calculating the center of each class using equation 5.1 and and the distance between each two classes using equation 5.2.

$$c_i = \frac{1}{N} \sum_{x \in X_i} x \tag{5.1}$$

where c_i is the center for class i, x_i is set of training data included in class i, and N is the number of elements in class i.

$$d_{ij} = ||c_i - c_j|| \tag{5.2}$$

where $d_{ij} = d_{ji}$ is the distance between center of class i and class j. At this step all the classes are separated from each other.

Step 2 Finding the smallest distance between class i and all other classes using equation 5.3.

$$l_i = \min_{j=1,\ldots,N, i \neq j} d_{ij} \tag{5.3}$$

based on calculated l_is, we put associated classes into the same cluster.

Step 3 Repeating from step 1 ($N-2$ times) until we have only two clusters.

Step 4 Using an SVM to separate two clusters from step 3.

Step 5 Checking if each cluster has exactly two or has more than two classes. If it is exactly tow classes, we train an SVM to separate these two classes. If it has more than two classes, we consider each class in a separated cluster and repeat from step 3 to 5.

Figure 5.9 shows how above algorithm works in each step for our EEG data. According to the algorithm, the mean of all data samples in each class is calculated first. This mean will be the representative of the whole group. For example, imagine the average of all samples from group /a/ (or group (1) in the figure) returns the vector $\bar{x}_1$. From this point to the end of first iteration, vector $\bar{x}_1$ represents group (1). Doing this for all groups of data, we would have vector $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$, and $\bar{x}_5$ representing classes a/, /e/, /i/, /o/, and /u/ respectively. Second, the distance of vectors will be calculated and compared. The minimum distance determines the closest groups. As shown in the figure 5.9 groups (4) and (5) are found as the closest ones among all possible pairs. Merging them into one group, creates group (6). At this point, there are only four groups of data instead of five groups to be separated, which are groups (1), (2), (3), and (6), as shown in figure 5.9b.

Second iteration of algorithm results in vectors $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, and $\bar{x}_6$ as representatives of groups (1), (2), (3), and (6). Calculating the distances of all possible pairs and comparing them to each other, groups (2) and (3) achieve the minimum distance. So that, they will be placed into one single cluster as group (7). This turns the original 5-class classification problem into a 3-class classification problem as demonstrated in figure 5.9c.

Although the original problem is simplified into a new one with fewer groups of data, still we need to do another iteration of the algorithm to achieve a binary class classification problem. Computing the mean of all samples in each of our new groups (1), (6), and (7) creates averaged vectors of $\bar{x}_1$, $\bar{x}_6$, and $\bar{x}_7$. Among all possible pairs

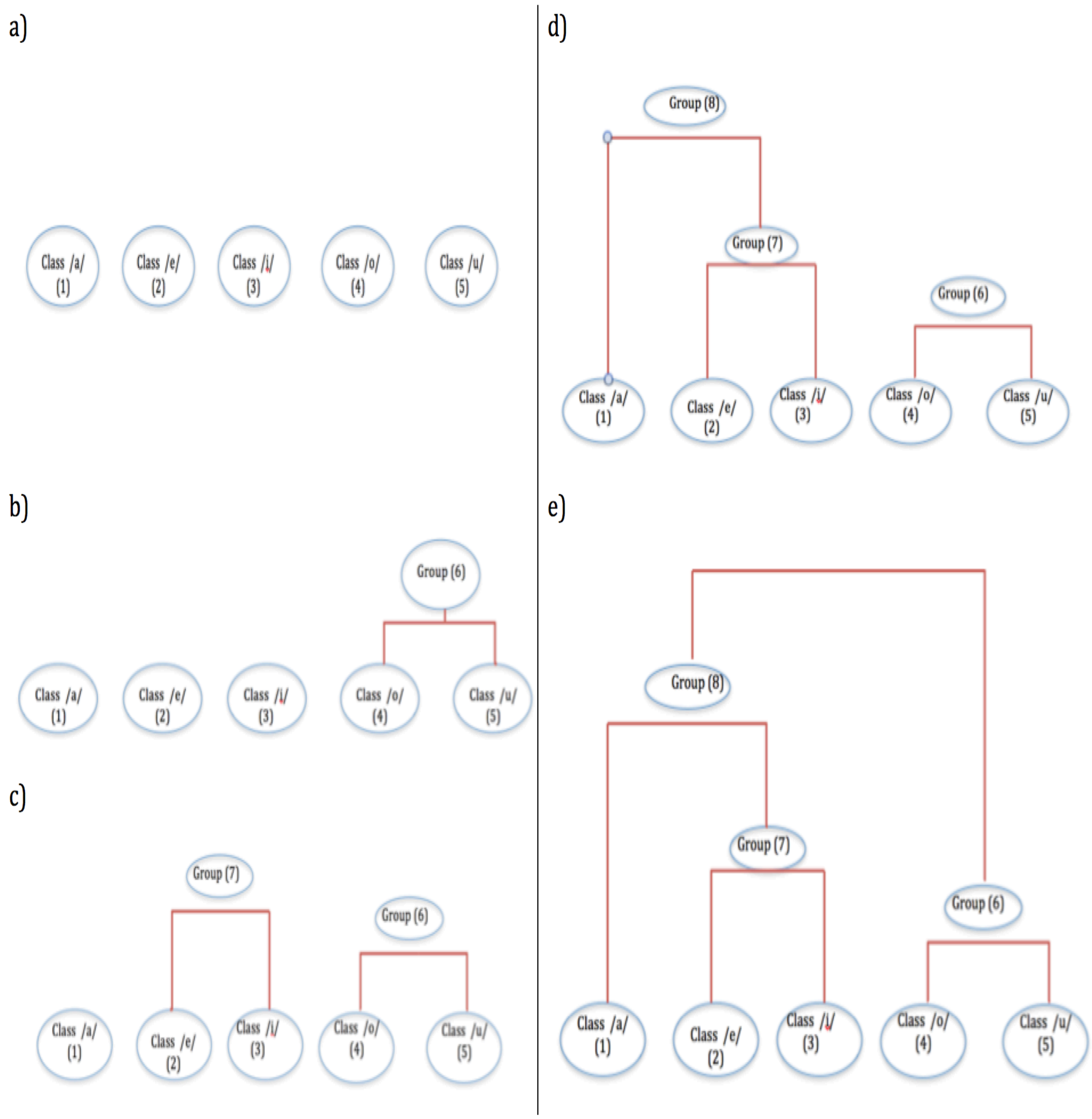

Fig. 5.9: a) groups of training data, b)first iteration of applying the tree algorithm creates a new group (6), c) second iteration of applying the tree algorithm creates the group (7), d) third iteration of the tree algorithm applying on groups (7), (3), and (6) creates a new group (8), e) finally there is only two groups of (6) and (8) to classify.

of this step, the minimum distance goes to the distance between groups (1) and (7). Putting the sample data of group (1) and (7) in a new group (8) (see figure 5.9d), reduces the number of classes from three to two.

At this point there are only two groups of (6) and (8) to be separated. The tree algorithm can be stopped from doing another iteration, since we reached the goal of having a binary classification problem and there is no need to go over the tree algorithm for another iteration. A decision tree has been created in such a way that all original groups sit on the leaves of the tree and each parent node can have only two child nodes. Also, at each level of the tree only one parent node can create its children. Running the algorithm in MatLab gives the tree shown in figure 5.10.

