Events prediction in Electroencephalographic signals

Department of Biomedical Engineering
University of Strathclyde
Glasgow
United Kingdom

Thesis submitted for the Degree of Doctor of Engineering by

Ange Guillaume Tano

Supervisors:
Dr Heba Lakany
Prof. Bernard A. Conway
Prof. John Soraghan (EEE)

Copyrights

Ange Guillaume Tano

Signed: Date:

Forecasting future events is often like

searching for a black cat

in an unlit room,

that may not even be there.

Steve Davidson in The Crystal Ball.

Acknowledgement

I have been extremely fortunate to be supervised by Dr Heba Lakany , Prof. Bernard A. Conway and Prof. John Soraghan during my EngD. Their support and guidance have been invaluable. I would like to thank all the staff members at the Department of Biomedical Engineering at the University of Strathclyde and particularly the staff member of the Doctoral Training Centre (DTC) for the excellent research facilities without which my Engineering Doctorate (EngD) would not have been possible.

A special mention must go to Dr Campbell Reid and Dr Sylvie Coupaud for critically reading this thesis and to Dr David Halliday for an insightful introduction of the multitaper and the multi wavelet theories.

The Neurophysiology laboratory at the Department of Biomedical Engineering at the University of Strathclyde has been a forum of both random and constructive discussions in a very pleasant atmosphere. My thanks therefore go to postgraduate students and friends Dr Catherine MacLeod, Dr Alejandra Aranceta Garza, Dr Harry Ellangowan, Dr Bahman Nasseroleslami, Bilal Nasser, Niall McKenzie, Pauline Axford, Jamie O'Reilly, Lijo Varughese, Syahrull Hi-Fi Syam and Chi Hsu Wu for their time and useful feedbacks. My gratitude is due to postgraduate students Sibani Mohanty and Radhika Menon for helping with the subjects preparation during the experiments.

Many thanks to the volunteers who completed the experiments for their time and patience throughout such a long and exhausting experimental task. I would also like to thank the technical staff members of the Department of Biomedical Engineering, John MacLean and Stephen Murray.

I would finally like to thank my family for their constant support and special thanks to my wife and my two sons who have terribly missed me during this long process, you can finally stop asking: Are you going to the University today?

Abstract

Brain Machine Interface (BMI) or Brain Computer Interface (BCI) technologies provide the prospect of regenerating or replacing functions lost due to motor disabilities. BCIs connect the brain to a computer which translates the electrical activity of the brain into commands used to control external devices, hence allowing people with motor disability to control their external environment through a non-muscular communication channel. A BCI operates by transforming electrophysiological signals, known as Electroencephalogram (EEG) signals, from the user into device commands under an operating protocol. The protocol initialises and defines the nature of the communication (i.e., discrete or continuous). It also determines the strategy which underpins the generation of the signals used by the system (i.e., what triggers the changes within the EEG signals). While protocols involving brisk and very constrained movements have been widely explored in BCI studies, more natural movements have barely been considered. The choice of protocols that entail non-realistic movements emanates from the generation of well understood neuronal correlates modulated by the execution of such movements, leading to a lack of freedom in the design of BCI protocols.

The development of algorithms translating EEG signals into commands that control external devices, a task termed as event detection in the present research, are a central part of any BCI system. However, most of the time, complex methods are used and most event detection in BCI development is devoted to the optimisation of such methods. Furthermore, the methods require extensive training of the user, putting a mental load on the user. The present study aims to investigate simple, but powerful, event detection methods requiring minimal training and the use of a protocol involving natural movements. Scalp EEG data was recorded from nine participants using natural hand movements. In particular, self-paced reaching hand movements were considered. The data was investigated in a pseudo time frequency domain using continuous wavelet coefficients. Methods using

wavelet modulus maxima, the Mahalanobis distance, and bootstrapping of the Mahalanobis distance were developed for event detection. The data was analysed over a frequency range from $0.1\,\mathrm{Hz}$ to $25\,\mathrm{Hz}$, covering the Slow Cortical Potentials (SCP), Mu and Beta frequency bands.

The results showed that the method using wavelet modulus maxima was able to predict reaching hand movements onset in the SCP, Mu and Beta frequency bands about $1\,\mathrm{s}$ before movement onset and yielded an maximum average prediction rate of approximately 80%. The Mahalanolobis distance and the bootstrap methods were able to predict reaching hand movements initiation in the SCP band about $1\,\mathrm{s}$ before movement onset with a maximum prediction rate of approximately 70%. The study has demonstrated that human voluntary movements can be predicted approximatively $1\,\mathrm{s}$ prior to movement onset with a good prediction rate. The study may contribute to the understanding of the planning and the control of human voluntary movements. Furthermore, the present research may contribute in designing advanced assistive devices in general and in particular may contribute in improving BCI systems design. Finally, the results may encourage the use of natural movements during BCI protocols design aiming to predict movement initiation and the monitoring of the mental state of BCI users.

Contents

List of Figures

List of Tables

List of Algorithms

Glossary

$\lfloor \cdot \rfloor$ The integer part. 53

μ The mean of the normal distribution. 63, 89

σ The variance of a distribution. 63

β_2 The kurtosis of a distribution. 63, 89

$E[\cdot]$ The expected value operator. 64

$N(\mu, \sigma)$ normal distribution with mean μ and variance σ. 93

C is the letter used to represent central electrodes. 15

F is the letter used to represent frontal electrodes. 15

inion is the point located between the nose and forehead. 15

O is the letter used to represent occipital electrodes. 15

P is the letter used to represent parietal electrodes. 15

T is the letter used to represent temporal electrodes. 15

Acronyms

AC Alternative Current. 72

ALS Amyotrophic Lateral Sclerosis. 22, 152

ANN Artificial Neural Networks. iii, 61, 62, 66

AP Action Potential. 7, 13

AR Auto Regressive model. ii, 37

ART Adaptive Resonance Theory. 62

BBI Brain-to-Brain Interface. 25

BCa Bias corrected and accelerated. 54, 55, 56

BCI Brain Computer Interface. i, ii, v, xi, 1, 3, 4, 19, 20, 21, 22, 23, 24, 25, 28, 58, 59, 60, 66, 75, 100, 144, 148, 152, 153, 157, 158, 160

BEB bootstrap estimate of bias. 49

BESE bootstrap estimate of standard error. 49, 51

BMI Brain Machine Interface. i

BPF Band Pass Filter. 35

BSS blind source separation. 46

c.d.f cumulative distribution function. 52

CAR Common Average Reference. 35, 69, 83, 148

CBI Computer Brain Interface. 25, 28

CI confidence interval. 50, 51, 52, 53, 54, 55, 56

CNS Central Nervous System. 6, 7, 11

CWT Continuous Wavelet Transform. 40, 83, 84, 85, 86, 87, 93, 145, 148, 153, 157

DC Direct Current. 72, 74, 145

DFT Discrete Fourier Transform. 37

dpss Discrete Prolate Slepian Sequence. 38

ECG Electro Cardiogram. 29

EcoG electrocorticograghy. 13, 23

ED Eucludian distance. 63, 65, 93

EEG Electroencephalogram. i, ii, vii, viii, 3, 4, 5, 12, 13, 15, 16, 18, 19, 21, 23, 24, 25, 27, 28, 29, 30, 33, 34, 35, 36, 37, 42, 43, 44, 45, 46, 56, 58, 60, 66, 67, 68, 69, 70, 72, 74, 75, 77, 79, 82, 83, 85, 88, 90, 92, 93, 98, 99, 109, 111, 141, 144, 145, 147, 148, 149, 152, 153, 157, 161

EM Expectation Maximisation. 61

EMG Electromyogram. 10, 29

EOG Electrooculogram. 70, 77

EPSP Excitatory Post synaptic Potential. 13

ERD Event Related Desynchronization. 22, 152

ERP Event Related Potential. 18, 19, 148

ERS Event Related Synchronization. 22

FFT Fast Fourier Transform. 37

FIR Finite Impulse Response. 35

fMRI Functional magnetic resonance imaging. 12, 13, 23

FN False Negative. 100

FP False Positive. 100

FT Fourier Transform. 37, 40

GMM Gaussian mixtures model. 57

HMM Hidden Markov Models. 58

HPF High Pass Filter. 35

I.R. Inferior Rectus. 70, 71

ICA Independent Component Analysis. 46, 47

IIR Infinite Impulse Response. 35, 145

IPSP Inhibitory Post synaptic Potential. 13

ITR Information Transfer Rate. xi, 99, 100, 134, 135, 137, 138, 143, 150, 158

kNN k Nearest Neighbours. 61

L Laplacian matrix. 33, 34

L.R. Lateral Rectus. 70, 71

LD Linear Discriminant. 59, 63

LDA Linear Discriminant Analysis. 59, 63

LE Lipschitz exponent. 88, 89, 108, 109

LFP Local Field Potential. 149

LPF Low Pass Filter. 35

M.R. Medial Rectus. 70, 71

MD Mahalanolobis distance. iv, v, viii, ix, x, xi, 59, 60, 63, 65, 93, 94, 95, 97, 109, 125, 126, 127, 128, 129, 130, 134, 137, 141, 142, 143, 149, 179, 180, 181, 182, 183, 184, 185, 186

MEG Magnetoencephalogram. 12, 13

MI Primary motor cortex. 8, 9, 25

MLP Multi-Layer Perceptron. 62

MN Motor neurone. 7

MND Motor neurone disease. 11

MRA multiresolution analysis. 40

MRCP Motor Related Cortical Potential. 1, 18, 69, 75, 144

MS Motor System. 11

NIRS Near Infrared Spectroscopy. 24

PCA Principal Component Analysis. iv, 42, 45, 46, 47, 63, 65, 83, 85, 86, 93, 113, 148, 149

PMA primary premotor areas. 76

PSP Post Synaptic Potential. 13

RBF Radial Basis Function. 62, 63

RP Readiness potential. 10, 11, 144, 161

S.O. Superior Oblique. 71

S.R. Superior Rectus. 70

SCP Slow Cortical Potentials. i, 19, 21, 22, 85, 116, 149, 179, 180, 181, 182, 183, 184, 185, 186

SMA premotor areas. 9, 76

SNR Signal to Noise Ratio. 28

SOM Self Organising Maps. 62

SSVEP Steady-state Visual Evoked Potentials. 22

STFT Short Time Fourier Transform. 37, 40, 41, 84

SVD Singular value decomposition. 45

TMS Transcranial Magnetic Stimulator. 25, 28

TN True Negative. 100

TP True Positive. 100

TPR True Positive Rate. v, viii, ix, 116, 117, 118, 119, 127, 128, 131, 143, 149, 150, 152, 179, 184, 185, 186

TTD Thoughts Translation Device. 21

VEP Visual Evoked Potentials. 19, 21

WMM Wavelet Modulus Maxima. iv, viii, ix, x, xi, 41, 87, 108, 113, 115, 116, 117, 118, 119, 134, 135, 137, 141, 142, 143, 179, 180, 181, 182, 183, 184, 185, 186

WT Wavelet Transform. 40, 41, 84

Chapter 1

Introduction

> My hand moves because certain forces - electric, magnetic, or whatever 'nerve-force' may prove to be - are impressed on it by my brain. This nerve-force, stored in the brain, would probably be traceable, if Science were complete, to chemical forces supplied to the brain by the blood, and ultimately derived from the food I eat and the air I breathe.
>
> Lewis Carroll, from Sylvie and Bruno,1890

The introductory quotation by Lewis Carroll suggested at a very early stage, in 1890, a possible correlation between voluntary movement and brain activity. In a historical breakthrough in 1929, the German psychiatrist Hans Berger discovered a technique to record the electrical activity of the brain called electroencephalogram (Haas, 2003). The possibility to record brain activity was a turning point in understanding the functional mechanism of the brain. The first evidence that, neuronal activity triggering voluntary movement precedes the conscious decision to move was provided by Deecke and his co-worker in 1969 (Deecke et al., 1969). Studies by Soon et al. (2008) clearly distinguish voluntary movement into three aspects, namely, what to move, when to move and whether to move or not, providing the possibility to put a veto on the execution of voluntary movements. The understanding of the neuronal mechanisms triggering voluntary movement is of great importance as such understanding can help subjects with motility deficiency to interact with their external environment.

