Advanced Signal Processing Solutions for Brain-Computer Interfaces: From Theory to Practice

Golnar Kalantar

A Thesis

in

The Department

of

Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Quality Systems Engineering) at

Concordia University

Montréal, Québec, Canada

June 2018

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Golnar Kalantar**

Entitled: **Advanced Signal Processing Solutions for Brain-Computer Interfaces: From Theory to Practice**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Quality Systems Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Name of the Chair

_____ External Examiner
Dr. Name of External Examiner

_____ Examiner
Dr. Name of Examiner One

_____ Supervisor
Dr. Yong Zeng

Approved by _____
, Chair
Department of Concordia Institute for Information Systems Engineering

_____ 2018

_____
Amir Asif, Dean
Faculty of Engineering and Computer Science

Abstract

Advanced Signal Processing Solutions for Brain-Computer Interfaces: From Theory to Practice

Golnar Kalantar

As the field of Brain-Computer Interfaces (BCI) is rapidly evolving within both academia and industry, the necessity of improving the signal processing module of such systems becomes of significant practical and theoretical importance. Additionally, the employment of Electroencephalography (EEG) headset, which is considered as the best non-invasive modality for collecting brain signals, offers a relatively more user-friendly experience, affordability, and flexibility of design to the developers of a BCI system. Motivated by the aforementioned facts, the thesis investigates several venues through which an EEG-based BCI can more accurately interpret the users' intention. The first part of the thesis is devoted to development of theoretical approaches by which the dimensionality of the collected EEG signals can be reduced with minimum information loss. In this part, two novel frameworks are proposed based on graph signal processing theory, referred to as the GD-BCI and the GDR-BCI, where the geometrical structure of the EEG electrodes are employed to define and configure the underlying graphs. The second part of the thesis is devoted to seeking practical, yet facile-to-implement, solutions to improve the classification accuracy of BCI systems. Finally, in the last part of the thesis, inspired by the recent surge of interest in hybrid BCIs, a novel framework is proposed for cuff-less blood pressure estimation to be further coupled with an EEG-based BCI. Referred to as the WAKE-BPAT, the proposed framework simultaneously processes Electrocardiography (ECG) and Photoplethysmogram (PPG) signals via an adaptive Kalman filtering approach.

Acknowledgments

This thesis would not have been possible without the guidance, collaboration, and encouragement of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of my research work. It is a pleasure to convey my gratitude to them all in my humble acknowledgment.

First and foremost, I would like to thank my supervisor, Professor Arash Mohammadi, and express my great appreciation for his guidance, patience, and support during these two years of my Master studies. The door to Prof. Mohammadis office was always open whenever I ran into a trouble spot or had a question about my research or writing. I am grateful for all the encouragements during my Master's degree and opportunities he sought to provide for me.

The journey of master's studies would have not been the same without the thoughtful comments during each of the meetings I had with Professor Amir Asif. His deep knowledge and astute vision in science has always been an inspiration to me and my work. Also, I would like to specially thank Professor Nizar Bouguila, for teaching me how to employ quality methodologies in all aspects of my life, a lesson I have carried with me ever since my first semester at Concordia University, throughout my research work.

I gratefully acknowledge my fellow colleagues, Sourav Kumar Mukhopadhyay and Tim A. Maloney, for their constant help and support and active interest.

I must express my very profound gratitude to my parents, their prayers and positive energy has been a great spiritual support for me not to give up, ever. Words cannot express how thankful I am for their wise counsel, sympathetic ear, and all the sacrifices they made for my success.

I would like to offer my special thanks to two of the most important people in my life, first,

Muhammad Nasir Shafique for being the realization of truth and honesty, keeping an eye on my well-being even if he was not physically present, and making sure I value my potentials and I am on my way of greater good. Second, Mahsa Mirgholami, for standing by my side regardless of what I was going through, and never stopped reassuring me to pave my desired path and reach the milestones. Mahsa and Nasir provided me with unfailing support and continuous encouragement throughout the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

Acknowledgments iv

List of Figures ix

List of Tables xi

Abbreviation 1

1 Thesis Overview 3

1.1 Outline . . . 3

1.2 Thesis Contributions . . . 4

1.3 Organization of the Thesis . . . 7

2 Background and Literature Review on Brain Computer Interfacing 8

2.1 Brain Computer Interfaces: Why and How? . . . 8

2.1.1 Introduction to the BCIs . . . 9

2.2 Electroencephalography (EEG) . . . 16

2.3 Data Dimensionality Reduction . . . 21

2.3.1 Principal Component Analysis (PCA) . . . 22

2.3.2 Singular Value Decomposition (SVD) and Percentage Root-mean-square Difference (PRD) . . . 23

2.4 Feature Extraction . . . 25

2.4.1 Common Spatial Patterns . . . 27

2.5 Classification 29

3 Graph-based Frameworks for Spatio-Temporal Filtering and Dimensionality Reduction in MI EEG-based BCIs 37

3.1 Introduction To Graph Signal Processing 38

3.2 The GD-BCI Framework 40

3.2.1 Framework Outline 40

3.2.2 Pre-Processing 41

3.2.3 Graph-based Spatio-Temporal Filtering 42

3.2.4 Tangent Space Mapping and Feature Extraction 45

3.2.5 Classification 47

3.2.6 Simulations 47

3.2.7 Conclusion 49

3.3 GDR-BCI: Dimensionality Reduction of EEG Signals via Functional Clustering and Total Variation Measure 50

3.3.1 Framework Outline 50

3.3.2 Defining the Adjacency Matrix 51

3.3.3 Total Variation Graph and Selection of best Clusters: Dimensionality Reduction 53

3.3.4 Feature Extraction and Classification 54

3.3.5 Simulations 55

3.3.6 Conclusion 56

3.4 Summary 56

4 Practical Solutions to Improve the Performance of EEG-based BCI systems 59

4.1 Progressive Fusion of Multi-rate MI Classification for BCIs 60

4.1.1 Progressive Multi-rate MI Classification Outline 60

4.1.2 Experimental Implementation of the Multi-rate MI Classification 62

4.1.3 Conclusion 66

4.2 Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs . . 67

4.2.1 Trimming Framework Outline and Simulation 68
4.2.2 Conclusion 69
4.3 Summary 70

5 Applications of Hybrid BCIs and WAKE-BPAT Framework for Blood Pressure Estimation 72

5.1 Introduction to Hybrid BCIs 72
5.2 WAKE-BPAT: Wavelet-based Adaptive Kalman Filtering for Blood Pressure Estimation via Fusion of Pulse Arrival Times 75
5.2.1 Introduction 75
5.2.2 Problem Formulation 78
5.2.3 The Kalman Filter 79
5.2.4 Proposed WAKE-BPAT 83
5.2.5 Simulation and Results 89
5.2.6 Conclusion 91
5.3 Summary 92

6 Summary and Future Research Directions 93

6.1 Summary of Contributions 93
6.2 Future Work 96

Bibliography 98

List of Figures

Figure 2.1 An example of 10-20 setting of EEG electrodes placement [30]. 19

Figure 2.2 The problem of Underfitting and Overfitting. 33

Figure 2.3 An example of possible and optimal hyperplanes [50] 36

Figure 3.1 Block diagram of the proposed graph-based spatio-temporal filtering framework for brain-computer interface (GD-BCI). 40

Figure 3.2 Functional clustering of human brain [60]. 51

Figure 3.3 Sparsity pattern of the EEG electrodes (2-dimensional projection). 53

Figure 4.1 (a) Emotiv Epoc headset. (b) Electrode placement and activation. 61

Figure 4.2 Arduino microcontroller used to turn BCI classifications into external actions like moving motors or controlling LEDs. 62

Figure 4.3 (a) The visual stimulus shown to the subject as 'Left Hand'. (b) The visual stimulus shown to the subject as 'Right Hand'. 63

Figure 4.4 Scatter plots obtained from two Progressive Filters trained based on 90 and 110 epochs, respectively. 65

Figure 4.5 Classification results obtained from implementation of the proposed progressive and multi-rate framework based on data collected via Emotiv headset. 66

Figure 4.6 (a) The accuracies achieved via adding the trimming step to the conventional CSP algorithm. (b) The accuracies achieved via the conventional CSP algorithm. . 69

Figure 5.1 The main ECG waves within a heartbeat. 77

Figure 5.2 (a) Noisy ECG signal. (b) Denoised ECG signal. (c) Noisy PPG signal. (d) Denoised PPG signal. 84

Figure 5.3 (a) Denoised ECG. (b) *QRS-coef* data. (c) Histogram analysis of the *QRS-coef*. (d) Amplitude-band where the population of coefficients is maximum. (e) Modified *QRS-coef* data. (f) Detected R-peaks. 85

Figure 5.4 The DWT-decomposition-tree of the PPG signal. 86

Figure 5.5 (a) Filtered PPG. (b) FD-PPG signal where marked samples are the ones within the threshold value. (c) Detected MSP and systolic-peaks. 87

Figure 5.6 (a) The PPG-onset detection technique. (b) The Gen-1 Device for PPG recordings developed recently by Marefat and Mohseni*et al.* [92]. 88

Figure 5.7 Estimated versus the actual BP. (a) Based on [89], i.e., Model 3. (b) Based on the proposed WAKE-BPAT and Model 1. 90

List of Tables

Table 3.1 Average reconstruction error obtained using various data smoothening methods for training datasets. 43

Table 3.2 Accuracy performance for predicting two classes and the corresponding standard deviation obtained using the proposed GD-BCI framework with two graph construction methods: PG and VPG 48

Table 3.3 Performance comparison of the proposed GD-BCI method in two-class classification problem with that provided by [54] and [59]. 49

Table 3.4 Accuracy comparison of the proposed GDR framework coupled with two different classifiers and with two CSP features. 57

Table 3.5 Similar to Table 5.1 except that four CSP features are utilized; (a) with 100 training trials; (b) with 60 training trials, and; (c) with 200 training trials. 58

Table 4.1 Performance of different models based on real experimental data sets. 64

Table 4.2 Accuracy comparison of the proposed trimming framework, (a) The accuracies for classifier trained with 168 training trial, and; (b)The accuracies for classifier trained with 196 training trial . . . 71

Table 5.1 Estimated BP versus the actual BP based on the proposed WAKE-BPAT. 87

Abbreviation

Abbreviation	Description
BCI	Brain-Computer Interfaces
EEG	Electroencephalography
SNR	Signal to Noise Ratio
fMRT	functional Magnetic Resonance Imaging
MEG	Magnetoencephalography
ECoG	Electrocortiography
SCP	Slow Cortical Potentials
ERP	Event Related Potential
SSVEP	Steady-State Visual Evoked Potentials
ERN	Error-related Negative
fNIRS	functional Near-Infrared Spectroscopy
PET	Positron Emission Tomography
MI	Motor Imagery
ECG	Electrocardiography
PPG	Photoplethysmogram
PCA	Principal Component Analysis
SVD	Singular Value Decomposition
PRD	Percentage Root-mean-square Difference
CSP	Common Spatial Patterns
LDA	Linear Discriminant Analysis

QDA	Quatratic Discriminant Analysis
SVM	Support Vector Machines
GSP	Graph Signal Processing
TVG	Total Variation Graph
hBCI	Hybrid Brain-Computer Interface
WT	Wavelet Transform
KF	Kalman Filter
BP	Blood Pressure
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
PAT	Pulse Arrival Time
MSP	Maximum Slope Point
FD-PPG	First Derivative of the filtered PPG

Chapter 1

Thesis Overview

1.1 Outline

Inside our heads, weighing about 1.5 kg, is an astonishing living organ consisting of billions of tiny cells. It enables us to sense the world around us, to think and to talk. The human brain is the most complex organ of the body, and arguably the most complex thing on earth. Since the original demonstration that electrical activity generated by ensembles of cortical neurons can be employed directly to control a robotic manipulator, research on brain-computer interfaces has experienced impressive growth. Brain-computer interface is a device that translates neural activity of the brain into commands driving a machine. Such a system consists of three major parts:

(1) A device to record the natural activity of the brain. The nature of these recordings can impose certain constraints on the implementation and potential capabilities of the system;

(2) An effector, which is controlled by the neural signals. The effector can be anything from a visual signal to, e.g., complicated robotic or prosthetic system, and;

(3) An algorithm that analyzes and interprets the neural signals as control commands. This algorithm links the other two parts together. It determines which features of the recorded neural activity will be employed, and therefore, should be produced by the user, and which control commands can be created from the activity.

Since the latter item is the part which plays the main role in correct interpretation of thoughts-into-commands, the major and fundamental challenges in the field of brain-computer interfaces have been declared and argued for this module. To be more specific and in particular, EEG-based BCI systems often suffer from two main problems: (i) Low spatial resolution, though high temporal resolution; and, (ii) Practical approaches through which the final accuracy of users' intention translation increases for BCI applications. The first propounded problem is more prominent while dealing with large datasets collected via EEG headsets, making attempts to extract the most informative parts of the data and to discard the rest, which calls for competent techniques to reduce the size of collected data (signals) with minimal loss of information. The second issue mentioned requires creative solutions to ensure the classifier is not defined too general, and is adaptive and robust enough in respect to each subject's data.

1.2 Thesis Contributions

Inspired by the stated issues, I have made some contributions [1–6] during my thesis research work as briefly outlined below:

(1) **The GD-BCI Framework [1]**: This framework is a new graph-based approach, proposed to spatiotemporally filter the data by taking into account both geometrical structure of the EEG channels and the correlation between the EEG signals. The end goal is to identify the pattern of the brain activity using a robust method for pre-processing, processing, and classification of the EEG signals, with the main focus on data dimensionality reduction. The proposed approach seeks for the most significant feature vectors required for better classification of EEG signals, therefore, adaptively selects them through spectral decomposition of the data using the graph Laplacian matrix. Then, the tangent space mapping method is applied to vectorize the dimensionally reduced matrices and map them onto Euclidean space. After that, the linear support vector machine algorithm is employed for classification. Experiments are conducted on Dataset IVa from BCI Competition III, including data from five different subjects consisting of the right hand and right foot motor imagery actions, and the results show that the GD-BCI framework provides higher classification accuracy as compared to its

counterparts.

- **Pros and Cons**: The GD-BCI framework proposes a graph-based approach which brings about a higher classification accuracy, and configures the graph in a fashion that the impact of active regions of the brain on the signals is taken into account. However, the impressive accuracies come with the price of an exhaustive search for a pair of constants required for adjustment of the graph, which is not preferable in real-world applications of BCI systems.

(2) **The GDR-BCI Framework [2]**: The GDR-BCI framework, is similar in nature to the GD-BCI, and it capitalizes on the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject. Initial functional clustering of EEG electrodes is built by designing a separate adjacency matrix for each identified functional cluster. A collapsing methodology is developed based on total variation measures on graphs, i.e., the overall model will eventually be reduced (collapsed) into two functional clusters. The experimental results based on the same Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to its counterparts

- **Pros and Cons**: The proposed framework offers two main superiorities over its state-of-the-art counterparts and the GD-BCI: (i) First, the resulting dimensionality reduction is subject-adaptive and respects the brain plasticity of subjects, and; (ii) Second, the proposed methodology identifies active regions of the brain during the motor imagery task, which can be used to re-align EEG electrodes to improve accuracy during consecutive data collection sessions. The only drawback this framework is the longer time of processing as compared to its counterparts, as it includes the total variation graph loop that evaluates each cluster.

(3) **The Progressive Fusion of Multi-rate MI Classification for BCIs [3]**: This framework addresses the issues arose in the case of limited number of training trials at hand. This approach consists of two filters running in parallel namely: (i) *The Progressive Filter*: An efficient filter

that performs both feature extraction and classification steps based on the set of *all* arriving epochs to re-train progressively over time. (ii) *The Active Filter*: A simplified CSP-based feature extraction approach running online based on pre-trained classifiers, i.e., a lighter version of the Progressive Filter that runs faster than its counterpart. The proposed framework is evaluated both based on dataset IVa from the BCI competition III, and through real data collected via the Emotiv Epoc headset.

- **Pros and Cons**: The proposed framework is a great solution for practical BCI applications, especially those that function in a real-time manner. However, depending on the capabilities of the implemented BCI system, this approach might be consuming relatively more energy/memory space, which would not be optimal with respect to its application.

(4) **Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs [4]**: This method is proposed to readjust the recorded epochs in a manner that most informative parts of the signals are extracted and the segments of the epochs which do not include the response of the subjects to the stimuli would be discarded. This approach is robust to the different natures of the recorded datasets, and readjusts the data in a way that the most informative time interval within each epoch would be processed and further on, classified.

- **Pros and Cons**: The implementation of the proposed method on the dataset IVa from the BCI competition III shows great potential for this approach, moreover, the processing time that this technique adds to an original code is very small. However, for the approach to be more robust regarding the final trimmed interval, it is important to take into account additional methods to ensure the classifier is of low bias and low variance at the same time, which might lead to an increase in the processing time.

(5) **The WAKE-BPAT Framework [5]**: The motivation behind this work was the recent urgency to design continuous and cuff-less blood pressure (BP) monitoring solutions, not only for the purpose of hybrid BCIs, but more importantly to prevent, detect, and treat hypertension. The WAKE-BPAT framework is a novel wavelet-based feature extraction algorithm coupled with

an adaptive and multiple-model Kalman filtering. This framework provides accurate and dynamic BP estimates by extraction and fusion of different pulse arrival time (PAT) features. Experimental evaluations of WAKE-BPAT based on a real dataset collected via Gen-1 device confirms the superiority of the proposed framework in comparison to its counterparts.

- **Pros and Cons**: The proposed approach exhibits great potential, the processing steps provide a solid robust framework to estimate the blood pressure through simultaneous ECG and PPG signals. However, the BP estimation methods have not been excessively tested as a module of hybrid BCI system and therefore, the effectiveness of such a system is to be investigated furthermore.

1.3 Organization of the Thesis

To provide the relevant context, the rest of the thesis is organized as follows:

- Chapter 1 provides an overview and a summary of important contributions made in the thesis.
- Chapter 2 presents a comprehensive background and literature review on brain-computer interfaces and their applications, as well as their modules and the corresponding processing methods.
- Chapter 3 considers the problem of dimensionality reduction. Different graph-based dimensionality reduction methods are proposed, introducing the GD-BCI and GD-BCI frameworks.
- Chapter 4 introduces two practical solutions to increase the accuracy of classification of BCI systems.
- In Chapter 5, a background is provided on hybrid BCI systems, and a cuffless blood pressure estimation framework, called the WAKE-BPAT, is proposed.
- Chapter 6 concludes the thesis and provides some directions for future work.

Chapter 2

Background and Literature Review on Brain Computer Interfacing

2.1 Brain Computer Interfaces: Why and How?

How do people with severe motor disabilities and/or speech problems manage to perform the activities of their daily lives? You may have seen someone using a set of push buttons on a computer or tablet that speaks for him/her, or those who use specialized physical devices, such as a wheelchair or a robotic arm, to help them move around and do accomplish the tasks of their day. These devices are known as Augmentative and Alternative Communication technologies (AAC), aiding those who suffer from motor disabilities or severe speech problems to improve their quality of daily life, possibly without requiring a caregiver. Despite the success and public acceptance of the conventional AAC, there have been people with certain need and requirements, due to their severe condition, who are not able to take advantage of these means of communication. For instance, those who are totally paralyzed, or "locked-in", are restricted from both verbal and non-verbal communication, even though they are conscious and alert [7]. The inability of communicating, neither emotions and thoughts nor physical needs of one, calls for a technology capable of a deeper level of communication and of reaching out to the thoughts of the impaired users [8]. This is essentially the origin of the recent upsurge in the field of Brain-Computer Interfaces (BCIs).

2.1.1 Introduction to the BCIs

The BCI systems, while requiring no peripheral muscular activity, enable a user to use solely his/her brain activities to send commands to an electrical device. BCI can be considered as a system for which the input is the brain activity and the output is a set of device control signals, therefore, the BCI system itself functions as a translator, measuring specific features of the brain signals. Jacques Vidal [9] was the first researcher who proposed the term "brain-computer interface" in 1973, when he presented a system that could interpret brain signals into computer control signals. BCI technology initially used to be unattractive for serious scientific investigation due to false assumptions about its applications. The general public, as well as academia, often used to reject the idea of successfully deciphering thoughts or intentions by means of brain activity in the past, as strange and remote. Hence, investigation in the field of brain activity has usually been limited to the analysis of neurological disorders in a clinical setting or the exploration of brain functions in laboratories. In contrary, during the past two decades, experimental research into BCIs has expanded significantly, with promising results presented for healthy people and few successful and practical controlled clinical outcome studies for patients. BCIs are starting to prove their efficacy as assistive and rehabilitative technologies in patients who suffer from severe motor impairments. Moreover, recently, several fruitful developments and expansions of its market for both healthy and unhealthy people have emerged. This sparked progress is driven by the advancements in terms of effectiveness and increase in the number of available technologies to record and process brain signals.

The BCI systems generally share the same principles, i.e. the detected brain signals are amplified and recorded, then filtered, smoothed, and classified according to relevant characteristics (e.g., sensorimotor rhythms over the motor cortex). After processing and decoding of the brain signals, the output of the BCI can be used to control the movement of a prosthesis, orthosis, wheelchair, robot or cursor, or to direct electrical stimulation of muscles [10]. Prior to naming the potential/active practical areas of the BCI technology, it is of great importance to clarify the difference between a tool, in this case a BCI, and an application. A tool in the present context is a device which is specified by the manner in which it performs its function, and applies to a wide variety of applications. Its effectiveness and ease of use evaluates a tool's performance, and its function remains the same,

regardless of the purpose it is serving. On the other hand, an application is primarily described by the purpose it serves, and its evaluation focuses on how well it fulfills serving the target purpose, while it may also be described regarding the tool it employs.

Having the above definitions in mind, it can be said that BCIs are tools that record and analyze brain activities, such as, Electroencephalography (EEG) signals, which will be discussed later in details in Subsection 2.2. Moving a cursor, selection among two or more possible choices shown to the subject of the study, or controlling a robot, are some examples of BCIs employed as suitable and responsive tools to use. BCI applications are widely spread in various fields of research and medical industry. Generally speaking, BCI applications can be classified into the following six main categories:

(1)*Medical*;
(2)*Games and Entertainment*;
(3)*Educational and Self-Regulation*;
(4)*Neuroergonomics and Smart Environment*;
(5)*Neuromarketing and Advertising*, and;
(6)*Security and Authentication*.

These applications are mostly in experimental research state and not all of them have been well-established to be operable by general public. While working on either of these areas, the primary concern of BCI developers must be the needs and priorities of the anticipated user, and researchers must guard against the tendency to approach the parameters of the tools and their applications as an abstract design exercise. BCI development requires an optimized design with well-defined objectives, which should be based on not only a complete technical study, but also a thorough and comprehensive behavioral analysis that essentially addresses the needs, desires, and incentives of the users and their possible caregivers. Satisfying these conditions requires collaborative interactions with the users, who must know how to use the technology and be persuaded that the technology is both useful and safe. Also, essential is the cooperation of the relevant health care professionals, who must be persuaded that the risk/benefit ratio is favorable, that the technology is safe and useful, and that it is equal or superior to available alternatives.

In regards to the acquisition of brain activity and the modalities to do so, BCIs can be classified into the following two main categories:

- **Invasive BCIs**

  Invasive BCIs are those that involve surgical implantation of electrodes, or multi-electrode grids in the brain. These systems are intended to measure patterns of neurons' activities in order to enable the researchers to decode behaviourally relevant information from the acquired data. The reason for such a risky and expensive intervention is to gain high signal-to-noise ratio (SNR) electrical responses, recorded directly from the brain. This is in contrast to the signals recorded from over the scalp, which is usually contaminated by the high amount of noise, and these recordings are of low amplitude due to the nature of human scalp, resulting in low SNR. However, there is not enough evidence to justify this brutal operation, and moreover, extensive work on brain plasticity [11] has shown that a plastic change in the adult nervous system through learning is possible, if the respective neuronal circuit participates functionally in the physiological tasks of that circuit. This implies that even if the advantages of invasive BCI would overweigh its disadvantages, since the function map of the brain is subject to change in response to the learning processes the person is exposed to, this approach will not remain favorable in the long run.

  In regards to implementation of Invasive BCIs, there are a few, yet of utmost importance concerns that must be addressed prior to any practical experiment.

  (1) *Possible locations of implanted electrodes, number of electrodes to implant, and the nature of the signals to record*: In most BCI applications and the majority of the cases in which BCI is used as a tool, the motor cortex of the brain is an obvious choice for recording the signals; reasons being direct relevance to motory tasks and the relatively better accessibility compared to other motor areas of the brain. In order to determine and identify the appropriate locations for implantation, functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and other functional imaging techniques are admittedly helpful [24]. Moreover, the number of electrodes to implant has a direct relation with the location of the electrodes, the minimum SNR required for the study, and the functional use of the signals, i.e., the purpose of the study/application and the rate of information transfer to fulfill the specifications of the procedure.

(2) *User groups who might be best suited for implanted electrodes and the stability of the recordings*: Due to the severity operation required, it makes perfect sense to only proceed with invasive BCIs for patients with extreme conditions and needs. Patients who are locked-in might benefit from invasive BCI technology if it is relatively safe and effective. Selected individuals with stroke, spinal cord injury, limb prostheses and other conditions might Also benefit. It is of utmost importance to keep in mind that the individuals' preferences play a key role in specifications and decisions about implantable systems, in other words, a substantial functional advantage over the conventional non-invasive systems must be proved to justify the implantation of the invasive electrodes. Extensive researches on non-human subjects [12] has shown that stable recording has been maintained over months, and in selected instances over the years. Positive results regarding human subjects have also been observed sparsely in different research groups around the world, however, this field is still in its infancy, requiring more research and proven reliable outcomes. Also, prior to any implantation, tissue acceptance of the microelectrode has to be ensured.

(3) *The ethical issues that must be considered in implanting recording electrodes in human volunteers*: It is perfectly clear that the patients must be informed of the risks and potential dangers of the operation. Also, all the potential benefits of the system need to be clearly and forcefully explained to the volunteers, especially because volunteers with severe conditions tend to overestimate the benefits of the BCI, and they must be aware of exact aid and service they will be provided with. An ethicist should be involved in the earliest phases of any human research developing or testing invasive BCI methods.

Invasive BCIs are not the focus of this thesis, however, it is worth naming the following five main types of brain activity that can be measured with invasive BCIs: (i) Local field potentials (LFPs) [13]; (ii) Single-unit activity (SUA) [14]; (iii) Multi-unit activity (MUA) [15]; (iv) Electrocorticographic oscillations recorded from electrodes on the cortical surface (electrocorticography, ECoG) [16]; and, (v) Calcium channel permeability [17].

- **Non-Invasive BCIs**

Noninvasive BCIs are implemented without any sort of surgical implantation, as they enable recording the brain signals from the external surface of the scalp. These systems are the most widely researched BCIs due to their minimal risk and the relative convenience of conducting studies and recruiting volunteers to participate in the study. Noninvasive interfaces are able to detect seven types of brain signals.

(1) *Slow cortical potentials*

Slow cortical potentials (SCPs) are shifts in the cortical electrical activity lasting from several hundred milliseconds to several seconds. These shifts might be initiated and triggered by an external event, or induced by self. Their moderating impact on information processing has been demonstrated in numerous studies, such as in [18].

(2) *P300 Event Related Potential*:

The P300 (P3) wave is an Event-Related Potential (ERP) component elicited in the process of decision making. These waves' occurrence does not link to the physical attributes of a stimulus, but to a person's reaction to it. In other words, the P300 is known to reflect processes involved in stimulus evaluation or categorization. When recorded by EEG, it surfaces as a positive deflection in voltage with a latency (delay between stimulus and response) of roughly 250 to 500 ms.