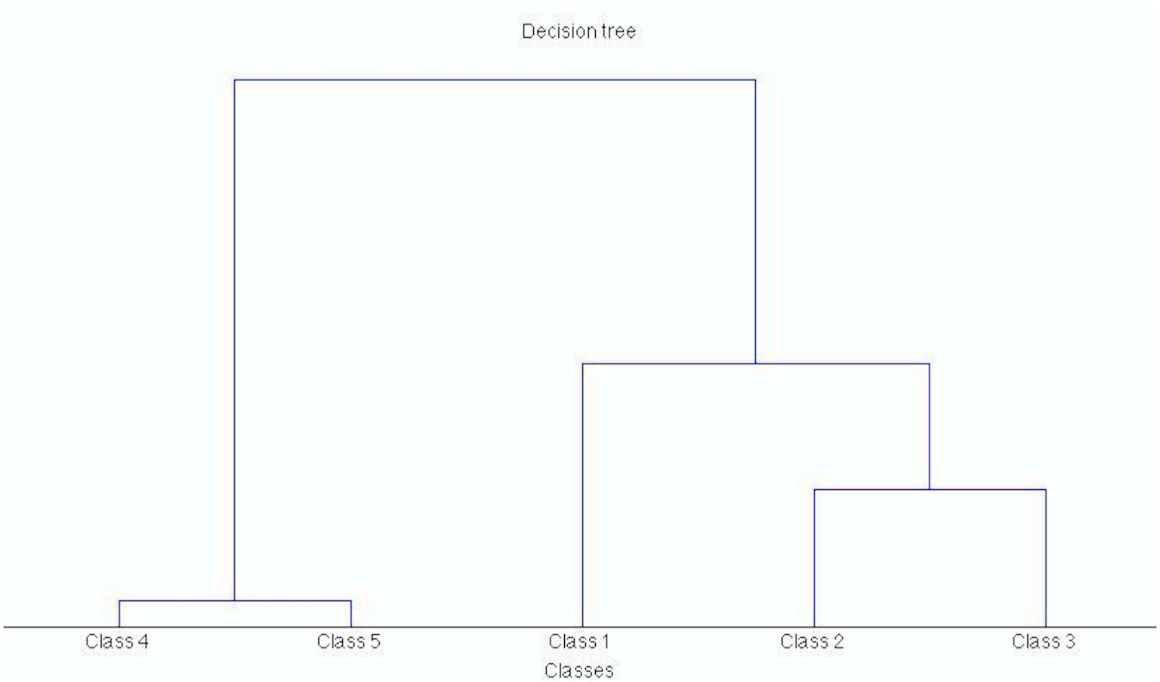

Fig. 5.10: Grouping classes based on their centers' distance.

Having only one parent node at each level of the tree, means there is a binary class problem at each level. Starting from the root node of tree, we can train an SVM classifier to separate groups (6) and (8). Next level is where an SVM should be train to divide groups (1) and (7). The third SVM classifier is trained for separating groups (2) and (3). Finally groups (4) and (5) will be used to train the classifier SVM4. This is demonstrated in figure 5.11.

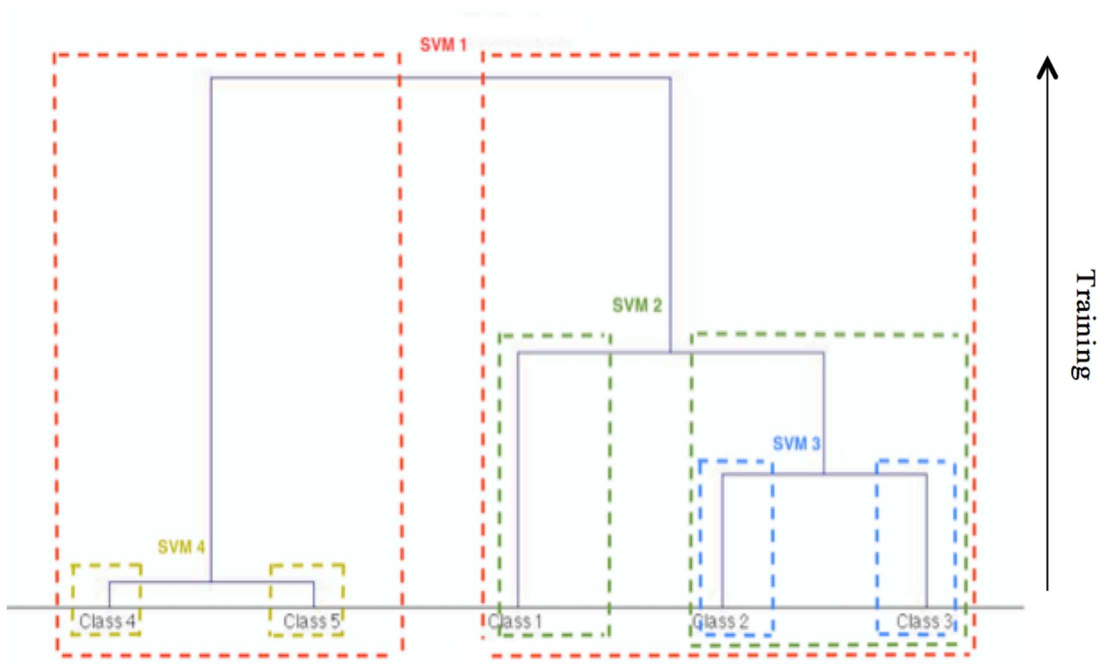

Fig. 5.11: SVM training for groups of classes.

5.4 Classifying an unknown data

Having all the SVM classifiers trained, the classification of a new data should be easy. At this point we have trained four SVMs to separate groups at each level. As shown in figure 5.12, to predict the class of a new set of data, it is required to move from top to bottom [100], Take x_k as a new EEG data set that wants to sit in one of five classes of /a/, /e/, /i/, /o/, or /u/. Starting from the top of the tree SVM1 should be activated to place the new data set into either group (6) or (8). At the second step, depending on the prediction of SVM1 one of SVM2 or SVM4 will be activated. If we imagine SVM2 is chosen to be activated, depending on the detection of this classifier, x_k will be placed in either group (1) or group (7). If group (1) is picked, the process will terminate and x_k is going to class (1) or /a/. If not, SVM3 starts to detect whether x_k is in group (2) or (3).

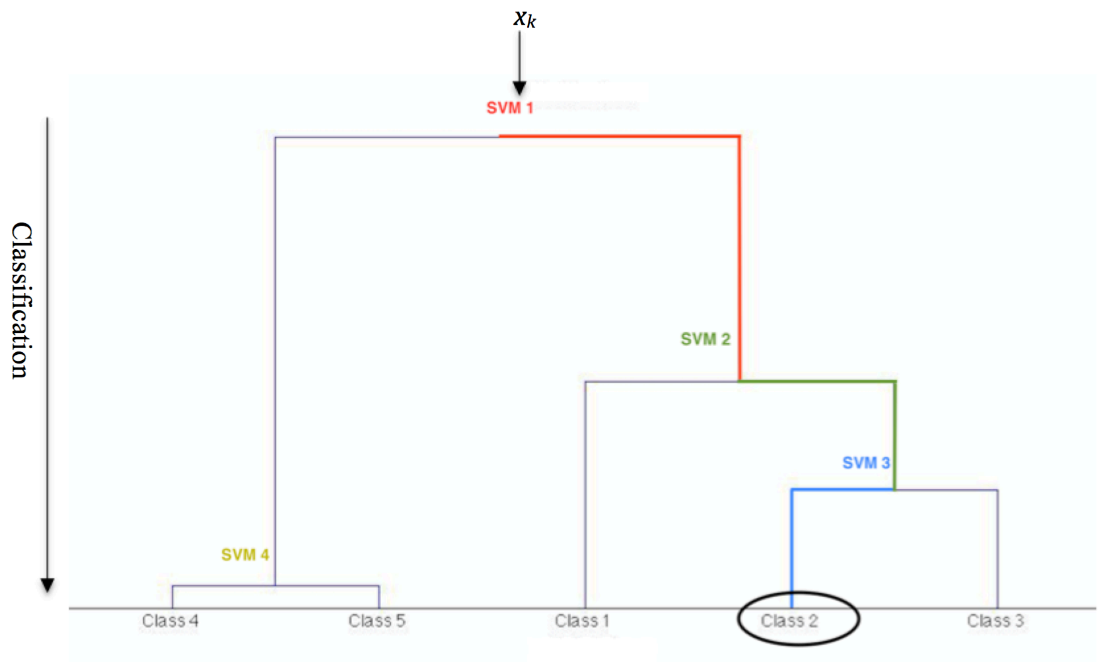

Fig. 5.12: Classifying the new input dataset x_k.

The process of EEG signals recording was repeated in 11000 trials. This number of trials were split into two groups of training and testing. The training set contain 90% of the whole data and testing set contain 10% of the whole data. The confusion matrix shown in figure 5.13 demonstrates the result of classification and table 5.1 shows the details of the confusion matrix.