Involvement of muscles during movement is very crucial as muscles are the effectors of those movements. Motor diseases can affect muscles making it impossible or difficult for certain individuals to use muscles to perform motor activities, hence unable to communicate with their external environment. Extensive research has been conducted in the field of Brain Computer Interface. The aim of Brain Computer Interface technology is to activate movement output directly at the cerebral level, establishing a new communication channel without involving muscle activity. The technology can enable people with motor disability to bypass muscles involvement when interacting with their environment. BCI systems operate by translating neuronal activity underpinning voluntary movements execution into commands that can be used to communicate without muscles involvement. In the translation of the brain activity into commands, predicting the initiation of movement onset is a central part during the implementation of any BCI system; and this task in termed 'event detection'. Event detection relies on two important factors. First, a control signal containing features related to the daily activities of the subject is needed. Secondly, mathematical algorithms are needed to transform the features of the control signal into commands. There are various type of BCI. However, the BCI considered in the present research use motor related tasks and are called movement related BCIs. Motor related Brain Computer Interfaces are the most widely studied class of BCIs (Pfurtscheller, 2006; McFarland et al., 2006). Understanding movement related neuronal correlates known as Motor Related Cortical Potential (MRCP) can contribute to the implementation of efficient motor related BCI systems. Various type of movements are used to design protocols to obtain control signals for BCI design. However, most of the protocols used in BCI design are too constrained to provide realistic communication. Movements used in BCI design protocols cannot accommodate a real life communication. The primary motivation of the present research is to investigate event detection from the perspective of self-initiated or self-paced centre out reaching movements. The study explores the possibility of detecting events in unconstrained hand movements in healthy subjects. Event detection using these particular type of movements could be relevant to develop assistive devices for rehabilitation purpose.

1.1 What is an event?

The term event can in general be regarded as some thing that happens during a certain period of time. In this respect, natural disasters such as earthquakes and virus outbreaks can be regarded as events. However, the term event is related in the present study to the prediction of the initiation of voluntary movement and in this case the latency of the event is correlated with the duration of the underpinning neural correlates of the movement. Compared to the events mentioned previously such as earthquakes which can last some few seconds, voluntary movement onset is a very fast event (lasting some milliseconds). Predicting relatively long lasting events such as earth quacks is a very difficult task, and even more laborious is the task of predicting fast events such as hand movement onset which requires complex feature extraction and mathematical algorithms. The task of detecting event in the present research will be performed on time series data making the definition of an event as something that occurs over a certain period of time relevant; as this definition can be used to differentiate an event from an outlier. The latter represents data point deviating from the time series data and is considered as noise. During the investigation of event such as hand movement initiation, defining an appropriate latency for this type of events can prevent any mismatch between desirable events and outliers.

1.2 Thesis statement

Event detection in EEG data is of central interest in the implementation of BCI systems. Various methods such as statistical and machine learning methods are often used for the purpose. However, the techniques used to date focus on machine learning methods which require intensive training. Most of the work in event detection is devoted to the development and optimisation of the underlying models rather than event detection itself. In the process of designing protocols to record EEG signals needed for event detection, the underlying neurophysiology knowledge imposes a rigid constraint on the experimental design so that movements used to design BCI systems are not

natural; and this can be a severe limitation for BCI systems designed for rehabilitation purpose. In the present thesis, it is hypothesised that changes in EEG modulated by means of slow voluntary reaching hand movements contain unique features to predict reaching hand movement initiation time.

The thesis is organised into eight chapters. Following the introduction, background neurophysiology and previous work in event detection in non EEG and EEG signals are presented in chapters two and three. The third chapter also describes previous work in BCI systems, the state-of-the-art BCIs and a brief overview of direct communication between two brains. Chapter four provides an explanation of data recording and the pre-processing of the data. In chapter five, the methods and algorithms used to process the data are presented. Chapters six and seven present an interpretation and a critical evaluation of the results. Finally, in chapter eight, a conclusion of the thesis is provided followed by a future orientation of the present research.

1.3 List of contributions

The present thesis explores the possibility to detect voluntary movement initiation time using realistic voluntary hand movement. This exciting journey encompasses challenges of diverse natures. To the best of the author's knowledge, those of the challenges that are believed to be original and regarded as significant engineering and scientific contributions are listed below.

Contribution to knowledge

- Application of a new protocol to record self-initiated natural voluntary reaching hand movements for events detection.
- Development of a novel algorithm to detect movement initiation by comparing statistical distances (Mahalanolobis distance) between baseline and EEG activity signal.

- Development of a novel algorithm to detect movement initiation by comparing bootstrap distributions of the distances between baseline and activity EEG signal.

- Development of a novel algorithm to detect movement initiation by thresholding Lipschitz coefficients density function from continuous wavelets moduli maxima.

Contribution to engineering

- Design of a device to be used to perform centre out reaching and grasping movements into three directions (left, right and forward).

- Design of a cable to mechanically synchronise EEG recording and Vicon motion capture systems.

- Data recording technique integrating EEG recording and motion capture system (Vicon) for the first time.

Chapter 2

Background

In this chapter the terminology of the anatomy of the human Central Nervous System (CNS) and how its different parts communicate to generate movements are explained. Neuronal disorders and their impacts on the communication paths of the CNS are explained. The chapter also explores the various brain waves in general and particularly explains the generation of human electric brain activity and its different harmonics. The various techniques to record brain waves are investigated and finally, changes in brain waves due to movements are explored.

2.1 Neuroanatomy and motor control

2.1.1 Movements

Movements can be classified into non-voluntary movements (reflexes), voluntary movements and a mixture of both type of movements known as rhythmic motor patterns. This study focuses on voluntary movements which are the most complex types of movements for it aims to detect voluntary movement initiation. Voluntary movements differ from other type of movements as they are purposeful and goal oriented. In addition, voluntary movements are subject to learning and they require planning (Krakauer and Ghez, 2000). Voluntary movements which are considered here are those generated internally and are the result of the motor system transmitting commands to muscles

to initiate movements.

2.1.1.1 Motor neurones

A Motor neurone (MN) is the basic unit responsible for transmitting electrical information through the nervous system. MNs have a single axon and dendrites emerging from the cell body. There are upper and lower MNs. The upper MNs have their cell bodies in the motor cortex and their axons projected in the spinal chord to activate the lower motor neurones. The lower MNs have their cell bodies in the spinal chord and their axons innervate muscles and glands. MNs carry commands from the brain or the spinal chord to the muscles. MN convey electrical signals called Action Potential (AP)s which constitute the signals used by the brain to receive and transmit information to the muscular system. APs are generated from the changes in the flux of Sodium and Potassium ions through the axon's cell membrane. This current flow is controlled by ion channels in the cell membrane (Kandel et al., 2000).

2.1.2 The motor system

The CNS has three hierarchical levels which are the cerebral cortex, the brain stem and the spinal cord. The cerebral cortex sends neuronal signals either directly or through the brain stem to the spinal cord to generate voluntary movements. The brain stem modifies neuronal information from the motor cortex to control movements. The spinal cord is responsible for reflex movements. Voluntary movements result from a complex coordination of neuronal activities from the highest to the lowest hierarchical levels of the motor system as shown in Figure 2.1.

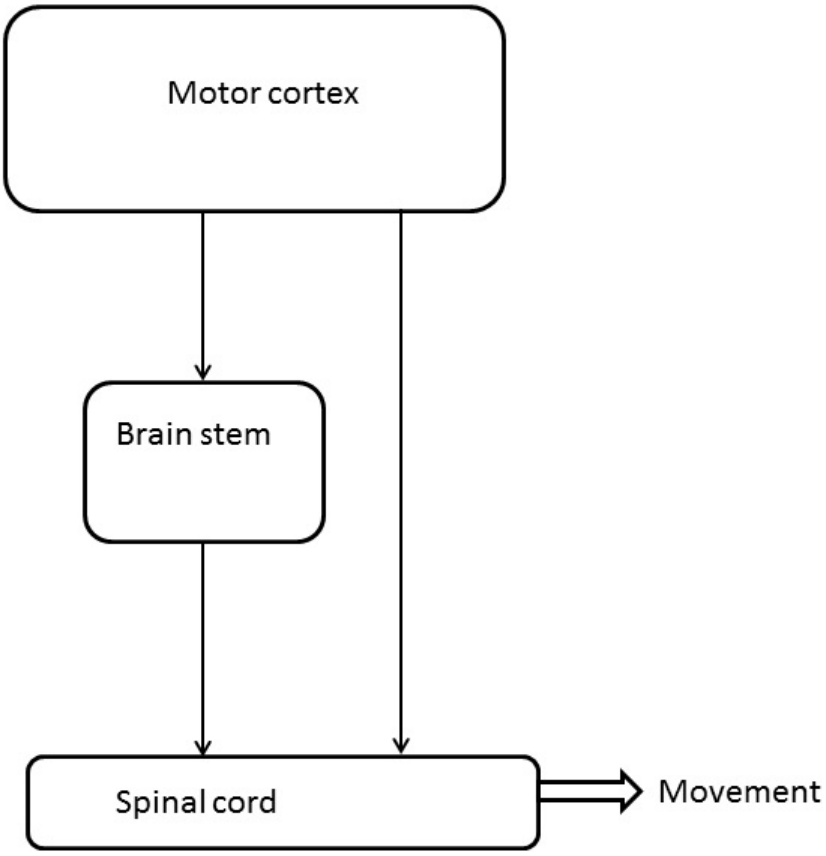


Figure 2.1: Interactions between the different hierarchical levels of the motor system to initiate voluntary movement. Neuronal information can be sent directly from the motor cortex to the spinal cord using the cortico spinal tracts or the information can be sent from the motor cortex to the spinal cord through the brain stem.

2.1.3 The motor cortex

The motor areas of the cerebral cortex are subdivided into a primary motor area and several pre-motor areas (Figures 2.2). The motor areas consist of neurones which project to the spinal cord through the brain stem, forming a communication path called the cortico spinal tracts. The Primary motor cortex (MI), also known as area 4 of Brodmann, provides the descending signals to execute movements. Cortico spinal tracts from MI directly project in the spinal cord. It was experimentally shown that electrical stimulation of different parts of MI produces movements of muscles on the opposite side of the body. These studies led to a topographical organisation of the entire head and body (Penfield and Rasmussen, n.d.; Penfield and Welch, 1951). The somatotopic organisation of

area 4 or MI is shown in Figures 2.2.

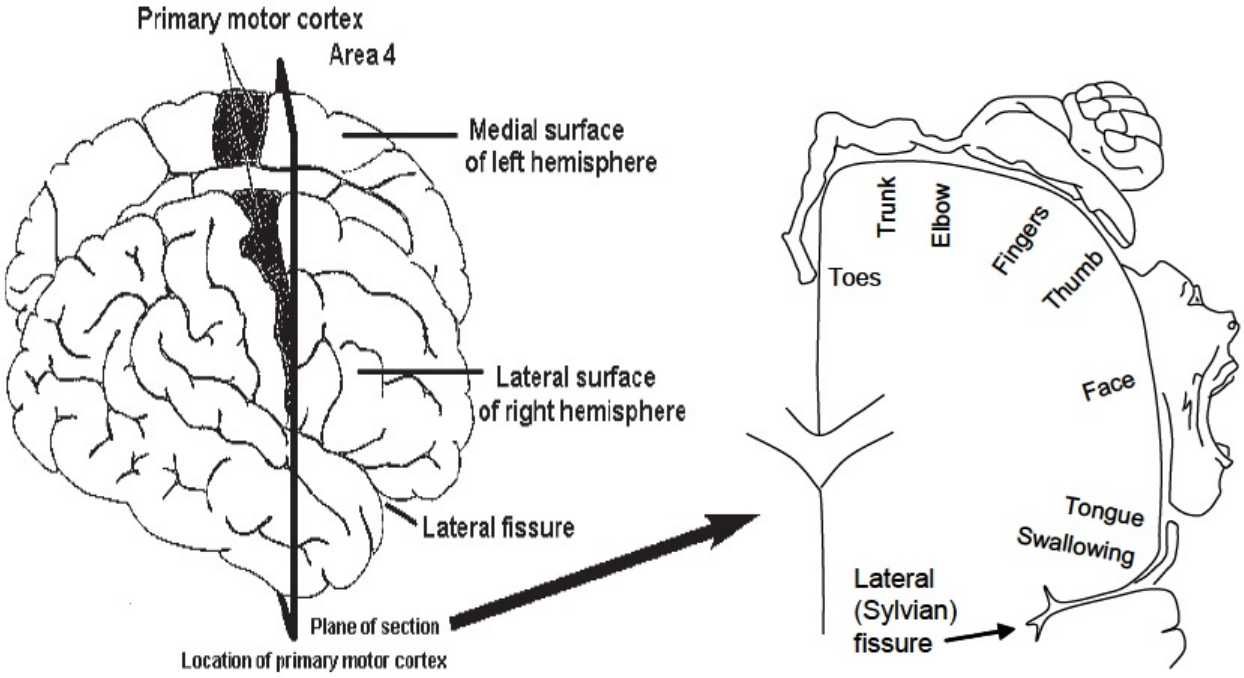


Figure 2.2: Human motor map. Left:Lateral and medial surfaces of MI. A frontal section is cut through MI. Right:The somatotopic organisation of area 4 or MI. Adapted from [www.neuroanatomy.wisc.edu/coursebook/motor1.pdf]

Electrical stimulation of premotor areas (SMA) (area 6) also elicited movements. However, the stimulation current has to be higher compare to the excitation of MI. Furthermore, movements elicited by exciting area 6 are more complex involving multiple joints such as reaching and grasping. Excitation inputs for MI and SMA and movements elicited are different. This suggests that the two motor areas control different aspect of voluntary movements.