(3) *Steady-State Visual Evoked Potentials*:

The Steady-State Visual Evoked Potentials (SSVEPs) are natural responses of the brain to visual stimulation at specific frequencies. The brain starts generating electrical signals at the frequencies ranging from 3.5 Hz to 75 Hz, or multiples of them, when the retina is excited by a visual stimulus at the same frequencies. SSVEPs are useful in research because of the excellent signal-to-noise ratio and relative immunity to artifacts [19].

(4) *Error-related Negative Evoked Potentials*:

The Error-related Negative (ERN) is a sharp negative going signal which begins about the same time an incorrect motor response begins, (response-locked event-related potential), and typically peaks from 80-150 milliseconds (ms) after the erroneous response begins (or 40-80 ms after the onset of electromyographic activity). The ERN is observed

after errors are committed during various choice tasks, even when the participant is not explicitly aware of making the error.

(5) *Blood-oxygen-level Dependent Contrast Imaging*:

The firing of neurons causes a need for more energy to be provided quickly. Through a process called the hemodynamic response, blood releases oxygen to the active neurons at a greater rate than to inactive ones. This causes a change of the relative levels of oxyhemoglobin and deoxyhemoglobin (oxygenated or deoxygenated blood) that can be detected by their magnetic properties. This event can be measured by fMRI method.

(6) *Cerebral Oxygenation Changes*:

The Near-Infrared Spectroscopy (NIRS), as a methodology for functional neuroimaging, is based on the fact that unlike visible light, near-infrared light (wavelength from 700 to 1000 nm) easily passes through biological tissues and is mainly absorbed by few chromophores like hemoglobin with different absorption spectra for their oxygenationdeoxygenation states. The activation of brain regions causes an increased oxygen metabolic rate and to initial deoxygenation of the tissue which is followed by increased regional cerebral blood flow (rCBF). These metabolic changes enable researchers to study emotional and cognitive tasks of the brain via measurements of functional NIRS [20].

(7) *Sensorimotor rhythms*

A Sensorimotor Rhythm (SMR) is a brain wave, the oscillatory idle rhythm of synchronized electric brain activity. These brain waves appear in the recordings over the sensorimotor cortex via modalities such as EEG. For most individuals, the frequency of the SMR is in the range of 13 to 15 Hz, in cortical regions outside of the motor strip. These frequencies relate to relaxed attention such as reading or engaging in a relaxing hobby such as knitting.

Out of all the types above of noninvasive BCI signals, the focus of this thesis is on *Sensorimotor Rhythms*. Such BCIs have been used with relatively satisfactory success, for instance, Wolpaw and McFarland published the results of their study in 2004 [23], in which they showed that patients with locked-in syndrome or high spinal cord lesions were able to

use sensorimotor rhythms to control cursor movements or select letters or words from a computer menu. EEG-based BCIs with the focus on sensorimotor rhythms will be fully discussed throughout the remainder of the thesis. This completes an outline of invasive and noninvasive BCIs.

As previously stated, BCIs fall into the category of communication and control systems and therefore, a BCI has an input, an output, and a translation algorithm that converts the former to the latter. BCI input consists of a particular feature (or features) of brain activities and the methodology used to measure that features. BCIs may focus on frequency-domain features (Spectral), time-domain features (Temporal), or the features measured in respect to the location of the events taken place on the scalp (Spatial). These features will be fully described and discussed later in Section 2.4. Each BCI uses a particular algorithm to translate its input into output control signals. Due to the high number of available techniques, the translation algorithm might include linear or nonlinear models, or neural network, to name a few, or a hybrid combination of them. In many cases, BCIs incorporate continual adaptation of important parameters to key aspects of the input provided by the user in order to improve the accuracy of capturing the intention and its translation corresponding to the cognitive learning curve the user takes while using the BCI system. BCI outputs can be cursor movements, letter or icon selection, controlling a robot arm, or another form of device control, and provide the feedback that the user and the BCI can use to adapt to optimize communication.

In addition to the three main components of a BCI, as a system, it has other distinctive characteristics as well, which may be the reference of BCI's evaluation, or comparison to its peer systems. These include a BCI's response time, speed and accuracy, the information transfer rate, type and extent of user training required (which can be very crucial in terms of the target population of the intended BCI application), appropriate user population, and required feedback technique by which the user would remain motionless and well-understood by the system.

The focus of this thesis, although studying and practically experimenting input and output of a BCI (Chapter 4), is mostly on the translation algorithms of a BCI. A BCI translation algorithm is essentially a series of computations, arranged and formed in a fashion to most optimally turn BCI system input features derived by the signal processing stage into actual device control commands [24]. The diversity in translation algorithms is due in part to diversity in BCIs intended

real-world applications. Nevertheless, in all cases the goal is to maximize performance and practicability for the chosen application.

2.2 Electroencephalography (EEG)

The BCI systems, as the translators of the brain activities, evidently require at least one modality to detect the user's intention. Either flavors of these activities, spontaneous in the absence of external stimuli, or evoked brain activity, as specific patterns elicited in the presence of external stimuli, can be captured by either "electrophysiological" or "hemodynamic"-based modalities. When information is exchanged between the electrochemical transmitters of the neurons, a set of electrophysiological activities is generated. The neurons generate ionic currents which flow within and across neuronal assemblies. This flow, once large enough, causes electrical and magnetic fields, which can be measured by means of Electroencephalography (EEG), Electrocorticography (ECoG), and Magnetoencephalography (MEG). On the other hand, the hemodynamic response, is a process in which the blood releases glucose to active neurons at a greater rate than in an area of inactivity. This chemical change in the blood can be monitored by neuroimaging methods such as functional Magnetic Resonance Imaging (fMRI), functional Near Infrared Spectroscopy (fNIRS), and Positron Emission Tomography (PET) [25].

Hans Berger, a German psychiatrist, was the first to systematically study the electrical activities of the human brain and to invent EEG [26]. His invention brought about a revolution and to date, a considerable number of research results related to EEG-based BCIs have been reported in various international journals, covering biomedical engineering, clinical neurology, neuroscience, and neurorehabilitation, using EEG headsets. The reason for this popularity is that EEG, although owning a few yet significant drawbacks, outweighs other modalities due to its minimal risk and the relative convenience of conducting studies, as well as other technical distinct advantages, which will be discussed further in this section.

EEG as a medical tool has become a routine clinical practice to distinguish epileptic seizures from other types of spells, such as psychogenic non-epileptic seizures, syncope (fainting), subcortical movement disorders, and migraine variants, to differentiate "organic" encephalopathy or

delirium from primary psychiatric syndromes such as catatonia, to serve as an adjunct test of brain death, to prognosticate, in certain instances, in patients with coma, and to determine whether to wean anti-epileptic medications. However, EEG has been even more trending in research-related practices, especially and extensively in neuroscience, cognitive science, cognitive psychology, neurolinguistics, and psychophysiological research. The reason behind this trending interest, despite the relatively poor spatial sensitivity of EEG, is the several advantages it possesses over other its counterparts, as briefly outlined below.

EEG hardware is significantly and considerably more affordable by comparison with most other techniques. Also, immobility of modalities such as fMRI, PET, or MEG, limit the flexibility of experiment design and require a more complex, therefore costly, arrangements and setting at the data collection venue, while EEG sensors can be placed anywhere on the scalp not requiring any specific ambient conditions to work at. Moreover, EEG recordings hold a very high temporal resolution, on the order of milliseconds rather than seconds, thus, for clinical and research settings, EEG is commonly recorded at sampling rates above 250 Hz and up to 2000 Hz. Nowadays, modern EEG data collection systems are capable of recording at sampling rates above 20,000 Hz if desired. EEG, being absolutely silent while recording, enables researchers to not only study the responses to auditory stimuli, but also to investigate and track the brain changes during different phases of life, e.g., EEG sleep analysis can indicate significant aspects of the timing of brain development, including evaluation of adolescent brain maturation. Additionally, EEG, as a powerful tool to detect covert processing (i.e., processing that does not require a response), is non-invasive and can be used in subjects who are incapable of making a motor response. In contrast to all the useful advantages, EEG also possesses disadvantages that researchers must take into account before adopting this technique of recording as the tool by which they aim to answer the question of their study. The first drawback of EEG recordings is a poor spatial resolution on the scalp as compared to techniques such as fMRI, and in order to compensate for this downside, intense interpretation is required just to hypothesize what areas are activated by a particular response. The quality of EEG signals is affected by scalp, skull, and many other layers as well as background noise. Noise is key to EEG, insofar as it reduces the SNR and therefore the ability to extract meaningful information from the recorded signals.

All in all, considering the pros above and cons, EEG has always been a strong candidate for

BCI developments, especially since changes in cortical electrical activity resulting from mental stimulation occur faster than the accompanying changes in hemodynamics. Many studies, namely, [27] and [28], have reported successful utilization of EEG-based BCIs, particularly enabling able-bodied users to generate fast and reliable control signals. In the interesting work of Moghimi *et al.* [29], 39 studies reporting EEG-oriented BCI assessment by individuals with disabilities were identified in the past decade and investigated, which shows EEG has been the most commonly and widely used modality in BCIs.

In the interest of standardizing the placement of EEG electrodes, an internationally recognized method is established, known as "10–20 system". The system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The distances between adjacent electrodes are either 10% or 20% of the total front-back or right-left distance of the skull; that is the reason behind the name of this system. Each site has a letter to identify the lobe, and a number to identify the hemisphere location. "F" stands for frontal, "T" for temporal, "C" for central, "P" for parietal, and "O" for occipital. Even numbers refer to the electrodes placed on the right hemisphere, and the odd numbers are allocated to the electrodes located on the left hemisphere. To measure the distances and segmenting them, four anatomical landmarks are used: first, the nasion which is the point between the forehead and the nose; second, the inion which is the lowest point of the skull from the back of the head and is generally indicated by a prominent bump; the preauricular points are anterior to the ears. Considering these main instructions, an example of a 10-20 system EEG electrode placement is elucidated in Figure 2.1.

To understand the studies employing EEG as a tool for recording brain signals, it is considerably important to be familiar with the EEG rhythmic activities and patterns. These activities are divided into frequency bands. These designations arose because rhythmic activity within a certain frequency range was noted to have a certain distribution over the scalp or a certain biological significance. The first frequency band commonly is considered to begin from 1 Hz and the last one is capped by 40 Hz, whereas activity below or above this range is likely to be artifactual. The following categories, provide an overview of the location and range of various meaningful and distinguishable brain signal patterns for data analysis.

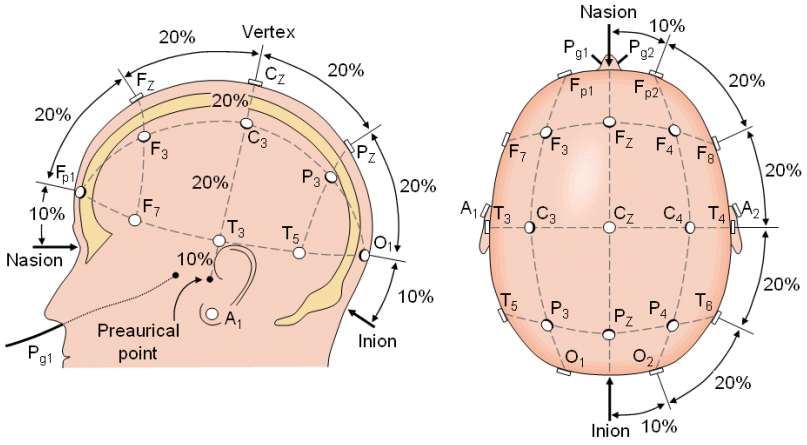


Figure 2.1: An example of 10-20 setting of EEG electrodes placement [30].

- **Delta Patterns**

  These patterns are ranged between 1–4 Hz with highest amplitudes and slowest waves. These signals are usually captured from frontal (in adults) and posterior (in children) part of the scalp. Delta waves are known as slow-wave sleep in adults and also are commonly observed in babies.

- **Theta Patterns**

  These patterns are ranged between 4 – 7 Hz. Theta signals are mostly found at the locations of the scalp that are not related to the task at hand, therefore, they are normally considered as "idling" waveforms. This category of signals has also been found to spike in situations where a person is actively trying to repress a response or action. Moreover, this range has been associated with reports of relaxed, meditative, and creative states.

- **Alpha Patterns**

  These patterns are ranged between 7–13 Hz. These signals are often generated at the posterior regions of the head (both sides) and are higher in amplitude on the dominant side. That is the reason why Hans Berger named this EEG activity as the "Alpha Wave" or the "posterior basic rhythm". This category is famously known as resting/relaxing state, due to its generation while the eyes are closed. It is interesting to know that one of the ways for EEG researchers and engineers to test their application, is to ask the subjects to close their eyes and relax;

the frequency content observed during this time has to be mostly focused at Alpha rhythms. In addition to the basic posterior rhythm, there are other normal alpha rhythms such as the mu rhythm (alpha activity in the contralateral sensory and motor cortical areas) that emerges when the hands and arms are idle.

- **Beta Patterns**

  These patterns are ranged between 14 – 30 Hz, and although being of high interest for capturing as a response to stimuli, this category is of low amplitude, multiple and varying frequencies are often associated with active, busy or anxious thinking and active concentration. Low amplitude of Beta pattern makes it notably susceptible to contamination by artifacts and noise. The location of signal generation on the scalp is at both sides, symmetrically distributed, and most evidently towards the frontal side of the head. These waves are generated while the subject is actively calm and is focused and highly alert, thinking, especially, it is the dominant rhythm in patients who are alert or anxious or who have their eyes open, namely, the locked-in patients.

- **Gamma Patterns**

  These patterns are seen at frequencies between 30–100 and are generated at the Somatosensory cortex of the brain. This category of waveforms is displayed during cross-modal sensory processing (a perception that combines two different senses, such as sound and sight). Gamma patterns are thought to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function. Also Gamma is appeared during short-term memory matching of recognized objects, sounds, or tactile sensations.

- **Mu Patterns**

  These patterns are ranged between 8–13 Hz and partly overlapped with other frequencies. They are generated at the Sensorimotor cortex of the brain, and represent rest-state motor neurons.

In awake people, primary sensory or motor cortical areas often display 8–12 Hz EEG activity

when they are not engaged in processing sensory input or producing motor output. Computer-based analyses reveal that idling waves are distinguished from each other by location, frequency, and/or relationship to concurrent sensory input or motor output. These idling patterns are usually associated with 18-26 Hz (range of Beta) rhythms. While some Beta patterns are harmonics of Mu patterns, some are separable from them by topography and/or timing, and thus are independent brain signal features [22]. Beta and Mu waves are associated with those cortical areas most directly connected to the brains normal motor output channels, therefore, are leader choices for EEG-based BCIs which enable the subject to command movements to the BCI, without any peripheral muscle movement.

As the last not least concept outlined before closing the discussion on EEG, it is important to introduce the mental process, Motor Imagery (MI). MI is one of the most popular and widely used techniques for the BCI systems to be efficient, as they are supposed to be highly accurate and capable of well-interpretation by the quickest pace possible, regardless of the limitations of the end-user. The MI is defined as mere imagination of a limb movement, with no actual movement or peripheral (muscle) activation [31]. This mental execution of a movement is known to induce brain activity in the same way performing an actual movement brings about the firing of neurons in the brain [32]. According to this view, the main difference between performance and imagery is that in the latter case execution would be blocked at some corticospinal level [33]. The variation in brain activity is quantified from an Electrophysiological recording by EEG during the MI task. In MI-based BCIs, patients often receive visual or kinesthetic feedback in order to promote the brain response to the MI task. Increasing the accuracy of BCIs using this mental process is the focus of my thesis throughout the next chapter.

This completes a brief discussion on BCIs, EEG, and essential knowledge for understanding the applications of these two. Next, I will discuss the techniques and methods employed to process data for an EEG-based BCI system.

2.3 Data Dimensionality Reduction

Through the recent advances in statistics, signal processing, machine learning, and information theory, the size and the number of random variables, storing information regarding an ongoing study,

are experiencing significant growth. As a first step to proceed on the data understanding, researchers and engineers are often confronted with the problem of a proper and/or optimal selection of information among numerous patterns in a set of data. The superficial dimensionality of data, or the number of individual observations constituting one measurement vector, is often much greater than the intrinsic dimensionality, the number of independent variables underlying the significant non-random variations in the observations [34]. Therefore, to conclude the features and/or the attributes of a set of data, the first step required is to reduce the dimensionality of the data. Dimensionality reduction is defined as the process of reducing the number of random variables under consideration by obtaining a set of key variables. Dimensionality reduction not only allows us to shift the focus of processing on the more informative variables, but also results in, generally noticeably, speeding up the processing and classification phase.

In comparison with other sorts of biomedical signals, EEG is considered to be excessively complicated for an untrained observer to understand. Raw EEG signals are also extremely burdensome to be directly processed, mostly, as a consequence of the high temporal resolution of EEG technique and the spatial mapping of functions onto different regions of the brain and electrode placement. Hence, prior to applying other processing methods to extract features and classify the recorded data, dimensionality reduction is commonly used to determine a reduced feature set, including only (more or less) the data needed for further quantifications, in respect to a negligible loss of information during this process.

Principal Component Analysis (PCA) [35] and Singular-Value Decomposition (SVD) [36] are known to be well-established methods for the purpose of dimensionality reduction. Percent Root-mean-square Difference (PRD) [37] is also a method based on SVD which has been well-regarded in Electrocardiography (ECG) processing and applies to EEG signals. This subsection serves to provide an overview of the techniques above.

2.3.1 Principal Component Analysis (PCA)

The PCA is one of the most traditionally used techniques in EEG signal processing and BCI module design [25]. This technique is a leader choice method both for dimensionality reduction and feature extraction among its similar techniques in many fields of computer science as well. In

the PCA, we seek to represent a given *n*-dimensional data in a lower-dimensional space. This will reduce the degrees of freedom; reduce the space and time complexities. The objective is to represent data in a space that best expresses the variation in a sum-squared error sense. The PCA functions significantly better if an estimation of the number of independent components is known apriori.

The basic approach in principal components is theoretically rather simple. First, the *n*-dimensional mean vector $\boldsymbol{\mu}$ and $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ are computed for the full data set. Next, the eigenvectors and eigenvalues are computed, and as the eigenvalues $\{\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n\}$ are sorted in a descending order, so are the eigenvectors $\{\boldsymbol{e_1}, \boldsymbol{e_2}, ..., \boldsymbol{e_n}\}$ accordingly. Subsequently, a subset of eigenvectors associated with the largest eigenvalues is chosen. In practice, this is done by looking at the spectrum of eigenvectors. Often there is a clear distribution within the spectrum, implying an inherent dimensionality of the subspace governing the signal. The other dimensions are noise. Form a $(m \times m)$ matrix $\boldsymbol{A}$ whose columns consist of the m eigenvectors. Preprocessing of the data is performed as follows

$$\hat{\boldsymbol{X}} = \boldsymbol{A}^T \times (\boldsymbol{X} - \bar{\boldsymbol{X}}), \tag{1}$$

where superscript T denotes transpose operator and $\bar{\boldsymbol{X}}$ is the matrix of PCA essentially rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information. The PCA has the distinction of being the optimal orthogonal transformation for keeping the subspace that has the largest variance.

2.3.2 Singular Value Decomposition (SVD) and Percentage Root-mean-square Difference (PRD)

Several (if not all) data dimensionality reduction techniques are based on the decomposition of a key matrix, into its eigenvectors and eigenvalues. The reason behind this approach is the fact that through a canonical representation, there can be information retrieved that are just not obvious to a researcher by a glance at the original matrix, information such as: the underlying probability distribution of the matrix; similarities of random variables; the dependencies of random variables

onto each other; and, many other useful key information [38]. In linear algebra, the SVD is a factorization of a real or complex matrix. SVD is proposed as the generalization of the conventional eigendecomposition of a positive semidefinite normal matrix (e.g., a symmetric matrix with positive eigenvalues) to any $(m \times n)$ matrix via an extension of the polar decomposition. It has several useful applications in signal processing and statistics, namely, computing the pseudoinverse, least squares fitting of data, multivariable control, matrix approximation, and determining the rank, range and null space of a matrix. However, in biomedical engineering, this approach can be the first step taken towards dimension reduction of the data at hand. To outline the method, suppose $\boldsymbol{X}$ is an $(m \times n)$ matrix which contains real or complex numbers. Then, there exists a factorization, called a singular value decomposition of $\boldsymbol{X}$, of the form

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{2}$$

where

- $\boldsymbol{U}$ is an $(m \times m)$ unitary matrix;
- $\mathbf{\Sigma}$ is a diagonal $(m \times n)$ matrix with non-negative real numbers on the diagonal;
- $\boldsymbol{V}$ is an $(n \times n)$ unitary matrix and $\boldsymbol{V}^H$ is the Hermitian (conjugate) transpose of $\boldsymbol{V}$.

Throughout the development of this thesis, I have been employing the SVD technique along with a well-regarded yet straightforward measure, PRD, to evaluate the optimality of the dimensionality reduction step. PRD is widely known for its noteworthy application for Electrocardiography (ECG) signal processing techniques, for instance in [39] and [40]. While illustrating notable results in ECG, during my thesis research work, PRD turned to be practically applicable for EEG signal processing, once applied alongside the dimensionality reduction step. PRD can be considered as a quality control measure, to address the main concern of the dimensionality reduction techniques, which is the preserving the target parts of the signals while removing the redundant and irrelevant information. PRD ensures that the signals matrices are reduced in size up to a certain amount of information loss. For instance, suppose we have an estimation of irrelevant data contained within the recorded signals, which is about $p\%$ of the entire data. PRD, utilizing SVD, decomposes the signals

Algorithm 1 SVD–PRD DIMENSIONALITY REDUCTION

Input: {*Original EEG (OEEG) signals* $\boldsymbol{X}$ *(Channels as the variables each containing same length of time series signals) and the Percentage of acceptable error* $p\%$.

Output: {*Minimum number of variables (channels) required to maintain the signals' quality in respect to the aforementioned error percentage.*

1: **Decompose OEEG by SVD** : $[\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}]$ = svd($\boldsymbol{X}$). Number of columns taken into account (k) in $\boldsymbol{U}$ and $\boldsymbol{V}$ is set to 1.

2: **Reconstruction Loop**:

• First k diagonal element(s) of $\boldsymbol{S}$ and first k column(s) from $\boldsymbol{U}$ and $\boldsymbol{V}$ are chosen to reconstruct the EEG Signal as $\hat{\boldsymbol{X}}$ ($\boldsymbol{REEG}$), as per Equation 2.

• The PRD is computed as follows

$$PRD = \sqrt{\frac{\sum_{i,j}(\boldsymbol{X_{i,j}} - \hat{\boldsymbol{X}}_{i,j})^2}{\sum_{i,j}(\boldsymbol{X}_{i,j})^2}} \times 100. \tag{3}$$

• **Evaluation**

If: $PRD \leqslant$ p%
break;
else:
$k = k + 1$;

3: **Number of Variables Required**: Final k will be the output of the function as the minimum number of variables (channels) required for a more accurate dimension reduction.

and begins reconstructing them variable by variable. Every time data of each variable is added, the reconstructed signals are evaluated and compared to the original one, and the error of compatibility is computed. The algorithm keeps adding variables until the error is qual or less than $p\%$. In order to clarify the way PRD functions as described, the Algorithm 2.3.2, presents the pseudocode of the PRD approach in details.

2.4 Feature Extraction

As discussed previously, different thinking activities result in different patterns of brain signals and these mental tasks are of utmost importance while designing an experiment or an application for BCIs. From this outlook, BCI is seen as a pattern recognition system that classifies each pattern into a class according to its features. To elucidate this matter more clearly, it is important to define a "feature". During a phenomenon which is under observation of a machine learning/pattern recognition outline, a feature is an individual measurable property or characteristic of that phenomenon

under study. In other words, a feature is supposedly a variable/attribute which together with other pearly chosen features can represent the data statistically and, if selected suitably, will contribute to rather accurate classification of the entire dataset. However, choosing informative, discriminating and independent features is a crucial step for the development of effective algorithms in pattern recognition, classification and regression. Having said that, "feature extraction" is the process of starting from an initial set of measured data and building the derived values (features), aiming to provide a suitable feature set that would be informative, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.

BCIs extract some features from brain signals that reflect similarities to a certain class as well as differences from the rest of the classes. The extracted features should be measured or derived from the properties of the signals which contain the discriminative information needed to distinguish their different types. However, the challenging issue of feature extraction in BCIs is that it is interwoven with the fact that the information of interest in brain signals is hidden in a highly noisy environment, and brain signals comprise a large number of simultaneous sources. A signal that may be of interest could be overlapped in time and space by multiple signals from different brain tasks. For that reason, in many cases, it is not enough to use simple methods such as a bandpass filter to extract the desired band power.

From a signal processing point of view, feature extraction is done after preprocessing and data dimension reduction, as an important step in the construction of any pattern classification and aims at the extraction of the relevant information that characterizes each class. These feature vectors are then used by classifiers to recognize the input unit with target output unit. The classifier's task is much more facilitated if it is to classify between different classes by looking at these features as it allows fairly straightforward to distinguish. However, before concatenated into a single feature vector, multiple features can be selected from different channels and from various time segments, although, it is not desirable to process high dimensional features. In several neuroimaging studies, the sample size, or the number of the subjects of the study, is often much less than the size of scanned samples. Therefore, the number of features greatly outnumbers the number of the subjects. This challenge is known as *"curse of dimensionality"* or *"small-n-large-p"* [41]. In order to choose the most suitable features, one may attempt to examine all the possible subsets for the features,

although, as the number of possibilities grows exponentially, this approach becomes more and more impractical and exhaustive method of search. Obviously, there are more efficient and optimal feature extraction methods to replace this exhaustive search.

Before moving forward to describing the feature extraction technique used in this thesis, it is worth viewing feature extraction and classification from a big picture. The classification step, as the final step of recognizing the patterns, can be done in three different ways: (i) Supervised; (ii) Unsupervised; and, (iii) Semi-Supervised, which will be discussed more in details throughout the next section. Note that the feature extraction techniques utilized for different parts of my thesis are all chosen with regards to a supervised classification method. In order to select the most appropriate classifier for a given BCI system, it is essential to clearly understand what features are used, what their properties are and how they are used. The next two subsections are allocated to provide a brief overview on Common Spatial Patterns (CSP) method.

2.4.1 Common Spatial Patterns

The Common spatial patterns is a particularly popular and powerful signal-processing technique used for feature extraction in EEG-based BCIs. Originally, CSP has been designed for the analysis of multichannel data belonging to 2-class problems. Nevertheless, some extensions for multiclass BCIs have also been proposed, e.g., [42], which is not the focus of this thesis, thus, is not explained.

The CSP, as a mathematical algorithm, computes spatial filters that aim at achieving optimal discrimination by separating a multivariate signal into additive subcomponents which have maximum differences in variance between two classes. Hence, the signal-to-noise ratio is increased and adverse effects of volume conduction is reduced [43]. In other words, CSP projects multichannel EEG signals into a subspace, where the differences between classes are highlighted and the similarities are minimized. It aims to make the subsequent classification much more effective. The main idea of the CSP approach is to employ a linear transform to project the multi-channel EEG data onto a low-dimensional spatial subspace. The rows of the projection matrix which serves to this goal consist of the associated weights of the channels. The CSP method is based on simultaneous diagonalization of the covariance matrices of both classes. The Algorithm 2 describes the steps of implementation of CSP.