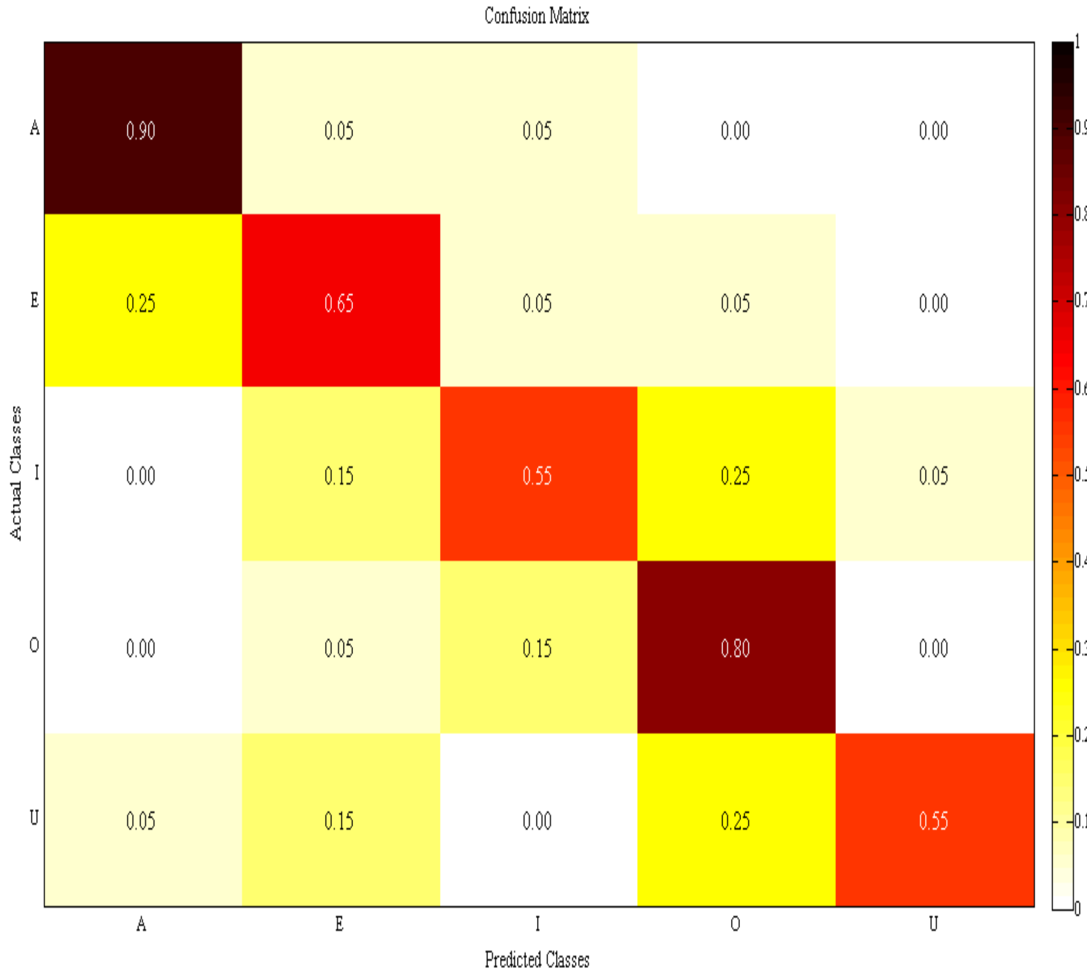

Fig. 5.13: Confusion matrix obtained for classification.

Table 5.1: Confusion matrix interpretation.

	Class A	Class E	Class I	Class O	Class U
Correctly Detected	90%	65%	55%	80%	55%
Incorrectly Detected	5% *in E* 5% *in I*	25% *in A* 5% *in I* 5% *in O*	15% *in E* 25% *in O* 5% *in U*	5% *in E* 15% *in I*	5% *in A* 15% *in E* 25% *in O*

6. SUMMARY AND FUTURE RECOMMENDATION

6.1 Summary

This research revolves around training and testing a non-invasive Brain Computer Interface (BCI) system. In this thesis, EEG signals captured from auditory area of brain were used. First, the process of brain EEG signals generation and recording is explained. Then, the process of noise removal, artifact rejection, feature extraction, system training, and system classification is elaborated in details.

This thesis starts with history and background of BCIs and introduces similar works in this field. It talks about the potentials of BCIs and applications. Besides, the shortcomings of current BCIs implementation is described. The main focus of this thesis is to recognize silent speech from EEG signals, which means the recognition of signals during a speech imagination in mind without actually speaking.

This work's EEG signals were taken from 20 subjects using a EEG recording system with 21 electrodes. This system is a product of NicoletOne (Natus, San Carlos, California) and placed in the National University of Colombia. All subjects were supposed to think about a specified case from 5 different cases that are thinking about /a/, /e/, /i/, /o/, and /u/. Signals were taken following a specific pattern. Taking these signals as inputs, the goal of this thesis is to realize if it is possible to train a non-invasive BCI or not.

It is required to know about the specifications of our training data (input signals) in order to be able to implement an algorithm to train a BCI. The input signals in this work are EEG signals from language area. These signals are produced when neurons are communicating and can be measured over scalp via different methods such as electrode based EEG headsets. In general, EEG signals are harmonic and they have a frequency band from less than $1Hz$ to around $50Hz$.

Since EEG signals are generated with a specific bandwidth, it is possible to filter the important information from EEG signals from the non-brain activities with different frequencies. We can get rid of a big portion of low frequency artifacts and high frequency noises using a Band Pass Filter (BPF). Here, an elliptic BPF with upper and lower cutoff frequencies of $2Hz$ and $50Hz$ were used. signals after filtering were much smoother and comparable. However, still they are in different ranges for different trials. This can be solved with a normalization in a specific interval. Here, all the signals were normalized between -1 and 1. After normalization, signals are ready to compare, so that it is time move forward to the next step which is feature extraction.

After reviewing several methods of feature extraction, we decided to use the power spectrum of signals as their feature. In order to calculate power spectral density and meanwhile reduce the size of our input data, we used periodogram, which in generally means breaking the time series data into small pieces, taking power spectral density in each piece separately, and then attach all pieces again. applying periodogram returns the feature space in a half size that is being used for training and classification of data.

The last but not the least part of this thesis is to train a classifier and finally test the performance of the whole system. Here, ninety percent of data were used for training and ten percent for testing. In most recent literatures, support vector machines has showed the best performance among other classifiers. Although, it is originally designed for binary classification problems, there are some strategies to use support vector machines for multi-class classification problems. Three example are one-against-all, one-against-one, and decision tree methods. In this thesis, decision tree approach has been used.

The use of decision tree makes it possible to solve a multi-class classification problem with less support vector machines rather than one-against-all and one-against-one. Considering N as number of classes, the one-against-all and one-against-one approaches require N and $N(N-1)$ classifiers respectively. The decision tree lets us to solve the same problem with at most $N-1$ classifiers. There are four types of decision trees that are different in their method of classes separation. In this project, decision tree type two, in which each classifier at each node separates some class from the remaining classes using the Euclidean distance, is used.

The algorithm for decision tree used our training data to identify how it is possible to reduce the number of classes by grouping them in same clusters. The original classes are grouped based on the closeness of the center of every class to all other classes. At the end of this step, classes four and five will be grouped in the same cluster. Similarly classes two and three are put in the same cluster. repeating this process, there will be two clusters only, which defines a new binary classification problem. At this point, support vector machines can be trained.

Once the training of support vector machine classifiers is done, the testing can be started. Having a new input data and moving from top of the tree to bottom, a trained support vector machine classifier identifies that the new data belongs to which group at each step. The system repeats this procedure until the input data is place in one of five original groups.

The experiment result show that class /a/ has the highest recognition accuracy with ninety percent of data correctly detected. Second place goes to class /o/ with eighty percent of accuracy. After that is class /e/ which has sixty-five percent of accuracy. This class looks like to be mistaken with class /a/ for 25 percent of test data. Finally, the system seem to have the lowest classification performance for classes /i/ and /u/ with fifty-five percent accuracy. Both groups look like to be highly mistaken with class /o/.

6.2 Future recommendations

This work can be expanded and improved in several ways. Different parts of the design can be modified depending on the final goal and application. Fortunately, the algorithm works for any number of classes, so is usable for a different dataset.

However, there might be some modifications required for inputs with different natures and characteristics. Any BCI is designed for a final application. Depending on the nature of the data that is being used, goals, and applications, different types and levels of modifications are required. Some areas that this work can be used in are in the following.

To develop this work in the line of speech recognition, it is required to collect larger set of data. besides, Each step of this thesis is focused on only one method. In order to achieve a higher accuracy in a shorter running time there can be some manipulations in each step. Bellow are some examples of future works that can improve this work:

- Capturing more EEG signals in different subjects' situations and room's conditions: this might decrease the accuracy of recognition , but will eventually have a higher accuracy when having an unknown data, which in turn helps to have a system for a wider ranges of applications. This is because the system can recognize more features of an unknown data if it is learned from a bigger set of data with more variety.

- Applying a higher level preprocessing like filtering: In this work a BPF was used. It can be efficient if a higher level filter would be replaced. For example spatial filters (like ICA) can perform as a signal source detection and sometimes a feature selection method. Spatial filters have been widely used for EEG signals. However, the performance of different types of filters need to be tested on the experimental data.

- Trying correlation and normalization on the data: there are several trials within and between subjects. The states, feeling, environment, and condition of subjects might change unexpectedly in every EEG recording. This change affects the signals we are taking through sensors. Correlation and normalization help to eliminate these effects. To reject the undesired changes in different trials for the same subject we can take a correlation function. On the other hand, normalization helps to reject person-to-person data variations.