It was shown experimentally that damage in premotor areas results in a more severe motor impairment compare to a damage in the primary motor areas. Furthermore, legions in premotor areas cause an inability to control the planning of hand kinematic in experiments with monkeys. These experiments show that premotor areas are involve in planning of voluntary movements (Kandel et al., 2000).

In the 1960s Edward Evarts showed a correlation between the activities of single neuron with spe-

cific motor behaviors in active monkeys. Evarts found that individual neuron in the primary motor cortex areas neurons are maximally activated during movement and the changes in neuronal activity begin some 100 ms or more before the onset of movement (Evarts, 1968; Evarts and Tanji, 1976). With his colleague Jun Tanji, he found that the activities in neurones change when an animal is preparing to execute a movement (i.e. waiting for a signal to move in a predetermined direction). These results pave the way to the hypothesis that there is causality between primary motor neuronal activity and motor behaviour (Kurata and Tanji, 1985; Tanji and Kurata, 1985). This hypothesis was verified by Eberhard Fetz and co-workers in 1970. They found that neurons in the primary motor cortex project to muscles through motor neurons, called corticomotorneuronal cells, establishing that neuronal activity in the primary motor cortex that precedes a voluntary movement directly influences the muscles used in the movement Figure 2.3.

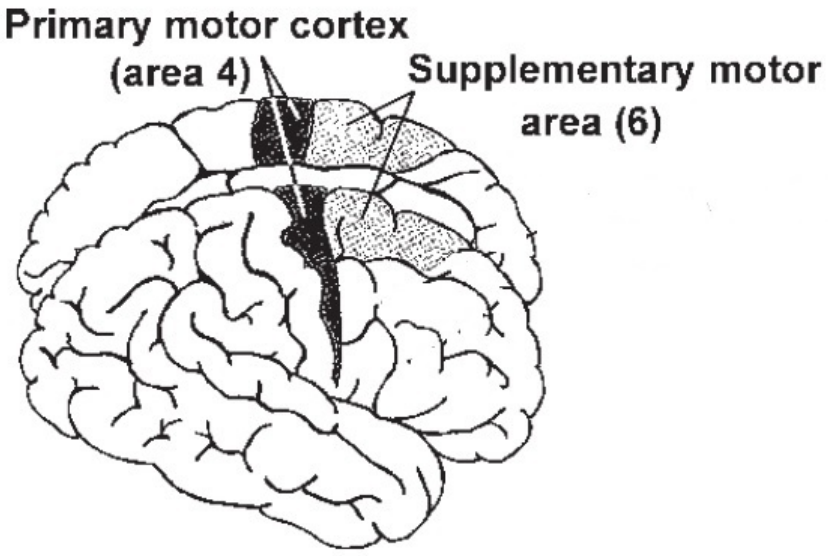


Figure 2.3: Primary motor cortex (area 4) and supplementary motor area (area 6) of the cerebral cortex. adapted from [www.neuroanatomy.wisc.edu/coursebook/motor1.pdf]

2.1.4 Movement initiation

The performance of self-paced movements are preceded by slow negative electrical change (Deecke et al., 1969). These slow potentials termed Bereitschaft potential or Readiness potential (RP) pre-

ceded voluntary movements initiation by up to one second or more suggesting that brain activity was taking place before movement onset. Libet then investigated when the conscious wish to perform a voluntary movement appear. During his experiment, RP were recorded during self-initiated sudden flick of the wrist in sessions of 40 trials each. The time associated with the awareness to move was reported and the Electromyogram (EMG) of activated muscle during movement initiation was also recorded (Libet et al., 1983a; Libet, 1993; Libet et al., 1999). RPs of the average of each session (40 trials) were considered and subjects reported that during few of the trials in each session they were planning their movements, while their movements were not planned during the remaining trials. The average RPs recorded shows an appearance of RPs at about $1050\,\mathrm{ms}$ and $550\,\mathrm{ms}$ before the beginning of the motor act and the RP were termed as type I (RP I) and type II (RP II) readiness potentials respectively. The timeline of the different events precedes the voluntary movements and the awareness of the intention to move and the initiation of muscle contraction indicating movement onset is shown in Figure 2.4.

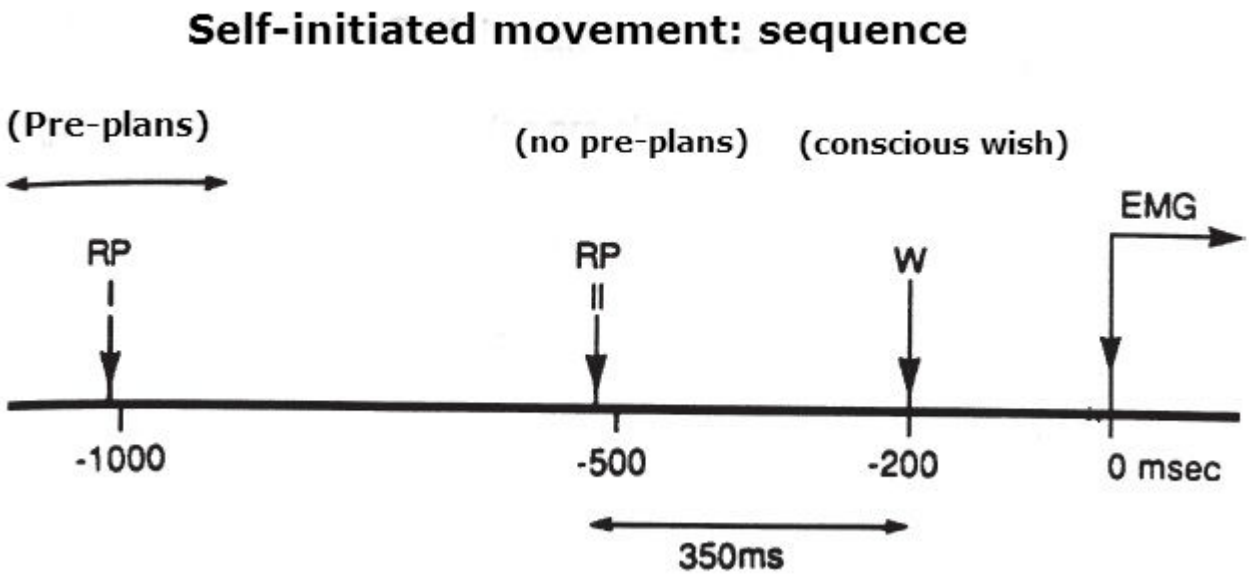


Figure 2.4: A schematic representation of the sequence of cerebral and subjective events that precede self-initiated movement. The time corresponding to $0\,\mathrm{ms}$ *is the activation of* EMG *of activated muscle. RP I (-1050 ms) indicates pre-planned reported RP while RP II (-550ms) indicates non planned reported RP.* W *(-200ms) some 350ms after RP II is the appearance of the subjective awareness to move (Conscious wish). Adapted from (Libet et al., 1999)*

2.1.5 Disorder of the motor system

The Motor System (MS) can be affected by tumours or injury. The disruption of the hierarchical elements of the MS and subsequent corticospinal tracts can lead to degenerative conditions termed as Motor neurone disease (MND) or neurodegenerative disorders. MND can lead to paralysis of various part of the human body. The paralysis depends on the part of the damaged CNS. A damage of one side of the motor area of the brain leads to the paralysis of contralateral side of the body and is known as Hemiplegia. A damage of the spinal cord can either result in the paralysis of both legs and partially the trunk known as paraplegia or in the paralysis of the whole trunk, the arms and the legs known as quadriplegia or tetraplegia. MND generally affects voluntary muscles activity and usually leaving intact, the intellectual faculties (Woolsey, 2005; Davis, 2005). MND can lead to a complete loss of of all voluntary muscle control so affected subjects can completely be locked in their bodies, unable to communicate with their surrounding environment. Motor rehabilitation research mainly endeavours to provide such subjects with alternative communication and control means through assistive technology (Wolpaw et al., 2002b; Lebedev and Nicolelis, 2006). In the absence to repair MND damage, lost functions can be regenerated through non muscular communication channels by connecting the brain to a computer in order to translate brain signals into commands to interact with the external world, a technique termed Brain computer Interface (Wolpaw et al., 2002b).

2.2 Neurophysiology

2.2.1 Brain activities

The activities generated by a normal brain are versatile and Various brain imaging techniques are used to record and visualise each type of brain activity. The magnetic fields of the brain are recorded using a technique known as Magnetoencephalogram (MEG). Metabolism and variation in blood flow can result in brain activity that is measured using a technique termed Functional magnetic resonance imaging (fMRI). Finally, electrical brain activity is measured using a technique called

electroencephalography. Important features for a brain imaging technique is to render brain activity images that are temporarily and spatially well localised (Fisch and Spehlmann, 1999; Nicolelis, 2001; Wolpaw et al., 2002b). Among the brain imaging techniques mentioned in this section, EEG is the most used method. The main reason is that EEG recording equipments are cheaper to manufacture and more portable due to their moderate size compared to other imaging equipment (Fisch and Spehlmann, 1999; Wolpaw et al., 2002b; Stanley et al., 1999; Vaughan et al., 2003). The recording methods of the brain activity are classified into invasive or non-invasive techniques. The next section provides the different brain activity measuring techniques.

2.2.2 Brain activity measuring techniques

The main brain activity measuring types are the invasive and non invasive methods. The invasive technique measures the brain activity inside the brain by mean of surgery. The non-invasive method measures the brain activity on the surface of the scalp. The invasive methods include Multi-electrode arrays and electrocorticograghy (EcoG). The non-invasive methods include , fMRI, MEG and EEG. The brain imaging techniques recording the brain's magnetic field such as MEG and fMRI result into signals with high spatial resolution and poor temporal resolution. However, these methods are very expensive and technically demanding. Furthermore, the equipment is not portable and is sensitive to noise (Vaughan et al., 2003; Kennedy and Bakay, 1998; Wolpaw et al., 2006). The recording of EEG is technically less demanding. Despite its poor spatial resolution compared to methods such as fMRI and MEG, the electric field measurement is the most used brain activity for communication and control.

2.2.3 Generation of EEG

The cerebral cortex contains cells called apical pyramidal cells arranged parallel to each other and perpendicular to the skull. Information is transmitted between the pyramidal neurones through axons and neuronal junctions called synapses in the form of action AP. Ionic currents which consist

of ions flow in or out of cellular membrane can be measured in the brain and surrounding tissues. The direction of the ionic current is the same as that of positive ions flow and the opposite direction of negative ions flow. At the synapse, the cell membrane is excited by neurotransmitters resulting into a Post Synaptic Potential (PSP). The ionic current associated with the PSP can be negative or positive. Positive current associated with excitatory Excitatory Post synaptic Potential (EPSP) will depolarise the cell membrane and cause the flow of positively charged ions inside the membrane. The extra cellular medium becomes negatively charge causing a sink. At a distant part of the neurone, the positively charged ions will leave the membrane, causing a positive extra cellular medium called a source. Negative current is associated with Inhibitory Post synaptic Potential (IPSP). For IPSP, the extra cellular medium near the synapse will be positively (source) and the extra cellular medium at a distant part of the synapse will be negatively charged (sink). Figure 2.5 shows the synaptic activity recorded as EEG form the contribution of both IPSP and IPSP potentials. A single neurone will then act like a dipole. However, the potential from a single dipole is too weak to be detected at the surface of the scalp. when a population of pyramidal cells are excited synchronously, the sum of their potentials can be detected as EEG at the surface of the scalp (da Silva and Van Rotterdam, 1999).