Algorithm 2 COMMON SPATIAL PATTERNS STEP BY STEP (FOR FEATURE VECTORS OF SIZE 2)

Input: {***(a)*** *Original EEG (OEEG) signals as a tensor of size $N_c \times N_t \times N_e$; N_c being the number of EEG channels, N_t denotes the number of time samples within each trial of performing the task by the subject, and N_e represents the number of times subject performed the task;* ***(b)*** *Labels of OEEG, that indicate to which class each trial belongs.*}

Output: {*Feature Vectors, ready to be classified*}

1: **Covariance Matrices**: Computing the sample covariance matrix corresponding to each trial $\boldsymbol{X_i}$ as follows:

$$\boldsymbol{C_i} = \frac{\boldsymbol{X_i}\boldsymbol{X_i}^T}{\text{Trace}(\boldsymbol{X_i}\boldsymbol{X_i}^T)} \tag{4}$$

The superscript T indicates the transpose of the matrix.

2: **Class Distinguisher Loop**: Trial by trial, each $\boldsymbol{C_i}$ is assigned to its corresponding class, using the labels.

3: **Class Averages**:

• The covariance matrices of each class are averaged resulting in $\bar{\boldsymbol{C}}_{\mathbf{1}}$ and $\bar{\boldsymbol{C}}_{\mathbf{2}}$.

• The composite spatial covariance $\boldsymbol{C}_c$ is computed as follows.

$\boldsymbol{C}_c = \bar{\boldsymbol{C}}_{\mathbf{1}} + \bar{\boldsymbol{C}}_{\mathbf{2}}$.

4: **Whitening Matrix**:

• The composite spatial covariance is decomposed to its eigenvalues and eigenvectors , in respect to the average covariance matrix of the first class, i.e.,

$[\boldsymbol{EVec}, \boldsymbol{EVal}] = eig(\bar{\boldsymbol{C}}_{\boldsymbol{c}}, \bar{\boldsymbol{C}}_{\mathbf{1}})$.

• Whitening matrix is composed as follows.

$\boldsymbol{W}(:, 1) = \boldsymbol{EVec}(:, 1)$;

$\boldsymbol{W}(:, 2) = \boldsymbol{EVec}(:, end)$;

5: **Finalizing the feature vectors**:

• The Whitening matrix is applied to each trial of the original data, i.e.,

$\boldsymbol{Z}_i = \boldsymbol{W}^t \times \boldsymbol{X}_i \times \boldsymbol{X}_i^t \times \boldsymbol{W}$.

• Feature vector corresponding to each trial's $\boldsymbol{Z}$ is computed as follows.

$$\boldsymbol{f}_i = \log \frac{\text{diagonal}(\boldsymbol{Z}_i)}{\text{trace}(\boldsymbol{Z}_i)}. \tag{5}$$

The aforementioned algorithm is constructed for the case where it is preferred to utilize feature vectors each including two elements. In the case where more elements are required, at Step 4, symmetrically, an equal number of eigenvectors are retrieved and put together to build the whitening matrix. This completes a brief outline over feature extraction. Throughout the next chapters, the additional feature extraction methods will be described. Next section is allocated to explanation of classification, the last step of processing data in any desired BCI.

2.5 Classification

As stated previously, the primary goal of a BCI is to translate the intent of a subject directly into control commands for a computer application, a neuroprosthesis, or any other external device. In BCIs which take advantage of training data to build a model for translation of the features (supervised learning), users are provided with instructions on how to perform a task as a response to a stimulus. Thereafter, an often significant number of trials are required to calibrate a BCI and prepare it for successful further analysis and interpretation. In most existing BCIs, this identification relies on a classification algorithm, i.e., an algorithm that aims at automatically estimating the class of data as represented by a feature vector. This objective can be accomplished by a statistical analysis of a calibration measurement in which the subject performs well-defined mental acts, such as imagined movements [44].

Classification is defined as the problem of statistically identifying to which of a set of categories a new observation belongs. This problem attempts to learn the relationship between a set of feature variables and a target variable of interest. Since many practical problems can be expressed as associations between feature and target variables, this provides a broad range of applicability of this model [45]. In the importance of clarification of the classification terminology, it is essential to define a few terms used throughout this thesis:

- **Trial & Epoch**: During a run/experiment with an EEG-based BCI, the subject is asked to perform the task related to the stimulus a certain number of times, each of these observations is called a trial. The set of signals recorded during a trial is called an epoch.

- **Label**: In the event of knowing the intent of the subject associated with a trial beforehand, this information is considered as a label.

- **Training Data/Trials**: The training data, as a part of collected data, consists of labeled trials utilized to construct a classification model. These trials together are also called the training data/dataset.

- **Test Data/trials**: The test trials, the rest of collected data, are those that are evaluated using the generated model via the training trials, and the classification model assigns a label to each

unlabeled test trial.

- **Classification Accuracy**: In the case that the grand truth, i.e. labels of the test data, is available, it is compared against the estimated label and overall performance of the classification model is reported as the classification accuracy, often in percentage.

Based on the stated definition, a *classifier* is defined as a function or an algorithm that maps every possible input available in the calibration (training) dataset to a finite set of decisions. In other words, given a set of training data points along with their associated training labels, a classifier determines the class label for an unlabeled test instance. Classification algorithms use the extracted features as independent variables to define boundaries between the different targets in the feature space. Building classification algorithms which are traditionally calibrated by users using a labeled dataset, are also known as supervised learning.

Classification algorithms can be developed via either offline, online or both kinds of sessions. The offline session involves the examination of datasets after the experiment is carried out. The statistics of the data may be estimated from observations across entire sessions and long computational processes may be performed. While offline data analysis is valuable regarding studying the behavior of brain signals and the effectiveness of the BCI processing algorithm, it does not address real-time requirements. However, online sessions provide a means of BCI system evaluation in a real-world environment. The data is processed causally and for higher efficiency of the processing algorithm, it is implemented as a closed-loop system. Evidently, online analysis can yield solid evidence of BCI system performance, and that is why offline simulation and cross-validation can be valuable methods to develop and test new algorithms [25].

Regardless of the model selected as to define the classifier, there are four main steps to construct a suitable and responsive classification model for the desired BCI system, as briefly outlined below:

(1) **Choosing a model**: There are several models of classifiers suggested and developed by researchers over time, some are well-suited for image processing, some other are known to work best for text-based data, and likewise for many other applications, also, the chosen model needs to be compatible with the feature vectors. A few classification models that are known to work best for BCI systems are described further in this section.

(2) **Training**: Using the training data, the parameters of the chosen model are determined (learned). In this step, the training epochs and their labels are employed to incrementally improve the classification models' ability to predict to which class the desired test epoch belongs. The training process is initialized with some random values for the parameters of the model and then attempting to predict the output with those values. Then, the model continues to learn and tunes its parameters to improve the comparison results, trial by trial until it finds the best fitting model. In the interest of determining how many trials out of the entire experiment should be utilized as for the training purpose, a conventionally common rule of thumb is to split into training-test by 70%–30% or 80%–20%; certainly, much of the applicability of this convention depends on the size of the source dataset.

(3) **Evaluation**: Once the training is complete, the trained model has to be evaluated. This is where that dataset that we set aside earlier comes into play. Evaluation enables the designer of the BCI to test the constructed model against data that has never been used for training, referred to as the test data. This is meant to be representative of how the model might perform in the real world. The evaluation result is generally reported as "validation accuracy" in percentage.

(4) **Prediction**: Prediction is the step where the question about the subject's intent for each trial is answered. The trained and evaluated model, if performing acceptably, are applied to the test dataset, and for each trial, a label is predicted by the classifier. The comparison between the grand truth and the estimated labels indicates the effectiveness of the classifier.

Training an algorithm and evaluating its statistical performance on the same data yields an overoptimistic result. Therefore, it is beneficial to know how a model would perform when it is applied to new data, beforehand. However, the test dataset is set aside to assess the predictive performance of the models and to judge how they perform outside the samples the parameters are learned from. In this case specifically, it is noteworthy to analyze the capability to accurately translate the intent of the BCI user. In most real applications, only a limited amount of data is available, which leads to the idea of splitting the data: Part of data (the training sample) is used for training the algorithm, and the remaining data (the validation sample) are used for evaluating

the performance of the algorithm. The validation sample can play the role of "new data". "Cross-validation" is the approach which is mainly used to serve this purpose and to estimate how accurately a predictive model will perform in practice. The cross-validation technique used throughout this thesis is called "k-fold cross-validation". In k-fold cross-validation, the original training dataset is randomly partitioned into k equal-sized subsets. Out of the k subsets, a single subset is preserved as the validation data for evaluation of the model, and the remaining $k-1$ subsets are used as to train the classification model and tune the parameters. The cross-validation process is then repeated k times, equal to the number of the folds, with each of the k subsets used exactly once as the validation data. The k results from the folds can then be averaged to serve as a single performance estimation. This way, all observations are used for both training and validation, and each observation is used for validation exactly once.

Classifiers commonly face the main problem related to the pattern recognition task, the problem of bias-variance tradeoff. This problem represents the natural trend of the classifiers towards a high bias with low variance and vice versa. Stable classifiers are characterized by high bias with low variance, while unstable classifiers show high variance with low bias. To achieve the lowest classification error, bias and variance should be low simultaneously. A set of stabilization techniques such as the combination of classifiers or regularization can be used to reduce the variance. In other words, low-performance classifiers may occur as a result of not proper fitting of the classification model, maybe because, the model is too simple to describe the target, or maybe the model is too complex to express the target. *Underfitting* and *Overfitting* both are issues that lead to poor predictions on new data sets, as shown in Fig. 2.2.

Underfitting occurs when a statistical model cannot capture the underlying trend of the data. Intuitively speaking, underfitting occurs when the model or the algorithm does not fit the data well enough. Specifically, underfitting occurs if the model or algorithm shows low variance but high bias. Underfitting is often a result of an excessively simple model. On the other hand, overfitting occurs when a statistical model captures the noise of the data. Intuitively speaking, overfitting occurs when the model or the algorithm fits the data too well. Specifically, overfitting occurs if the model or algorithm shows low bias but high variance. Overfitting is often a result of an excessively complicated model, and it can be prevented by cross-validation. This completes a general overview

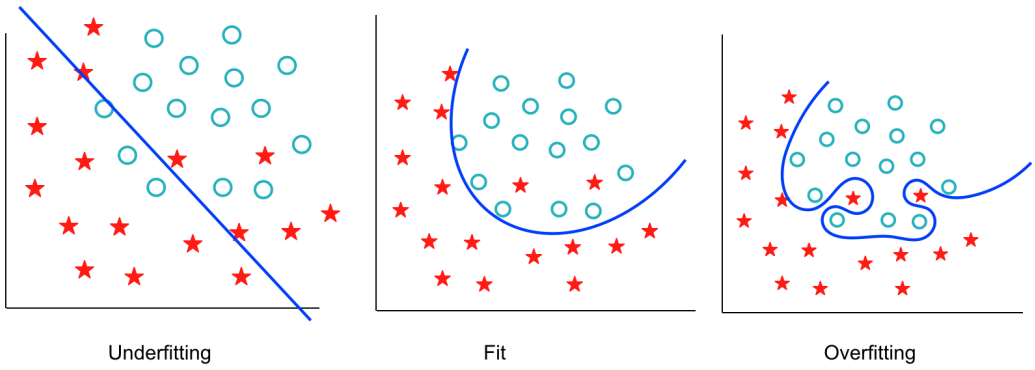


Figure 2.2: The problem of Underfitting and Overfitting.

of the concept of classification and its essential terms related to it. Throughout the thesis, I have taken advantage of classification models such as "Linear/Quadratic Discriminant" and "Support Vector Machines"; a brief outline of each of these models is provided below.

- **Linear and Quadratic Discriminant Analysis**:

  Linear Discriminant Analysis (LDA) is a simple classification model with rather acceptable accuracies, [46] for instance, without requiring highly complex computations. LDA is usually applied to classify patterns into two classes, although it is possible to extend the method to multiples classes [47]. For a two-class problem, LDA assumes that the two classes are linearly separable. In respect to this assumption, a linear discrimination function is defined in a fashion that it represents a hyperplane in the feature space in order to distinguish the classes. The class to which the feature vector belongs will depend on the side of the plane where the vector is found. The decision plane can be elucidated mathematically as follows

$$\boldsymbol{g(f_i)} = \boldsymbol{a}^T \boldsymbol{f}_i + \boldsymbol{b}_0, \tag{6}$$

  where, $\boldsymbol{a}$ is the weight vector, T is the transpose operator, $\boldsymbol{f}_i$ is the input feature vector extracted from the training sets, and $\boldsymbol{b}_0$ is the threshold. The input feature vector is assigned to one class or another on the basis of the sign of $\boldsymbol{g(f_i)}$. There are several methods to compute $\boldsymbol{a}$, e.g., $\boldsymbol{a}$ may be calculated as follows,

$$\boldsymbol{a} = \boldsymbol{C_c^{-1}}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \tag{7}$$

where the matrix $\boldsymbol{C_c}$ is given by

$$\boldsymbol{C}_c = \frac{\bar{\boldsymbol{C}}_\mathbf{1} + \bar{\boldsymbol{C}}_\mathbf{2}}{2}, \tag{8}$$

and, $\bar{\boldsymbol{C}}_\boldsymbol{i}$ is calculated in a similar manner as specified in Algorithm 2, and $\boldsymbol{\mu}_i$, the estimated mean of class i is calculated as follows

$$\boldsymbol{\mu}_i = \frac{1}{n} \times \sum_{i=1}^{n} \boldsymbol{F}_i, \tag{9}$$

where, $\boldsymbol{F}_i$ is a matrix containing n feature vectors of class i, as $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{f}_3, ..., \boldsymbol{f}_n \in \mathbb{R}^{N_e}$.

LDA serves to create a new variable, which would be a combination of the original predictors (features of the train set), which is accomplished through maximizing the differences between the features of the predefined classes. As a result, a discriminant score is a weighted linear combination of the predictors. The weights are estimated in a way that the differences between mean discriminant scores of each class have the maximum distance from all others. Generally, those predictors which have large dissimilarities between class means will have larger weights, at the same time weights will be small when class means are similar.

LDA assumes that the observations within each class are drawn from a multivariate Gaussian distribution and the covariance of the predictor variables are common across all levels of the responses (labels), however, Quadratic discriminant analysis (QDA) provides an alternative approach. While utilizing QDA, it is assumed that the measurements from each class are normally distributed. Unlike LDA, in QDA there is no assumption that the covariance of each of the classes is identical. In the case where the feature vectors consist of two classes, QDA seeks for surfaces separating the classes among conic sections (i.e. either a line, a circle or ellipse, a parabola or a hyperbola). In this sense, we can state that a quadratic model is a generalization of the linear model, and its use is justified by the desire to extend the classifier's ability to represent more complex separating surfaces. The QDA is very similar to the LDA except that because the covariance matrix is not identical, quadratic terms are also involved.

This approach allows for more flexibility for the covariance matrix, therefore, tends to fit the data better than LDA, in contrast, it has more parameters to estimate.

- **Support Vector Machines**:

  Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane into two parts wherein each class lay on either side. SVM is a classifier similar to LDA classifiers, however, in contrast, SVM selects the hyperplanes that maximize the margins, i.e., the distance between the nearest training samples and the hyperplanes [48]. Due to highly accurate results, SVMs have been successfully used in several BCI applications, namely, [35, 44, 49]. The majority of the similar classifiers use hyperplanes to separate classes, based on a flat plane within the predictor space. Whereas SVMs broaden the concept of hyperplane separation to data that cannot be separated linearly, by mapping the predictors onto a new, higher-dimensional space in which they can be separated linearly.

  The method is called as it is, for the support vectors, a subset of training points in the decision function, are lists of the predictor values, taken from cases that lie closest to the decision boundary separating the classes. Computationally, finding the best location for the decision plane is an optimization problem that makes uses of a kernel function to build linear boundaries through nonlinear transformations, or mappings, of the predictors. The intelligent component of the SVM algorithm, however, is that it locates a hyperplane in the predictor space which is stated regarding the input vectors and dot products in the feature space without ever representing the space explicitly. SVM chooses one particular solution: the classifier which separates the classes with maximal margin, as Fig. 2.3. The margin is defined as the width of the largest 'tube' not containing samples that can be drawn around the decision boundary.

  It should be noted that it is possible to create nonlinear decision boundaries, with only a low increase of the classifiers complexity, by implicitly mapping the data to another space, generally of much higher dimensionality [51].

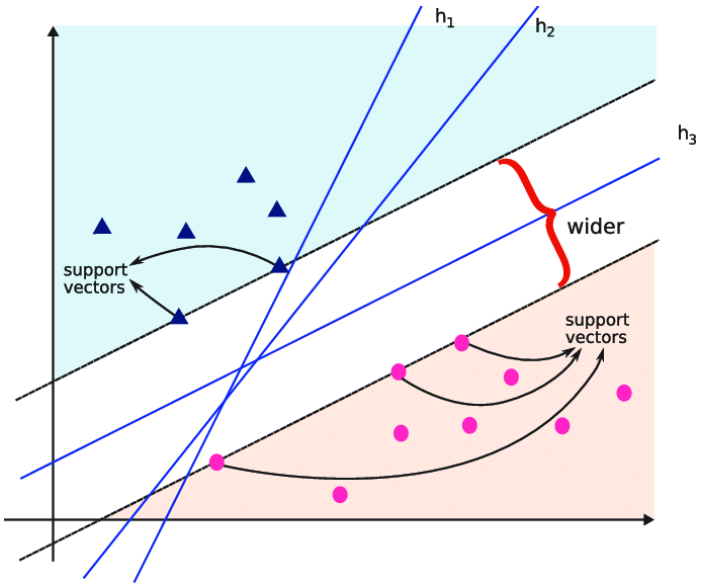


Figure 2.3: An example of possible and optimal hyperplanes [50]

This completes my overview of all the necessary concepts and methods for understanding BCIs, before I move on to describing my contributions. Next three chapters describe the frameworks I proposed and investigated during my research work.

Chapter 3

Graph-based Frameworks for Spatio-Temporal Filtering and Dimensionality Reduction in MI EEG-based BCIs

In this chapter, two frameworks, GD-BCI and GDR-BCI, are proposed to address the problem of processing EEG signals for motor imagery brain-computer interfaces. The goal is to identify the pattern of the brain activity using a robust method for pre-processing, processing, and classification of the EEG signals. To this end, the GD-BCI framework is a new graph-based approach, proposed to spatiotemporally filter the data by taking into account not only the geometrical structure of the channels/electrodes, but also the correlation between the EEG signals. The most significant feature vectors required for classification of EEG signals are adaptively selected through spectral decomposition of the data using the graph Laplacian matrix. The tangent space mapping method is then applied to bring the captured data into Euclidean space. In order to classify the dimensionally-reduced EEG signals, the linear support vector machine algorithm is employed. Experiments are conducted on Dataset IVa from BCI Competition III, including data from five different subjects

consisting of the right hand and right foot motor imagery actions. The results show that the proposed GC-BCI framework provides higher classification accuracy as compared to the other existing methods. However, as the impressive accuracies come with the price of an exhaustive search for a pair of constants required for adjustment of the graph, I sought for a solution to adaptively and automatically configure the graph, hence, the second framework the GDR-BCI is developed.

In the GDR-BCI, similar to the one proposed in GD-BCI, by capitalizing the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject, an initial functional clustering of EEG electrodes are built by designing a separate adjacency matrix for each identified functional cluster. Using a collapsing methodology based on total variation measures on graphs, the overall model will eventually be reduced (collapsed) into two functional clusters. The proposed framework offers two main superiorities over its state-of-the-art counterparts and the GD-BCI: (i) First, the resulting dimensionality reduction is subject-adaptive and respects the brain plasticity of subjects, and; (ii) Second, the proposed methodology identifies active regions of the brain during the motor imagery task, which can be used to re-align EEG electrodes to improve accuracy during consecutive data collection sessions. The experimental results based on the same Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to its counterparts.

3.1 Introduction To Graph Signal Processing

In mathematics, graph theory proposes a "graph" as a structure corresponding to a set of objects in which some pairs of the objects are in some sense "related". It consists of a set of vertices or nodes (objects) and a set of edges or connections indicating the presence of some of the interaction (relation) between the vertices. Therefore, a graph can be defined as an ordered pair of $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ comprising a set $\boldsymbol{V}$ of vertices, together with a set $\boldsymbol{E}$ of edges. Moreover, for a graph representing a real-life model, e.g., when a graph is an abstract representation of a network of sensors, the definition may be extended to $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{K})$, where a number is assigned to each edge, together summarized in matrix $\boldsymbol{K}$. Due to this simple yet powerful and flexible structure, graphs offer the ability to model massive amounts of data and complex interactions among them in a systematic,

organizable, and processing-friendly manner. Graphs are generic data representation forms that are useful for describing the geometric structures of data domains in numerous applications, including social, energy, transportation, sensor, and neuronal networks. The weight associated with each edge in the graph often represents the similarity between the two vertices it connects. The connectivities and edge weights are either indicated by the physics of the problem at hand or inferred from the data. On the other hand, growth of interest is widely observed in efficient signal processing techniques for representation, analysis and processing of large datasets (big data) emerging in various fields and applications. These datasets share common traits: their elements are related to each other in a structured manner, e.g., through similarities or dependencies between data elements, and the conventional methods to deal with them are inadequate. Hence, as a solution to address this issue, the field of Graph Signal Processing (GSP) [52] has emerged to merge computational analysis of the aforementioned signals/data with the graph theory.

Biological networks have proved to be a popular application domain for graph signal processing, with recent research works focusing on the analysis of data from systems known to have a network structure, especially, the human brain. The growing number of publications studying brain activity or brain network features from a GSP perspective, namely, [53–56], indicates that these are promising applications in this field. Despite recent advances in the GSP field, however, its application for EEG processing is still in its infancy.

As described in Chapter 2, EEG is widely used to capture brain waves and due to volume conduction, unprocessed EEG signals are known to have poor spatial resolution and a rather blurred image of the brain activity is often obtained from multichannel EEG signals due to low SNR. The field of GSP makes it possible to non-invasively infer the anatomical connectivity of distinct functional regions of the cerebral cortex via utilization of a regular/weighted graph with the vertices corresponding to different EEG channels. The connectivity between the nodes of the graph (channels of the EEG headset) can also be expressed through the physical distance between the electrodes. During my research work in this thesis, I employed two GSP models for the purpose of dimensionality reduction. The explanation of each and the final results are presented in the following sections.

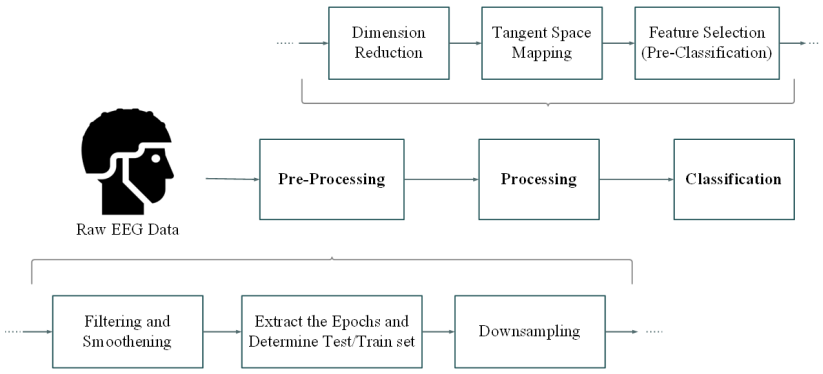


Figure 3.1: Block diagram of the proposed graph-based spatio-temporal filtering framework for brain-computer interface (GD-BCI).

3.2 The GD-BCI Framework

In this framework, a new graph-based approach for spatiotemporal filtering and feature selection in motor imagery brain-computer interfaces is proposed, referred to as the GD-BCI.

3.2.1 Framework Outline

The GD-BCI framework is realized by taking into account not only the geometrical structure of the electrode channels, but also the correlation coefficients obtained from the EEG signals. In brief, the graph similarity and Laplacian matrices are computed, the spatiotemporal filtering is performed in the graph Fourier transform domain, which is obtained in a spectral decomposition of the graph which is later applied to the EEG signals for dimensionality reduction purpose. A tangent space mapping technique is employed to project data from the Riemannian to the Euclidean domain. The PCA method is applied for selecting the most significant features for the classification task. Finally, the linear SVM is used to solve a two-class classification problem.

We consider supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by $\boldsymbol{X}_i \in \mathbb{R}^{N_{\text{ch}} \times N_{\text{t}}}$, for $(1 \leq i \leq N_{\text{Trial}})$, where N_{Trial} is the total number of trials used for processing; N_{ch} is the number of EEG channels (electrodes), and; N_{t} is the number of time samples collected from each electrode in one trial. The training dataset is denoted by $\{(\boldsymbol{X}_i, l_i)\}$, for $(1 \leq i \leq N_{\text{Trial}})$, where l_i represents the label corresponding to the ith trial, e.g., l_i could be "right foot" or "right hand". Parameters rTR and rTS are the numbers of eigenvectors required for the

dimensionality reduction process of training and test sets, respectively. For vector $\boldsymbol{X}_i$, the sample covariance matrix is defined as

$$\boldsymbol{C}_i = \frac{1}{N_{\mathrm{t}} - 1} \big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big) \big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big)^T, \tag{10}$$

where $\boldsymbol{\mu}_i$ is the column-wise mean of $\boldsymbol{X}_i$. The Eq. 10 above and the equation used in Step 2 of the Algorithm 2.4.1 for computing the covariance are equivalent. The proposed GD-BCI framework consists of the following main tasks, (as shown in Fig. 3.1): (i) Pre-processing; (ii) Spatio-temporal filtering; (iii) Mapping to tangent space; (iv) Feature selection; and (v) Classification. Below and in each sub-section, the aforementioned tasks are described in details.

3.2.2 Pre-Processing

Before processing EEG signals for classifying MI tasks, typically, a pre-processing step is required. At this stage, bandpass filtering is applied to extract specific frequency contents of the signal, containing brain patterns regarding the motory tasks. This step is then followed by downsampling in order to reduce the load of processing for the algorithm. The signal used for processing is extracted from a specific period of each trial time interval, i.e., as per the instructions given by the data collectors, the time interval during which the subject is responding to a stimulus shown. This step is conventionally done by selecting a predefined time interval after a visual cue and selecting one sample value out of n samples. In order to take into consideration the most of subject's response to each stimulus, I have tested the use of other methods for data smoothing in advance to the downsampling step, namely, simple averaging, simple moving average, weighted moving average and moving median. More specifically, instead of choosing randomly one value, i.e., random selection, within the time interval between the two consecutive visual cues, I apply the above methods. It is experimentally, as can be viewed in Table 3.1, found that using the weighted moving average as a smoothening filter provides a lower reconstruction error after dimensionality reduction stage and results in better classification accuracy.

The results in the first column of Table 3.1 are achieved by decomposing each trial utilizing SVD, and then reconstructing the matrix of trials using 41 eigenvectors corresponding to the 41

highest eigenvalues, and finally, measuring the difference (error) between the original and rebuilt signals in percentage. The number 41 is chosen considering the method employed in [54] and the downsampling approach they had is presented as **RS**. The second column, shows the measurement of similarity of the original and reconstructed signals on the scale of 0–1. For this purpose, the signals are reconstructed using a certain number of eigenvectors, corresponding to a measure known as 'expected variance'. It can be proved [57] that when the covariance of a matrix is computed and then decomposed to its eigenvectors and eigenvalues, the eigenvector with the largest eigenvalue will be the direction in which the most variance occurs. Consequently, if all the eigenvectors are put together, the entire variance in the target matrix can be explained. This means, as a surrogate for measuring the absolute value of variance explained, the simple fraction below can be an indicator of the variance covered by the reconstructed matrix. For a covariance matrix of size ($N_{\text{ch}} \times N_{\text{ch}}$), Expected Variance is defined as follows

$$EV = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{N_{\text{ch}}} \lambda_i} \times 100, \tag{11}$$

where the largest r eigenvalues, and their associated eigenvectors, are put together for the purpose of reconstruction of the data matrix, and λ_i is an eigenvalue retrieved from the decomposition of the covariance matrix. Having that said, the second column of Table 3.1 is composed by first, determining the number of eigenvectors/eigenvalues (r) required to cover $EV \geq 80\%$, and then the matrix is reconstructed in the same manner as for the first column of Table 3.1, only by r eigenvectors instead of 41. The last column of Table 3.1 shows the number of eigenvectors required to ensure PRD of 2% or less. In view of this, I use this filter in the smoothening filter block of our proposed framework. Next, I present my proposed graph-based dimensionality reduction technique.