- Extracting more features: performance of a system can be improved dramatically if more and more features are send to machine learning steps (classification in this project). Any of other features introduced in this document (and many others that are not mentioned here) can be extracted as a new set of features. For the EEG data, w can define the averaged PSD in each of theta, alpha, beta, and gamma frequency bands and take them as separate features. Note that the number of features should be consistent with the number of trials. Generally, it is recommended to have the number of trials at least 7 times more than the number of feature. To deal with this problem, a new step of feature selection can be added to the algorithm.

- Adding a feature selection step: it would significantly improve the performance of the system if we extract many relevant features and then select only a number of the most effective ones. This step should be added between feature extraction and classifier training. There has been many methods offered in the literature for this purpose.

The algorithm proposed in this research can be used for other purposes like in control of a robot, a wheelchair, or any moving object. However, it is recommended to use a different set of data as inputs to achieve higher accuracy. Instead of vowels, we can use left/right hand or left/right foot movements in trials. Movement imagination EEG signals are much better distinguishable rather than vowels EEG signals. Furthermore, it is more convenient for the user to memorize what imagination will be translated to which direction movement.

In addition, this work can be expanded in feeling detection. Using the same algorithm, it is easy to design a system for emotion detection. Here, like the previous case, it is better to use a different set of data which are recorded in trials where the subjects express a feeling. Detection of love, sympathy, pride, shame, fear, anger, and many more can be done using the algorithm of this thesis.

LIST OF REFERENCES

LIST OF REFERENCES

[1] L. Haas, "Hans Berger (1873–1941), Richard Caton (1842–1926), and Electroencephalography," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 9–9, 2003.

[2] H.-R. Wiedemann, "Hans Berger," *European Journal of Pediatrics*, vol. 153, no. 10, pp. 705–705, 1994.

[3] J. Wolpaw and E. W. Wolpaw, *Brain Computer Interfaces: Principles And Practice*. Oxford University Press, 2012.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain Computer Interfaces For Communication And Control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[5] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain Computer Interface Systems: Progress And Prospects," *Expert Review of Medical Devices*, vol. 4, no. 4, pp. 463–474, 2007.

[6] S. Silvoni, A. Ramos-Murguialday, M. Cavinato, C. Volpato, G. Cisotto, A. Turolla, F. Piccione, and N. Birbaumer, "Brain Computer Interface In Stroke: A Review Of Progress," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 245–252, 2011.

[7] D. Xiao, Z. Mu, and J. Hu, "Classification Of Motor Imagery EEG Signals Based On Energy Entropy," *International Symposium on Intelligent Ubiquitous Computing and Education*. IEEE, pp. 61–64, 2009.

[8] M. A. Lebedev and M. A. Nicolelis, "Brain Machine Interfaces: Past, Present And Future," *TRENDS in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.

[9] S. Siuly and Y. Li, "Improving The Separability Of Motor Imagery EEG Signals Using A Cross Correlation-Based Least Square Support Vector Machine For Brain Computer Interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 526–538, 2012.

[10] R. Kus, D. Valbuena, J. Zygierewicz, T. Malechka, A. Graeser, and P. Durka, "Asynchronous BCI Based On Motor Imagery With Automated Calibration And Neurofeedback Training," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 823–835, 2012.

[11] W. K. Tam, K.Y. Tong, F. Meng, and S. Gao, "A Minimal Set of Electrodes For Motor Imagery BCI To Control An Assistive Device In Chronic Stroke Subjects: A Multi-Session Study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 6, pp. 617–627, 2011.

[12] R. C. Panicker, S. Puthusserypady, and Y. Sun, "Adaptation In P300 Brain Computer Interfaces: A Two Classifier Co-training Approach," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2927–2935, 2010.

[13] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A Speedy Hybrid BCI Spelling Approach Combining P300 And SSVEP," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 473–483, 2014.

[14] H. Gollee, I. Volosyak, A. J. McLachlan, K. J. Hunt, and A. Graser, "An SSVEP-Based Brain Computer Interface For The Control Of Functional Electrical Stimulation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1847–1855, 2010.

[15] C. Jia, X. Gao, B. Hong, and S. Gao, "Frequency And Phase Mixed Coding In SSVEP-Based Brain Computer Interface," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 200–206, 2011.

[16] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain Computer Interfaces, A Review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[17] M. Teplan, "Fundamentals Of EEG Measurement," *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002.

[18] S. Waldert, T. Pistohl, C. Braun, T. Ball, A. Aertsen, and C. Mehring, "A Review On Directional Information In Neural Signals For Brain Machine Interfaces," *Journal of Physiology-Paris*, vol. 103, no. 3, pp. 244–254, 2009.

[19] D. J. McFarland, C. W. Anderson, K. Muller, A. Schlogl, and D. J. Krusienski, "BCI Meeting 2005 - Workshop on BCI Signal Processing: Feature Extraction And Translation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 135–138, 2006.

[20] M. Safari, A. Kordi, M. Oladazimi, F. Shiman, and F. M. Vaneghi, "A Review Of Most Current Feature Extraction Methods For EEG Signal Processing," *International Conference on Computer and Automation Engineering, 4th (ICCAE 2012).* ASME Press, 2012.

[21] F. Lotte, M. Congedo, A. Lécuyer, and F. Lamarche, "A Review Of Classification Algorithms For EEG-Based Brain Computer Interfaces," *Journal of Neural Engineering*, vol. 4, 2007.

[22] P. R. Kennedy and R. A. Bakay, "Restoration Of Neural Output From A Paralyzed Patient By A Direct Brain Connection," *Neuroreport*, vol. 9, no. 8, pp. 1707–1711, 1998.

[23] G. B. Stanley, F. F. Li, and Y. Dan, "Reconstruction Of Natural Scenes From Ensemble Responses In The Lateral Geniculate Nucleus," *The Journal of Neuroscience*, vol. 19, no. 18, pp. 8036–8042, 1999.

[24] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. A. Nicolelis, "Real Time Prediction Of Hand Trajectory By Ensembles Of Cortical Neurons In Primates," *Nature*, vol. 408, no. 6810, pp. 361–365, 2000.

[25] D. M. Taylor, S. I. H. Tillery, and A. B. Schwartz, "Direct Cortical Control Of 3D Neuro-Prosthetic Devices," *Science*, vol. 296, no. 5574, pp. 1829–1832, 2002.

[26] S. Musallam, B. Corneil, B. Greger, H. Scherberger, and R. Andersen, "Cognitive Control Signals For Neural Prosthetics," *Science*, vol. 305, no. 5681, pp. 258–262, 2004.

[27] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A New Approach For EEG Feature Extraction In P300-Based Lie Detection," *Computer Methods and Programs in Biomedicine*, vol. 94, no. 1, pp. 48–57, 2009.

[28] L. Sarmiento, P. Lorenzana, C. Cortes, W. Arcos, J. Bacca, and A. Tovar, "Brain Computer Interface (BCI) With EEG Signals For Automatic Vowel Recognition Based On Articulation Mode," *Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC), 5th ISSNIP-IEEE*. IEEE, pp. 1–4, 2014.

[29] X. Pei, D. L. Barbour, E. C. Leuthardt, and G. Schalk, "Decoding Vowels And Consonants In Spoken And Imagined Words Using Electro-Corticographic Signals In Humans," *Journal of Neural Engineering*, vol. 8, no. 4, p. 046028, 2011.

[30] L. Wang, X. Zhang, X. Zhong, and Y. Zhang, "Analysis And Classification Of Speech Imagery EEG For BCI," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 901–908, 2013.

[31] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent Speech Interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270–287, 2010.

[32] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic Brain Mapping," *Signal Processing Magazine, IEEE*, vol. 18, no. 6, pp. 14–30, 2001.

[33] W. O. Tatum IV, *Handbook Of EEG Interpretation*. Demos Medical Publishing, 2014.

[34] J. Malmivuo and R. Plonsey, *Bio-Electromagnetism: Principles And Applications Of Bioelectric And Biomagnetic Fields*. Oxford University Press, 1995.

[35] V. Botella-Soler, M. Valderrama, B. Crépon, V. Navarro, M. Le Van Quyen, and O. Sporns, "Large Scale Cortical Dynamics Of Sleep Slow Waves," *PloS One*, vol. 7, no. 2, p. e30757, 2012.

[36] I. E. Wickramasekera, "On Attempts To Modify Hypnotic Susceptibility: Some Psychophysiological Procedures And Promising Directions," *Annals of the New York Academy of Sciences*, vol. 296, no. 1, pp. 143–153, 1977.