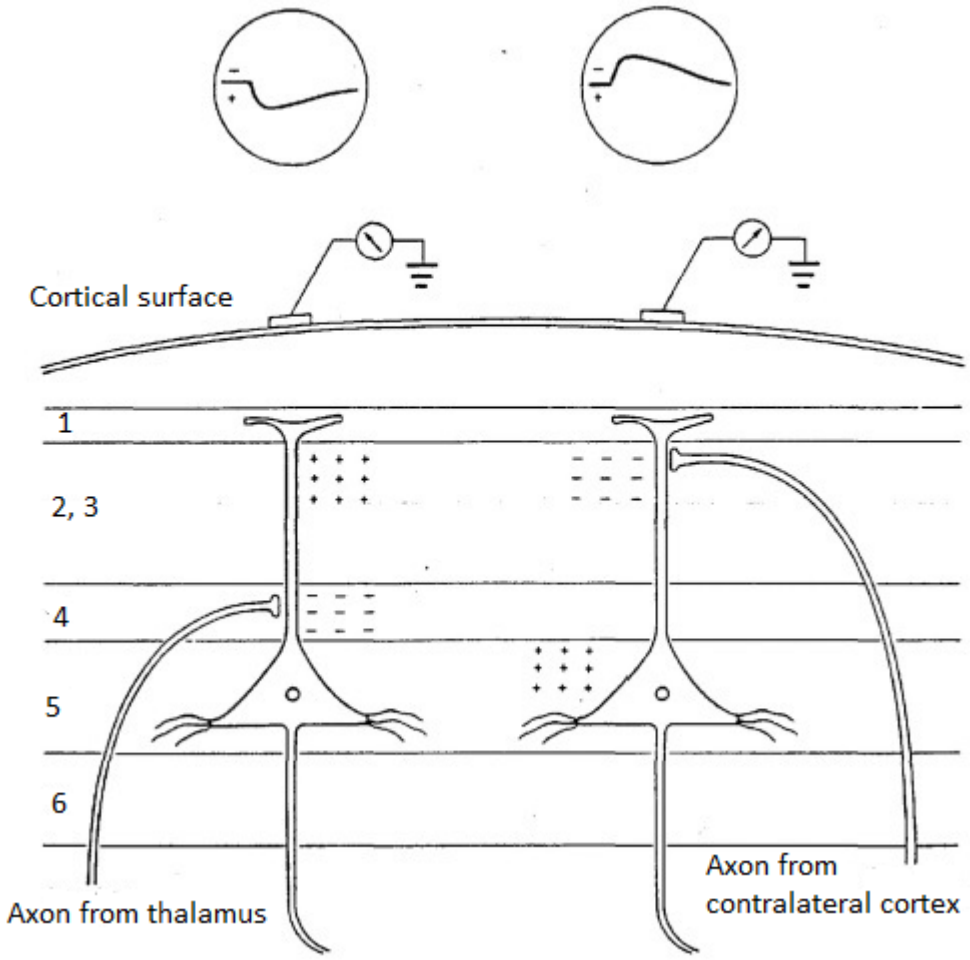


Figure 2.5: Generation of EEG signals. Binding of neuro transmitters at synaptic cleft create dipoles between the different parts of the parallel pyramidal cells vertical to the skull. The sum of these dipoles can be detected at the surface of the skull. Adapted from (Olejniczak, 2006)

2.2.4 EEG recording

In his pioneering work, Hans Berger recorded the first human EEGs in 1924 (Haas, 2003). The recorded brain waves reflect the activity in the brain and has amplitude in the range of hundreds of micro Volts. To measure the EEG, electrodes have to be placed at appropriated locations at the surface of the scalp. Various type of electrodes including dry electrodes and electrodes placements are available. However, in this section the electrode placement known as the international 10-20 electrodes placement system is described (Jurcak et al., 2007) as any other electrode placement

derives from the international 10-20 standard. According to this standard electrode placement, the distances between adjacent electrodes are either 10% or 20% of the total distance from the back of the head called inion to the front of the head called or the total distance from one side of the head to another. Each electrode location is characterised by a letter (identifying the lobe) and a number (identifying the hemisphere). Hence, electrodes located in the occipital, temporal, parietal, central and frontal lobes are represented by the letters O, T, P, C and F respectively. The letter Z represents the electrode in the midline. Even numbers (2, 4, 6, 8) and odd numbers (1, 3, 5, 7) refer to the right and left hemispheres respectively, in increasing order from the central electrode position (Z) to the peripheral positions. According to this notation, electrode Cz represents the electrode at the midline while electrode F3 represents the frontal electrode in the left hemisphere. Other standard positions systems such as the international 10-10 and 10-5 systems, with more dense electrode locations directly derive from the 10-20 system (Jurcak et al., 2007). The 10-20 standard electrodes placement is shown in Figure 2.6.

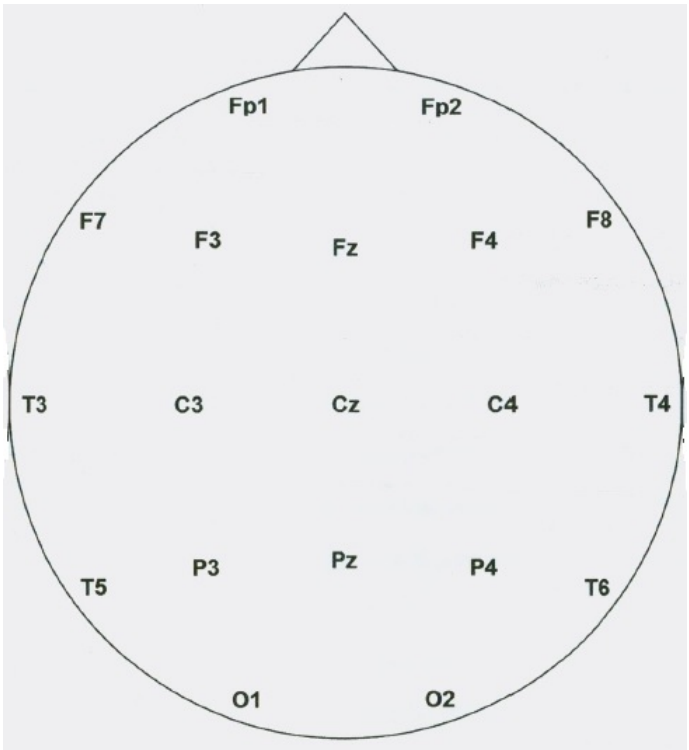


Figure 2.6: The 10-20 electrodes placement. Adapted from http://www.hksccm.org/files/Presentations/CCM_-_Basic_EEG.pdf

2.2.5 EEG oscillations

Brain activity recorded at the surface of the scalp oscillates at various frequencies and are named according to different frequency ranges. The brain waves are named according to five overlapping frequency bands in the literature. In the present study, the following five frequency ranges will be adopted:

- Alpha rhythm $7\,\mathrm{Hz}$ - $13\,\mathrm{Hz}$
- Beta rhythm $12\,\mathrm{Hz}$ - $30\,\mathrm{Hz}$
- Theta rhythm $4\,\mathrm{Hz}$ - $7\,\mathrm{Hz}$
- Delta rhythm $<4\,\mathrm{Hz}$
- Gamma rhythm $>31\,\mathrm{Hz}$

The Alpha wave, commonly known as Berger's waves are generally observed in relaxed awake subjects with closed eyes. The Alpha waves are often present as normal activity in awake adults and are characteristic for relaxed and alert state of conscientiousness. The Alpha waves are induced by eyes closed and suppressed by eyes opened. The Alpha waves have a high amplitude from $40\,\mu\mathrm{V}$ to $100\,\mu\mathrm{V}$ peak to peak and are mainly observed in the occipital region (Fisch and Spehlmann, 1999; Neidermeyer, 1999).

The Beta waves appear in wakefulness with open eyes. The Beta waves are often associated with active, busy or anxious thinking and active concentration and these activities are present when the eyes are closed or opened. The Beta waves have lower amplitude than the Alpha waves and are often recorded from the centre and front electrodes (Fisch and Spehlmann, 1999; Kozelka and Pedley, 1990).

The Delta waves are associated with moderate or deep sleep in adult subjects and can be interpreted as brain dysfunction if observed in a different state (Fisch and Spehlmann, 1999; Teplan, 2002).

Theta waves are associated with drowsiness in adult subjects. These waves are normal waves

observed with young adulthood, childhood and adolescence. The Theta waves are the main brain activity observed in the occipital and central regions in young subjects (Teplan, 2002).

The Gamma waves are mostly observed in high mental activity such as problem solving. These rhythms are also seen during consciousness and fear (Fries et al., 2007; Fisch and Spehlmann, 1999; Teplan, 2002). In this section, the brain's electrical activity is described according to the different frequency ranges and their topographic map is provided. In the next section, changes to the brain wave due to stimuli (both exogenous and endogenous) are presented.

2.2.6 EEG dynamics

It was mentioned in the previous subsection that EEG signals can be measured at particular frequency bands of interest. EEG signal recording can also be associated with internal or internal stimuli. Changes resulting from such stimuli are said to be time locked and are termed Event Related Potential (ERP). The ERPs are usually weak signals and are not distinguishable in the time domain due to random noise. They are traditionally processed by recording a sufficient number of repeats also called epochs which are coherently averaged to cancel the random noise. When changes in EEG is due to motor stimulus (actual or imagined), the resultant signal is termed as MRCP (Niedermeyer and da Silva, 2005). The most common MRCPs are the Mu rhythms associated with real or imagined movements observed in the motor area in the $13\,\mathrm{Hz}$ - $30\,\mathrm{Hz}$ frequency band.

Chapter 3

Literature review

This chapter explores the literature that is relevant in the detection of events in electroencephalographic signals. The chapter starts by discussing BCI systems which are the main application target of event detection. The relevance of the different brain waves and their use in the development of BCI systems are discussed. The very new concept of the communication between two brains is presented. The various pre-processing of brain signals and methods to extract useful features and the necessity to explore new methods are explained. Finally the different approaches to detect events with their advantages and limitations and the need to explore new event detection techniques are presented.

3.1 Type of BCI systems

BCIs can be classified as initiative and passive systems. Initiative BCI systems require subjects to generate self-regulated brain waves such as Beta or Mu rhythms. For passive BCI systems, subjects generate control brain signals as a response to a stimulus. Stimuli can be auditory, visual, tactile or movements (reel or imagined) and The resultant brain signal is called a ERP (Pfurtscheller, 2004). BCI systems can be described as dependant or independent. A dependent BCI does not use the brain's normal output pathways to carry the message, but activity in these pathways is needed to

generate the brain activity such as EEG that does carry it. A dependent BCI is merely an alternative method for detecting the message carried in the brain's normal output pathways. An independent BCI does not depend in any way on the brain's normal output pathways. The message is not carried by peripheral nerves and muscles, and, furthermore, activity in these pathways is not needed to generate the brain activity such as EEG that does carry the message. Independent BCIs provide a new output pathways to the brain and are therefore of great importance for people with severe neuromuscular disabilities. (Wolpaw et al., 2002b). Dependent BCIs are relevant for rehabilitation purpose where it is essential to regenerate or augment motor function. On the other hand, independent BCIs are important for communication purpose where subjects are completely locked in their bodies.

BCIs can also be categorised according to their mode of operation. For instance synchronous BCIs generate commands as a response to a given external cue. Synchronous BCIs have the limitation that they do not provide a natural way of communication. For asynchronous BCIs, output commands are generated spontaneously, hence providing a more natural way to interact with the user's external environment. During synchronous BCIs communication, the time when a command output is generated is irrelevant as it is dictated by the protocol while during an asynchronous BCI operation, a detection of the output command is crucial. The detection of the output command is termed as event detection and is the aim of the present dissertation. BCIs can be described according to the way the signals they use were recorded. Invasive BCIs use brain waves recorded by implanting electrodes in the brain by means of surgery. Non invasive BCIs use brain activity recorded at the surface of the scalp. Another alternative way to classify BCIs is according to the principles used to generate their control signals (Guger et al., 2013; Allison et al., 2007). BCIs are classified according to the methods they use to modify the brain activity and this method is known as paradigm. The most common mental strategies used to implement BCI system to date are Visual Evoked Potentials (VEP), SCP, P300 evoked potentials, Mu and Beta rhythms and cortico neuronal action potentials.

3.2 Brain signals used in BCIs

Various type of brain signals are used in BCI systems. This section provides the most used brain signals, their neurophysiological origins and their applications.

3.2.1 Visual evoked potentials

During this mental strategy, visual stimuli recorded from the scalp over the visual cortex are used as a control signal. EEGs are generated by presenting a 8×8 grid of symbols to a subject who selects a symbol of choice. Subgroups of the 64 symbols undergo an equivalent red or green alternation or a fine red/green check pattern alternation 40-70 times per second. Each symbol is included in several subgroups and the entire set of subgroups is presented several times. Each subgroup's EEG amplitude about $100\,\mathrm{ms}$ after the stimulus is computed and compared to a EEG template already established for the user. From these comparisons the system determines the symbol that the user is looking at. Another method to use VEPs is to present the subject with buttons on a screen flashing at different frequencies. The user looks at a button and the system determines the photic driving response over the visual cortex. When this frequency matches that of a button, the system concludes that the user wants to select it (Vaughan et al., 1996; Sutter, 1992).

3.2.2 Slow cortical potentials

SCPs are slow negative potentials generated in the motor cortex during movements execution. Negative SCPs are related to movements while positive SCPs are associated with reduced cortical activation. People can learn to control SCPs and thereby control movement of an object on a computer screen. A BCI using this method is the Thoughts Translation Device (TTD) (Kübler et al., 1999).