3.2.3 Graph-based Spatio-Temporal Filtering

The aforementioned weighted graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{K})$ consists of a finite set $\boldsymbol{V}$ of vertices (electrode channels) and a finite set $\boldsymbol{E}$ of edges with the corresponding weights $[k_{pq}]_{n \times n} \in \boldsymbol{K}$. The

Table 3.1: Average reconstruction error obtained using various data smoothening methods for training datasets.

	FF: rTR = 41	EV: 80%	PRD: 2%
	AA		
RS	4.32	0.90	69
SA	4.21	0.87	68
SMA	4.21	0.86	68
WMA	4.08	0.83	67
MM	4.33	0.91	69
	AW		
RS	3.39	0.68	59
SA	3.27	0.66	57
SMA	3.27	0.65	57
WMA	3.12	0.62	55
MM	3.39	0.69	59
	AL		
RS	3.63	0.68	58
SA	3.51	0.66	57
SMA	3.50	0.63	57
WMA	3.37	0.62	55
MM	3.62	0.69	58
	AY		
RS	4.20	0.68	64
SA	4.06	0.65	63
SMA	4.06	0.65	63
WMA	3.90	0.62	62
MM	3.19	0.69	64
	AV		
RS	4.39	0.83	67
SA	4.34	0.81	66
SMA	4.34	0.81	66
WMA	3.27	0.79	65
MM	3.41	0.85	67

weights k_{pq} can be defined as a function of proximity between vertices (electrodes) p and q, as

$$\boldsymbol{K}_{\text{PG}} = \exp\left(-\frac{\boldsymbol{D}(p,q)^2}{2\sigma_d^2}\right), \tag{12}$$

where p and q are the electrode positions, and $\boldsymbol{D}(p,q)$ denotes the distance between the two electrodes via the following equation

$$D_{pq} = \sqrt{{X_{pq}}^2 + {Y_{pq}}^2}. \tag{13}$$

In this framework, in order to take into account the dependencies of the data captured at each electrode, a weight matrix is proposed which is a function of both the electrode proximity and correlation coefficients obtained from the EEG signals.

$$\boldsymbol{K}_{\text{VPG}} = \exp\left(-\frac{\boldsymbol{D}(p,q)^2}{2\sigma_d^2}\right).\exp\left(-\frac{(1-\|\boldsymbol{\rho}(p,q)\|)^2}{2\sigma_\rho^2}\right), \tag{14}$$

where σ_d and σ_ρ specify the amount of exponential decay rate, and

$$\boldsymbol{\rho}(p,q) = \frac{c_{pq}}{\sqrt{c_{pp}c_{qq}}}, \tag{15}$$

obtained using the elements of the covariance matrix $\boldsymbol{C}$, given in Eq. (11). Accordingly, the degree matrix $\boldsymbol{D}$ is defined using the weight matrix as

$$\boldsymbol{D} = \text{diag}\left\{\sum_q k(1,q), ..., \sum_q k(n,q)\right\}. \tag{16}$$

The graph Laplacian matrix is derived from $\boldsymbol{K}$ and plays an important role in describing the underlying structure of the graph signal. The graph Laplacian and its normalized version are defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{K}$ and $\boldsymbol{L}_{\text{normal}} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{K}\boldsymbol{D}^{-1/2}$, where $\boldsymbol{I}$ is the identity matrix. Spectral graph theory studies the graph properties in terms of eigenvalues and eigenvectors associated with the Laplacian matrix of the graph. The set of eigenvectors of $\boldsymbol{L}_{\text{normal}}$ constitute the basis function for the underlying signal defined on graph, and its eigenvalues are known as the corresponding graph frequencies. The eigen decomposition of the real and symmetric normalized Laplacian is given by

$$\boldsymbol{L}_{\text{normal}} = \sum_i \boldsymbol{\lambda}_i \boldsymbol{u}_i \boldsymbol{u}_i^T, \tag{17}$$

where $\{\boldsymbol{\lambda}_i\}_{i=1,\ldots,n}$ is the set of eigenvalues and $\{\boldsymbol{u}_i\}$ the set of orthogonal eigenvectors used for dimension reduction.

Let $\boldsymbol{U}$ contain the $\boldsymbol{L}$'s first r eigenvectors corresponding to the first r eigenvalues of $\boldsymbol{L}$ sorted in ascending order. The proposed spatiotemporal filtering on graph spectral theory employs matrix $\boldsymbol{U}$ to represent the EEG signals with a lower number of features $\mathbf{F}_r$ as given by

$$\mathbf{F}_r = \boldsymbol{U}_r^T \boldsymbol{X}. \tag{18}$$

It should be noted that the first r eigenvectors correspond to the first r low-frequency basis functions in graph spectral domain. To perform dimensionality reduction, the value of r can be adaptively determined for different subjects using the PRD method. The spatiotemporal filtering stage is followed by mapping the data from the existing manifold to the Euclidean space. To this end, tangent space mapping method is employed as a bridge operation to enable us to treat the data transferred to Euclidean space as vectors. This mapping method is discussed next.

3.2.4 Tangent Space Mapping and Feature Extraction

A 'manifold' is a topological space that locally resembles Euclidean space near each point. In other words, each point of an n-dimensional manifold has a neighborhood that is homeomorphic to the Euclidean space of dimension n. Likewise, a Riemannian manifold is defined as a smooth manifold with a smooth section of the positive-definite quadratic forms on the tangent space. Having that said, it is known that the sample covariance matrices belong to the Riemannian manifold of the symmetric and positive definite matrices. However, several significant and commonly used state-of-the-art methods of machine learning, and specifically, classification techniques, such as those introduced in Chapter 2, are mostly designed to be applied to datasets in the Euclidean space. Given this, we employ the tangent space mapping technique [58] and the vectorizing approach that follows it as the feature extraction method for the GD-BCI framework and then project data to Euclidean space as vectors to proceed with the classification step.

To explain the utilized approach, let a Riemannian manifold $\mathcal{S}(n)$ be a space of $(n \times n)$ symmetric positive definite matrices given by $\mathcal{S}(n) = \left\{\boldsymbol{S} \in \mathcal{M}(n), \boldsymbol{S}^T = \boldsymbol{S}\right\}$, where $\mathcal{M}(n)$ is the space of

all square real matrices. The set of all the matrices is denoted as $\mathcal{C}(n) = \left\{\boldsymbol{C} \in \mathcal{S}(n), \boldsymbol{u}^T \boldsymbol{C} \boldsymbol{u} > 0\right\}$, which is not Euclidean. Tangent space mapping provides a Euclidean tangent space, $\boldsymbol{T}_Q \boldsymbol{C}(n)$ at the point Q, which approximates the aforementioned Riemannian manifold through the following steps:

- Compute the set of sample covariance matrices for each trial as given in Eq. (10).
- Compute the mean Riemannian distance as

$$\bar{\boldsymbol{C}} = \frac{1}{N_{\text{Trial}}} \sum_{i=1}^{N_{\text{Trial}}} \boldsymbol{C}_i. \tag{19}$$

- Compute the map $\mathbf{s}_i$ from $\boldsymbol{C}$ to $\boldsymbol{T}_Q \boldsymbol{C}(n)$ as

$$\mathbf{s}_i = \text{Upper}\left(\log\left(\bar{\boldsymbol{C}}^{\frac{-1}{2}} \boldsymbol{C}_i \bar{\boldsymbol{C}}^{\frac{-1}{2}}\right)\right), \tag{20}$$

where Upper$(.)$ is used to weigh the upper triangular half of a matrix and vectorize it. In particular, it assigns 1 as the weight for main diagonal and $\sqrt{2}$ for off-diagonal entries. The resulting feature vectors are further trimmed and the most relevant ones are selected for classification purpose.

As per described in Chapter 2, the high dimensional features may lead to poor classification performance. This is due to the fact that a large number of irrelevant features not only degrades the generalization of the model, but also imposes a computational cost. Given this and in order to determine the most significant feature vectors of $\mathbf{s}_i$ in the tangent space mapping process, which are maximally related to the desired classes, we use a graph-based feature selection method similar to what is given in Eq. (18). To this end, a weighted graph is built from $\mathbf{s}_i$ and the first 10 eigenvectors of the corresponding Laplacian matrix are selected. The elected eigenvectors are then fed into the classifier as its input. In addition, for $PRD = 2\%$, eigenvectors are adaptively selected for each subject. It should be noted that using the PRD method, one can adaptively compact feature vectors for a better classification result.

3.2.5 Classification

In order to classify the selected feature vectors as representative of the right hand or right foot MIs, we employ the support vector machine (SVM) algorithm. The SVM uses training feature vectors to learn a decision boundary that separates these two classes by projecting data into a higher dimensional space using a kernel function. Once the decision boundary is learned, the SVM determines the class membership of a newly-observed feature vector according to the side of the boundary that the vector falls.

3.2.6 Simulations

The proposed framework is benchmarked on the dataset IVa from the BCI competition III taken from http://www.bbci.de/competition/iii/. The EEG positioning is based on $10-20$ standard system. The dataset is composed of EEG recordings of 118 electrodes. The experiment is a classical cue-based MI paradigm in which each of 5 subjects, namely, AA, AL, AV, AW, and AY, perform 280 trials of the right hand and right foot MIs. In the pre-processing step, the EEG signals are bandpass filtered in the frequency band $[8 - 30]$ Hz by a 5th order *Butterworth* filter. The time interval is restricted to the segment located from 0.5s to 4s after the cue. The weighted moving average filter is then applied to smoothen the data.

In order to obtain the most significant features for each subject, PRD is used to adaptively determine the required number of eigenvectors from which data can be reconstructed with a predefined error. The corresponding size of datasets for different subjects is given in Table 3.2.

Table 3.2 illustrates the classification accuracy and its corresponding standard deviation averaged over 400 runs, obtained using the proposed GD-BCI method using two graph construction method, namely, physical graph (PG) and value-physical graph (VPG). It is seen from this table that proposed a method using GD-BCI-VPG outperforms its GD-BCI-PG counterpart by almost 10%. This is due to the fact that the GD-BCI-VPG is built using the electrode channel proximity (PG case) as well as the correlation coefficients of the EEG signals, i.e., spatiotemporal filtering of the EEG signals.

We now compare the performance of the proposed GD-BCI method to that obtained from the

Table 3.2: Accuracy performance for predicting two classes and the corresponding standard deviation obtained using the proposed GD-BCI framework with two graph construction methods: PG and VPG

Subject AA (168 Train + 112 Test)			
PRD<2.5% rTR=rTS=60		PRD<4.15% rTR=41, rTS=43	
VPG	PG	VPG	PG
82.90±1.23	80.99±1.64	76.29±1.61	74.38±2.14
Subject AL (224 Train + 56 Test)			
PRD2% rTR=59, rTS=56		PRD<3.5% rTR=41, rTS=39	
VPG	PG	VPG	PG
97.37±0.30	97.62±0.23	97.62±0.27	97.38±0.31
Subject AW (56 Train + 224 Test)			
PRD<2% rTR=rTS=58		PRD<3.3% rTR=rTS=41	
VPG	PG	VPG	PG
92.58±2.82	91.30±2.75	93.70±2.21	90.83±3.14
Subject AV (84 Train + 196 Test)			
PRD<2% rTR=rTS=66		PRD<4.29% rTR=41, rTS=42	
VPG	PG	VPG	PG
65.79±2.97	66.87±2.69	66.99±2.83	64.21±3.32
Subject AY (28 Train + 252 Test)			
PRD<2% rTR=rTS=66		PRD<4% rTR=41, rTS=38	
VPG	PG	VPG	PG
83.14±5.76	68.82±5.67	80.29±4.64	84.32±4.61

other existing methods, in terms of the classification accuracy. Table 3.3 gives the comparison results of the proposed method using VPG and PG and that provided by [54] and [59], when the constant or adaptive number of features are selected. It is seen from this table that the proposed method provides higher classification accuracy for various subjects as compared to those yielded by [54] and [59]. In addition, the standard deviation of the classification accuracy obtained using the proposed method is lower than that provided by [54].

Table 3.3: Performance comparison of the proposed GD-BCI method in two-class classification problem with that provided by [54] and [59].

	GD-BCI-PG		GD-BCI-VPG		[54]	[59]
	PRD: rTR = 41	PRD: 2%	PRD: rTR = 41	PRD 2%	rTR = 41	rTR = 10
AA	74.38± 2.14	80.99± 1.64	76.29± 1.61	82.90± 1.23	81.43± 10.9	74.1
AL	97.38± 0.31	97.62± 0.23	97.62± 0.27	97.37± 0.30	97.50± 2.98	98.2
AW	90.83± 3.14	91.30± 2.75	93.70± 2.21	92.58± 2.82	98.57± 0.79	77.7
AV	64.21± 3.32	66.87± 2.69	66.99± 2.83	65.79± 2.97	69.29± 5.56	59.2
AY	84.32± 4.61	68.82± 5.67	80.29± 4.64	83.14± 5.76	93.93± 4.30	80.6
Average	82.22± 2.70	81.12± 2.60	82.98± 2.31	84.35± 2.69	88.14± 4.90	78.0

3.2.7 Conclusion

In this framework, a new dimensionality reduction and feature selection technique are proposed for analyzing EEG signals obtained from an EEG-based BCI during MI tasks. The GD-BCI has been established by leveraging the recent advances in the field of graph signal processing. The proposed method is composed of an efficient graph-based dimensionality reduction technique followed by tangent space mapping of the EEG signals to the Euclidean space. Experiments have been conducted on a set of EEG signals obtained from the BCI competition. The results have shown that the proposed method produces encouraging results providing high recognition accuracy for two motory tasks classification. However, all these great advantages come with an undesirable price: to determine the constants σ_d and σ_ρ to configure the best-suited graph for each subject's data, an exhaustive search is required. The trial and error solution to find the proper σ_d and σ_ρ is not preferable in a real-world application. Therefore, I sought for an adaptive and systematic approach as explained in the next section, to adjust the graph for each subject's dataset without excessive trial and error.

3.3 GDR-BCI: Dimensionality Reduction of EEG Signals via Functional Clustering and Total Variation Measure

In this framework, similar to the one proposed in GD-BCI, the observations obtained from the EEG channels, a non-uniformly distributed sensor field, are taken into account in a manner that a representation graph is formed using geographical distances between sensors to form connectivity neighborhoods. However, by capitalizing on the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject, an initial functional clustering of EEG electrodes was formed by designing a separate adjacency matrix for each identified functional cluster.

3.3.1 Framework Outline

This framework contributes to detection of main EEG electrodes that capture the intention of the subject due to the exposed stimulus. This purpose is achieved through an uncomplicated yet straightforward and systematic method which separates the EEG electrodes in a desired number of clusters. These clusters are formed under a biological constraint and then collapsed intelligently to identify the two clusters that best spread over the active part of the brain during the motor imagery task. In order to better perceive the impact of our proposed GSP-based approach for processing EEG signals, widely used feature extraction/classification algorithms are coupled with the proposed framework deliberately. The incentive of the proposed method roots back in the tendency of reducing the dimension of the dataset before processing it and without losing crucial information, and also, in order to address challenges that previously proposed graph-based methods typically face, i.e, formerly graph transforms applied to EEG datasets for dimensionality reduction were formed empirically through trial and error for determining/tuning the constants required for selecting the number of neighbors of each node. Via a change in the perspective of graph-based EEG signal processing, I rather focus on deciding which nodes are capturing the intended activity, and thereafter, are showing most relative activity to the purpose of the BCI, at the time of the experiment. It is important to note that this approach is subject-adaptive, technique-adaptive, applicable to multiclass experiments, and has practical advantages for experiments with limited means of measurement at

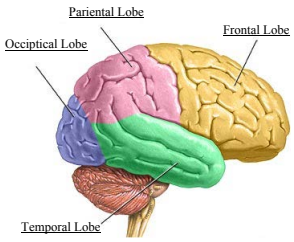


Figure 3.2: Functional clustering of human brain [60].

hand.

The electrodes of EEG headsets are often distributed over the subject's skull in a consistent geometrical structure, i.e., the arrangement or location of the electrodes of EEG recording device. Hence, the neighborhood of each channel and their distances from other channels can be treated as the required information for constructing a graph signal.

The GD-BCI framework involves a graph defined as a triplet $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{K})$, in which $\boldsymbol{K}$ is a weighted adjacency matrix. Intuitively speaking, each element of $\boldsymbol{K}$ represents the weight of the corresponding element in $\boldsymbol{E}$. The matrix $\boldsymbol{K}$ is formulated as an exponential term including the Euclidean distance as in (12) where as stated previously, the Euclidean distance between any two channels is computed via the Eq. (13). The matrix $\boldsymbol{K}_{\text{VPG}}$ takes into account the physical distance, as well as the correlation of the data in each trial [61]. However, the main obstacle in the way of constructing $\boldsymbol{K}_{\text{VPG}}$ above, is how to determine the threshold the constants σ_d and σ_ρ. During the previous framework, the neighborhood and the structure of the graph are highly sensitive to the values of σ_d and σ_ρ. This challenge inspired us to investigate an alternative approach to define the graph and reduce the dimension of the EEG data. Inspired by Reference [62], we decided to define our graph $\boldsymbol{G}$ as a twin of $\boldsymbol{G} = (\boldsymbol{V}; \boldsymbol{A})$ in which the matrix of nodes, i.e., $\boldsymbol{V}$, is similarly defined as the set of EEG electrodes. However, the second term $\boldsymbol{A}$ is the adjacency matrix. In this work, $\boldsymbol{A}$ is constructed via the Euclidian distances of electrodes from each other as described in detail further on.

3.3.2 Defining the Adjacency Matrix

As is described in Reference [62], in the case of measurements from a non-uniformly distributed sensor field (in our context, EEG channels), data recorded from each sensor is a separate time series.

In their example, a representation graph was constructed using geographical distances between sensors and considering L nearest sensors as the connectivity neighborhood. However, motivated by existence of different regions of the human brain with different functionality as shown in Fig. 3.2, and the fact that these functions might have various intensities of activity due to the status of the person and her/his concentration, I formed the idea of grouping the electrodes and proposing an individual corresponding adjacency matrix for each cluster independent from the others. In this fashion, the problem of determining/extracting the constants (σ_d, and σ_ρ) through trial and error is tackled, as the biological function of brain regions would be the constraint of the neighborhood for forming the required graphs.

Fig. 3.3 illustrates the sparsity of the electrodes (projected onto 2-dimension) of the EEG headset used to collect the dataset and the groups (clusters) assigned to this structure. A number of electrodes considered in each cluster is defined as:

- Cluster{1} = [1 2 3 4 5 6 7 8 9];
- Cluster{2}=[11 12 17 18 19 25 26 27 28 35 36 37 45 46];
- Cluster{3} = [10 14 15 16 23 24 33 34];
- Cluster{4} = [13 20 21 22 29 30 38 39];
- Cluster{5} = [31 32 42 50 59 67 68];
- Cluster{6} = [40 41 49 56 58 76 77];
- Cluster{7} = [43 44 51 52 60 61 69 70];
- Cluster{8} = [47 48 56 57 64 65 74 75];
- Cluster{9} = [53 54 55 62 63 71 72 73 81 82 90 91 92 99 100];
- Cluster{10} = [78 79 80 86 87 88 89 98];
- Cluster{11} = [83 84 85 93 94 95 96 101], and;
- Cluster{12} = [97 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118];

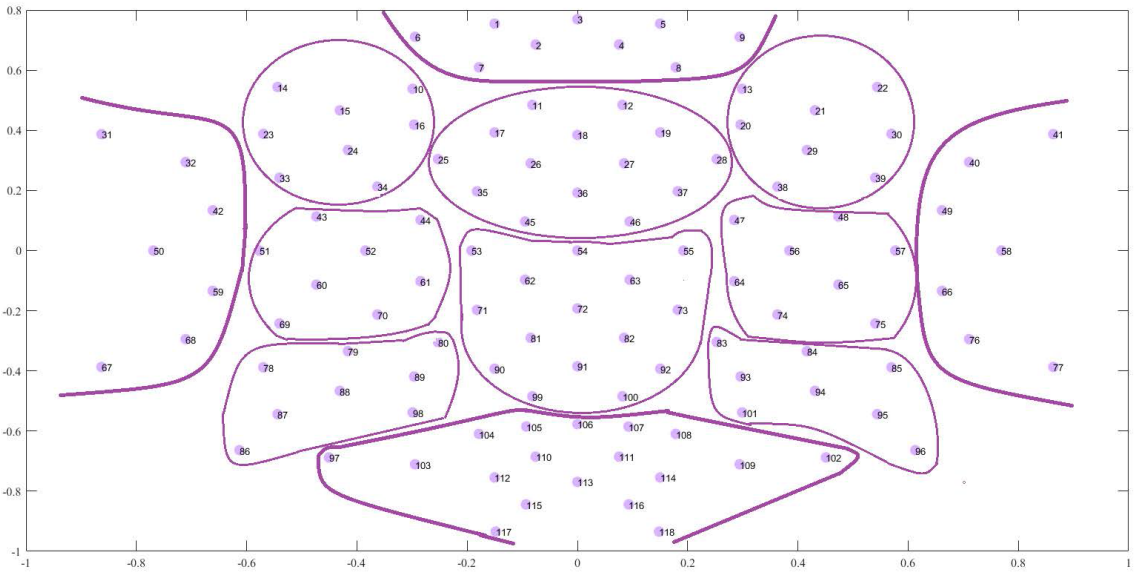


Figure 3.3: Sparsity pattern of the EEG electrodes (2-dimensional projection).

whereas each cluster is detached and defined separately, each of them can be considered as an isolated graph, $\boldsymbol{G}_{(c_i)} = \{\boldsymbol{V}_{(c_i)}, \boldsymbol{A}_{(c_i)}\}$, for $(1 \leq i \leq 12)$. The adjacency matrix for the i-th cluster is formed as a square matrix of Euclidean distances the nodes of the i-th cluster see each other from. It is worth mentioning that at the beginning of this work, a general adjacency matrix was considered to model the connectivity of each electrode to all others within the structure of the headset. However, the accuracies were not satisfying, as such the results are not reported in the thesis.

3.3.3 Total Variation Graph and Selection of best Clusters: Dimensionality Reduction

In Reference [63] an interesting concept is introduced referred to as the Total Variation Graph. This concept denotes the signal variations on graphs and is formulated as follows.

$$\mathrm{TV}_G(\boldsymbol{s}) = \frac{1}{\|\boldsymbol{s}\|_2^2} \left\| \boldsymbol{s} - \frac{1}{|\lambda_{\max}|} \boldsymbol{A}\boldsymbol{s} \right\|_2^2. \tag{21}$$

Intuitively speaking, Eq. (21) evaluates the variation of the graph from a time sample to another. Term $\boldsymbol{s}$ is the vector of values produced from all the nodes of the graph at a certain time sample; Term $\boldsymbol{A}$ is the aforementioned adjacency matrix, and; $|\lambda_{\max}|$ is the largest-magnitude eigenvalue of

matrix $\boldsymbol{A}$, which satisfies the condition $|\lambda_{\max}| \geq |\lambda_m|$, for $(0 \leq m \leq N_{\text{ch}} - 1)$. We employed this concept to extract the features and feed a LDA classifier with the features generated via Total Variation Graph. In other words, the set of data collected from each cluster of electrodes was separately fed to Total Variation Graph model (Eq. (21)), and the vector of features for each trial (epoch) would be from the size of (number of time samples $\times$ 1).

The same sub-steps of feature extraction and classification are carried out for all the clusters, and the accuracies of the classification are then compared to each other. The two clusters with maximum accuracy are selected and their electrodes are used for the final feature extraction (through CSP) and classification with three different kernels of classifiers.

It is important to note that this step of this framework has two main superiorities over similar works. First, the dimensionality reduction is subject-adaptive and respects the brain plasticity of subjects, by finding the two clusters matching the most active regions of the brain, in regards to the task the subject was asked to do. Therefore, the proposed approach is not limited to motor imagery and can be employed within the processing kernel of BCIs using other techniques as well. The second advantage is that the proposed approach can be taken to another level of implementation in practical experiments. In the case that the EEG cap would have the location of electrodes, yet not all/enough electrodes would be available to record the data, using steps outlined in Subsections 3.3.2 and 3.3.3, the researchers can find out which regions of the brain are active during the time that the subject is performing the tasks. Then, they can place the electrodes in other regions close to the active area and increase the accuracy of their data collection. This advantage adds flexibility to practical experiments regardless of the technique being used.

3.3.4 Feature Extraction and Classification

As mentioned previously, the feature extraction method is a conventional CSP. The reason for using this technique of feature extraction is the fact that this method is one of the most known and well-regarded techniques of extracting features from 2-class datasets in EEG signal processing. The target of this framework is to keep the non-graph part of the approach simple and as common as possible, so that the effect of the previous two steps could be observed and validated more clearly. As for the classification, two well-known classifiers, i.e., LDA and QDA are used with 5-fold validation

technique.

3.3.5 Simulations

For the purpose of evaluating the proposed graph-based dimensionality reduction (GDR) framework, the dataset IVa from BCI Competition III was employed. This dataset includes the information regarding the location of electrodes with respect to the center of the scalp. An order five Butterworth bandpass filter is used in the pre-processing step to filter each raw of the dataset and then a moving average function is utilized to smoothen the values. Next, the time series were down-sampled by the factor of 10, meaning that only one time sample is selected out of each batch of 10 samples.

As for the feature extraction section, we considered two different scenarios of using two or four eigenvectors to construct the whitening matrix, and the impact of the increase in the size of whitening matrix can be easily noticed by comparing the table of results. Tables 5.1(a)-(c) illustrate the accuracy of classifications of features extracted by a whitening matrix constructed by two eigenvectors, i.e., the ones corresponding to the max and min eigenvalues, with two different classifiers. Table 3.4(a) illustrates the classification accuracies obtained with 100 training samples (50 per class). Table 3.4(b) shows the accuracies for the classifiers trained with only 60 samples (30 per class), which is suitable for BCIs with a limited number of available training trials. Table 3.4(c) shows the classification accuracies with only 200 samples (100 per class), which is a conventional number of training trials for this dataset. Based on these results, it is observed that the proposed approach is significantly and effective for motor imagery classification problem. Besides, improved accuracies through both classifiers clearly reflect the positive impact of our proposed GDR framework. Tables 3.5(a)-(c) exhibit the accuracies for the scenarios of extracting features with a whitening matrix of 4 eigenvectors, where other items are kept similar to the scenario in Table 3.4. These results again elaborate on the effectiveness of the proposed GDR framework and illustrate its great potential for improving the overall motor imagery classification accuracy when coupled with more complex feature extraction methods.

3.3.6 Conclusion

As a novel graph-based dimensionality reduction, (GDR) framework for processing EEG signals is proposed based on functional clustering of EEG electrodes and a collapsing step via a total variation on graphs. The proposed methodology initially forms connectivity neighborhoods and constructs a global block-diagonal graph representation of the EEG channels. By capitalizing on the fact that functionality of different connectivity neighborhoods varies, a collapsing step is introduced based on total variation measures on the graph, to reduce the overall graph model into two functional clusters. The experimental results based on Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to the other existing methods.