[37] M. E. Sabourin, S. D. Cutcomb, H. J. Crawford, and K. Pribram, "EEG Correlates Of Hypnotic Susceptibility And Hypnotic Trance: Spectral Analysis And Coherence," *International Journal of Psychophysiology*, vol. 10, no. 2, pp. 125–142, 1990.

[38] B. R. Cahn and J. Polich, "Meditation States And Traits: Eeg, erp, and neuroimaging studies." *Psychological Bulletin*, vol. 132, no. 2, p. 180, 2006.

[39] J. Williams, D. Ramaswamy, and A. Oulhaj, "10 Hz Flicker Improves Recognition Memory In Older People," *BMC Neuroscience*, vol. 7, no. 1, p. 21, 2006.

[40] J. Williams, "Frequency Specific Effects Of Flicker On Recognition Memory," *Neuroscience*, vol. 104, no. 2, pp. 283–286, 2001.

[41] T. Nomura, K. Higuchi, H. Yu, S. I. Sasaki, S. Kimura, H. Itoh, M. Taniguchi, T. Arakawa, and K. Kawai, "Slow Wave Photic Stimulation Relieves Patient Discomfort During Esophagogastroduodenoscopy," *Journal of Gastroenterology and Hepatology*, vol. 21, no. 1, pp. 54–58, 2006.

[42] H. C. Ossebaard, "Stress Reduction By Technology? An Experimental Study Into The Effects Of Brain Machines On Burnout And State Anxiety," *Applied Psychophysiology and Biofeedback*, vol. 25, no. 2, pp. 93–101, 2000.

[43] R. P. Brenner, R. F. Ulrich, D. G. Spiker, R. J. Sclabassi, C. F. Reynolds, R. S. Marin, and F. Boller, "Computerized EEG Spectral Analysis In Elderly Normal, Demented And Depressed Subjects," *Electroencephalography and Clinical Neurophysiology*, vol. 64, no. 6, pp. 483–492, 1986.

[44] T. Egner and J. H. Gruzelier, "EEG Biofeedback Of Low Beta Band Components: Frequency Specific Effects On Variables Of Attention And Event Related Brain Potentials," *Clinical Neurophysiology*, vol. 115, no. 1, pp. 131–139, 2004.

[45] E. John, L. Prichep, W. Kox, P. Valdes-Sosa, J. Bosch-Bayard, E. Aubert, M. Tom, L. Gugino *et al.*, "Invariant Reversible QEEG Effects Of Anesthetics," *Consciousness and Cognition*, vol. 10, no. 2, pp. 165–183, 2001.

[46] M. H. Munk, P. R. Roelfsema, P. König, A. K. Engel, and W. Singer, "Role Of Reticular Activation In The Modulation Of Intracortical Synchronization," *Science*, vol. 272, no. 5259, pp. 271–274, 1996.

[47] I. S. Kotchetkov, B. Y. Hwang, G. Appelboom, C. P. Kellner, and E. S. Connolly Jr, "Brain Computer Interfaces: Military, Neurosurgical, And Ethical Perspective," *Neurosurgical Focus*, vol. 28, no. 5, p. E25, 2010.

[48] T. N. Lal, T. Hinterberger, G. Widman, M. Schröder, N. J. Hill, W. Rosenstiel, C. E. Elger, N. Birbaumer, and B. Schölkopf, "Methods Towards Invasive Human Brain Computer Interfaces," *Advances in Neural Information Processing Systems*, 2004, pp. 737–744.

[49] T. Yanagisawa, M. Hirata, Y. Saitoh, H. Kishima, K. Matsushita, T. Goto, R. Fukuma, H. Yokoi, Y. Kamitani, and T. Yoshimine, "Electro-Corticographic Control Of a prosthetic arm in paralyzed patients," *Annals of Neurology*, vol. 71, no. 3, pp. 353–361, 2012.

[50] X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw, and G. Schalk, "Spatiotemporal dynamics of electro-Corticographic High Gamma Activity During Overt And Covert Word Repetition," *Neuroimage*, vol. 54, no. 4, pp. 2960–2972, 2011.

[51] M. Lopez-Gordo, D. Sanchez-Morillo, and F. P. Valle, "Dry EEG Electrodes," *Sensors*, vol. 14, no. 7, pp. 12 847–12 870, 2014.

[52] H. Jasper, "Report Of The Committee On Methods Of Clinical Examination In Electroencephalography," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 370–375, 1958.

[53] S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski *et al.*, "Independent Component Analysis Of Electroencephalographic Data," *Advances in Neural Information Processing Systems*, pp. 145–151, 1996.

[54] A. Subasi and M. I. Gursoy, "EEG Signal Classification Using PCA, ICA, LDA and Support Vector Machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.

[55] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined Optimization Of Spatial And Temporal Filters For Improving Brain Computer Interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[56] A. Delorme and S. Makeig, "EEGLAB: An Open Source Toolbox For Analysis Of Single Trial EEG Dynamics Including Independent Component Analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.

[57] A. Subasi, "EEG Signal Classification Using Wavelet Feature Extraction And A Mixture Of Expert Model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[58] S. Sanei and J. A. Chambers, *EEG Signal Processing*. John Wiley & Sons, 2013.

[59] A. Aarabi, R. Grebe, and F. Wallois, "A Multistage Knowledge Based System For EEG Seizure Detection In Newborn Infants," *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2781–2797, 2007.

[60] T. M. Vaughan, J. R. Wolpaw, and E. Donchin, "EEG Based Communication: Prospects And Problems," *IEEE Transactions on Rehabilitation Engineering*, vol. 4, no. 4, pp. 425–430, 1996.

[61] L. A. Farwell and E. Donchin, "Talking Off The Top Of Your Head: Toward A Mental Prosthesis Utilizing Event Related Brain Potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[62] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain Computer Interface: A New Communication Device For Handicapped Persons," *Journal of Microcomputer Applications*, vol. 16, no. 3, pp. 293–299, 1993.

[63] J. R. Wolpaw, D. Flotzinger, G. Pfurtscheller, and D. J. McFarland, "Timing Of EEG Based Cursor Control," *Journal of Clinical Neurophysiology*, vol. 14, no. 6, pp. 529–538, 1997.

[64] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG Based Brain Computer Communication," *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 6, pp. 444–449, 1994.

[65] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods Of EEG Signal Features Extraction Using Linear Analysis In Frequency And Time Frequency Domains," *ISRN Neuroscience*, vol. 2014, 2014.

[66] M. K. Kiymik, A. Subasi, and H. R. Ozcalık, "Neural Networks With Periodogram And Autoregressive Spectral Analysis Methods In Detection Of Epileptic Seizure," *Journal of Medical Systems*, vol. 28, no. 6, pp. 511–522, 2004.

[67] O. Faust, R. Acharya, A. Allen, and C. Lin, "Analysis Of EEG Signals During Epileptic And Alcoholic States Using AR Modeling Techniques," *IRBM*, vol. 29, no. 1, pp. 44–52, 2008.

[68] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification Of EEG Signals Using The Wavelet Transform," *13th International Conference on Digital Signal Processing DSP 97*, vol. 1. IEEE, pp. 89–92, 1997.

[69] M. Akin, "Comparison Of Wavelet Transform And FFT Methods In The Analysis Of EEG Signals," *Journal of Medical Systems*, vol. 26, no. 3, pp. 241–247, 2002.

[70] D. Cvetkovic, E. D. Übeyli, and I. Cosic, "Wavelet Transform Feature Extraction From Human PPG, ECG, And EEG Signal Responses To ELF PEMF Exposures: A Pilot Study," *Digital Signal Processing*, vol. 18, no. 5, pp. 861–874, 2008.

[71] I. Daubechies, "The Wavelet Transform, Time-Frequency Localization And Signal Analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[72] S. Soltani, "On The Use Of The Wavelet Decomposition For Time Series Prediction," *Neurocomputing*, vol. 48, no. 1, pp. 267–277, 2002.

[73] M. N. Kousarrizi, M. Teshnehlab, M. Aliyari, A. Gharaviri *et al.*, "Feature Extraction And Classification Of EEG Signals Using Wavelet Transform, SVM And Artificial Neural Networks For Brain Computer Interfaces," *International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS'09.* . IEEE, 2009, pp. 352–355.

[74] C. Valens, "A Really Friendly Guide To Wavelets," *Last Date Accessed: 11/05/2015: http://perso.wanadoo. fr/polyvalens/clemens/wavelets/wavelets.html*, 1999.

[75] I. Daubechies *et al.*, *Ten Lectures On Wavelets.* SIAM, 1992, vol. 61.

[76] A. Tsymbal, S. Puuronen, M. Pechenizkiy, M. Baumgarten, and D. W. Patterson, "Eigenvector-Based Feature Extraction For Classification." *FLAIRS Conference*, 2002, pp. 354–358.