3.2.3 P300 evoked potentials

Routine or frequent visual stimuli typically produce in the brain signal over the parietal cortex a positive peak at approximatively $300\,\mathrm{ms}$. In BCI implementation using this method, the subject is presented with a 6×6 letter matrix, in which, in short intervals, one of the rows or one of the columns of the matrix is flashed. Fixation of the user to one letter/item elicits a VEP component called P300. A BCI system based on P300 evoked potential was developed by Donchin (Farwell and Donchin, 1988).

3.2.4 Mu and Beta rhythms

In awake people, primary sensory or motor cortical areas display $8\,\mathrm{Hz}$ to $12\,\mathrm{Hz}$ brain activity when they are not engaged in processing sensory input or producing motor output. This activity is called the Mu rhythm when focused over the somatosensory or motor cortex. These Mu rhythms are usually associated with Beta rhythms in the $18\,\mathrm{Hz}$ to $26\,\mathrm{Hz}$ frequency range. Some Beta rhythms are harmonic with the Mu and some are independent from Mu rhythms. Both Mu and Beta rhythms are associated with those cortical areas which are directly connected to the brain's normal motor output; hence, they present a potential mean for BCI based communication. Movement or preparation of movement is typically accompanied by a decrease or an increase in the Mu and the Beta rhythms. The decrease of the MU and Beta rhythms before movement is termed Event Related Desynchronization (ERD) whereas the increase of the rhythms after movement in called Event Related Synchronization (ERS).

3.2.5 Steady-state visual evoked potentials

Steady-state Visual Evoked Potentials (SSVEP) generally are considered as a continuous visual cortical response evoked by repetitive stimuli with a constant frequency on the central retina. As a nearly sinusoidal oscillatory waveform, the SSVEP usually contains the same fundamental frequency as the stimulus and some harmonics of the fundamental frequency (Regan, 1977).

3.3 Brain computer interface

BCIs provide an alternative communication channel to people with mobility disability such as stroke or Amyotrophic Lateral Sclerosis (ALS). BCI systems are implemented using various signals as reported in section 3.2. Spelling devices were implemented using P300 (Donchin et al., 2000). A thought translation device was designed based on SCP (Niels et al., 2000; Hinterberger et al., 2004) and sensory motor rhythms BCI systems were designed by (Wolpaw and McFarland, 1994; Pfurtscheller et al., 2006). Both SCP and sensory motor based BCI systems require a long term training of the subjects to generate specific brain patterns (Wolpaw and McFarland, 2004), putting a mental load on the subjects and it was clinically reported that patients have difficulties to cope with long term BCI training (Birbaumer, 2006). An alternative strategy to develop BCI systems alleviating the mental load on the subjects is the use of natural movements. External cued hand movements were used to detect finger tapping (with either hands) movements intention (Bai et al., 2007; Blankertz et al., 2003) and wrist flexion (*Prediction of human voluntary movement before it occurs*, 2011) before the movements occur. Only one self-paced reaching hand movements were detected during movement planning phase (Lew et al., 2012).

3.4 BCIs state of the art

Up to the early 2000s, only the five research groups led by Jonatan Wolpaw in Albany, Niels Birbaumer in Tubingen, Gert Pfurcheler in Graz, Jose del Millan in Martigny and the Berlin BCI research group were the most prominent BCI research groups. The trend of BCI research has grown exponentially today and about 200-300 laboratories are involved in BCI research and almost each BCI research laboratory is equipped with a BCI system (Guger et al., 2013; Allison et al., 2007). The medical engineering company, g.tec[1] each year organises a BCI award. The main objective of the annual award is to promote excellence in the field of BCI and to make BCI systems more

[1] g.tec is a leading provider of BCI research equipments.

powerful, more intelligent and more applicable. During the BCI award in 2010, 57 BCI projects were submitted by the participants. Table 3.1 shows the properties and certain application areas of the submitted BCI projects. The properties of the submitted projects show that real-time BCIs are the most dominant implementation (65.2%). Motor imagery is the most used mental strategy to control BCI (40.4%) followed by P300 (29.8%) strategy. EEG is by far the most recording technique (75.4%) followed by fMRI and EcoG (3.5%) (see Table 3.1(a)). The overview of BCI applications shows that most BCIs are developed for spelling (19.3%) followed by general control (17.5%). However, 12.3% of submitted projects concern developing BCI platform or improvement of technology table 3.1(b) (Guger et al., 2013).

Table 3.1: BCI strategies and application of the submissions to the 2010 BCI Award [Rearranged from (Guger et al., 2013)]

(a) BCI strategy

BCI strategy	% (N=57)
Real-time BCI	65.2
Off-line algorithms	17.5
P300	29.8
SSVEP	8.9
Motor imagery	40.4
EEG	75.5
fMRI	3.5
ECoG	3.5
NIRS	1.8

(b) BCI applications

BCI application	% (N=57)
Stroke	7.0
Spelling	19.3
Wheelchair/robot	7.0
Internet/VR	17.5
Control	17.5
Platform/Technology	12.3

Another trend in BCI implementation is the importance of the use of cues in most of the proposed BCI protocols. In their study, Alonso-Valerdi et al. (2015) investigated the importance of cues as a means to detect more easily, motor imagery BCI control tasks. They found that brain patterns contributing in detecting control tasks are located approximately $250\,\mathrm{ms}$ after cue onset and

the detection accuracy of cues delay task detection do not depend on the type of cue used during the protocol. The study showed that the classification accuracy of movement imagery depends on the perception and the cognition of the cue in use rather than the cue itself.

In their study, Koo et al. (2015) presented a hybrid BCI system composed of an Near Infrared Spectroscopy (NIRS) and EEG recording where the NIRS part of the system detects the imagery task and the EEG part classifies the task as left or right movement imagination. The hybrid system uses a cue to detect the movement initiation by presenting a visual cue to instruct the participant to move to the right and the left direction. This paradigm is not a self-paced strategy as claimed by the authors as it uses visual cues to indicate the participants in which direction to move. In this hybrid online BCI system, the motor imagery detection is based on a NIRS system while the motor imagery classification is based on an EEG system.

Salvaris and Haggard (2014) investigated the detection of movement intentions in a precued delayed response task. The task consisted in asking the participants to perform right or left hand finger tapping. The precue indicated the participants which hand to use and the GO cue instruct the execution or the imagination of the movement. Three tasks were considered during the experiments, namely a movement execution, movement imagery and a free choice task where the subjects were allowed to follow or to ignore the precue instructions. The major finding of the studies by Salvaris and Haggard (2014) was that free choices movement intentions were less well detected compared to instructed intentions. The research suggested that the poor decoding of free choices intentions were due to the fact that intentions of free movement are subject to change compared to instructed intentions. The author further argued that during the free choice paradigm, there was no difference in the detection accuracies whether the subjects follow the precue or merely ignore it. The study by Salvaris and Haggard (2014) showed the use of cues in self-paced BCI protocols as key elements to control the timing of the event during self-paced control tasks. Even though the results are based on free choice intentions, those intentions are controlled by a precue and a Go cue and therefore the cues can not be totally dissociated from the movements.

3.5 Brain-to-Brain interface

By using the BCI technique, the brain is able to communicate in a more direct way without using muscular channels. A more advanced step in such communication involved demonstration of bidirectional Computer Brain Interface (CBI) and cortical spinal communication in a monkey (Shanechi et al., 2014). Communication through brain waves has taken a step further by connecting a brain to another brain for communication purpose, a technique termed Brain-to-Brain Interface (BBI). BBI was demonstrated by connecting the brains of two rats through an internet connection (Pais-Vieira et al., 2013). During the experiment, cortical microelectrode arrays were used to record the neural activity of encoder rats performing a motor task to identify a visual stimulus and make a choice between one of two levers to get a reward. The MI neural activity of the encoder rat is recorded and transmitted to the MI of the decoder rat, resulting in the decoder rat mimicking the encoder rat, leading to a direct transfer of information between the two brains of the encoder and decoder rats. The BBI systems mentioned so far use invasive techniques and are mostly limited to animal models. A non-invasive BBI was developed by transmitting information from a human brain to a rat brain. During the procedure, the EEGs in the human brain were recorded and transmitted to the part of the MI of the rat controlling its tail (Yoo et al., 2013).

A direct brain-to-brain communication using a BCI from the sender side and a Transcranial Magnetic Stimulator (TMS) on the receiver side was demonstrated during a pilot study. For the participant playing the sender's role, EEG signals were recorded. Motor imagery of the right hand movement was used by the sender to learn to control the vertical motion of a cursor on a computer screen. The subject controlled the cursor by increasing or decreasing the Mu rhythms. The computer translated the power in the Mu band to vertical movement of a cursor, which provided visual feedback to the sender. Specifically, the decrease in power that accompanied right hand motor imagery was mapped to upward movement of the cursor, while a lack of suppression in the mu band caused downward cursor movement. This information was encoded and transmitted to the receiver system.

On the receiver side, the upper left region of the cortex, was stimulated using the TMS coil until an ideal position was found to stimulate the motor region that controls the extensor carpi radialis and the intensity that causes an upward jerk of the hand in determined. Notice that, because this muscle extends the wrist, it produces an upward movement of the hand. During the experiment, the sender's intention to move the hand caused the TMS machine to fire a single pulse causing the hand to press down a touchpad (Rao et al., 2014).

The transmission of information between two brains separated by a long distance is theoretically expected from quantum entanglement. In quantum physics, entanglement is a physical phenomenon during which pairs or groups of particles interact. According to this principle, the measurement of the quantum state of each particle in the group is correlated with the others, irrespective of their distance. For instance considering two electrons which are very far apart, measuring the spin of one electron gives the spin of the second electron according to the entanglement principle. The rational of the generalisation between physics variables and biological (mental) variables is discussed in (Filk and Römer, 2011). By analogy with physical systems, it is assumed that entanglement can occur if the two brains involved are complementary parts of a single system and their interactions at a distance are studied. To demonstrate BBI between two brains separated by a long distance, five couples with long time friendship were used to test interaction between two human brains (Tressoldi et al., 2014). The sender received a sequence of silence-signal events of two and half and one minute, respectively, and the receiver was asked to mentally connect to his or her partner. The brain activities in the sender and the receiver brains were analysed to fine the coincidence between the sequence of events between the two brains. It was found that the overall coincidence percentage of correct coincidence was 78%. Furthermore, there was evidence of a robust statistically significant correlation (mainly in the Alpha band) in all but the Beta EEG frequency bands. The experiment was repeated on pairs in two laboratories located approximatively 190 Km from each other. Figure 3.1 reports the alpha band normalised power spectrum values recorded in fourteen channels of the EEG activity of a selected pair where the sender and receiver power spectrum values are denoted by T and R respectively. The figure shows a strong correlation in the Alpha

bands both for the silence and signal events (Tressoldi et al., 2014).

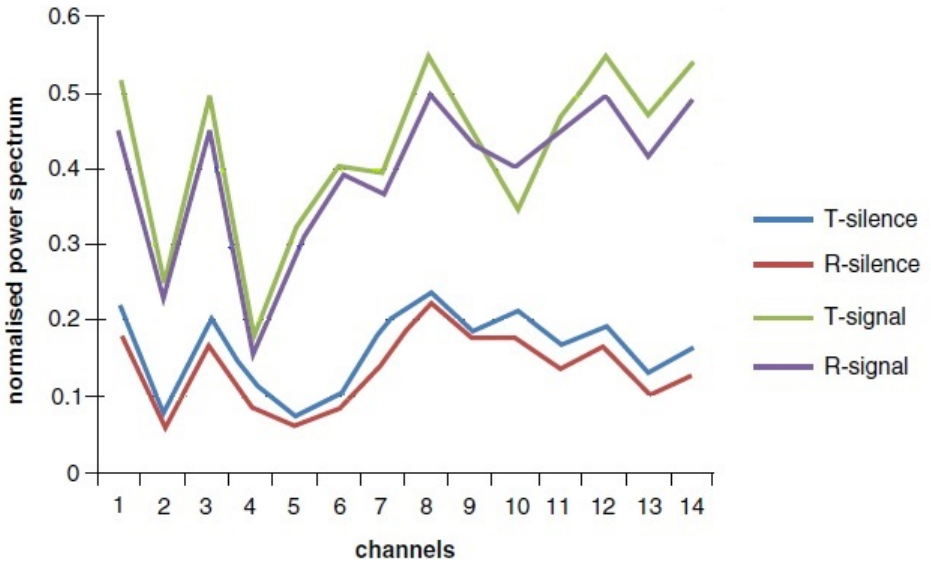


Figure 3.1: Normalised power values of the alpha band recorded on fourteen channels of EEG activity related to the silence and signal events for a selected pair of subjects. T-silence and R-silence represent the sender and receiver's power values related to silence events while T-signal and R-signal represent the sender and receiver's power values related to signal events. Rearranged from [(Tressoldi et al., 2014)]

The first brain to brain communication between two human brains was achieved using a BCI system from the sender side and a CBI [2] on the receiver side. During the direct brain to brain communication, the BCI system from the sender's side uses motor imagery (imagination of hand movement) to encode brain activity in a binary (0 and 1) form. The subject is shown a computer screen with two targets. An 'up, right' target representing 1 and a 'down, right' target representing 0. The motor imagery task is associated with a ball moving from left to right to randomly hit one of the targets. Hitting 'up, right' encodes the brain activity into binary 1 and hitting 'down, right' encodes the brain activity into binary 0 and the generated binary sequence is transmitted immediately to the CBI system to the receiver's side. The CBI system consists of a robotised TMS. The TMS system is used to localise a hotspot in the right occipital area of the receiver subject, generating a light flash with an appropriated stimulation intensity. The same stimulation results in

[2] A CBI is a system that connects a computer to the brain so that the brain interprets (decodes) information from the computer. This operating mode is the reverse of a BCI operating mode where the computer interprets information from the brain.

no flash when the TMS coil is rotated by 90°. When a bit stream is received by the CBI system, the robotised TMS system delivers a series of stimulations in such a way that the coil is oriented to produce a flash when the transmitted bit is 1 or the coil is oriented to produce no flash when the transmitted bit is 0.