3.4 Summary

Throughout this chapter, the methods and techniques suggested in graph theory and graph signal processing were applied to the processing core of an MI EEG-based BCI system. Two frameworks were proposed, GD-BCI and GDR-BCI and the effectiveness of them is well-elaborated via the results obtained from several implementation studies. This completes my research work on developing the theoretical approaches to improve the performance of MI EEG-based BCI systems.

Table 3.4: Accuracy comparison of the proposed GDR framework coupled with two different classifiers and with two CSP features.

(a)

With 100 Training Trials

	CSP with 2 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	84.14	0.43	85.00	53.89	83.98	1.27	87.00	56.11
	GDR + CSP	85.00	1.00	87.00	81.00	81.20	1.76	86.00	82.78
AW	Conventional CSP	91.84	0.43	93.00	66.67	92.06	0.72	94.00	66.11
	GDR + CSP	89.47	0.80	91.00	81.11	89.82	1.12	93.00	83.89
AV	Conventional CSP	76.08	1.35	80.00	52.22	67.86	0.20	73.00	54.44
	GDR + CSP	72.30	0.75	74.00	64.44	72.72	0.17	77.00	65.00
AL	Conventional CSP	94.55	0.55	95.00	86.11	93.94	0.79	96.00	88.33
	GDR + CSP	86.14	0.99	89.00	91.11	86.75	1.32	90.00	91.67
AA	Conventional CSP	78.54	1.14	82.00	51.11	73.43	1.25	77.00	50.56
	GDR + CSP	75.60	1.09	79.00	64.44	75.21	1.77	79.00	66.67

Average	Conventional CSP	62.00	**Average**	Conventional CSP	63.11
	GDR + CSP	76.42		GDR + CSP	78.00

(b)

With 60 Training Trials

	CSP with 2 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	88.71	1.48	91.67	53.64	88.52	1.36	91.67	55.00
	GDR + CSP	59.95	2.47	65.00	50.91	54.68	1.45	58.33	47.27
AW	Conventional CSP	85.86	1.87	90.00	47.27	77.48	1.37	81.67	47.73
	GDR + CSP	82.67	0.82	83.30	55.91	81.58	1.68	85.00	55.00
AV	Conventional CSP	82.69	1.70	86.67	56.36	74.99	2.95	81.67	55.91
	GDR + CSP	82.67	0.82	83.30	55.91	62.66	1.88	68.33	54.09
AL	Conventional CSP	80.20	1.52	85.00	49.55	79.12	2.16	85.00	57.27
	GDR + CSP	85.97	1.01	88.33	89.55	90.94	1.34	93.33	91.82
AA	Conventional CSP	81.32	1.49	85.00	56.36	76.03	1.34	80.00	54.55
	GDR + CSP	78.71	1.58	81.67	71.36	75.99	1.54	80.00	68.64

Average	Conventional CSP	52.64	**Average**	Conventional CSP	54.09
	GDR + CSP	64.73		GDR + CSP	63.36

(c)

With 200 Training Trials

	CSP with 2 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	74.66	0.44	76.00	56.25	71.55	0.75	73.50	52.50
	GDR + CSP	85.96	0.69	88.00	86.25	82.09	0.28	83.00	85.00
AW	Conventional CSP	67.96	0.52	70.00	45.00	68.20	1.32	70.50	47.50
	GDR + CSP	88.01	0.42	89.00	88.75	88.21	0.44	89.00	88.75
AV	Conventional CSP	66.56	0.88	68.50	50.00	55.74	1.03	59.50	50.00
	GDR + CSP	72.31	0.88	74.50	65.00	71.25	1.48	75.50	63.75
AL	Conventional CSP	93.81	0.43	94.50	90.00	91.53	0.56	93.00	91.25
	GDR + CSP	91.84	0.32	92.50	93.75	92.42	0.21	93.00	95.00
AA	Conventional CSP	78.97	0.64	81.50	57.50	65.69	0.67	67.50	55.00
	GDR + CSP	78.94	0.81	81.00	68.75	78.31	0.95	81.00	68.75

Average	Conventional CSP	59.75	**Average**	Conventional CSP	59.25
	GDR + CSP	80.50		GDR + CSP	80.25

Table 3.5: Similar to Table 5.1 except that four CSP features are utilized; (a) with 100 training trials; (b) with 60 training trials, and; (c) with 200 training trials.

(a)

With 100 Training Trials

	CSP with 4 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	95.08	0.88	97.00	47.78	94.46	1.00	97.00	57.22
	GDR + CSP	83.02	1.40	86.00	85.00	83.80	1.35	88.00	80.00
AW	Conventional CSP	96.04	0.85	98.00	51.67	96.32	98.00	0.66	68.33
	GDR + CSP	94.18	1.17	96.00	78.33	94.17	0.7.	96.00	82.22
AV	Conventional CSP	78.91	1.63	84.00	51.11	78.46	2.18	85.00	53.89
	GDR + CSP	70.79	1.66	76.00	65.00	69.59	2.15	76.00	66.67
AL	Conventional CSP	96.90	0.36	98.00	53.89	97.38	1.15	99.00	87.75
	GDR + CSP	92.35	0.58	94.00	93.89	90.57	1.06	92.00	95.56
AA	Conventional CSP	86.80	1.02	89.00	49.44	85.49	1.31	89.00	45.56
	GDR + CSP	81.39	0.92	84.00	68.89	81.11	1.12	84.00	65.56

Average	Conventional CSP	50.78	**Average**	Conventional CSP	62.55
	GDR + CSP	78.22		GDR + CSP	78.00

(b)

With 60 Training Trials

	CSP with 4 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	94.43	0.86	96.67	45.45	94.93	0.33	95.00	52.27
	GDR + CSP	84.02	2.17	90.00	82.27	82.60	2.65	88.33	77.27
AW	Conventional CSP	93.96	0.83	95.00	47.73	91.08	1.41	95.00	45.00
	GDR + CSP	83.66	2.25	88.33	58.64	87.88	1.36	91.67	55.91
AV	Conventional CSP	85.49	2.48	91.67	52.73	87.62	2.64	91.67	51.82
	GDR + CSP	75.06	1.86	78.33	65.00	70.38	2.18	75.00	55.00
AL	Conventional CSP	90.78	1.59	95.00	45.00	87.92	3.20	95.00	55.45
	GDR + CSP	94.60	0.89	95.00	89.55	92.49	1.75	95.00	92.73
AA	Conventional CSP	86.89	1.82	90.00	47.27	84.49	1.49	88.33	52.27
	GDR + CSP	71.97	2.53	78.33	76.36	69.54	2.70	75.00	73.64

Average	Conventional CSP	47.64	**Average**	Conventional CSP	51.36
	GDR + CSP	74.36		GDR + CSP	70.91

(c)

With 200 Training Trials

	CSP with 4 Eigen Vecors	LDA				QDA			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Conventional CSP	94.49	0.67	96.50	45.00	93.12	0.71	94.50	73.75
	GDR + CSP	86.34	0.73	88.00	86.25	86.00	0.62	87.50	90.00
AW	Conventional CSP	92.76	0.35	93.50	57.50	91.60	0.62	93.00	88.75
	GDR + CSP	88.34	0.52	89.50	80.00	86.58	0.97	89.00	80.00
AV	Conventional CSP	73.49	1.33	76.50	51.25	67.85	1.97	72.00	47.50
	GDR + CSP	76.76	0.98	79.00	67.50	74.63	1.30	78.50	63.75
AL	Conventional CSP	94.91	0.49	96.00	50.00	93.05	0.29	94.00	95.00
	GDR + CSP	92.04	0.19	92.50	93.75	92.31	0.60	93.50	97.50
AA	Conventional CSP	84.19	0.72	86.00	50.00	75.65	0.71	78.00	61.25
	GDR + CSP	81.27	0.59	83.00	73.75	79.40	0.91	81.50	72.50

Average	Conventional CSP	50.75	**Average**	Conventional CSP	73.25
	GDR + CSP	80.25		GDR + CSP	80.75

Chapter 4

Practical Solutions to Improve the Performance of EEG-based BCI systems

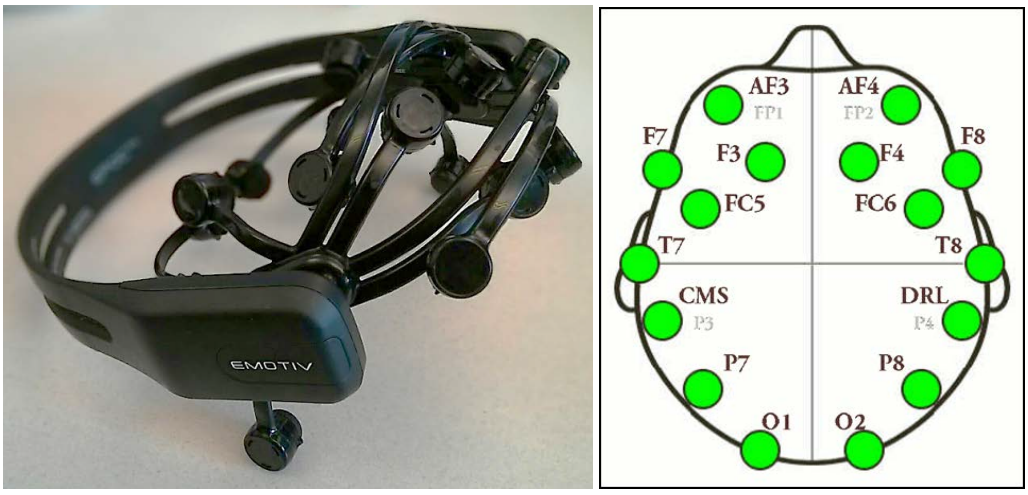
As stated in Chapter 2, the techniques and methods proposed in theory to serve the end goal of advancing the BCI systems, as well-structured and proven to be highly effective as can be, are not the sole approach mitigate the real-world problems that BCI systems struggle with during implementation and utilization. Therefore, should a BCI researcher desire to truly move the edges of knowledge in this field, her/his primary task is to get involved in practical experiments and implementations of BCIs, however basic and pilot, while developing the theoretical methods and more complicated approaches of these systems. This way, challenges such as the comfort of the subject/patient/end-user of the BCI, the ethical issues, the feasibility and practicability of a designed system, and so many other similar aspects of implementing a BCI system will have a real sense and substantial meaning to the aforesaid researcher. Motivated by this outlook, this chapter is allocated to the practical experiments I carried out for my thesis and the solutions provided and tested. The first solution is *Progressive Fusion of Multi-rate MI Classification for BCIs* to address the issues arose in the case of a limited number of training trials at hand. The second solution is *Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs* to readjust the recorded epochs in a manner that most informative parts of the signals are extracted and the segments of the epochs which do not include the response of the subjects to the stimuli would be discarded.

4.1 Progressive Fusion of Multi-rate MI Classification for BCIs

The section presents a practical implementation of an EEG-based BCI system developed based on the Emotiv EPOC headset [64]. The focus of this approach is on development/implementation of a synchronous BCI system, i.e., EEG signals are analyzed during pre-defined periods of time, initiated by an interface. The objective is to research promising ideas in the design, development, and implementation of signal processing technologies that contribute to the advancement of robust, real-time, and adaptive EEG-based BCI system. In particular, the developed EEG-based framework consists of two filters running in parallel namely: (i) *The Progressive Filter*: An efficient filter that performs both feature extraction and classification (CSP is employed) steps based on the set of all arriving epochs to re-train progressively over time. (ii) *The Active Filter*: A simplified CSP-based feature extraction approach running online based on pre-trained classifiers, i.e., a lighter version of the Progressive Filter that runs faster than its counterpart. After each trial and during the rest period, the Active Filter produces the classification results and communicates its decision to the next processing module in the BCI pipeline (e.g., a connected Arduino microcontroller). In the meantime, the Progressive Filter incorporates new EEG epochs to adopt and re-train. Once the Progressive Filter is trained, it checks its output with the Active Filter and if the two are not in consensus, it updates the models of the Active Filter, i.e., the coupling of the two filters to improve active classification performance. The proposed framework is evaluated both based on dataset IVa from the BCI competition III, and through real data collected via the Emotiv Epoc headset.

4.1.1 Progressive Multi-rate MI Classification Outline

In an EEG-based BCI, as explained previously, several channels (sensors) are used to pick the potential differences detected at the scalp produced from combined activity of several neural impulses. Once the signal is recorded, the next steps involve pre-processing the recorded data, extracting discriminating features, and training a classifier to detect different MI classes, similar to the approaches above. The following framework is motivated by the fact that throughout a session of EEG recording, as a cognitive process, subjects become habituated towards the stimuli triggering



(a) (b)

Figure 4.1: (a) Emotiv Epoc headset. (b) Electrode placement and activation.

their response. Therefore, in the case that there is a limited number of training trials available to initiate the classification model, the BCI system fails to adapt itself to the ongoing changes happening in the patterns of data. Having that said, the proposed framework consists of two parallel pathways for classification, among which, one of them functions as an observer. The observer takes action once the recorded epochs can improve the performance of the previous main classifier.

The proposed progressive and multi-rate classification framework is developed based on the CSP feature extraction method and consists of two filters, referred to as the Active Filter and the Progressive Filter, which are partially coupled at the consensus epochs based on the individual classification results. The Active Filter is a simple MI classification feature extraction method (e.g., with two eigenvectors taken into account for the construction of whitening matrix) which forms the CSP features and uses a pre-trained classification model to assign them to the two MI classes. The Active Filter produces results at the end of each epoch while the Progressive Filter uses several consecutive epochs for re-training and adaptation. In other words, after each trial and during the rest period, the Active Filter produces the classification results and communicates its decision to the next processing module in the BCI pipeline (e.g., a connected Arduino microcontroller). In the meantime, the Progressive Filter, in an offline fashion, performs the processing, adaptation, and re-training tasks and it couples its models with the Active Filter in case the two filter is not in

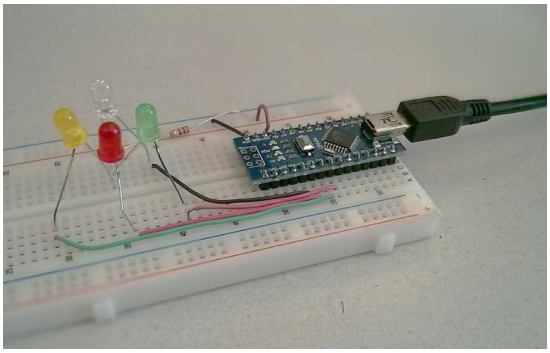


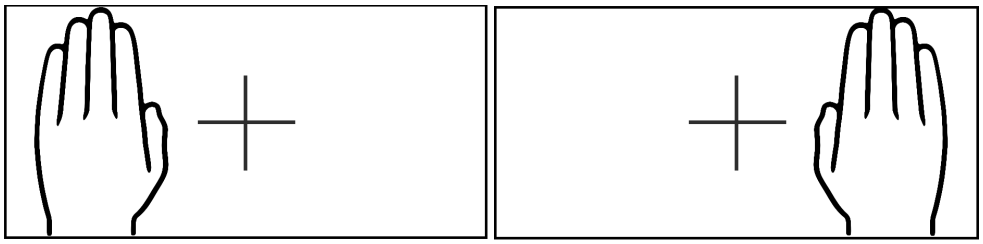
Figure 4.2: Arduino microcontroller used to turn BCI classifications into external actions like moving motors or controlling LEDs.

consensus. This completes the outline of the proposed multi-rate MI classification framework, next, the development of the experimental setup and report results obtained from the implementation of the proposed framework will be discussed.

4.1.2 Experimental Implementation of the Multi-rate MI Classification

In this subsection, the proposed progressive and multi-rate MI classification framework and its practical implementation using the Emotiv Epoc headset is further elaborated. Fig. 4.1 illustrates the headset and electrode placements. EMOTIV company describes their headset as a high resolution, multichannel, portable system designed for practical research applications [64, 66]. It has 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) and two reference points (P3 and P4). Sampling is done sequentially at 2048 Hz internally, but the signal sent to the computer is already processed and reduced to 128 Hz with a resolution of 14 bits (1 LSB = 0.51V), a bandwidth of 0.2 - 45Hz, and filtered with digital notch filters at 50Hz and 60Hz. The built-in filters are digital 5th order sinc filters. The input dynamic range is 8400 μV. The connection to the computer is wireless over the 2.4GHz band, and uses a proprietary protocol. The headset itself is wireless and powered with a LiPoly battery with 12 hours working time per full charge. Signal quality is determined internally with the proprietary system and impedance measurements. Several recent research studies [67, 68] used the Emotiv headset reporting promising results which were the motivation to use it to implement and develop the proposed framework as described further on.

Several online and offline experimental scenarios were performed to develop/implement the



(a) (b)

Figure 4.3: (a) The visual stimulus shown to the subject as 'Left Hand'. (b) The visual stimulus shown to the subject as 'Right Hand'.

proposed multi-rate classification framework. In the first scenario reported below, I briefly outline different experimental setups developed/implemented and observations made that led to the setup used in Scenarios 2 and three where the proposed framework was implemented/tested.

Scenario 1: In the experiments, originally, MI was considered during trials that last 3-5 seconds with 1-3 seconds break between trials and then the following two classes were used: (i) Forward arrow $\rightarrow$ stimuli provided to imagine lifting one's right hand, and; (ii) Backward arrow $\leftarrow$ stimuli provided to imagine lifting one's right foot continuously during each trial. At first, a total of 200 trials was performed by each subject and data was collected with the headset tilted back to approximately cover C3 and C4 motor cortex area as suggested by [69]. From the feedback obtained from the subjects, it is concluded that the experiment was too long and headset placement was also uncomfortable. Therefore, we reduced the total trials and performed two sessions of 30 trials with an intermission to make the experiment easier on the subjects. The headset was also worn normally and the experiments were done in the middle of an empty conference room, far from other electronics (other than the laptop). Besides, feedback (shown in Fig. 4.2) was tried out by the subject, using an Arduino microcontroller which is used to turn BCI classification results into the external lighting of different-colored LEDs. This feedback is removed later on to keep the setup simpler. Finally, the classes were changed to "left and right hand" movements. The primary attempt, originally, was to avoid this to prevent dyslexic confusion of the stimuli, yet, the undertaken effort was to simplify the cognitive process such that the cues are more simply linked with the action. As a result, the stimuli were chosen as black and white line drawings of a left or right hand offset on the left or right side of a central cross, Fig. 4.3.

Table 4.1: Performance of different models based on real experimental data sets.

	Model 1	Model 2	Model 3	Model 4	Model 5
CSP(2d) LDA	52.00%	62.50%	66.67%	50.00%	20.00%
CSP(2d) QDA	54.00%	60.00%	70.00%	55.00%	30.00%
CSP(4d) LDA	48.00%	47.50%	50.00%	60.00%	80.00%

Scenario 2: Before applying the proposed framework to real datasets collected from Emotiv headset, in this scenario, the proposed algorithm was used to classify BCI Competition III IVa datasets. The experiment is performed based on the following steps: (i) Raw data is filtered to remove the DC gain and to pick the information within 7-30Hz. Thereafter, the filtered data is chopped into epochs and then smoothed using the weighted moving average method by a window size of 10 time samples. After this step, the dataset is downsampled to keep one sample out of each batch of 10; (ii) The CSP method is used to extract four features, M first trials were selected to train the classifier; (iii) A Quadratic Discriminant (QD) classifier by 10-fold cross validation is trained; (iv) The trained classifier is used by the Active Filter to classify incoming epochs, as the Progressive Filter was collecting the new epochs to update and re-train itself; (v) After L epochs, the Progressive Filter, re-trained by $M + L$ trials and fused with Active Filter to be used as the new Active Filter for classifying the next $L1$-epochs. Steps (ii) to (v) are repeated for each L-epoch batch under the condition of improving the classification accuracy. As an observation, the dataset of subject "aa" in BCI Competition II-Iva, was evaluated by the aforementioned algorithm, by setting $M = 90$ and $L = 20$. The QD classifier is trained by the accuracy of 93.3% and classifies the next 20 epochs, as the Active Filter, by the accuracy of 55%. Thereafter, Progressive Classifier is updated using 110 epochs, trained classier's accuracy is 98.2%, and is fused with the Active classifier, which classifies the next 30 epochs by the accuracy of 60%. The same procedure is done to train the Progressive Filter by the accuracy of 93.6% using 130 epochs. The updated Active Filter then classifies next 40 new epochs by the accuracy of 65%. Scatter plots of these two Progressive Filters are shown in Fig. 4.4.

Scenario 3: In this scenario, the same proposed framework is implemented based on the experimental setup described in Scenario 1. The first ten epochs are incorporated to train a model, hereafter referred to as Model 1, using CSP feature extraction, and LDA model for classification. Only two

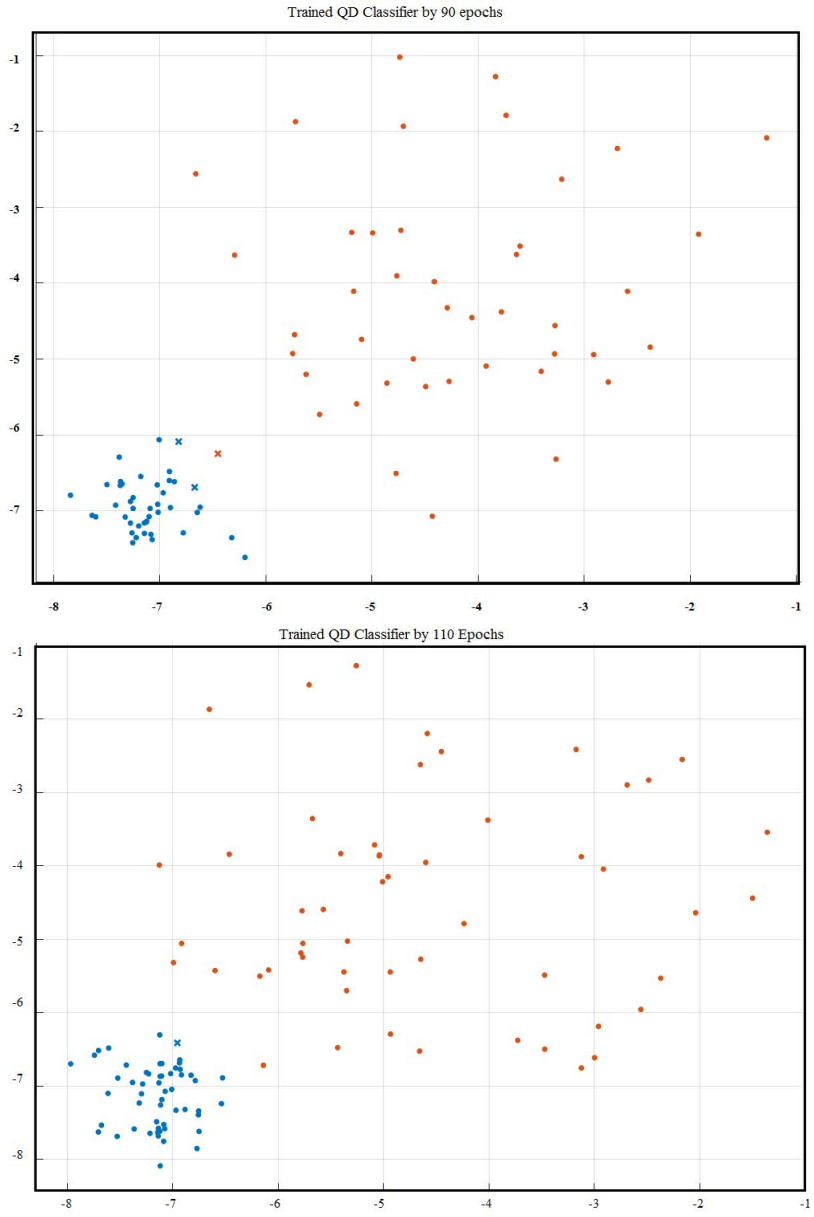


Figure 4.4: Scatter plots obtained from two Progressive Filters trained based on 90 and 110 epochs, respectively.

dimensions of the CSP are used. The resulting trained model is employed by the Active Filter for online analysis of the following epochs, i.e., categorizing them as they come in. While the Active Filter performs online classification based on Model 1, the Progressive Filter uses the next ten epochs, i.e., it trains another model (referred to as Model 2) based on twenty epochs and after the 20th epoch starts classifying in parallel to the Active Filter which uses Model 1. This process was repeated

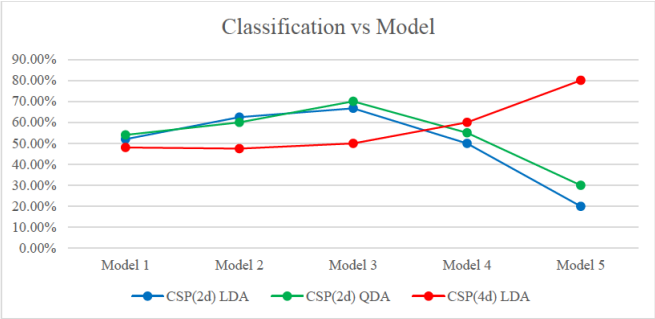


Figure 4.5: Classification results obtained from implementation of the proposed progressive and multi-rate framework based on data collected via Emotiv headset.

until Model 5 is trained. Fig. 4.5 and Table 4.1 illustrate the results. It is observed that using more training data could result in providing improved performance. In practice, the online model (the Active Filter) should be updated to the best performing model available at the Progressive Filter. The same procedure was repeated by using a QDA classifier, and then again for both LDA and QDA using four dimensions from the CSP. The results show a peak return from training in LDA and QD for 2d CSP, before more training data actually decreases the model's performance. Although 4d CSP with LDA starts out with lower performance as compared to its counterparts, it produces improving results as more epochs are taken into account. These results are significantly promising with further improvement/investigation which is one of the objectives of our future research in this direction.

4.1.3 Conclusion

In this section, a novel progressive and multi-rate EEG-based MI classification framework is proposed, which consists of two separate filters. The Emotiv Epoc headset is used to develop and implement the proposed framework. The two filters (referred to as the Active Filter and the Progressive Filter) are partially coupled at the consensus epochs based on the individual classification results, i.e., the Active Filter (a simple MI feature extraction algorithm which uses pre-trained classification models) produces results at the end of each epoch while the Progressive Filter uses several epochs to re-train. The experimental implementations of the proposed framework indicate its potential for improving the performance of real-time MI classification.

4.2 Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs

As previously stated, one of the most popular and commonly used techniques to satisfy the requirement for an effective and efficient BCI is Motor Imagery, which is defined as mere imagination of a limb movement, with no actual movement or peripheral (muscle) activation. However, as much as this field outlines a promising framework, the researchers dealing with MI commonly face two types of challenges. The first challenge is to deal with **different comprehension of subjects from "imagination of the movement"**. Some subjects imagine repeating the movement during each epoch, while some others might execute the mental imagination of the activity only once, and not necessarily within consistently equal time intervals after the stimulus is shown. In order to tackle this obstacle of various reactions, implementation of methods in which the classifier is trained subject by subject has been adopted by the researchers of the field. These approaches are adaptive to the nature of the datasets collected, for instance, should the subject react to the stimulus right after he/she sees it, the classifier of the dataset collected from this subject is trained to read the epochs' information right after the marker. The second challenge is the fact that through a cognitive process, the brain of the subject learns to decrease the motor concentration while doing the same task. Hence, the amplitude of the signals within each epoch tends to descend over time. As a clarification to the case propounded, an actual example and the observation corresponding to that follows. Through experiments done for the research work described in the previous section, the subjects were asked to fill in surveys after they participated in each session of the experiment. The outcome of these surveys was critical: it was difficulties for them to keep their concentration all through the long experiments and they were not able to consistently do the tasks they were instructed for. The experiment was extremely tiring on the subjects, and, for instance, 100 trials would be too long to keep the subjects focus. The substantial difference of their performance was proved once the long experiments were split into sets of runs with much less number of trials: through feedbacks and computational analysis on the data collected from them, much better performance and more informative datasets were observed. Altogether, as the latency of the overall system (from the headset, wireless communication and software) is unknown, and more importantly, delay in human response

is inevitable, this question was raised: *where it is best to start time sample within the time samples of each epoch?* To answer this question, the following approach was designed and tested.