[77] E. D. Übeyli, "Analysis Of EEG Signals By Implementing Eigenvector Methods/Recurrent Neural Networks," *Digital Signal Processing*, vol. 19, no. 1, pp. 134–143, 2009.

[78] E. D. Übeyli, E. Derya, "Analysis Of EEG Signals By Combining Eigenvector Methods And Multiclass Support Vector Machines," *Computers in Biology and Medicine*, vol. 38, no. 1, pp. 14–22, 2008.

[79] K. Fukunaga, *Introduction To Statistical Pattern Recognition.* Academic Press, 2013.

[80] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-Based Learning Algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.

[81] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987.

[82] J. Shlens, "A Tutorial On Principal Component Analysis," *ArXiv Preprint ArXiv:1404.1100*, 2014.

[83] M. H. Hayes, *Statistical Digital Signal Processing And Modeling.* John Wiley & Sons, 2009.

[84] S. A. Awang, M. Paulraj, and S. Yaacob, "Analysis Of EEG Signals By Eigenvector Methods," *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES).* IEEE, 2012, pp. 778–783.

[85] R. Duda, *Pattern classification.* John Wiley & Sons, 2012.

[86] A. K. Jain, R. P. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[87] A. K. Jain and B. Chandrasekaran, "Dimensionality And Sample Size Considerations In Pattern Recognition Practice," *Handbook of Statistics*, vol. 2, pp. 835–855, 1982.

[88] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* , vol. 23, no. 2, pp. 228–233, 2001.

[89] G. McLachlan, *Discriminant Analysis And Statistical Pattern Recognition.* John Wiley & Sons, 2004, vol. 544.

[90] H. Abdi, "Discriminant Correspondence Analysis," *Encyclopedia of Measurement and Statistics*, pp. 270–275, 2007.

[91] V. N. Vapnik and A. J. Chervonenkis, "Theory Of Pattern Recognition," *Nauka*, 1974.

[92] V. Vapnik, *The Nature Of Statistical Learning Theory.* Springer Science & Business Media, 2013.

[93] V. N. Vapnik and S. Kotz, *Estimation Of Dependences Based On Empirical Data.* New York: Springer-verlag, vol. 40, 1982.

[94] P. E. Gill, W. Murray, and M. H. Wright, "Practical Optimization," *Academic Press*, 1981.

[95] W. Kuo, H.-H. Lin, Z. Xu, and W. Zhang, "Reliability Optimization With The Lagrange Multiplier And Branch-And-Bound Technique," *IEEE Transactions on Reliability*, vol. 36, no. 5, pp. 624–630, 1987.

[96] C. Huang, L. Davis, and J. Townshend, "An Assessment Of Support Vector Machines For Land Cover Classification," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 725–749, 2002.

[97] F. Melgani and L. Bruzzone, "Classification Of Hyper-Spectral Remote Sensing Images With Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[98] B. Kijsirikul and N. Ussivakul, "Multi-class Support Vector Machines Using Adaptive Directed Acyclic Graph," *Proceedings of International Joint Conference on Neural Networks (IJCNN 2002)*, pp. 980–985, 2002

[99] F. Takahashi and S. Abe, "Decision-Tree-Based Multi-class Support Vector Machines," *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02*, vol. 3. IEEE, pp. 1418–1422, 2002.

[100] Y. Bennani and K. Benabdeslem, "Dendogram-Based SVM For Multi-Class Classification," *CIT. Journal of Computing and Information Technology*, vol. 14, no. 4, pp. 283–289, 2006.

[101] T. Lajnef, S. Chaibi, P. Ruby, P. E. Aguera, J. B. Eichenlaub, M. Samet, A. Kachouri, and K. Jerbi, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *Journal of neuroscience methods*, pp.94–105, 2015.

APPENDICES

A. DECISION TREE CONFIGURATION

In order to find a configuration for the decision tree we relied on test and trials on different distance metrics. Euclidean distance showed the best performance in terms of running time and accuracy of classification. the different distance metrics that were tried are seuclidean distance (standardized Euclidean distance), cosine (One minus the cosine of the included angle between vectors), spearman rank (One minus the sample Spearman's rank correlation between observations), and hamming distance (the percentage of coordinates that differ). See figures A.1 to A.8.

Let's consider the total accuracy of a BCI system as the average of the correctly detected values in each class. Table A.1 will be obtained according to confusion matrices for different methods of distance calculation (Euclidean distance, Seuclidean distance, Cosine distance, Spearman distance, Hamming distance). The correctly detected values are equal to the diagonal of the confusion matrix.

Table A.1: Confusion matrix interpretation for different methods of distance calculation.

	Detection Accuracy
Euclidean Distance	70%
Seuclidean Distance	67%
Cosine Distance	65%
Spearman Distance	65%
Hamming Distance	65%

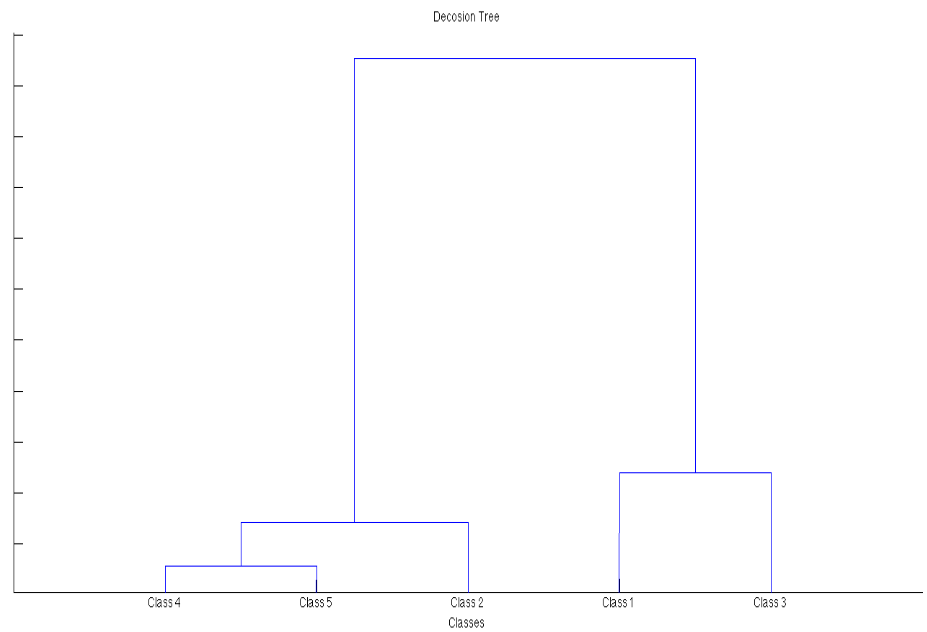

Fig. A.1: Decision tree using seuclidean distance method.

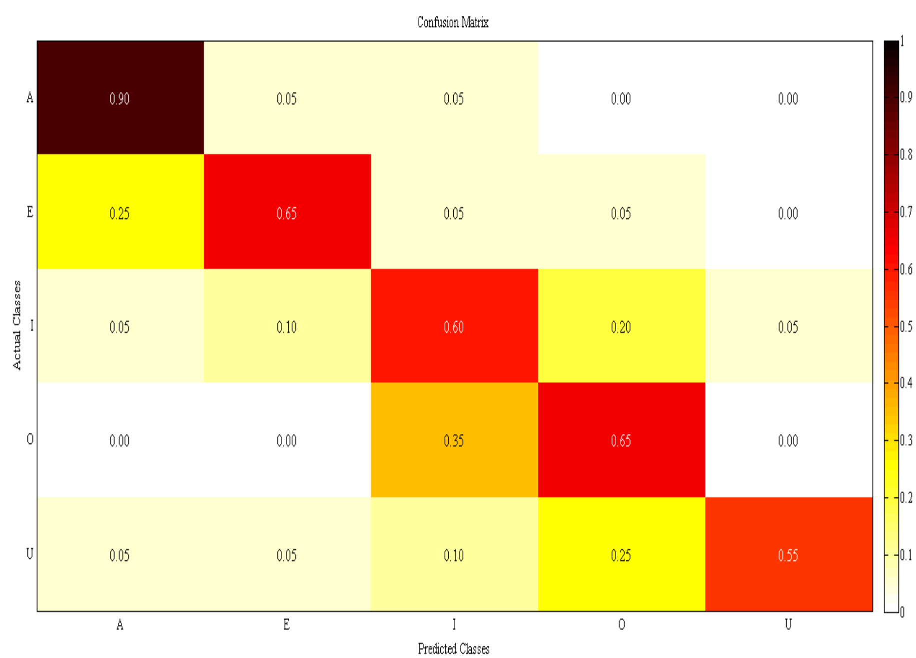

Fig. A.2: Classification result using the tree with seuclidean distance method.