Having described EEG signals, changes they undergo and their use in BCIs and BBIs systems, the next section will deal with traditional methods used to pre-process EEG signals to remove some noise and improve the Signal to Noise Ratio (SNR) for further processing.

3.6 EEG preprocessing

Various techniques are used to pre-process the recorded EEG signals in order to increase its quality. The general purpose of these techniques is to reduce artefacts due to unwanted phenomena such as eye blink, swallowing, various undesirable muscular artefacts, and the influence of neighbouring electrodes.

3.6.1 Spatial filters

Brain activity recorded at a given location is always in reference to another location called a reference location. Ideally, the potential at a specific location should measure only the activity at that site. The reference location should be silent, for any form of activity at that site will influence the resultant activity at the recorded site. One solution is to choose an inactive location such as the nose, ear lobes or the mastoids. This choice of reference sites proved to be problematic, as electrodes close to the reference sites will result in a very weak potential difference, for their potentials are close to that of the reference sites. To overcome this issue, non cephalic reference can be used instead (Lehtonen and Koivikko, 1971). Unfortunately, such reference sites are susceptible to interference from muscular (EMG) or heart (Electro Cardiogram (ECG)) activity. In the next sections, techniques using distant cephalic location as a reference are presented. This technique is known generically as spatial filtering. The term spatial filtering refers to the enhancement of the

EEG quality based on the spatial location of the electrodes used to record the signal.

3.6.1.1 Surface Laplacians

Surface Laplacian filters are mainly used to improve the spatial resolution of the EEG signal. The Laplacian is the second derivative of the EEG potential at the location of interest. It can then be interpreted as the divergence or the fading of the electric field at the surface of the scalp with respect to the radial sources immediately below the considered location (Nunez et al., 1994; Zhou, 2012). The Laplacian is evaluated by combining the value at that location with the values of a set of surrounding electrodes. When the source is well localised at the recording location, the Laplacian can be derived by using the nearest neighbour (adjacant) electrodes, and this type of Laplacian is called the small Laplacian. When the calculation is based on the next nearest neighbour electrodes, it is called the large Laplacian (see Figure 3.2). When the source signal is well localised at the recording site, the small Laplacian is a good approximation, while when the source is less localised or varies in location over time, the large Laplacian is expected to provide a better approximation of the Laplacian (McFarland et al., 1997).

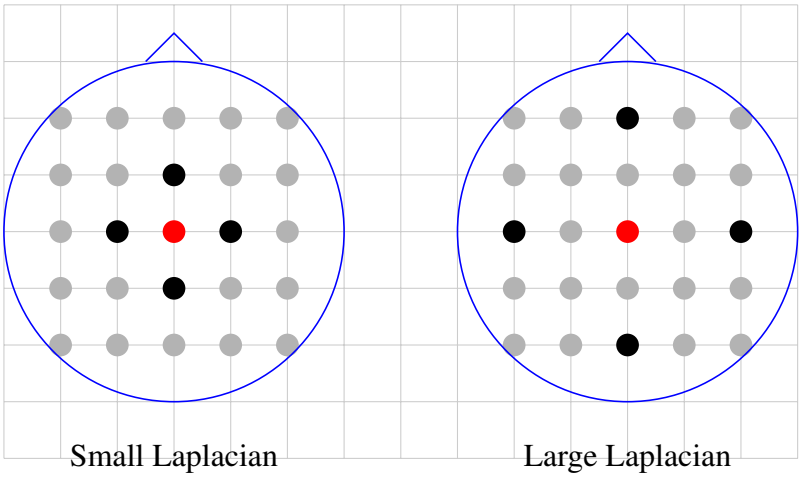


Figure 3.2: The different spatial filtering using the small and large Laplacians. In the figure, the red electrode location is measured with respect to the average of the black electrode positions.

The small and large Laplacians only consider the nearest or next nearest neighbour electrodes in order to compute the Lapalcian at a given location irrespective of the electrode configuration.

However, an high spatial resolution can be achieved by using a large electrode configuration and an alternative approximation of the Laplacian considers all the neighbour electrodes of the electrode configuration. Using such a large number of electrodes covering the full head, the surface of the scalp become curved, having a consequence on the application of the surface Laplacian in general. The new approximation of the Laplacian will be illustrated using the 3×3 electrode configuration shown in Figure 3.3. To apply the surface Laplacian on EEG data, it is assumed that the scalp surface is approximatively flat. This assumption is true when few electrodes are considered, leading to a definition of the surface Laplacian of a potential V in a rectangular coordinate system (x,y) by Equation [3.1] as

$$L(V) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2}. \qquad [3.1]$$

When it is further assumed that the recording sites are regularly sampled at discrete locations (assumed to be a grid), the surface Laplacian at a central location (i, j) of the grid $L_{s_{i,j}}$, can be estimated by averaging the nodes $\varphi_{k,l}$ at the direct neighbours of (i, j) by Equation [3.2] (Bradshaw and Wikswo Jr, 2001; Carvalhaes and de Barros, 2015)

$$L_{s_{i,j}} \approx \frac{1}{d^2} \left(\varphi_{i-1,j} + \varphi_{i+1,j} + \varphi_{i,j-1} + \varphi_{i,j+1} - 4\varphi_{i,j} \right). \qquad [3.2]$$

A schematic representation of the topology considered is the present estimation of the surface Laplacian is shown in Figure 3.3.

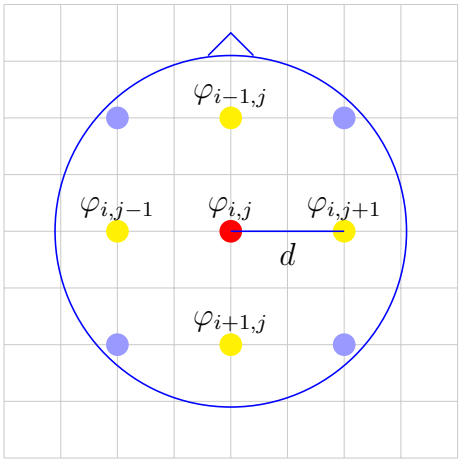


Figure 3.3: A schematic representation of electrode grid to illustrate the computation of the surface Laplacian. The red position represents the position at which the Laplacian is estimated while the positions in yellow are the neighbour positions considered is the Laplacian computation. d is the distance between the position at which the Laplacian is estimated and its neighbouring sites.

Equation [3.2] is known as Hjorth's approximation and has a limitation as it can only approximate the surface Laplacian of a central node (Hjorth, 1975). The evaluation of all electrode sites can be obtained using Figure 3.4. Using this electrodes arrangement to evaluate the Laplacian of the 'top, left' node in Figure 3.3, which is at position $(1, 1)$, the nodes along the 'top, right' side $\{(1, 2), (1, 3)\}$ and those along the left side $\{(2, 1), (3, 1)\}$ of the grid will be considered in the computation of the Laplacian. When evaluating the small and large Laplacians at this location, only nodes $\{(1, 2), (2, 1)\}$ or nodes $\{(1, 3), (3, 1)\}$ will be considered respectively.

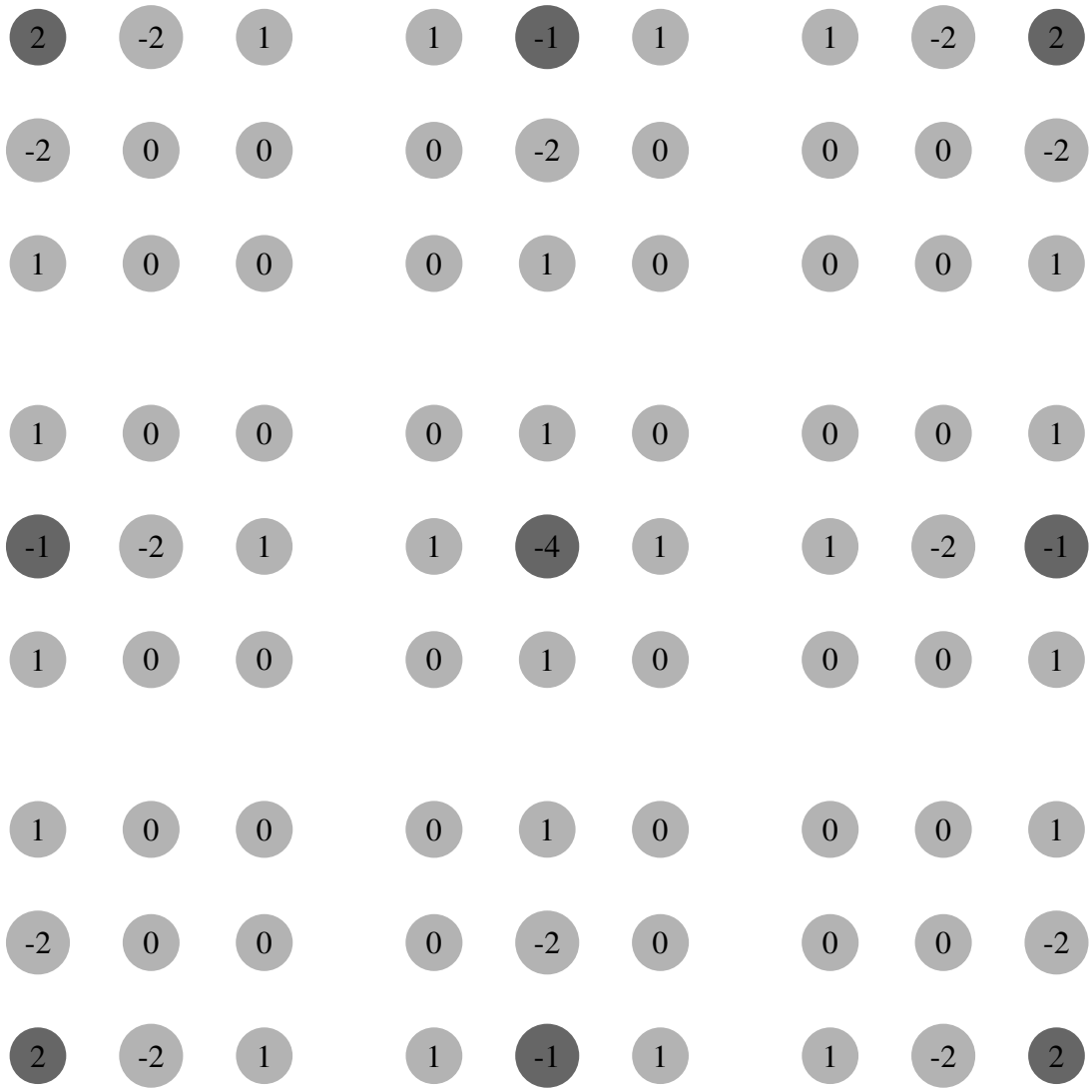


Figure 3.4: The geometric placements and weights to approximate the surface laplacian at each electrode location on a 9×9 *grid. The dark nodes are the locations at which the surface Laplacian is approximated while the grey positions with non zero weights are the neighbouring positions used in the computation of the Laplacian. The positions not considered in the Laplacian calculation have weights 0. In the figure, the surface Laplacians are evaluated row wise.*

The Laplacian approximation using Figure 3.4 can be represented in a matrix form called the Laplacian matrix (L). The Laplacian matrix can be calculated for each electrode configuration and does not require EEG data for its computation. L is constructed by evaluating row by row the surface Laplacians in Figure 3.4. When considering the 3×3 electrode grid in Figure 3.3, L is a 9×9 matrix which row corresponds to the surface Laplacian of the given electrode locations. As

in the computation of the surface Laplacian, each location has four neighbouring electrodes, L is a sparse matrix with only five non zero elements on each row. The locations not contributing to the Laplacian calculation are assigned zero weights. The Laplacian matrix of the electrode grid considered in the present section can be written in matrix form as

$$\mathbf{L} = \frac{\mathbf{1}}{\mathbf{d^2}} \begin{pmatrix} 2 & -2 & 1 & -2 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & -2 & 0 & 0 & 1 & 0 \\ 1 & -2 & 2 & 0 & 0 & -2 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & -2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -2 & 0 & 0 & 2 & -2 & 1 \\ 0 & 1 & 0 & 0 & -2 & 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 & -2 & 1 & -2 & -2 \end{pmatrix}.$$

The first row of L corresponds to the weights used to approximate the surface Laplacian at the 'top, left' node in the electrode grid in Figure 3.3. The second row of L corresponds to the weights used to approximate the Laplacian at the second node of the first row of the grid. Thus, each row of L corresponds to the approximation of the surface Laplacian of each location of electrode along the rows of the electrode grid. For instance the node at the 'top, left' of the grid located at position $(1, 1)$ has neighbouring positions $(1, 2)$, $(1, 3)$ on the upper side of the grid and positions $(2, 1)$, $(3, 1)$ on the left side of the grid. Leading to a surface Laplacian $(L_{s_{1,1}})$ at this node defined by Equation [3.3]

$$L_{s_{1,1}} \approx \frac{1}{d^2} \left(\varphi_{1,3} - 2\varphi_{1,2} + \varphi_{3,1} - 2\varphi_{2,1} + 2\varphi_{1,1} \right). \qquad [3.3]$$

The various methods to evaluate the Laplacian can be used as spatial filters to enhance the quality of the EEG. However, these methods have some limitations. The small and large Laplacians require

the EEG sources to be highly localised which in turn requires the use of source localisation to ensure such a condition to confidently apply these spatial filtering techniques. On the other hand, the application of the general Laplacian relies on the assumption that the head model is flat, an assumption which is inappropriate when a dense electrode set covering the full head, resulting in a curved head surface.