4.2.1 Trimming Framework Outline and Simulation

To answer the raised question regarding the most informative time interval within each epoch, straightforward processing was considered. As the main goal was to investigate whether or not a trimmed epoch would contribute to better classification accuracy, the following modules were considered.

- **Preprocessing**: Raw dataset of BCI Competition III-IVa were filtered via an order five *Butterworth* filter, to remove the DC gain and to pick the information within 7-30Hz. Thereafter, the epochs were arranged and then smoothed using the weighted moving average method by a window size of 10 time samples. Then, the dataset is downsampled to keep one sample out of each batch of 10. Finally, within each epoch, there are 350 time samples.

- **Feature Extraction**: The CSP method is used to extract features. Two eigenvectors are taken for the purpose of construction of the whitening matrix. Also, in respect to the rule of thumb mentioned in Chapter 2, both scenarios of splitting the dataset into training and test trials are implemented: once with 60% of the trials for training, which is 168 trials out of 280, and the remaining 112 trials were tested by the trained classifier. Another run, 196 trials (70% of the trials) were put aside for training the classifier and that leaves 84 trials in the test dataset.

- **Classification**: Both LDA and QDA models with 5-fold cross validation were utilized to classify the test sets.

These three steps, without any dimensionality reduction technique applied or complexity in the algorithm were deliberately adopted for the sole purpose of enabling a better exhibition of the effects of the trimming step, which comes before feature extraction. The trimming step is elaborated in Algorithm 3.

The accuracies achieved from implementation of this approach are presented in Tables 4.2. As the tables and the Fig. 4.6 illustrate, trimming step impressively boosts the performance of the conventional CSP-based algorithm.

Algorithm 3 TRIMMING THE EPOCHS

Input: $\{$*Original EEG signals* $\boldsymbol{X}\}$

Output: $\{$*EEG signals* $\hat{\boldsymbol{X}}$ *which epochs are trimmed based on the best time of start* $\hat{t}\}$

1: **Sampling the trainset**: Half of the $\boldsymbol{X}_{train}$ is considered as a sample of the training set $\bar{\boldsymbol{X}}_{train}$, to investigate the best time of start.

2: **Best Time of Start Loop**:

Step = number of time samples corresponding 0.1s

For: $\hat{t}$ = 1:(Number of time samples between time = 0 to 1s)/Step

• Epochs of $\boldsymbol{X}_{train}$ are adjusted as $\bar{\boldsymbol{X}}_{train}$ = $\boldsymbol{X}_{train}$(:, $\hat{t}$:end, :);

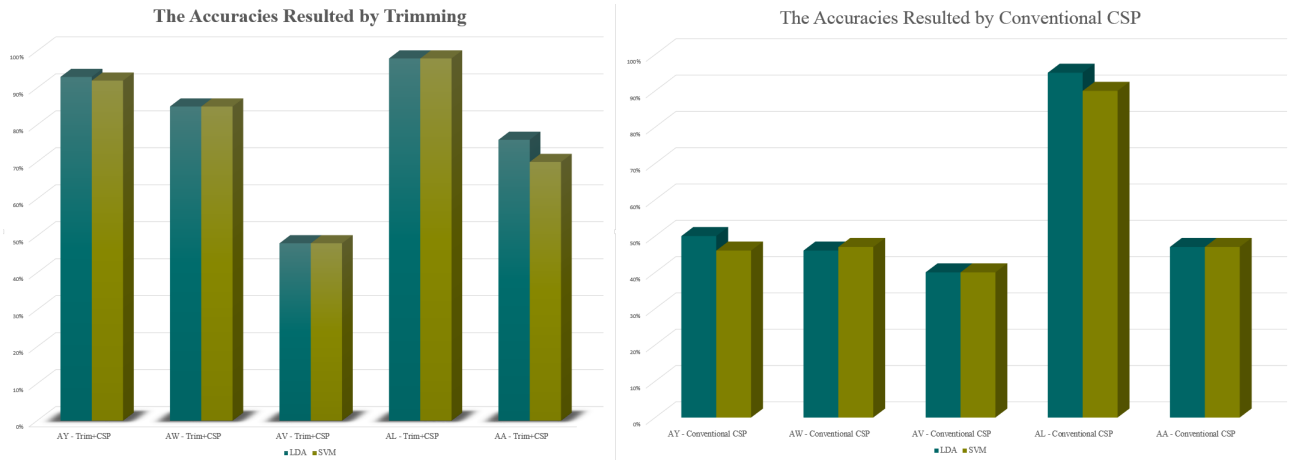
• CSP is applied to $\bar{\boldsymbol{X}}_{train}$ and then LDA is trained and the training accuracy is stored.

End For.

3: **Finding the best** $\hat{t}$: The stored accuracies are evaluated, the $\hat{t}$ corresponding to the best $\hat{t}$ determines $\hat{t}_{best}$.

4: **Preparing the final** $\bar{\boldsymbol{X}}$:

Final $\bar{\boldsymbol{X}}$ = $\bar{\boldsymbol{X}}$(:, $\hat{t}_{best}$:end, :);



(a) (b)

Figure 4.6: (a) The accuracies achieved via adding the trimming step to the conventional CSP algorithm. (b) The accuracies achieved via the conventional CSP algorithm.

4.2.2 Conclusion

The proposed algorithm aims to help BCI developers who focus on MI EEG-based systems to better evaluate and estimate the intentions of the subject. The accuracy of classification in MI EEG-based BCIs can be significantly enhanced through trimming the epochs. This method is specifically useful for those BCIs used for Locked-in patients due to the way it enriches the processing algorithm and takes the burden of performance concentration off the patients' shoulders. Implementation of

this algorithm is highly recommended for any EEG-based study in order to prevent the possible exhaustion of the subjects of the study. The proposed trimming mechanism significantly outperforms its counterpart regarding achievable performance.

4.3 Summary

Throughout this chapter, the practical experiments I carried out for my thesis and the solutions provided and tested were elucidated in details. The first solution was "Progressive Fusion of Multi-rate MI Classification for BCIs" which aimed to address the issues arose in the case of a limited number of training trials available at the initiation phase of a BCI. The second solution was "Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs" during which an additional step was suggested in order to trim the recorded epochs in a manner that most informative parts of the signals are extracted and the segments of the epochs which do not include the response of the subjects to the stimuli would be discarded. Both frameworks show notable and impressive impact on the performance of MI EEG-based BCI systems. This completes the discussion on BCI systems and frameworks to enhance the end results of such systems.

Table 4.2: Accuracy comparison of the proposed trimming framework, (a) The accuracies for classifier trained with 168 training trial, and; (b)The accuracies for classifier trained with 196 training trial .

(a)

Subjects	168 Train (60% - 40%)	LDA				SVM			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Trim	91%	0.5	92%	93%	94%	0.6	95%	92%
	CSP	65%	0.7	67%	50%	66%	1.7	70%	46%
AW	Trim	93%	0.2	94%	85%	93%	0.4	94%	85%
	CSP	59%	0.8	62%	46%	60%	0.9	62%	47%
AV	Trim	60%	0.9	62%	48%	58%	1.0	61%	48%
	CSP	57%	0.7	59%	40%	56%	1.8	60%	40%
AL	Trim	96%	0.1	97%	98%	96%	0.2	97%	98%
	CSP	93%	0.4	94%	95%	95%	0.4	96%	90%
AA	Trim	88%	0.5	90%	76%	90%	0.5	90%	70%
	CSP	0.68	0.6	70%	47%	0.68	1.0	71%	47%

Average			Average		
	Trim + CSP	80%		Trim + CSP	66%
	Conventional CSP	56%		Conventional CSP	54%

(b)

Subjects	196 Train (70% - 30%)	LDA				SVM			
		Train			Test	Train			Test
		Mean	STDV	Best		Mean	STDV	Best	
AY	Trim	90%	0.5	92%	92%	95%	0.4	95%	93%
	CSP	65%	0.6	66%	55%	66%	1.5	69%	51%
AW	Trim	94%	0.2	95%	93%	94%	0.6	96%	93%
	CSP	94%	0.3	95%	90%	93%	0.4	94%	92%
AV	Trim	61%	0.6	63%	40%	63%	1.3	66%	39%
	CSP	58%	0.8	61%	46%	58%	1.0	60%	46%
AL	Trim	97%	~0	97%	100%	97%	0.3	97%	100%
	CSP	93%	0.3	94%	98%	95%	0.3	95%	94%
AA	Trim	88%	0.3	89%	68%	89%	0.4	90%	63%
	CSP	0.71	0.4	72%	52%	0.7	1.0	72%	56%

Average			Average		
	Trim + CSP	78%		Trim + CSP	78%
	Conventional CSP	68%		Conventional CSP	68%

Chapter 5

Applications of Hybrid BCIs and WAKE-BPAT Framework for Blood Pressure Estimation

5.1 Introduction to Hybrid BCIs

When we talk about 'what the user wants', the discussion is usually about a direct *control* of an application, in other words, the question is 'how can a user control the environment and make changes to it intentionally, in a real-time fashion?'. Understandably, the primary requirement for realizing such a purpose is for the user to be able to interact with the environment. Traditional interfaces making this desire come true were devices such as keyboards, joysticks, and remote controllers, however, nowadays intentional interactions with the environment is feasible through more human-like interaction modalities, including gestures, facial expressions, gaze behavior, body language, and physiological information, via which user is enabled to control an environment or an application. As stated before, BCI systems utilize neurophysiological signal to establish direct communication between human brain and computing devices without the involvement of neuro-muscular pathways. Performance of a non-invasive BCI is highly dependent on the separability of cognitive task related features extracted from cerebral electrophysiological signals (e.g., EEG)

acquired during psycho-cognitive tasks undertaken by the user. The EEG has highly non-stationary and non-linear dynamic characteristics and, therefore, current BCI systems suffer from limited MI task detection accuracy. Nevertheless, the successful real-time (online) MI EEG-based BCI systems suffer from the limited number of classes (motory tasks), and this issue is not arose only because of the limitations of the processing methods, but also, because in BCIs based on imagined movements approximately 20% of users do not exhibit BCI performance adequate enough for effective control, a phenomenon called 'BCI illiteracy' [70, 71], this problem has been reported with other major BCI approaches as well.

To bridge the gap described above, novel approaches have been proposed to address these issues in current BCI studies, by combining a BCI system with other system(s) that utilize neurological signals, physiological signals, and/or external signals. This is the birthplace of systems called Hybrid BCIs or hBCIs. A typical hybrid BCI is, (i) composed of one BCI and another system, and; (ii) supposedly, able to achieve specific goals better than a conventional system. A hybrid BCI can either use two different brain signals (e.g. electrical and hemodynamic signals), one brain signal (e.g. EEG) associated with two mental strategies (motor imagery and spatial visual attention, or one brain signal and another input. Such an additional input can be a physiological signal like the electrocardiogram (ECG) or a signal from an external device such as an eye gaze control system [72]. According to [73], there are four criteria for a hBCI to be fulfilled for it to achieve specific goals better than a conventional BCI.

(1) The system must rely on activity recorded directly from the brain;

(2) At least one recordable brain signal, which can be intentionally modulated, must provide input to the BCI (electrical potentials, magnetic fields or hemodynamic changes);

(3) The signal processing must occur online and yield a communication or control signal, and;

(4) The user must obtain feedback about the success or failure of his/her efforts to communicate or control.

Hybrid BCI systems can either have more than one input whereby the inputs are typically processed simultaneously or operate two systems sequentially, where the first system can function as

a “brain switch” or as a “selector”. A brain switch is a BCI system designed to detect only one brain state (brain pattern) in the ongoing brain activity, as the system is then followed by another modality. The purpose of a brain switch is to ensure no output is produced when the user does not intend to communicate.

In the rehabilitation literature review done by Decety *et al.* [74], it is suggested that just MI of exercises (i.e. involving motor tasks) causes significant changes in the cardiopulmonary autonomic responses consisting of alterations in heart rate, blood pressure, respiration rate, and blood oxygen content. This means that central processes, such as, for example, motor preparation, mental simulation, stimulus anticipation and translation, can result in a cardiac response. Therefore, such a cardiac response detected in the ongoing ECG signal, it can be used as either a switch on/off for a BCI, or to follow a brain switch. In respect to this fact, there are several studies proposing effective results of hBCIs using ECG signals. In [75], a hybrid BCI is introduced by combining ECG with EEG, which uses newly developed bispectrum based features and concludes that the fusion of ECG with traditional EEG-based BCI does enhance the performance of resulting hBCI. As an another example, in [76], a P300 speller BCI’s performance is evaluated using the features obtained from EEG and ECG signals. Although some research studies have shown promising results in this field, the exhaustive review of Choi *et al.* in [77], shows that out of 74 journal articles on fruitful hBCIs, only 2 of them are employing ECG as the second modality, therefore, the field of hBCI, and more specifically, ECG-based hBCIs is still so new that no topic area within this domain can be considered mature. Moreover, to this date, the majority of the reported studies in this field are focused on heart rate, as cuff-less blood pressure estimation through ECG signal processing is yet to be more investigated and improved. Motivated by this, I sought for a solution for improvement of cuff-less blood pressure estimation, which can be employed in a hybrid BCI system in future. The framework suggested for this objective is described next.

5.2 WAKE-BPAT: Wavelet-based Adaptive Kalman Filtering for Blood Pressure Estimation via Fusion of Pulse Arrival Times

This work is motivated by recent urgency to design continuous and cuff-less blood pressure (BP) monitoring solutions, not only for the purpose if hBCIs, but more importantly to prevent, detect, and treat the hypertension. In this regard, a novel wavelet-based feature extraction algorithm coupled with an adaptive and multiple-model Kalman filtering framework is proposed, referred to as the WAKE-BPAT. This framework provides accurate and dynamic BP estimates by extraction and fusion of ***different pulse arrival time (PAT)*** features. In particular, a wavelet transform and histogram analysis-based robust and high-accurate R-peak detection algorithm is proposed without incorporation of any pre-defined thresholds. This in combination with high-quality photoplethysmogram (PPG) characteristic points obtained from signal recordings of a recently developed PPG device (Gen-1), are used for BP estimation, which is modeled as a hybrid state-space model with structural uncertainties to fuse different PAT features in an adaptive fashion. Experimental evaluations of WAKE-BPAT based on a real data set collected via Gen-1 device confirms the superiority of the proposed framework in comparison to its counterparts.

5.2.1 Introduction

Blood pressure (BP) is a crucial hemodynamic parameter that varies between two pressure levels in each heartbeat, called the Systolic BP (SBP) and the Diastolic BP (DBP). Hypertension, which is also known as High Blood Pressure (HBP), is defined as a medical condition in which arteries are experiencing a persistently elevated blood pressure, and is the cause for at least 45% deaths due to heart disease, and 51% of deaths due to stroke [78]. The HBP is usually referred to as the silent killer, as it does not show up significant symptoms. However, long-term high blood pressure is a principal risk factor for coronary artery disease, stroke, heart failure, peripheral vascular disease, vision loss, and chronic kidney disease [79]. An individual is called Hypertensive, if their SBP or DBP reaches more than 140 or 90 mmHg respectively, at rest [80]. The BP measurements, and in particular, continuous BP measurements are great means of retrieving invaluable information about subjects' health conditions in order to prevent, detect, evaluate, and early start of treatment of

hypertension [81]. Conventionally, cuff-based instruments are used to determine the BP, which are by nature discontinuous means of measurement, time consuming to use, and also cause discomfort and inconvenience in case of many repetitions.

The aforementioned drawbacks of cuff-based BP monitoring have resulted in a recent surge of interest [82–90] to develop novel and innovative signal processing solutions for continuous BP monitoring. A potential surrogate of BP which is able to perform BP measurements non-invasively and continuously is the Pulse Arrival Time (PAT), which is defined as the time for the pulse to travel from the heart to a peripheral site. The PAT is considered as a notably practical solution for ambulatory BP monitoring due to being readily acquired by wearable devices. Ahmad *et al.* [91] showed that a significant correlation exists between the BP and the PAT, however, this correlation depends on several parameters, which vary among different individuals. The end goal throughout this work is to investigate BP estimation through ECG and PPG signals [92], using the PAT method.

Prior to a brief review of previous related works in this field, it is vital to explain the triplet deflections in a typical ECG signal, known as QRS complex. Interpreting the ECG signals includes assessment of the morphology (appearance) of the waves. The basic pattern of the ECG is logical, electrical activity towards an ECG lead causes an upward deflection, electrical activity away from a lead causes a downward deflection, and depolarization and repolarization deflections occur in opposite directions. The QRS complex represents the depolarization (activation) of the ventricles. To be able to detect the Q, R, and S waves (Fig. 5.1), and in combination, the QRS complex of each heartbeat, one must know that the P wave is a small deflection wave that represents atrial depolarization, and appears as the first curve in the signal recorded during a heartbeat, and is then followed by the QRS complex. Normal Q waves represent depolarization of the interventricular septum and appear as a downwards deflection in the signal, followed immediately by a peak, known as R. The R wave reflects depolarization of the main mass of the ventricles, hence, it is the largest wave. Finally, the S wave is any downward deflection occurred right after the R wave. Before the next heartbeat signal follows, the rapid ventricular repolarisation is reflected by a wave named T wave. The graphical appearance and the time intervals related to the QRS complex are well-regarded features, tremendously used for medical diagnosis and also in biosignal processing.

To date, there are numerous proposed methods in the literature for detection of QRS-complex

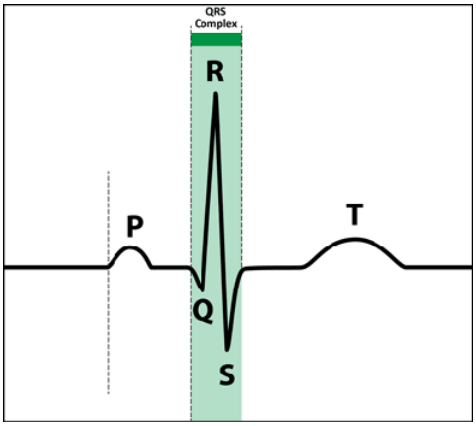


Figure 5.1: The main ECG waves within a heartbeat.

and R-peak. Recently, a derivative and adaptive threshold-based algorithm is proposed by Khamis *et al.* in [93] for the detection of QRS complex. A quadratic filter-based ECG enhancement and QRS detection technique is proposed by Phukpattaranont in [94]. However, the detection performance of such methods is reliant on heuristically determined threshold values, that are either static or dynamic in time or frequency domains. Threshold-based detection approaches, however, are not generally suitable and/or applicable, particularly in presence of in-band noises. On the other hand, in this work, the first-generation (Gen-1) device which is very recently developed by Marefat and Mohseni*et al.* [92] is utilized for collecting PPG signals. The Gen-1 device performs minimally invasive, muscle-based recording of the PPG signal in the reflective mode. Finally, different linear [86] and non-linear models [81, 84] have been developed in the literature to estimate the BP from a computed PAT feature. While most of the model-based BP estimations from PAT are static algorithms in nature, recently dynamic BP estimation via Kalman filtering (KF) [85] is proposed, however, fixed/known parameters are used based on a single first-order scalar Markov model and a single extracted PAT feature.

The research work presented in this section is a novel wavelet-based feature extraction algorithm coupled with an adaptive and multiple-model Kalman filtering framework (referred to as the WAKE-BPAT). As the resulting pattern of the PPG wave obtained from Gen-1 device is of high quality, high signal-to-noise ratio (SNR), and is smooth, a derivative and threshold-based technique is developed for extraction of main features from the PPG signals in the WAKE-BPAT framework. Due to incorporation of state-of-the-art PPG recording system (Gen-1 device [92]), performance of

continuous and automated BP-measurement relies heavily upon the accuracy of the feature extraction algorithm from ECG signal. A Wavelet Transform (WT) and histogram analysis-based robust and highly-accurate R-peak detection algorithm is proposed here. The novelty of the algorithm lies in its accuracy and simplicity. The algorithm does not use any threshold value for the detection of R-peaks. The third contribution of the paper is development of a novel adaptive, and multiple model [95] KF framework for BP estimation, which considers inherit structural uncertainties of the state and observation models, and fuses different PAT features in an adaptive fashion.

The rest of the section is organized as follows: Subsection 5.2.2 formulates the problem. The proposed WAKE-BPAT framework is developed in Subsection 5.2.4. Subsection 5.2.5 presents the experimental results. Finally, Subsection 5.2.6 concludes the paper.

5.2.2 Problem Formulation

As stated previously, the PAT is, typically, derived from ECG and PPG signals. The PPG is a non-invasive measurement technique that measures relative blood volume changes in the blood vessels. On the other hand, the ECG is the graphical representation of the heart's electrical activity, and the QRS-complex (Q, R and S waves are usually treated as a single composite wave known as the QRS-complex) is the most prominent feature of the ECG signal, which as stated previously, provides useful information about the depolarization of ventricular myocardium and indicates the start of ventricular contraction of the heart.

Commonly, the PAT is computed from the time interval between the R-peak of the ECG signal and a characteristic point of the PPG signal. Different features (characteristics points) can be extracted from the PPG signals among which the following three are typically used: (i) The on-set of the PPG; (ii) The peak of the PPG, and; (iii)The peak of the derivative of each PPG cycle i.e, the maximum-slope-point (MSP) of each PPG cycle. Once characteristic points are identified through feature extraction on both the ECG and PPG signals, the next step is to estimate the BP based on the extracted features. The BP estimation task depends on the model used to relate the BP to the

selected time difference such as the following

$$\text{Model 1:} \qquad \text{BP} = \alpha_1 \ln(\text{PAT}) + \beta_1 \tag{22}$$

$$\text{Model 2:} \qquad \text{BP} = \alpha_2 \text{PAT} + \beta_2 \tag{23}$$

$$\text{Model 3:} \qquad \text{BP} = \frac{\alpha_3}{\text{PAT}^2} + \beta_3, \tag{24}$$

where the model parameters are, typically, computed through a calibration step, which is performed based on couple of ground truth points and using least square (LS) approach.

5.2.3 The Kalman Filter

The Kalman filter is essentially a set of mathematical equations that implement a predictor-corrector type estimator that is optimal in the sense that it minimizes the estimated error covariance, when some presumed conditions are met. This method uses a system's dynamics model, known control inputs to that system, and multiple sequential measurements (such as from sensors) to form an estimate of the system's varying quantities (its state) that is better than the estimate obtained by using only one measurement alone. The KF generates an estimate of the state of the system as an average of the system's predicted state and of the new measurement. The estimation is done using a weighted average, as the weights are the means of realizing a certain purpose: the values with better (i.e., smaller) estimated uncertainty are rather more 'trustable'. As the covariance is a measure of the estimated uncertainty of the prediction of the system's state, it is employed for the purpose of computing the weights. The result of the weighted average is a new state estimate that lies between the predicted and measured state, and has a better estimated uncertainty than either alone. This described process is repeated at every time step, with the new estimate and its covariance informing the prediction used in the following iteration. In other words, the Kalman filter works recursively and requires only the last 'best guess' to calculate a new state, rather than the entire history of a system's state.

5.2.3.1. Kalman Filter Described

A brief and comprehensive definition of Kalman Filter which can help better understanding of

the way this filter functions. The KF is known to be an *efficient algorithm* for *estimating* the *state of a linear system* from a series of noisy measurements. This definition shows that KF comprises of two main blocks, *modeling of linear systems*, and *linear recursive estimation*. A discrete-time linear time-invariant dynamical system is a set of matrix equations of the form

$$\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{B}\boldsymbol{u}_i, \tag{25}$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{C}\boldsymbol{x}_i + \boldsymbol{D}\boldsymbol{u}_i. \tag{26}$$

where $\boldsymbol{x}_i$ as the state of the system at time i (which is a natural number), $\boldsymbol{y}_i$ is the output of the system, and $\boldsymbol{u}_i$ is the input of the system, as they are all real numbers. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ are all constants and as they do not change over time i, the system is time-invariant. The question here is how to lead the state $\boldsymbol{x}$ to become 'good', when only $\boldsymbol{y}$ is observed, and the degree of freedom at hand by which we can impact the state is $\boldsymbol{u}$. Obviously, the term 'good' depends on the context of the system; generally, in the field of control engineering, the objective is to drive $\boldsymbol{x}$ to 0. For the sake of simplification, the state-feedback controller is considered, which is easy to design, analyze, and implement, and for this controller $\boldsymbol{D} = 0$ and $\boldsymbol{C}$ is identity, which means that $\boldsymbol{x}$ is directly measured, and by choosing output

$$\boldsymbol{u}_i := \boldsymbol{K}\boldsymbol{x}_i, \tag{27}$$

we have

$$\boldsymbol{x}_{i+1} = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})\boldsymbol{x}_i. \tag{28}$$

However, more often it happens that the direct state is not observable, and only an estimation of the state is available. Thus, the controller's equation changes to

$$\boldsymbol{u}_i := \boldsymbol{K}\hat{\boldsymbol{x}}_i, \tag{29}$$

where $\hat{\boldsymbol{x}} = \boldsymbol{\xi}(\boldsymbol{y})$ is the estimator of the actual output. In real-world applications, however, the system is usually pushed around by random variables that stand for the noise that the linear system

is exposed to. Therefore, the system's model is revised as follows.

$$\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{B}\boldsymbol{u}_i + w_i, \tag{30}$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{C}\boldsymbol{x}_i + v_i, \tag{31}$$

where w_i and v_i are the random variables presenting the noise, often with a normal distribution.

The aforementioned adjective *efficient* for KF comes from the second block of its structure: linear recursive estimation. The linear recursive estimator estimates a constant vector from noisy measurements. In other words, the sequence of measurements as

$$\boldsymbol{y}_{i+1} = \boldsymbol{C}_i\boldsymbol{x} + v_i \tag{32}$$

are observed, the target is to estimate the $\boldsymbol{x}$. The measurement matrix $\boldsymbol{C}_i \in \mathbb{R}^{m\times n}$ and $\boldsymbol{R}_i := \boldsymbol{Cov}(v_i) \in \mathbb{R}^{m\times m}$ are known, and $\hat{\boldsymbol{x}}_i$ is the estimate of $\boldsymbol{x}$ after i measurements. Given a $\boldsymbol{y}_i$, the estimate is updated from $\hat{\boldsymbol{x}}_{i-1}$ to $\hat{\boldsymbol{x}}_i$. This is were the efficiency of this method is signified: the update occurs via Eq. (33) with no need to storing $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, ..., $\boldsymbol{y}_{i-1}$, or $\hat{\boldsymbol{x}}_1$, $\hat{\boldsymbol{x}}_2$, ..., $\hat{\boldsymbol{x}}_{i-2}$. The update equation (Eq. (33)) is a blend of the previous estimate, and the innovation, which is essentially a matrix of constants multiplied by a term that shows how wrong the previous estimate was.

$$\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_{i-1} + \boldsymbol{K}_i(\boldsymbol{y}_i - \boldsymbol{C}_i\hat{\boldsymbol{x}}_{i-1}), \tag{33}$$

where the matrix $\boldsymbol{K}_i \in \mathbb{R}^{n\times m}$, also known as the estimator gain, is picked in a way that mean-squared estimation error, or $\mathbb{E}||\boldsymbol{x} - \hat{\boldsymbol{x}}_i||$ is minimized ($\mathbb{E}(.)$ denotes expectation operator). To find the $\boldsymbol{K}_i$, three steps must be taken,

- **Step 1**: Let the estimation error $\boldsymbol{\epsilon}_i : \boldsymbol{x} - \hat{\boldsymbol{x}}_i$. Then, through simplification and rearrangement,

$$\boldsymbol{\epsilon}_i = (\boldsymbol{I} + \boldsymbol{K}_i\boldsymbol{C}_i)\boldsymbol{\epsilon}_{i-1} - \boldsymbol{K}_i v_i. \tag{34}$$

- **Step 2**: The covariance of both sides of the Eq. (34), considering that $\boldsymbol{R}_i := \boldsymbol{Cov}(v_i)$ and

$\boldsymbol{P}_i := \boldsymbol{Cov}(\boldsymbol{\epsilon}_i)$, resulting

$$\boldsymbol{P}_i = (\boldsymbol{I} + \boldsymbol{K}_i\boldsymbol{C}_i)\boldsymbol{P}_{i-1}(\boldsymbol{I} + \boldsymbol{K}_i\boldsymbol{C}_i)^T + \boldsymbol{K}_i\boldsymbol{R}_i{\boldsymbol{K}_i}^T, \tag{35}$$

where $\boldsymbol{P}_i$ is essentially speaks for the uncertainty of the quality of the estimate.