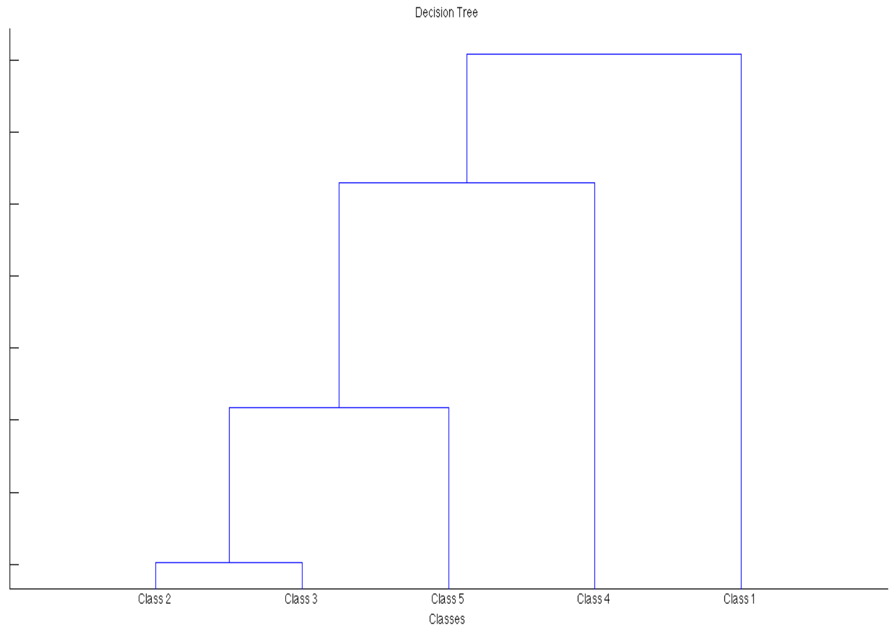

Fig. A.3: Decision tree using cosine distance method.

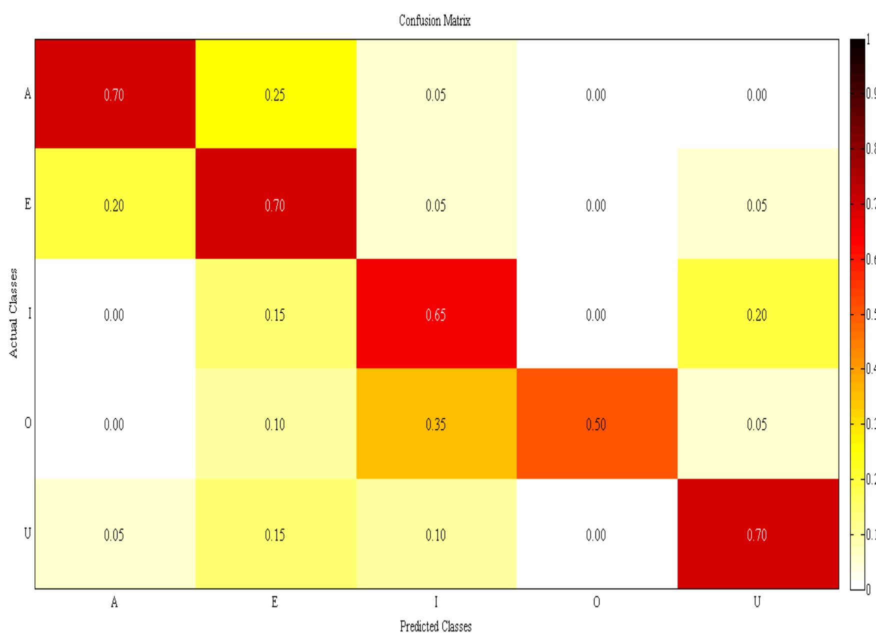

Fig. A.4: Classification result using the tree with seuclidean distance method.

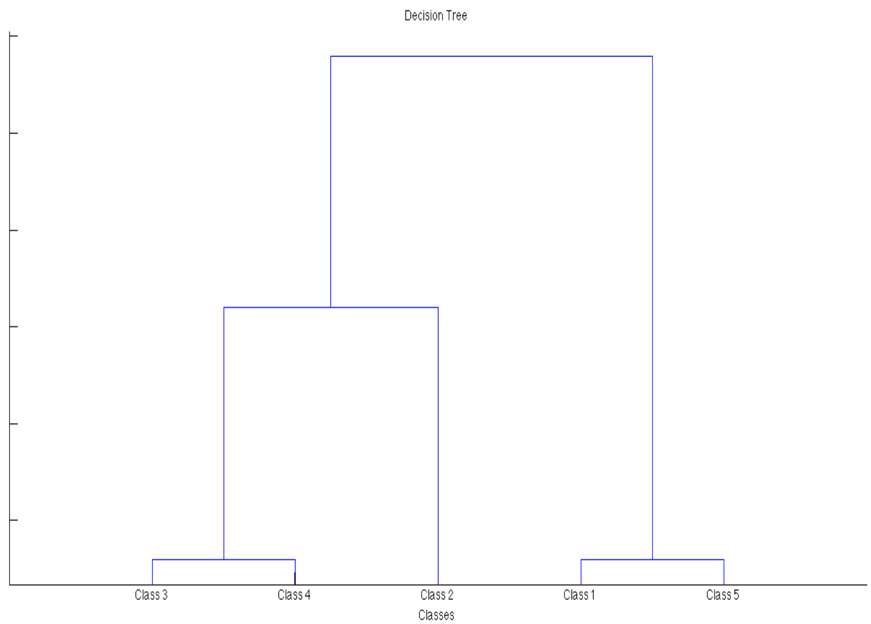

Fig. A.5: Decision tree using spearman distance method.

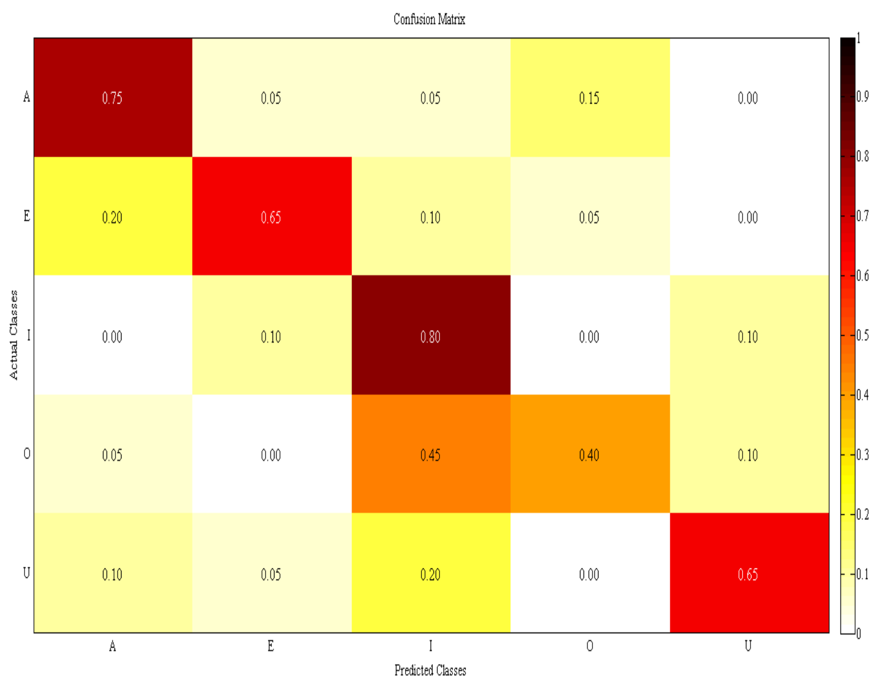

Fig. A.6: Classification result using the tree with spearman distance method.