3.6.1.2 Common average reference

The introduction of the Common Average Reference (CAR) method was motivated by the alleviation of any contribution of the reference site when measuring brain activity at a given site. To elaborate, it is assumed that any brain activity observed across the head at locations other than the measured site result from the reference site and should be subtracted from all the sites. This leads to a reference free recording method. Furthermore it was proven that the sum of potentials in a conductive sphere is equal to null (Bertrand et al., 1985). Based on the two previous assumptions, the average reference is mathematically calculated by subtracting the mean of all the potentials of all the sites from the potential of a given site. For a reference location with a potential V_{Ref}, the potential at location *n* is defined as $V_n - V_{Ref}$. The average reference is calculated by subtracting the average potentials of all locations and is given by Equation [3.4] as

$$\begin{aligned} V_n - V_{Ref} - \{(V_1 - V_{Ref}) + (V_2 - V_{Ref}) + \ldots + (V_2 - V_{Ref})\}/n = V_n - V_{Ref} \\ + \{V_{Ref} - [V_1 + V_2 + \ldots + V_n]/n\} = V_n - [V_1 + V_2 + \ldots + V_n]/n \end{aligned} \quad [3.4]$$

If the average reference assumption is met, the $V_1 + V_2 + \ldots + V_n = 0$ and in equation 3.4, only V_n remains. This clearly indicates, that the average reference consists in re referencing the data by defining a reference site which corresponds to zero. For such a reference point, any discrepancy between it and another site is due to the other site, so that the brain activity measured at that site contains the discrepancy (Dien, 1998; Osselton, 1965; Offner, 1950). The CAR method is a spatial filtering method which is computationally efficient and the method is suitable when a dense electrode configuration covering a large recording surface is considered.

3.6.2 Temporal filters

Brain waves recorded at the surface of the scalp are usually mixed with other signals which are not of interest. The unwanted signals termed as noise can be of technical or biological nature. Technical noise usually results from measurement equipment or other equipment in the vicinity of the recording premises. Noise from biological source can include sweat at the recording site, the interface between electrode and the recording sites, spontaneous alpha waves, the electrical activity of the heart, various unwanted movements during recording and ocular artefacts. These background noise signals have amplitude which are very high compared to the amplitude of the EEG signal of interest. Temporal filters are mainly used to remove the background noise in order to improve the EEG signal of interest. To select EEG components of interest, temporal filters such as Band Pass Filter (BPF), Low Pass Filter (LPF) and High Pass Filter (HPF) are used. BPFs are mostly used to select EEG components in specific frequency ranges. Other types of temporal filters such LPFs or HPFs are used to remove slow varying or high frequency EEG components respectively. Electrical equipment often generate $50\,\mathrm{Hz}$ power line noise during EEG recording. When the EEG is contaminated by $50\,\mathrm{Hz}$ noise during recording, the power line noise is removed from the recorded EEG signal using a dedicated type of temporal filters known as Notch filters. Most temporal filters are implemented using classic filter design methods such as Infinite Impulse Response (IIR) or Finite Impulse Response (FIR) filter implementations.

3.7 Feature extraction

Recorded brain waves result in time series which represent the evolution of the brain wave as a function of time. This representation of brain activity is often of little use if the brain activity of interest is buried in noise. To enhance the quality of the EEG signal of interest, signal processing methods are used to transform the signal from the original domain (time domain) into another domain which emphasises, the characteristic features of the EEG of interest, a technique termed as

feature extraction[3]. In the next subsections, the most important techniques used to remap EEG data to extract characteristic features are presented.

3.7.1 Fourier techniques

One of the most popular technique to map time series data from the time to the frequency domain is the Fourier Transform (FT). The FT is used to resolve a time domain signal into the frequency domain, showing the various frequency components of the signal. Although sufficient for some application, this transformation is of limited use when it comes to the location of the frequency components in time. To overcome this limitation, the signal can be segmented and the FT can be performed on those segments, a method termed the Short Time Fourier Transform (STFT). To perform the STFT, it is assumed that the length of the signal segment is sufficiently short to guarantee that the signal is stationary[4] of the signal over the segment. When applied to a discrete signal, the FT is known as the Discrete Fourier Transform (DFT) and the Fast Fourier Transform (FFT) is an efficient implementation of the DFT.

3.7.2 Autoregressive model AR

The AR model was first applied to EEG by Zetterberg (Zetterberg, 1969; Fenwick et al., 1969). In this model, the output depends on a linear combination of the previous inputs. When the output is a linear combination of p previous inputs, the model is said to be of order p and is denoted AR(p). The AR(p) model is mathematically described by Equation [3.5] as

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + \epsilon_t. \qquad [3.5]$$

[3]The present definition of the term feature extraction was solely adopted by the author to make a distinction between preprocessing and feature extraction, for the two notions are interchangeable in the literature.

[4]The properties of a time series such as the EEG generally change over time, hence it is non stationary. It is assumed that, when considering a short segment of the time series, its properties remain constant within the selected window, making it stationary.

Where ϵ_t is assumed to be random noise with zero mean and variance σ_ϵ^2. The model has been used extensively by various researchers (Wennberg and Isaksson, 1976) simulate to simulate EEG. For example an AR model of order six (AR(6)) was used to extract EEG features to classify alphabet letters (Jansen, Hasman and Lenten, 1981; Jansen, Bourne and Ward, 1981). The AR method has been extensively used to model EEG data. However, the method models EEG by evaluating current data values with a weighted sum of past data values. When corrupted by noise, previous data values might necessarily not be the best predictors of a current value.

3.7.3 Multi-taper spectral estimation

Various methods exist to represent neuronal activity and one particularly useful representation is the spectrum. In this section a spectrum with a Gaussian property is derived using a multi-taper spectral estimation approach (Percival and Walden, 1993). Given a time series X_n of length N, the general conventional estimate of its spectrum is defined by Equation [3.6] as

$$S(f) = |\frac{1}{N}\sum e^{(2\Pi i f n)} h_n X_n|^2 = |\tilde{X}(f)|^2 \qquad [3.6]$$

where h_n is called a taper which is generally a smooth function with falling edges. $X(f)$ is known as the Fourier transform of the sequence X_n

With the multi-taper approach several tapers are used to estimate an average spectrum. Various types of tapers are available and the family of tapers considered in the present section is a set of orthogonal tapers called slepian sequences. In particular prolate spherical functions called Discrete Prolate Slepian Sequence (dpss) providing a local eigenbasis in frequency space for finite length data sequences are described (Moore and Cada, 2004; Slepian, 1978). For a data segment of length N and a given frequency band W, the first $K = 2NW - 1$ sequences of these Slepian sequences are optimally concentrated in the frequency range $[-W, W]$. The general form of the multi-taper

estimate of the spectrum using the first K sequences is gigen by Equation [3.7] as

$$S^{MT}(f) = \frac{1}{K}\sum_{k=1}^{K} |\frac{1}{N}\sum e^{(2\Pi i f n)} u_n^k X_n|^2 = \frac{1}{K}\sum_{k=1}^{K} |\tilde{X_k}(f)|^2 \quad [3.7]$$

where $u_n^k, n = 1, 2, \ldots N$ is the k^{th} slepian sequences (Bokil et al., 2007). The spectrum estimated by Equation [3.7] assume that the time series X_n is stationary. Neurophysiological time series are not stationary. However, there is evidence that these time series are stationary over short segments. The general approach used to estimate the spectrum of a non stationary time series is to segment the time series into L disjoint segments and compute the spectrum over these segments. This results in time dependent spectrum $S(t, f)$ which represents spectral changes of the time series over time.

3.7.3.1 Choice of the bandwidth parameter

The choice of the bandwidth *W* parameter and the length of the time window *T* is critical to obtain an optimal time-frequency representation. The bandwidth parameter is used to tune the variance while the length of the time window controls the resolution of the estimated spectrum. As a rule of thumb, *W* is fixed as a small number and *T* is varied until a good resolution is obtained (Mitra and Pesaran, 1999). The bandwidth and the time widow parameters are conjointly referred to as the time bandwidth parameter *TW*. The parameters to estimate the spectrum are chosen such that the relationship between the time bandwidth *TW* and the Slepian sequences *P* satisfies the Equation [3.8]

$$TW = P - 1 \quad [3.8]$$

3.7.3.2 Confidence limit for the estimated spectrum

To estimate the time dependent frequency $S(t, f)$ using the multi-taper method, the signal was divided into L segments. In (Halliday et al., n.d.; Bloomfield, 2004), it is shown that the variance

of the time dependent estimate spectrum $S(t, f)$ can be approximated by Equation [3.9]

$$var\{S(t, f)\} \approx L^{-1}(S(t, f))^2 \quad [3.9]$$

where L is the number of segments of the disjoint segments used to estimate the spectrum. The variance estimated $var\{S(t, f)\}$ above is dependent of the frequency. This frequency dependence of the variance can be alleviated by using the logarithmic transform (Jenkins and Watts, n.d.). This transform is called the variance stabilizing transform as its application results in a non frequency dependent variance given by Equation [3.10]

$$var\{log_{10}(\tilde{S}(t, f)\} = (log_{10}(e))^2 L^{-1} \quad [3.10]$$

with the resulting estimate and 95 confidence limits at frequency f of

$$log_{10}\tilde{S}(t, f) \pm 0.851L^{-\frac{1}{2}}. \quad [3.11]$$

The multi-taper method is a powerful method to derive a time-frequency representation of a signal. The method results in a time-frequency representation spectrum with Gaussian property as the spectrum is evaluated by averaging multiple STFT spectra. One limitation of this method is the selection of the appropriated time bandwidth parameters and the number of tapers which is a trade off between the spectral resolution and the stability of of estimated spectrum. The term stability mean a spectrum with less variance.

3.7.4 Wavelet Transform

The Wavelet Transform (WT), also known as multiresolution analysis (MRA), is a technique used to map a signal in the time-frequency domain. Contrary to the FT where the decomposition uses sinusoidal waves as basis functions, the WT uses a localised function called the mother wavelet. To

resolve a signal in the time-frequency domain, the mother wavelet is scaled to match the oscillation of the signal across time. Various type of mother wavelets exist according to the application. The WT can be discrete where the time-frequency representation of the signal is evaluated only at discrete scales. The most traditional discrete WT approach uses dyadic scales where the scales are multiple values of 2^j where j represents the scale. Another type of Wavelet transform, known as the Continuous Wavelet Transform (CWT) uses continuous scale values.

Given a square integrable functions $s(t) \in L^2$ and a wavelet function $\psi(t) \in L^2$, the continuous wavelet transform W(a,b) Mallat (1989a) of the signal $s(t)$ is defined by Equation [3.12] as:

$$W(a,b) = \int_{-\infty}^{+\infty} s(t)\Psi^*_{a,b}(t)dt \qquad [3.12]$$

where $\Psi^*_{(a,b)}(t)$ represents the complex conjugate of $\psi_{a,b}(t)$.