- **Step 3**: To find the optimal estimator gain matrix $\boldsymbol{K}_i$,

$$\mathbb{E}||\boldsymbol{x} - \hat{\boldsymbol{x}}_i||^2 = \boldsymbol{Trace}(\boldsymbol{P}_i), \tag{36}$$

therefore, if the derivative of the $\boldsymbol{Trace}(\boldsymbol{P}_i)$ is set to 0, the following equation is achieved.

$$\boldsymbol{K}_i = \boldsymbol{P}_{i-1}{\boldsymbol{C}_i}^T(\boldsymbol{C}_i\boldsymbol{P}_{i-1}{\boldsymbol{C}_i}^T + \boldsymbol{R}_i)^{-1} \tag{37}$$

Hence, when a new observation $\boldsymbol{y}_i$ arrives, the linear recursive estimation algorithm computes the optimal $\boldsymbol{K}_i$ and updates the estimate.

With all that said, the KF essentially applies the linear recursive estimator to a discrete time linear system. To summarize the combination, suppose $\hat{\boldsymbol{x}}_0$ is the best guess for initial condition $\boldsymbol{x}_0$, and, similarly $\boldsymbol{P}_0$ is the best guess for the initial uncertainty. Thus, every time step (iterate) of the system, the KF propagates its guess forward in time according to the dynamics, through

$$\hat{\boldsymbol{x}}_{i+1} = \boldsymbol{A}\hat{\boldsymbol{x}}_i + \boldsymbol{B}\boldsymbol{u}_i, \tag{38}$$

assuming that there is no noise. That is because the noise is zero-mean, therefore, this assumption seems to be the most reasonable assumption the filter can make about how to propagate its estimate forward in time. The uncertainty of the estimation is also propagated forward in time,

$$\boldsymbol{P}_{i+1} = \boldsymbol{A}\boldsymbol{P}_i\boldsymbol{A}^T + \boldsymbol{Q}_i, \tag{39}$$

where $\boldsymbol{Q}_i$ is the covariance of w_i which is known. Everytime a measurement comes in, the estimate

gain matrix, the estimate and its uncertainty are updated as follows.

$$\boldsymbol{K}_i = \boldsymbol{P}_{i-1}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{P}_{i-1}\boldsymbol{C}^T + \boldsymbol{R}_i)^{-1} \tag{40}$$

$$\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_{i-1} + \boldsymbol{K}_i(\boldsymbol{y}_i - \boldsymbol{C}\hat{\boldsymbol{x}}_{i-1}) \tag{41}$$

$$\boldsymbol{P}_i = (\boldsymbol{I} + \boldsymbol{K}_i\boldsymbol{C})\boldsymbol{P}_{i-1}(\boldsymbol{I} + \boldsymbol{K}_i\boldsymbol{C})^T + \boldsymbol{K}_i\boldsymbol{R}_i{\boldsymbol{K}_i}^T \tag{42}$$

All three equations above, are simpler than their counterparts in linear recursive estimation algorithm, as the time-invariance of linear systems implies that $\boldsymbol{C}_i$ is replaced by $\boldsymbol{C}$. This completes my overview on Kalman Filter. Next I will describe the application of KF in BP estimation.

5.2.3.2. Kalman Filter's Application in WAKE-BPAT

I developed a KF-based algorithm for dynamical estimation of the BP values from PAT features. In this context, recently Reference [85] proposed a KF formulation where the BP constitutes the state variable and a simple random walk process, where the current value of BP variable is composed of the past value, is used to model BP dynamics. The observation model is constructed based on a single PAT feature resulting in the following state-space model to track the BP continuously

$$\text{State Model:} \quad \text{BP}(k) = \text{BP}(k-1) + w(k) \tag{43}$$

$$\text{Observation Model:} \quad \ln \text{PAT}(k) = \frac{1}{\alpha_1}\text{BP}(k) - \frac{\beta_1}{\alpha_1} + v(k), \tag{44}$$

where k denotes the time index, and $w(k) \sim \mathcal{N}(0, Q)$ and $v(k) \sim \mathcal{N}(0, R)$ represent the forcing terms and the observation noise, respectively. This completes a brief overview of the problem at hand. Next, we present the proposed WAKE-BPAT framework.

5.2.4 Proposed WAKE-BPAT

The proposed WAKE-BPAT framework consists of three main components namely: (i) Pre-processing; (ii) Feature extraction, and; (iii) BP estimation mechanism. Below, we describe the above mentioned components respectively and in details.

5.2.4.1. Pre-processing

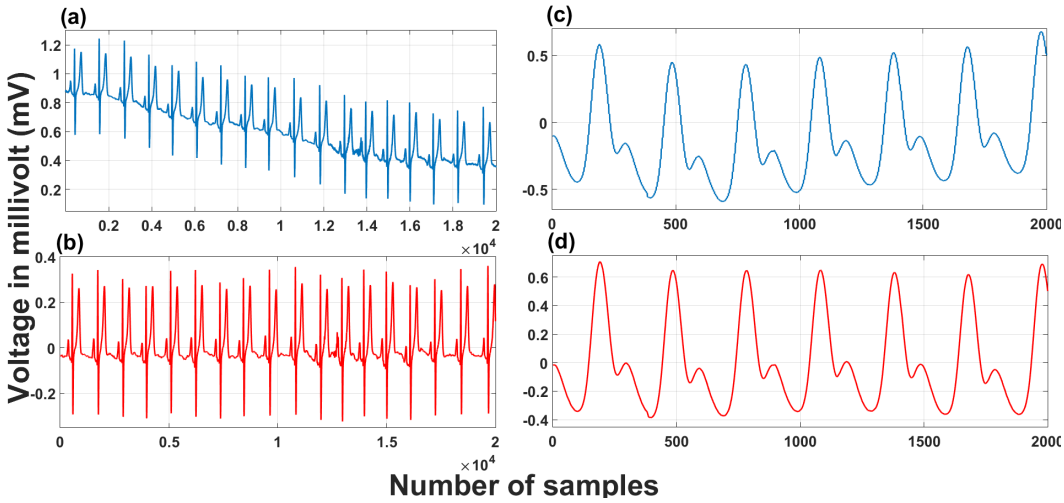


Figure 5.2: (a) Noisy ECG signal. (b) Denoised ECG signal. (c) Noisy PPG signal. (d) Denoised PPG signal.

At the time of acquisition, ECG signal often gets heavily contaminated by various high and low-frequency noises including the 50/60 Hz power line interference, electrosurgical noise, and baseline drift, which degrades the performance and the accuracy of automated ECG processing algorithms. Therefore, at first, the ECG signal has to be extracted from the background noise. The WT is a well-regarded technique which is able to effectively decompose a signal at various time-frequency resolutions and consequently, WT has been widely utilized for analyzing non-stationary signals such as the ECG.

The ECG signal is characterized by a periodic or quasi-periodic occurrence of various waves and segments having different frequency bands. Hence, WT is considered to be an excellent means for the analysis of ECG signals [96]. Assorted ECG-waves, segments, and also the noises come to be prominent at different frequency bands once subjected to the multi-resolution wavelet analysis. The discrete wavelet transform (DWT)-based ECG de-noising technique used in [97] is adopted in this framework. The Discrete Wavelet Transform decomposes the signal into mutually orthogonal set of wavelet, i.e., a wave-like oscillation with an amplitude that begins at zero, increases, and then decreases back to zero. The DWT provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. The wavelet can be constructed from a scaling function which describes its scaling properties. The scaling function, or

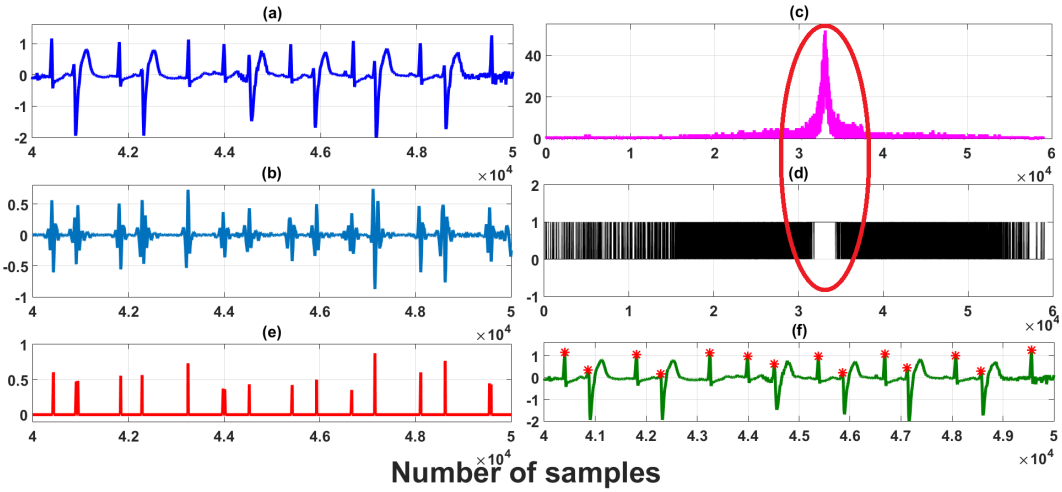


Figure 5.3: (a) Denoised ECG. (b) *QRS-coef* data. (c) Histogram analysis of the *QRS-coef*. (d) Amplitude-band where the population of coefficients is maximum. (e) Modified *QRS-coef* data. (f) Detected R-peaks.

mother wavelet, is created from

$$\phi(\boldsymbol{x}) = \sum_{k=-\infty}^{\infty} a_k \phi(s\boldsymbol{x} - k), \tag{45}$$

where a_k are the filter coefficients and s is a scaling factor (usually chosen as 2).

The clinical bandwidth of ECG signal lies between 0.05-100 Hz [98], and the signal is recorded at various sampling rates starting from 200Hz. To bring uniformity to the processing approach of ECG signals recorded at different sampling rates, the signal is re-sampled at 1KHz, and then the signal is decomposed using DWT by selecting the Biorthogonal 6.8 wavelet (bior6.8) as the mother wavelet function.

High and low-frequency noises are eliminated by discarding the corresponding detail and approximation wavelet-coefficients from the noisy signal. On the other hand, all the clinical signatures of PPG signal reside below 25 Hz [99], and therefore, the PPG signal is recorded at different sampling rates starting from 50Hz. Hence, the input PPG signal is also re-sampled at 1 KHz due to the same reason as in the case of ECG. The DWT-based signal denoising technique, which has been used for ECG is also used for PPG (Fig. 5.4), selecting the 'db8' wavelet function from the Daubechie's wavelet family [84]. An illustration of noisy and denoised ECG and PPG signals are shown in Fig 5.2.

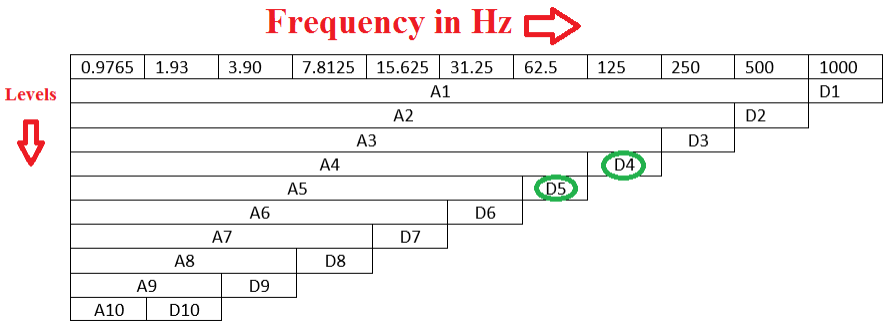


Figure 5.4: The DWT-decomposition-tree of the PPG signal.

5.2.4.2. Feature Extraction

The feature extraction component of the proposed WAKE-BPAT framework consists of two main tasks as explained below. *5.2.4.2.-A) The R-peak Detection* Since the detail-coefficients *D4* and *D5* of the wavelet transformed data contain most of the information-energy of the QRS-complex [97], these two coefficient-bands are selected for the purpose of R-peak detection. An array, which is denoted as *QRS-coef* is formed by adding the coefficients of the *D4* and *D5* (*QRS-coef* = *D4* + *D5*.) The local peaks of the data present in the *QRS-coef* array clearly indicate the QRS-complexes of the denoised signal. To eliminate the contribution of other waves and segments on the *QRS-coef* data, and to boost the detection accuracy of the R-peak points, histogram analysis of the wavelet coefficients present in the *QRS-coef* array is performed. Since most of the samples of an ECG-beat (one complete ECG cycle is considered as a beat) belong to the low-frequency non-QRS regions (T and P-peaks, ST-segment etc.), the peak of the histogram, and its surrounding samples values constitute the non-QRS regions. The amplitude-band where the population of coefficients is maximum is identified from the histogram analysis, and the corresponding amplitudes of those coefficients are made zero in the *QRS-coef* array. Finally, the local peaks are identified from the modified *QRS-coef* array, and the corresponding indices are marked as R-peak in the filtered signal. Figure 5.3 demonstrates the fact. The R-peak detection algorithm has been tested on a large number of ECG data files of different sampling rates, and the accuracy is found over 99.9%.

5.2.4.2.-B) Fiducial-point Detection from PPG Signal

First-derivative of the filtered PPG (FD-PPG) signal is calculated, and two different features are

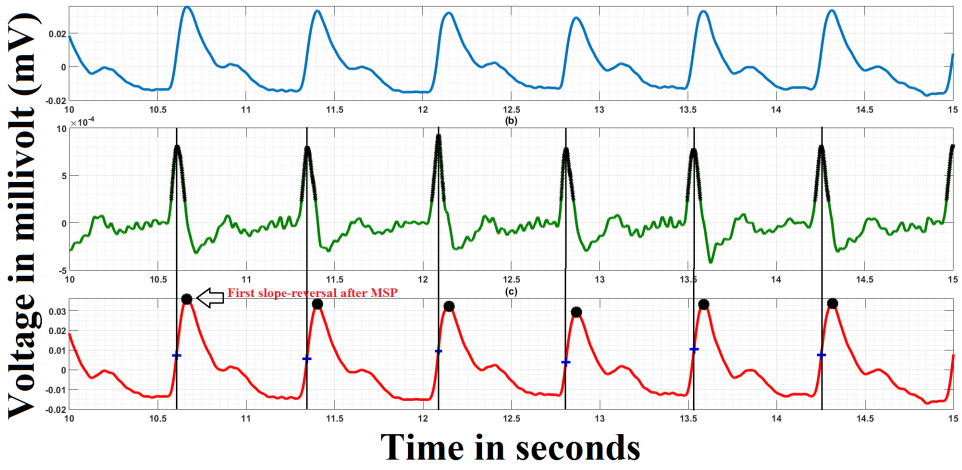


Figure 5.5: (a) Filtered PPG. (b) FD-PPG signal where marked samples are the ones within the threshold value. (c) Detected MSP and systolic-peaks.

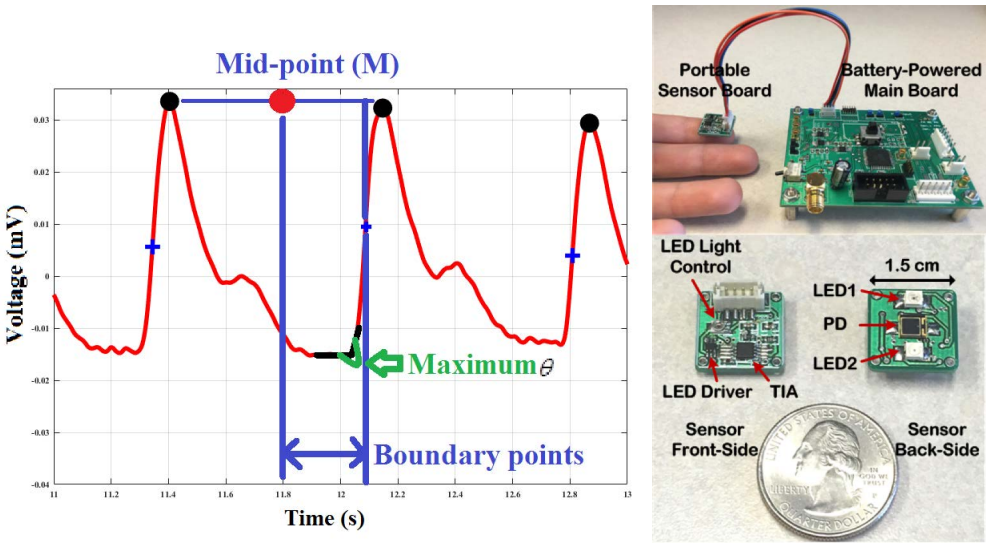
Table 5.1: Estimated BP versus the actual BP based on the proposed WAKE-BPAT.

Statistics	WAKE-BPAT	Proposed Features via Model 1	Reference [89]	Reference [87]
Mean Error	**2.67**	3.47	4.32	4.46
Standard Deviation	**2.51**	2.79	5.46	6.05
RMSE	**3.62**	4.41	5.52	5.74

extracted from the FD-PPG signal: (1) maximum-slope-point (MSP) of every PPG cycle, and (2) systolic-peaks. The maximum amplitude of the FD-PPG signal is found, and the indexes of those samples having an amplitude within 25% of the maximum are marked, and then the local-maximum amplitude within a sliding-window of width 0.25s [100] is identified and considered as the MSP of that PPG cycle. Now, traversing right in the time-domain PPG signal from the most recently detected MSP, the first slope-reversal event is identified as the systolic-peak. Figure 5.5 demonstrates the operations.

Now, the PPG-onset point is detected using a different method. First, the systolic-peak intervals in the filtered signal are divided into 2:1 ratio, i.e., the mid-point of the peak-to-peak interval is calculated, which is denoted as "M". Then, all the samples in between every M-point and the immediate next MSP are considered. Thereafter, among those considered samples, the maximum value of angle θ is found, and the corresponding index on the filtered PPG is marked as the PPG-onset point. Figure 5.6(a) shows the operations, while Figure 5.6(b) illustrates the Gen-1 Device.

5.2.4.3. BP Estimation Models



(a) (b)

Figure 5.6: (a) The PPG-onset detection technique. (b) The Gen-1 Device for PPG recordings developed recently by Marefat and Mohseni*et al.* [92].

We developed an adaptive KF-based algorithm for dynamical estimation of the BP values from features extracted in Section 3.2. Unlike the conventional approach of using a simple random walk to model BP evolutions over time, we use an Autoregressive (AR) process of order p (in the experiments we used $p = 4$) for relating the current BP estimates to its previous ($N > 1$) values, i.e.,

$$\text{BP}(k) = \sum_{i=1}^{p} a_i(k)\text{BP}(k-i) + w(k), \tag{46}$$

where a_i, for ($1 \leq i \leq p$), denotes the AR coefficients to be updated at each iteration. The evolution of the AR coefficients is modeled as

$$\boldsymbol{a}(k) \triangleq [a_1(k), \ldots, a_p(k)]^T = \boldsymbol{a}(k-1) + \boldsymbol{\nu}(k), \tag{47}$$

where superscript T denotes transpose operator, and $\boldsymbol{\nu}(k)$ is considered to follow a zero-mean and white Gaussian process with known covariance matrix. To recursively update the AR coefficients, a KF is performed based on Reference [101] where an instantaneous (static) estimate of the current BP based on Model 3 in Eq. (24) is used in the update step of the KF implemented for updating the AR coefficients.

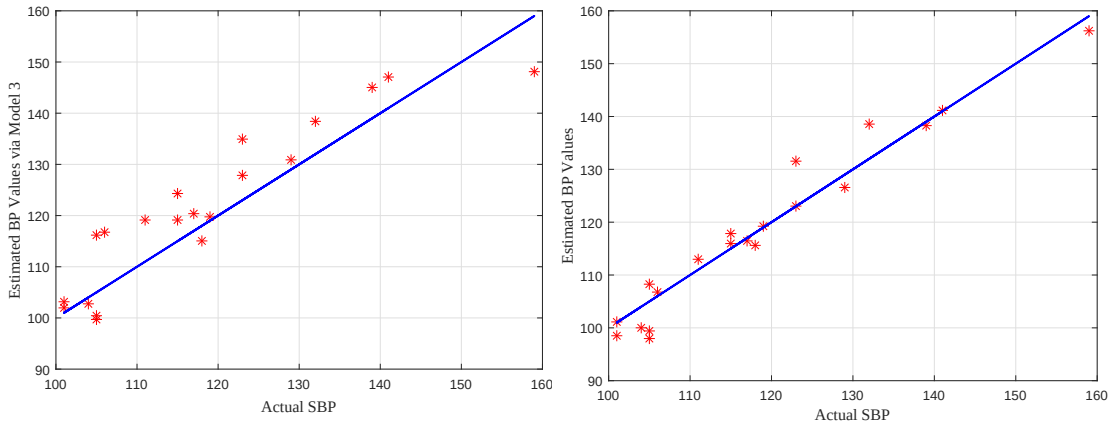
Instead of using a single observation model (such as the one introduced in Eq. (44)), a bank (combination) of ($N_f > 1$) different observation models is proposed to be employed and construct a hybrid state-space model for recursive estimation of the BP (in the experiments we used $N_f = 2$ based on the two characteristic points of the PAT features). In other words, I propose to consider a combination of observation models (PAT features) and fuse the estimation result based on each feature using adaptively computed weights. Intuitively speaking, the reason behind this scenario is that one feature might not be the best choice at all times and potentially using different measurement models would improve the performance. The observation model used in the WAKE-BPAT is, therefore, given by

$$y^{(l)}(k) = C_1^{(l)}\text{BP}(k) + C_2^{(l)} + v^{(l)}(k), \tag{48}$$

where superscript l, for ($1 \leq l \leq N_f$), refers to one of the candidate PAT features/models within the set of N_f considered multiple models, and $y^{(l)}(k)$ denotes its associated PAT measurement. For example, when Model 1 in Eq. (22) is included in the set of candidate models, the mode-matched terms in Eq. (48) are defined based on Eq. (44) as follows: $y^{(l)}(k) \triangleq \ln \text{PAT}(k)$, $C_1^{(l)} \triangleq \frac{1}{\alpha_1}$, and $C_2^{(l)} \triangleq -\frac{\beta_1}{\alpha_1}$. A KF is matched to each observation model l to form an updated BP estimate defined as $\hat{\text{BP}}^{(l)}(k) \triangleq \mathbb{E}\{\text{BP}(k)|\boldsymbol{Y}^{(l)}(k)\}$ where $\boldsymbol{Y}^{(l)}(k) = \{y^{(l)}(1), \ldots, y^{(l)}(k)\}$ is the set of all available observations upto and including the current iteration, and $\mathbb{E}\{\cdot\}$ denotes expectation operator. The mode-matched KFs are then fused through a collapsing step [95] which forms the optimal single Gaussian distribution in the mean-square error (MSE). Details of adaptive multiple model KF estimation is not included here due to lack of space, please refer to [95] and references therein for further details.

5.2.5 Simulation and Results

In this section, experimental results are presented based on a real data set collected from a healthy female volunteer. The BP variation is introduced by change in posture and exercise of the volunteer. The measured BP varied between 101 to 159 mmHg. The ECG signals are collected via a 3-lead ECG commercial device, while the PPG recordings are collected based on the Gen-1



(a) (b)

Figure 5.7: Estimated versus the actual BP. (a) Based on [89], i.e., Model 3. (b) Based on the proposed WAKE-BPAT and Model 1.

device from fingertip. As stated previously, Gen-1 device is very recently developed by Marefat and Mohseni *et al.*, which records the PPG signals in the reflective mode using a portable sensor board interfaced with a battery-powered main board for control and data processing. Please refer to Reference [92] for further details on the Gen-1 device. Finally, 20 reference BP recordings are measured by a cuff-based Omron 10 device. The results obtained based on all 20 measurements with Point 5 with BP equal to 101 mmHg, and Point 6 with BP equal to 141 mmHg are used for calibration via the LS approach. The PAT values are averaged over previous 10 epochs at each ground truth point. Four different BP estimation algorithms are implemented and compared for accuracy as follows: (i) The proposed WAKE-BPAT framework which provides dynamical estimates of the BP and uses the proposed features together with the proposed adaptive and multiple model KF; (ii) Instantaneous (static) BP estimation based on Model 1 and the proposed features; (iii) Instantaneous BP estimation based on [89], and; (iv) Instantaneous BP estimation based on [87].

Table 5.1, compares the accuracy of the above four estimation algorithms in terms of the mean error in absolute value, the standard deviation, and the root mean squared error (RMSE). It is worth mentioning that mean estimation error below 5 mmHg (in absolute value) with standard derivation of below 8 mmHg is the requirement set by the Association for the Advancement of Medical Instrumentation. It is observed that the proposed WAKE-BPAT framework provides significantly superior results in comparison to its counterpart based on previously developed features. In particular, the

mean error in absolute value is reduced approximately in half. At the same time, the effect of the proposed feature extraction algorithms is observed in the improved accuracy of Item (ii). This improvement can be attributed to the proposed histogram analysis of the wavelet coefficients, which not only helps removing the contribution of other waves, but also the presence of in-band noises, which in-turns serves to accurately and reliably detect R-peaks. The prime advantage of the proposed R-peak detection algorithm over others is that it does not require any threshold value for the estimation of the peaks. This algorithm is also potential to be effectively employed in a variety of applications including heart rate calculation, heart rate variability estimation, classification of ECG beats, or, as the motivation of this work, the hBCIs. Fig. 5.7 compares the estimation error results in absolute value versus the actual BP values computed based on Items (i) and (iii). It is observed the proposed WAKE-BPAT framework outperforms its counterpart and the estimated BP values are fairly close to their actual ground truth, which attests to the effectiveness of the proposed cuff-less and continuous BP estimation framework.

5.2.6 Conclusion

In this section, I proposed a novel framework for non-invasive and continuous estimation of the blood pressure (BP) from Pulse Arrival Time (PAT). The PAT is computed from the time interval between the R-peak of the ECG signal and a characteristic point of the PPG signal collected based on a recently developed PPG recording device (Gen-1). In particular, a wavelet-based feature extraction algorithm coupled with an adaptive and multiple-model Kalman filtering framework (referred to as the WAKE-BPAT) is proposed, which provides accurate BP estimates by extraction/fusion of different PAT characteristics. The WAKE-BPAT framework is evaluated based on a real data set, and it was shown that the proposed framework significantly outperforms its counterparts. Several potential future venues may be envisioned foe this work, for instance, investigation for using nonlinear filters instead of the KF.