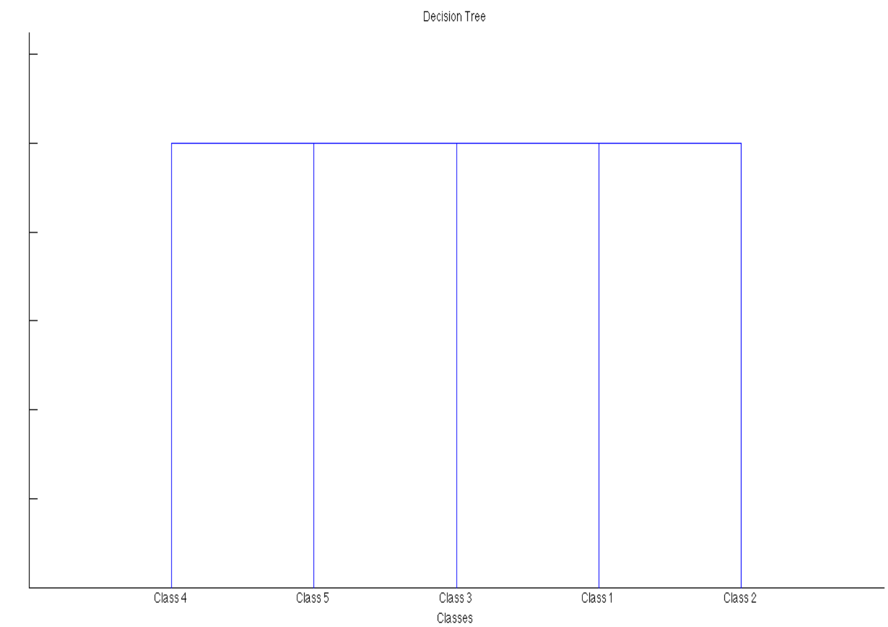

Fig. A.7: Decision tree using hamming distance method.

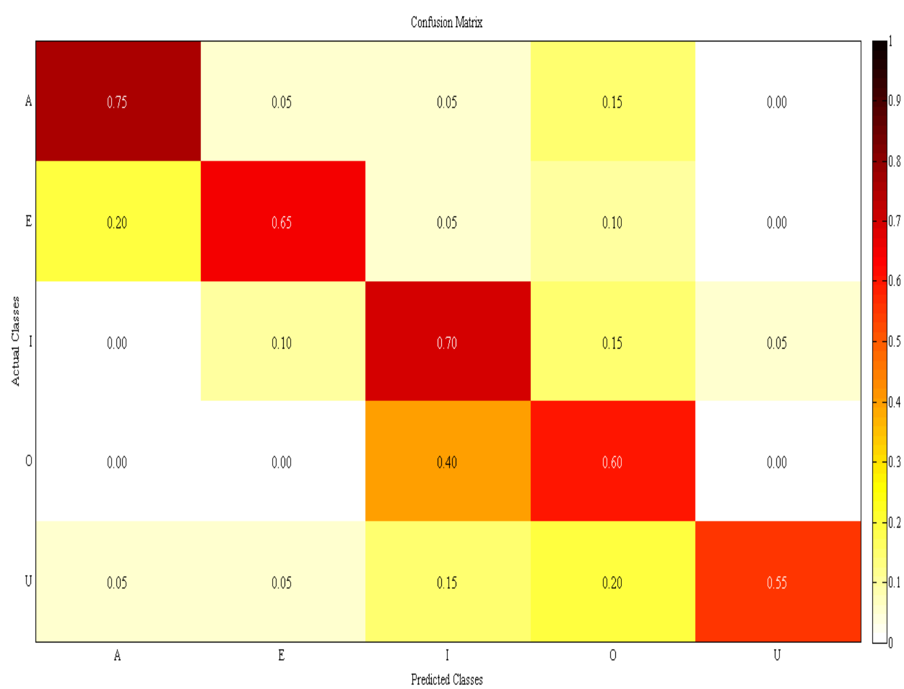

Fig. A.8: Classification result using the tree with hamming distance method.

B. MATLAB CODES

B.1 Gaussian wavelet

Here is the code to generate the Gaussian wavelet in figure 3.4:

```
clc; close all
[psi,x] = wavefun('gaus3',20);
plot(x,psi);
grid on
```

B.2 Preprocessing

In this step all we have is a bunch of raw signals, named in the form of mXXX.mat, with the information about their group. Taking these signals as inputs, preprocessing creates a group label for each of data, and saves a new matfile named alldata.mat which contain AllData and AllLabels. Then it separates every single experiments and saves them with their associated group labels in a new file as crop_data.mat. After that, it filters all of the signals (including signals from all experiments and all electrodes) using the function myfilter. See the code bellow:

```
%Preprocessing
a = [887, 899, 909, 921, 932, 946, 956, 973, 983, 993, 1003, 1024, ...
              1045,1065, 1075, 1086, 1096, 1116, 1126];

AllData = zeros(5 * length(a), 33000, 21);
AllLabels = zeros(5 * length(a), 1);
for k = 0:4
   for j = 1:length(a)
```

```
      load(['m' num2str(a(j)+k) '.mat'], ['m' num2str(a(j)+k)]);
      eval(['D = m' num2str(a(j)+k) '; clear m' num2str(a(j)+k) ';']);
      AllData(k*length(a)+j, :, :) = D(1:33000, 1:21);
      AllLabels(k*length(a)+j, 1) = k;
   end
end
save('alldata.mat', 'AllData', 'AllLabels')

load alldata.mat
for i =1:95
    for j=0:10
        for k=1:21
            crop_data(11*(i-1)+j+1,:,k)= ...
             AllData(i,(1750+3000*j):(2750+3000*j),k);
        end
        crop_label(11*(i-1)+j+1,1) = AllLabels(i,1);
    end
end
save('crop_data.mat', 'crop_data', 'crop_label');

load crop_data.mat
for i =1:1045
    for j=1:21
        filtered_data(i,:,j)=filter(myfilter,crop_data(i,:,j));
    end
end
filtered_label = crop_label;
save('filtered_data.mat', 'filtered_data', 'filtered_label');
```

```
%Function myfilter
function BPF = myfilter
Fs = 500;
Fstop1_b = 1.85;
Fpass1_b = 2.15;
Fpass2_b = 49;
Fstop2_b = 51;
Astop1_b = 80;
Apass_b  = 1;
Astop2_b = 80;
match_b  = 'both';
h_b  = fdesign.bandpass(Fstop1_b, Fpass1_b, Fpass2_b, Fstop2_b,...
    Astop1_b, Apass_b, Astop2_b, Fs);
BPF = design(h_b, 'ellip', 'MatchExactly', match_b);
```

B.3 Dimension reduction and feature extraction

The following function calculate the periodogram, normalize the PSD values, take the average of the normalized PSDs, and finally save the final data and their labels under the name ave_xx.mat. At the end, it plots figures 5.7 and 5.8.

```
function ave_pxx(x)


for i=1:1045
    for j=1:21
        [pxx(i,:,j),w(i,:,j)]= ...
            periodogram (x(i,:,j));
    end
end
```

```
for i=1:1045
    for j=1:21
        x_c(i,:,j) = ...
           (-1)*((pxx(i,:,j)-min(pxx(i,:,j)))*(1-(-1))/...
           (max(max(pxx(i,:,j)))-min(min(pxx(i,:,j))))+(-1));
    end
end

ave_pxx_data = zeros(1045,21);
for i=1:1045
    for j=1:21
        ave_pxx_data(i,j) = mean(x_c(i,:,j));
    end
end
ave_pxx_label = filtered_label;
save('ave_pxx', 'ave_pxx_data','ave_pxx_label')

ave_pxx_s12 = ave_pxx_data(1:95,:);
for k=0:4
ave_pxx_s1(k+1,:) = ave_pxx_s12(19*k+1,:);
end
figure;bar(ave_pxx_s1,'DisplayName','ave_pxx_s1');

for k=0:4
ave_pxx_s2(k+1,:)=ave_pxx_s12(19*k+2,:);
end
figure;bar(ave_pxx_s2,'DisplayName','ave_pxx_s2');
```

B.4 Classification training and testing

The following lines of codes take advantage of all of above steps to preprocess, dimension reduction, feature extraction, classifier training, and classifier testing for the data that has been used in this project. It returns all the plots and calculates the running time.

```
tic

load alldata.mat
figure;
plot(AllData(1,1:1000,1));
hold on

load filtered_data.mat
plot(filtered_data(1,:,1),'r');
hold off
x=filtered_data;
ave_pxx(x);

%train
load ave_pxx.mat
train_data = cell(1,5);
train_label = cell(1,5);
for i = 1:length(train_data)
    train_data{i} = ave_pxx_data(209*(i-1)+10:209*i,:);
    train_label{i} = ave_pxx_label(209*(i-1)+10:209*i,1);
end
[svmstruct,level] = Train_Multi_SVM(train_data,train_label);
```

```
%classify
label = [0 1 2 3 4];
test_mat = cell(length(label),1);
for i = 1:length(label)
    test_mat{i} = ave_pxx_data(209*(i-1)+1:209*(i-1)+20,:);
end

test_mat = cell2mat(test_mat);
[Class_test] = ...
 Classify_Multi_SVM(test_mat,label,svmstruct,level);
labels = ...
 [zeros(1,20),ones(1,20),2*ones(1,20),3*ones(1,20),4*ones(1,20)];
confusion_matrix(Class_test,labels,{'A','E','I','O','U'});
xlabel('Predicted Classes'), ylabel('Actual Classes');
title ('Confusion Matrix')

toc
```