5.3 Summary

Throughout this chapter, first I described hybrid BCI systems, which are BCI systems combined with other system(s) that utilize neurological signals, physiological signals, and/or external signals. This category of BCIs are composed of two different brain signals, one brain signal associated with two mental strategies, or one brain signal and another input such as ECG, and are supposedly, able to achieve specific goals better than a conventional system. Inspired by the fact that MI of exercises causes significant changes in the cardiopulmonary autonomic responses consisting of alterations in heart rate, blood pressure, respiration rate, and blood oxygen content, I sought for a framework to estimate BP through measurements of ECG and PPG signals. The WAKE-BPAT framework is a novel wavelet-based feature extraction algorithm coupled with an adaptive and multiple-model Kalman filtering. This framework provides accurate and dynamic BP estimates by extraction and fusion of different pulse arrival time (PAT) features. Experimental evaluations of WAKE-BPAT based on a real data set collected via Gen-1 device confirms the superiority of the proposed framework in comparison to its counterparts.

Chapter 6

Summary and Future Research Directions

The chapter concludes the thesis with a list of important contributions made in the dissertation and some proposed directions for future work.

6.1 Summary of Contributions

A list of the contributions of the thesis is as follows.

(1) **Adaptive Dimensionality Reduction Method using Graph-based Spectral Decomposition for Motor Imagery-based Brain-Computer Interfaces**: In this framework, referred to as the GD-BCI as a graph-based approach, I proposed steps as follows: (i) more systematic and robust preprocessing approaches were tried and the best-suited technique was chosen to smoothen the data; (ii) an uncomplicated technique, referred to as PRD, to make a more accurate estimation of the minimum size of the trial matrices; (iii) a graph-based spatiotemporal filter to reduce the dimensionality of the trial matrices, which takes into account both geometrical structure of the channels/electrodes and the correlation between the EEG signals; (iv) a tangent space mapping technique, to extract the features of the data matrices from the Riemannian manifold and thereafter, map the vectors onto Euclidean space; and, (v) classification of the feature vectors via a linear SVM classifier. The results of benchmarking this

framework on datasets IVa of BCI competition III prove the effectiveness of the proposed method.

(2) **Graph-based Dimensionality Reduction of EEG Signals via Functional Clustering and Total Variation Measure for BCI Systems**: In this framework, referred to as the GDR-BCI framework, similar to the GD-BCI, (i) a more systematic approach was chosen to smoothen the data; (ii) based on the fact that functionality of different connectivity neighborhoods varies based on the intensity of the performed activity and concentration level of the subject, an initial functional clustering of EEG electrodes is built by designing a separate adjacency matrix for each identified functional cluster; (iii) a collapsing methodology is used which is based on total variation measures on graphs, the overall model will eventually be reduced (collapsed) into two functional clusters; (iv) in order to better elaborate the effectiveness of the proposed method, CSP algorithm was chosen to extract the features; and, (v) linear and quadratic discriminant analysis models were employed to classify the data. The experimental results based on the same Dataset IVa from BCI Competition III show that the proposed method can provide higher classification accuracy as compared to its counterparts.

(3) **Progressive Fusion of Multi-rate Motor Imagery Classification for Brain-Computer Interfaces**: This framework consists of two filters running in parallel namely: (i) *The Progressive Filter*: An efficient filter that performs both feature extraction and classification steps based on the set of *all* arriving epochs to re-train progressively over time. (ii) *The Active Filter*: A simplified CSP-based feature extraction approach running online based on pre-trained classifiers, i.e., a lighter version of the Progressive Filter that runs faster than its counterpart. The classification model of Active Filter is updated every time the Progressive Filter's evaluation shows a possibility to increase the classification accuracy. The proposed framework is evaluated both based on dataset IVa from the BCI competition III, and through real data collected via the Emotiv Epoc headset.

(4) **Improving the Accuracy of MI EEG-based BCIs Through Trimming the Epochs**: This method proposes a readjustment in the recorded epochs via trimming the epoch signals in a manner that most informative parts of the signals are extracted and the segments of the

epochs which do not include the response of the subjects to the stimuli would be discarded. This approach was tested by adding the trim step to a conventional CSP-based framework, and the results successfully prove the positive effect of this technique.

(5) **Wavelet-Based Adaptive Kalman Filtering For Blood Pressure Estimation Via Fusion Of Pulse Arrival Times**: The WAKE-BPAT framework (i) filters ECG and PPG signals to eliminate the noise; (ii) extract the ECG and PPG features by a wavelet-based algorithm; and (iii) employs an adaptive and multiple-model Kalman filter to estimate the blood pressure. This framework provides accurate and dynamic BP estimates by extraction and fusion of different pulse arrival time (PAT) features. Experimental evaluations of WAKE-BPAT based on a real dataset collected via Gen-1 device confirms the superiority of the proposed framework in comparison to its counterparts.

As a final note, it is substantially important to explicitly point out the assumptions made throughout this thesis, for there can be amendments to relax them in future works.

(1) The vital biosignals processed and analyzed throughout the thesis research work, i.e., EEG, ECG, and PPG signals, are all assumed to be normally distributed by nature. In other words, the underlying reason for considering "covariance" as an effective statistical measure for retrieving the dependencies within the data is the Gaussianity of the data's distribution from probability and statistical point of view (second-order statistics). This assumption can be further relaxed/amended in the future works of my colleagues at I-SIP Lab, or myself, by considering higher orders of statistical distribution, and hence, more complicated explanation of the dependencies within the data.

(2) Throughout the development of theoretical frameworks for BCIs (GD-BCI and GDR-BCI), and similarly for designing the practical solutions (multi-rate filters and trimming technique), I utilized a 5–the order *Butterworth* filter. A few other filters were also tested with higher and lower orders, however, the lower orders brought about large transition in the frequency domain for cutting the bands before and after the desired frequency band. Moreover, higher orders caused undesirable impacts on the frequency contents around the cut-off frequencies. The overshoots and undershoots occurring because of the higher orders of the filter could have

been neglected if the processing of brain patterns was not greatly sensitive to effects of the filter, whereas, my studies showed otherwise. This problem calls for seeking for a solution to ensure the optimality of the band-pass filters order used for the preprocessing step.

(3) For any EEG-based BCI set up, in regards to the synchronization of the recordings and the markers of the stimuli, an implicit assumption exists, i.e., the logged marker precisely shows the actual time of the stimuli. This assumption is generally made because the experimental setup, including the computer(s), the cables, and any other device used for the purpose of brain activity collection, involves propagation delay by nature. This error is inevitable, although, it can be mitigated. It is noteworthy that if the software/application used for creating the data variables would provide the marking option, the propagation error it supposedly reduced.

(4) The dimensionality reduction methods proposed in my research work are an effort to ensure that innovative ideas like those suggested in GD-BCI and GDR-BCI are effective for the majority of currently-in-high-interest BCIs. In the future, optimization methods can be integrated with the proposed graph-based dimensionality reduction techniques, to ensure the optimal number of key variables required.

6.2 Future Work

This section discusses the possible future extensions to the proposed contributions throughout this thesis.

- **Graph-based EEG-based BCI Processing Frameworks**: The GD-BCI and GDR-BCI frameworks will be followed by:

  (1) Implementing the same framework on other datasets, such as datasets collected at I-SIP Lab;

  (2) The clustering step of GDR-BCI framework needs to be implemented automatically;

  (3) More complicated classification models shall be tried out, and;

  (4) The frameworks can be developed for multi-class BCIs.

- **The Practical Solutions for EEG-based BCIs**: Both of the approaches proposed in chapter 3 will be developed with

  (1) Implementing the same framework on other datasets, such as datasets collected at I-SIP Lab;

  (2) More flexible time interval selection algorithm will be developed, and the two techniques will be merged;

  (3) More complicated classification models shall be tried out, and;

  (4) The frameworks can be developed for multi-class BCIs.

- **The WAKE-BPAT Framework**: The framework has to be carried out for more subjects' data, and the possibility of employing non-linear Kalman Filter will be investigated.

Bibliography

[1] G. Kalantar, H. Sadreazami, A. Mohammadi, and Amir Asif, "Adaptive Dimentionality Reduction Method using Graph-based Spectral Decomposition for Motor Imagery-based Brain-Computer Interfaces," *IEEE Global Signal Processing Conference (GlobalSip)*, 2017.

[2] G. Kalantar and A. Mohammadi, "Graph-based Model of EEG Signals via Functional Clustering and Total Variation Measure for Brain Computer Interfacing," Accepted in *IEEE International Engineering in Medicine and Biology Conference (EMBC)*, 2018.

[3] T. Maloney, G. Kalantar, and A. Mohammadi, "Progressive Fusion of Multi-rate Motor Imagery Classification for Brain Computer Interfaces," in *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017.

[4] G. Kalantar, M. Mirgholami, Amir Asif, A. Mohammadi, "Improving the Performance of Motor Imagary-based EEG-based BCIs via an Adaptive Epoch Trimming Mechanism," Under submission to *IEEE Global Signal Processing Conference (GlobalSip)*, 2018.

[5] G. Kalantar, S.K. Mukhopadhyay, F. Marefat, P. Mohseni, and A. Mohammadi, "WAKE-BPAT: Wavelet-based Adaptive Kalman Filtering for Blood Pressure Estimation via Fusion of Pulse Arrival Times," in *2018 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)*.

[6] G. Kalantar and A. Mohammadi, "Analyzing the Effect of Bluetooth Low Energy (BLE) with Randomized MAC Addresses in IoT Applications," Accepted in *IEEE International Conference on Internet of Things (iThings)*, 2018.

[7] A. Kubler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, & N. Birbaumer, "Braincomputer communication: Unlocking the locked in", *Psychological bulletin*, 127(3), 358, 2001.

[8] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, & J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system", *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034-1043, 2004.

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

[10] B. Rebsamen, E. Burdet, C. Guan, H. Zhang, C. L. Teo, Q. Zeng,... & M. H. Ang Jr. "Controlling a wheelchair indoors using thought", *IEEE intelligent systems*, 22(2), 2007.

[11] D. V. Buonomano, & M. M. Merzenich, "Cortical plasticity: from synapses to maps", *Annual Review of Neuroscience*, 21(1), 149-186, 1998.

[12] P. R. Kennedy, R. A. Bakay " Activity of single action potentials in monkey motor cortex during long-term task learning", *Brain research*, 760(1-2), 251-254, 1997.

[13] R. D. Flint, Z. A. Wright, M. R. Scheid & M. W. Slutzky "Long term, stable brain machine interface performance using local field potentials and multiunit spikes", *Journal of neural engineering*, 10(5), 056005, (2013).

[14] C. Mehring, M. P. Nawrot, S. C. de Oliveira, E. Vaadia, A. Schulze-Bonhage, A. Aertsen, & T. Ball, "Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex", *Journal of Physiology-Paris*, 98(4-6), 498-506, (2004).

[15] A. K. Bansal, W. Truccolo, C. E. Vargas-Irwin, & J. P. Donoghue. "Decoding 3D reach and grasp from hybrid signals in motor and premotor cortices: spikes, multiunit activity, and local field potentials", *Journal of neurophysiology*, 107(5), 1337-1355, (2011).

[16] E. C. Leuth ardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, & D. W. Moran, ' A braincomputer interface using electrocorticographic signals in humans", *Journal of neural engineering*, 1(2), 63, (2004).

[17] K. B. Clancy, A. C. Koralek,R. M. Costa, D. E. Feldman, & J. M. Carmena, "Volitional modulation of optically recorded calcium signals during neuroprosthetic learning", *Nature neuroscience*, 17(6), 807, (2014).

[18] N. Birbaumer, L. E. Roberts, W. Lutzenberger, B. Rockstroh, & T. Elbert. "Area-specific self-regulation of slow cortical potentials on the sagittal midline and its effects on behavior", *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 84(4), 353-361 (1992).

[19] K. E. Misulis (Ed). "Spehlmann's evoked potential primer: visual, auditory, and somatosensory evoked potentials in clinical diagnosis", *Butterworth-Heinemann Medical*, 1994.

[20] M. J. Herrmann, A. Walter, A. C. Ehlis, & A. J. Fallgatter. "Cerebral oxygenation changes in the prefrontal cortex: effects of age and gender", *Neurobiology of Aging*, 27(6), 888-894 (2006).

[21] F. L. da Silva, "Neural mechanisms underlying brain waves: from neural membranes to networks", *Clinical Neurophysiology*, 79(2), 81-93, (1991).

[22] D. J. McFarland, L. A. Miner, T. M. Vaughan, & J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements", *Brain topography*, 12(3), 177-186, (2000).

[23] J. R. Wolpaw, & D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans", *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17849-17854, (2004).

[24] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, ... & T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting", *IEEE transactions on rehabilitation engineering*, 8(2), 164-173, 2000.

[25] L. F. Nicolas-Alonso & J. Gomez-Gil, "Brain computer interfaces, a review", *Sensors*, 12(2), 1211-1279, 2012.

[26] H. Berger. " On the electroencephalogram of man; Third report", *Electroencephalography and clinical neurophysiology*, Suppl-28., 1969.

[27] M. Cheng, X. Gao, S. Gao, & D. Xu. "Design and implementation of a brain-computer interface with high transfer rates", *IEEE transactions on biomedical engineering*, 49(10), 1181-1186, 2002.

[28] H. Serby, E. Yom-Tov, G. F. Inbar, "An improved P300-based brain-computer interface", *IEEE Transactions on neural systems and rehabilitation engineering*, 13(1), 89-98, 2005.

[29] S. Moghimi, A. Kushki, A. Marie Guerguerian, & T. Chau, "A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities", *Assistive Technology*, 25(2), 99-110, 2013.

[30] "Human electroencephalography book", *Chapter 10. Examination of biolectrical signals accompanying brain function (EEG)*, retrieved from http://ttktamop.elte.hu/online-tananyagok/physiology_practical/ch10s02.html, May 24th, 2018, 15:00.

[31] B. J. Edelman, B. Baxter, & B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks", *IEEE Transactions on Biomedical Engineering*, 63(1), 4-14, 2016.

[32] J. Kim, B. Lee, H. S. Lee, K. H. Shin, M. J. Kim, & E. Son, "Differences in brain waves of normal persons and stroke patients during action observation and motor imagery", *Journal of physical therapy science*, 26(2), 215-218, 2014.

[33] J. Decety, D. Perani, M. Jeannerod, V. Bettinardi, B. Tadary, R. Woods, J. C. Mazziotta & F. Fazio, "Mapping motor representations with positron emission tomography", *Nature*, 371(6498), 600, 1994.

[34] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks", *AIChE journal*, 37(2), 233-243, 1991.

[35] 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.

[36] P. Berg, & M. Scherg, " A multiple source approach to the correction of eye artifacts", *Electroencephalography and clinical neurophysiology*, 90(3), 229-241, 1994.

[37] P. Karthikeyan, M. Murugappan, & S. Yaacob, " ECG signal denoising using wavelet thresholding techniques in human stress assessment", *International Journal on Electrical Engineering and Informatics*, 4(2), 306, 2012.

[38] L. De Lathauwer, B. De Moor, & J. Vandewalle, " A multilinear singular value decomposition", *SIAM journal on Matrix Analysis and Applications*, 21(4), 1253-1278, 2000.

[39] R. Javaid, R. Besar, & F. S. Abas, "Performance evaluation of percent root mean square difference for ecg signals compression", *Signal Processing: An International Journal (SPIJ)*, 1-9, 2008.

[40] R. Benzid,F. Marir, & N. E. Bouguechal, "Electrocardiogram compression method based on the adaptive wavelet coefficients quantization combined to a modified two-role encoder", *IEEE Signal Processing Letters*, 14(6), 373-376, 2007.

[41] B. Mwangi, T. S. Tian, & J. C. Soares, "A review of feature reduction techniques in neuroimaging", *Neuroinformatics*, 12(2), 229 - 244, 2014.

[42] G. Townsend, B. Graimann, & G. Pfurtscheller, " A comparison of common spatial patterns with complex band power features in a four-class BCI experiment", *IEEE Transactions on Biomedical Engineering*, 53(4), 642-651, 2006.

[43] Y. Wang, S. Gao, & X. Gao, " Common spatial pattern method for channel selelction in motor imagery based brain-computer interface", *27th Annual international conference of the Engineering in medicine and biology society*, (pp. 5392-5395), 2005.

[44] F. Lotte, M. Congedo, A. Lcuyer, F. Lamarche, & B. Arnaldi, 'A review of classification algorithms for EEG-based braincomputer interfaces", *Journal of Neural Engineering*, 4(2), R1, 2007.

[45] C. C. Aggarwal (Ed), "Data classification: algorithms and applications", *CRC Press*, 2014.

[46] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, & G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate", *IEEE Transactions on Biomedical Engineering*, 51(6), 979-984, 2004.

[47] D. Garrett, D. A. Peterson, C. W. Anderson, & M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification", *IEEE Transactions on neural systems and rehabilitation engineering*, 11(2), 141-144, 2003.

[48] C. J. Burges, "A tutorial on support vector machines for pattern recognition" *Data mining and knowledge discovery*, 2(2), 121-167, 1998.

[49] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, & J. R. Wolpaw, "A comparison of classification techniques for the P300 Speller" *Journal of neural engineering*, 3(4), 299, 2006.

[50] N. L. da Costa, I. A. Castro, & R. Barbosa, "Classification of cabernet sauvignon from two different countries in South America by chemical compounds and support vector machines" *Applied Artificial Intelligence*, 30(7), 679-689, 2016.

[51] S. S. Keerthi, & C. J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel" *Neural computation*, 15(7), 1667-1689, 2003.

[52] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains", *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83-98, 2013.

[53] M. Mnoret, N. Farrugia, B. Pasdeloup, & V. Gripon, "Evaluating Graph Signal Processing for Neuroimaging Through Classification and Dimensionality Reduction" *arXiv preprint*, arXiv:1703.01842, 2017.

[54] T. Tanaka, T. Uehara and Y. Tanaka, "Dimensionality reduction of sample covariance matrices by graph Fourier transform for motor imagery brain-machine interface", *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 1-5, 2016.

[55] H. Hiroshi, T. Tanaka and Y. Tanaka, "Smoothing of spatial filter by graph Fourier transform for EEG signals", *IEEE Asia-Pacific Signal and Information Processing Association (APSIPA)*, pp. 1-8, 2014.

[56] D. W. Shattuck, & R. M. Leahy, "Automated graph-based analysis and correction of cortical volume topology", *IEEE transactions on medical imaging*, 20(11), 1167-1177, 2001.

[57] N. Hansen, "The CMA evolution strategy: a comparing review. In Towards a new evolutionary computation", *Springer Berlin Heidelberg*, pp. 75-102, 2006.

[58] A. Barachant, S. Bonnet, M. Congedo and C. Jutten, "Multiclass brain-computer interface classification by Riemannian geometry", *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 920-928, 2012.

[59] A. Barachant and S. Bonnet, "Channel selection procedure using Riemannian distance for BCI applications", *IEEE International Conference on Neural Engineering*, pp. 348-251, 2011.

[60] J. Hanaway, "The Brain Atlas: A Visual Guide to the Human Central Nervous System," *Bethesda, Md: Fitzgerald Science Press*, 1998.

[61] G. Kalantar, H. Sadreazami, A. Mohammadi, and Amir Asif, "Adaptive Dimentionality Reduction Method using Graph-based Spectral Decomposition for Motor Imagery-based Brain-Computer Interfaces," *IEEE Global Conf. Signal & Inf. Process.*, 2017.

[62] A. Sandryhaila, J.M.F. Moura, "Discrete Signal Processing on Graphs: Graph Fourier Transform," *IEEE ICASSP*, 2013, pp. 6167-6170.

[63] A. Sandryhaila, J.M.F. Moura, "Classification via Regularization on Graphs," *IEEE Global Conf. Sig. & Inf. Process.*, 2013, pp. 495-498.

[64] M. Duvinage, T. Castermans, M. Petieau, T. Hoellinger, G. Cheron, T. Dutoit, "Performance of the Emotiv Epoc Headset for P300-based Applications," *Biomed. Eng. Online*, vol. 12, p. 56, 2013.

[65] H. Banville and T. H. Falk, "Recent Advances and Open Challenges in Hybrid Brain-Computer Interfacing: A Technological Review of Non-invasive Human Research," *Brain-Comput. Interfaces*, vol. 3, no. 1, pp. 9-46, 2016.

[66] "EMOTIV Epoc - 14 Channel Wireless EEG Headset", *Emotiv. [Online]. Available: https://www.emotiv.com/epoc/*, [Accessed: 09-Jun-2017].

[67] M.S. Diab, Z. Hussain and S. Mahmoud, "Restoring Function in Paralyzed Limbs Using EEG," *2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Abu Dhabi, 2016.

[68] H. Boutani and M. Ohsuga, "Applicability of the Emotiv EEG Neuroheadset as a user-friendly input interface," *IEEE Int. Conf. Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 1346-1349.

[69] F. Carrino, J. Dumoulin, E. Mugellini, O. A. Khaled, and R. Ingold, "A Self-Paced BCI System to Control an Electric Wheelchair: Evaluation of a Commercial, Low-cost EEG Device," *ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, 2012, pp. 1-6.

[70] B. Z. Allison, E. W. Wolpaw & J. R. Wolpaw, "Brain-computer interface systems: progress and prospects", *Expert review of medical devices*, 4(4), 463-474, 2007.

[71] B. Z. Allison & C. Neuper, "Could anyone use a BCI? Brain-computer interfaces.", *Springer*, pp. 3554, 2010.

[72] T. O. Zander, C. Kothe, S. Jatzev, & M. Gaertner, "Enhancing human-computer interaction with input from active and passive brain-computer interfaces", *In Brain-computer interfaces, Springer, London*, pp. 181-199, 2010.

[73] G. Pfurtscheller, B. Z. Allison, G. Bauernfeind, C. Brunner, T. Solis Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper, & N. Birbaumer, "The hybrid BCI.", *Frontiers in neuroscience*, 4,3, 2010.

[74] J. Decety, M. Jeannerod, M. Germain, & J. Pastene, "Vegetative response during imagined movement is proportional to mental effort.", *Behavioural brain research*, 42(1), 1-5, 1991.

[75] S. Shahid, G. Prasad, & K. R. Sinha, "On fusion of heart and brain signals for hybrid BCI.", *5th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 48-52, 2011.

[76] T. Sato, Y. Okuyama, & M. Sakai, "Simulation study of a P300 speller for single-lead hybrid BCI.", *In Proceedings of SICE Annual Conference (SICE)*, 2017-2023, 2013.

[77] I. Choi, I. Rhiu, IY. Lee, M. H. Yun, & C. S. Nam, (2017). "A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives.", *PloS one*, 12(4), e0176674, 2017.

[78] "A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis", *World Health Organization*, Geneva, Switzerland, 2014.

[79] D.T. Lackland and M.A. Weber "Global Burden of Cardiovascular Disease and Stroke: Hypertension at the Core," *Canadian Journal of Cardiology*, 31(5), pp. 569-571, 2015.

[80] World Health Organization, "World Health Statistics 2015," *World Health Organization*, 2015.

[81] X. R. Ding, Y. T. Zhang, J. Liu, W. X. Dai, and H. K. Tsang, "Continuous Cuffless Blood Pressure Estimation using Pulse Transit Time and Photoplethysmogram intensity Ratio," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, 964-972, 2016.

[82] T.M. Seeberg *et al.*, "A Novel Method for Continuous, Noninvasive, Cuff-Less Measurement of Blood Pressure: Evaluation in Patients With Nonalcoholic Fatty Liver Disease," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1469-1478, 2017.

[83] M. Gao, H.M. Cheng, S.H. Sung, C.H. Chen, N.B. Olivier and R. Mukkamala, "Estimation of Pulse Transit Time as a Function of Blood Pressure Using a Nonlinear Arterial Tube-Load Model," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1524-1534, 2017.

[84] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 859-869, 2017.

[85] Q. Zhang *et al.* "Cuff-less Blood Pressure Measurement using Pulse Arrival Time and a Kalman Filter," *J. Micromech. Microeng.*, vol. 27, 2017.

[86] R. Mukkamala *et al.*, "Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879-1901, 2015.

[87] Y.L. Zheng, B.P. Yan, Y.T. Zhang and C.C.Y. Poon, "An Armband Wearable Device for Overnight and Cuff-Less Blood Pressure Measurement," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 7, pp. 2179-2186, 2014.

[88] V. Chandrasekaran, R. Dantu, S. Jonnada, S. Thiyagaraja, and K. P. Subbu, "Cuffless Differential Blood Pressure Estimation using Smart Phones," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1080-1089, 2013.

[89] M. Theodor *et al.* "Subcutaneous Blood Pressure Monitoring with an Implantable Optical Sensor," *Biomedical Microdevices*, vol. 15, no.5, pp. 811-820, 2013.

[90] Y. Chen, C. Wen, G. Tao, M. Bi, and G. Li, "Continuous and Noninvasive Blood Pressure Measurement: A Novel Modeling Methodology of the Relationship Between Blood Pressure and Pulse Wave Velocity," *Annals of Biomedical Engineering*, vol. 37, no. 11, pp. 2222-2233, 2009.

[91] S. Ahmad, S. Chen, K. Soueidan, I. Batkin, M. Bolic, H. Dajani, and V. Groza, "Electrocardiogram-assisted Blood Pressure Estimation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 608-618, 2012.

[92] F. Marefat, R. Erfani, K. L. Kilgore and P. Mohseni, "Minimally Invasive Muscle-based Recording of Photoplethysmogram Toward Chronic Implantation," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 388-391, 2016.

[93] H. Khamis, R. Weiss, Y. Xie, C. W. Chang, N. H. Lovell,and S. J. Redmond, "QRS Detection Algorithm for Telehealth Electrocardiogram Recordings," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1377-1388, 2016.

[94] P. Phukpattaranont, "QRS Detection Algorithm based on the Quadratic Filter," *Expert Systems with Applications*, 42(11), 4867-4877, 2015.

[95] A. Mohammadi and K. N. Plataniotis, "Improper Complex-Valued Multiple-Model Adaptive Estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1528-1542, 2015.

[96] E. B. Mazomenos, D. Biswas, A. Acharyya, T. Chen, K. Maharatna, J. Rosengarten, J. Morgan, and N. Curzen, "A Low-Complexity ECG Feature Extraction Algorithm for Mobile Healthcare Applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 459-469, 2013.

[97] S. Banerjee, and M. Mitra, "Application of Cross Wavelet Transform for ECG Pattern Analysis and Classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 2, 2014.

[98] W. J. Tompkins, "Biomedical Digital Signal Processing," *New Delhi: Prentice Hall of India Pvt. Ltd*, New Delhi, India, 2006.

[99] S. K. Mukhopadhyay, M. Omair Ahmad, and M.N.S. Swamy, "ASCII-Character-Encoding based PPG Compression for Tele-Monitoring System," *Biomedical Signal Processing and Control*, vol. 31, pp. 470-482, 2017.

[100] S.K. Mukhopadhyay, S. Mitra, M. Mitra, "ECG Feature Extraction Using Differentiation, Hilbert Transform, Variable Threshold and Slope Reversal Approach," *Journal of Medical Engineering and Technology*, vol. 36, no. 7, pp. 372-386, 2012.

[101] M. Arnold, X.H.R. Milner, H. Witte, R. Bauer and C. Braun, "Adaptive AR Modeling of Nonstationary Time Series by Means of Kalman Filtering," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 5, pp. 553-562, 1998.