In Equation [3.12], $\psi_{a,b}(t)$ is a shifted and scaled version of the mother wavelet $\psi(t)$. The variables a and b are the scale and shift parameters respectively. The wavelet transform $W(a,b)$ is a mapping of the signal $s(t)$ from the time domain to a scale space domain $s(a,b)$. The wavelet transform decomposes the time series into a certain number of scales corresponding to specific sub-bands. The local extrema of the moduli of the WT also known as WMM represent the positions of sharp transitions in the time series(Mallat, 1989b; Mallat and Hwang, 1992). The positions of the modulus maxima can be used to determine the change points in the time series. The WT and the STFT both provide a time-frequency representation of a time series with an important difference in the spectral properties of each method. Whilst the STFT time-frequency representation has a constant resolution, the WT time-frequency representation exhibits a good frequency resolution for slow varying time series (high scales) and good time resolution for time series with high frequency components (low scales). The length of the window used to perform the WT is crucial for the time-frequency representation as it determines the limit of the frequency component that can be analysed. For example, it is not possible to analyse frequencies below $1\,\mathrm{Hz}$ when using a data epoch of $1\,\mathrm{s}$. Furthermore, a short data window will result in a good time resolution and poor frequency resolution

and is *appropriate* for high frequencies (small scales) time-frequency resolution. A large time window will result in a poor time resolution and good frequency resolution and is *appropriate* for low frequencies (large scales) analysis.

3.7.5 Principal component analysis

PCA is a non-parametric method used to reduce the dimension of a complex data set into a lower dimension which often reveals hidden structure in the data set. PCA technique achieves the reduction of the dimension by simply projecting the original data onto a new basis called the transformation basis, which emphasises the maximum variations of the data. PCA is a very powerful method that can be used for event detection. However, it is used in the present report as a *means* to dimensionality reduction (Faloutsos et al., 1997; Parra et al., 1996). In the following sections, a brief explanation of the transformation of the initial basis and how to select the appropriate new basis in which the transformed data has maximum variation is provided.

3.7.5.1 Problem formulation

EEG data are recorded using a set of electrodes placed at different sites on the scalp. The EEG signals from each electrode are resolved into frequency bands resulting into high dimensionality and sometimes redundancy. In such a situation, there could be a twofold goal to use PCA. The PCA method could for example be used to reduce the number of channels by only selecting EEG channels those data have maximum variation. Another application of the PCA technique could be to select EEG frequency components with maximum variation when the EEG signal from a given channel is split in frequency bands.

3.7.5.2 Basis transformation

The main objective of PCA is the re-express a given data set in a new basis where the data have maximum variation. Consider the following data sample $(x_1, x_2, \ldots, x_m)$ of dimension m where

x_i represents an observation at a given time point. A data set of n such samples will be represented by a $m \times n$ matrix $\mathbf{X}$ defined as:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix}.$$

Consider a transformation that maps the original data matrix $\mathbf{X}$ to another $m \times n$ data matrix $\mathbf{Y}$ defined as:

$$\mathbf{PX} = \mathbf{Y}, \qquad [3.13]$$

where in Equation [3.13] the $m \times m$ transformation matrix $\mathbf{P}$ is defined as:

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,m} \\ p_{2,1} & p_{2,2} & \dots & p_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m,1} & p_{m,2} & \dots & p_{m,m} \end{bmatrix}.$$

Equation [3.13] shows that each element $\mathbf{Y_{i,j}}$ of the matrix $\mathbf{Y}$ is the dot product of the i^{th}row of the transformation matrix ($\mathbf{P}(\mathbf{i}, :)$) and the j^{th} column of the original data matrix ($\mathbf{X}(:, \mathbf{j})$). The rows of the matrix $\mathbf{P}$ represent the new basis vectors that transform the columns of $\mathbf{X}$ into the columns of $\mathbf{Y}$. The row vectors which form the new basis are called the principal components of $\mathbf{X}$. In the following section, a method is derived to find the rows of $\mathbf{P}$ along which the re-expressed data has maximum variation.

3.7.5.3 Finding the principal components

During the experimental protocol in the present studies, EEG data are recorded from et set of electrodes. Furthermore, the EEG signal from each electrode was resolved in frequency components. It is of interest to investigate whether the data recorded by the set of electrode could be represented by few electrode with maximum variation or whether the frequency components of data from each electrode could be represented by few frequency components with maximum variation. This can be done by probing how EEG signals from different electrodes (different spectral representations of EEG from the same electrode) vary together. To elaborate, consider n EEG observations from m electrodes (or n observations of EEG data from a given electrode resolved in m frequency components) represented in matrix $\mathbf{X}$ form as:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix}$$

where the rows of $\mathbf{X}$ represent the EEG observations in the time (frequency) domain and the colonms represent the number of electrodes (frequency components) respectively. The covariance matrix $\mathbf{C_X}$ of the data matrix $\mathbf{X}$ measures the degree of correlation between any two EEG signals (frequency components) in the data set $\mathbf{X}$ and is computed as:

$$\mathbf{C_X} = \mathbf{X}\mathbf{X^T}, \qquad [3.14]$$

where $\mathbf{X^T}$ denotes the transpose matrix of $\mathbf{X}$ and is written as:

$$\mathbf{X^T} = \begin{bmatrix} x_{1,1} & x_{2,1} & \dots & x_{m,1} \\ x_{1,2} & x_{2,2} & \dots & x_{m,2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,n} & x_{2,n} & \dots & x_{m,n} \end{bmatrix}$$

Equation [3.14] shows that each element $\mathbf{C_{Xij}}$ is the dot product of the i^{th} row $\mathbf{X(i,:)}$ of the original data matrix $\mathbf{X}$ and the j^{th} column $\mathbf{X(:,j)^T}$ of its transpose matrix $\mathbf{X^T}$. If $\mathbf{X}$ is a zero mean matrix, then the elements on the diagonal of the covariance matrix $\mathbf{C_X}$ represent the variance of each EEG signal (frequency component) from each electrode and the off diagonal elements are the covariance between any pair of EEG signal (frequency components). Large values at the diagonal indicates higher variation and low values indicate noise. While high off-diagonal values indicate high redundancy and low off-diagonal values indicate low redundancy. With this in mind, an optimal solution of Equation [3.13] will result in a transformed matrix Y those covariance matrix $\mathbf{C_Y} = \mathbf{YY^T}$ is diagonal i.e zero off diagonal elements (no redundancy). Furthermore, to project the data $\mathbf{X}$ on $\mathbf{Y}$, the PCA technique assumes that new basis vectors (the rows of the transformation matrix $\mathbf{P}$) are orthonormal i.e $P_iP_j = \delta_{i,j}$ ($\delta_{ij} = 1$ if $i = j$ or 0 otherwise).

Finding the principal components can be resumed as follows:

Given a $m \times n$ data matrix $\mathbf{X}$, find m orthonormal vectors $(\mathbf{p_1}, \mathbf{p_2}, \dots, \mathbf{p_m})$ which transform $\mathbf{X}$ into $\mathbf{Y}$ those covariance matrix $\mathbf{YY^T}$ is diagonal. It is known that under such conditions, the covariance matrix ($\mathbf{XX^T}$) of the original data $\mathbf{X}$ is the same as the covariance matrix ($\mathbf{YY^T}$) of the transformed data $\mathbf{Y}$ and that the rows of the transformation matrix P are the eigenvectors of the original data matrix $\mathbf{X}$. The transformation basis is defined by calculating the eigenvectors or the Singular value decomposition (SVD) of the covariance matrix of the original data $\mathbf{X}$ which consists in finding m orthonormal vectors in directions along which the original data $\mathbf{X}$ has maximum variation (Press et al., 2007; Bishop et al., 1995). The procedure to determine the transformation basis is described in algorithm 3.1.

Algorithm 3.1 PCA procedure

1: Organise the original data matrix X such that the rows represent the observations and the columns represent the samples.
2: Substrate each sample from its mean value i.e $\tilde{\mathbf{X}}(i,:) = \mathbf{X}(\mathbf{i},:) - \sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{x_i}$
3: Calculate the covariance matrix $\mathbf{C}_{\tilde{\mathbf{X}}}$ of the zero mean data matrix $\tilde{\mathbf{X}}$.
4: Compute the eigen vectors or the SVD of the covariance matrix $\mathbf{C}_{\tilde{\mathbf{X}}}$ of the zero mean data matrix

PCA is a useful technique to reduce the dimensionality of data by projecting the data on orthogonal directions with maximum variance. The method is also used to remove variance in data. The method projects the data on orthogonal directions which might not always be the most appropriated basis. When used on EEG data with data from specific channels, the transformed data does not belong to any of the original channels and this can be a disadvantage for applications where it is vital conserve the integrity of the data from each source.

3.7.6 Independent Component Analysis

Independent Component Analysis (ICA) is a blind source separation (BSS) method aiming at separating sources that are mixed and recorded by p sensors. The sources are called unobserved signals while the signals at the sensors are known as observed mixtures. In principle, the number of sensors can be greater than the number of sources, and vice versa.

A famous application of ICA is in the 'cocktail party problem', in which a listener is faced with the problem of separating the independent voices chattering at a cocktail party.

In the BSS approach, 'blind' refers to the total lack of an *a priori* knowledge of the source signals. ICA as a BSS application on biomedical signals makes the following assumptions. It is assumed that the sensors used to record the observed signals do not add any noise to the recorded signal. The source signals are linearly mixed. There are equal numbers of sources as recording sensors. The sensors recording the signal mixtures do not physically move during recording. Finally and

most importantly, the sources are statistically independent (James and Hesse, 2005). Under the previous assumptions, the ICA method can be determined mathematically by assuming at a given time point, *t*, a set of *p* observations $[x_1(t), x_2(t), \ldots, x_p(t)]^T$ and the existence of *n* statistically independent unobserved signals (sources) $[s_1(t), s_2(t), \ldots, s_n(t)]^T$ so that the observations are a linear combination of the independent sources $x_i(t) = a_{i1}s_1(t) + a_{i2}s_2(t) + \ldots + a_{in}s_n(t)$. This is represented by the linear equation system [3.15] in matrix form as

$$\mathbf{x}(\mathbf{t}) = \mathbf{AS}(\mathbf{t}), \qquad [3.15]$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_p(t)]^T$ and $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ are $p \times 1$ and $n \times 1$ column vectors respectively. **A** is called the mixing matrix and is a full rank $p \times n$ matrix. The ICA method consists in recovering the underlying sources using the observations by deriving a $n \times p$ separation matrix **W** such as $\hat{\mathbf{S}}(t) = \mathbf{W}\mathbf{x}(t)$ is a scaled and permuted version of $\mathbf{s}(t)$.

The statistical independence between the sources is measured by a non Gaussian nature of the sources. The ICA technique separates sources by optimising their non Gaussian nature, hence, Gaussian mixtures can not be separated using this method and the maximisation of non Gaussian distribution is used to implement ICA algorithms (Hyvärinen et al., 2004).

The ICA method can be used as a feature extraction method by using it to separate the data of interest from noise. The ICA method is a generalisation of the PCA technique for the source components in ICA need not be orthogonal. ICA is also often used for automatic ocular artefact rejection in biomedical signals (James and Gibson, 2003). One drawback of ICA applications is the selection of the number of sources when the number of sources is less than the number of recording sensors. As the number of sources is selected subjectively, estimating more sources than present often leads to overfitting the ICA model.

PCA technique is often used to overcome this issue by using the dominant principal components to reduce the dimensionality of the data. However, there is no guarantee that the subspace of the selected principal component corresponds to the variability of the signal of interest (James and

Hesse, 2005).

3.7.7 Bootstrap

The Bootstrap or sampling method is a way to determine the distribution of the data from a single sample. The method is performed by creating new samples of equal size by resampling the original sample with replacement.[5] The Bootstrap distribution is calculated by calculating the statistic of each resample and the distribution of these Bootstrap statistics is called the Bootstrap distribution (Hesterberg et al., 2005). In this section a method to devise bootstrap distributions for a single population is presented. The method is further extended to generate a bootstrap distribution that compare two different populations. The bootstrap methods are limited to non-parametric bootstrap models for the population distributions of interest are unknown and in this case, non parametric bootstrap approaches are preferred (Efron and Tibshirani, 1993; Davison and Hinkley, 1997).
In the following a random sample of observations from an unknown distribution F is considered and the distribution F is approximated from the sample. The power of the bootstrap method is that with few data, it is possible to estimate the sampling distribution with few assumptions on the distribution of the population. Another advantage of using this model is that statistics such as the difference between two populations can be easily evaluated and a confidence interval can be estimated easily, compared to more tedious methods from classical approaches.

3.7.7.1 Bootstrap distribution of one population

Consider the observation random data sample $\mathcal{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ with estimated parameter $\hat{\theta}$ and standard error $\hat{\sigma}$, from an unknown distribution F with unknown parameter θ and standard deviation σ. The bootstrap method to evaluate the bootstrap distribution of the sample population is performed following Algorithm 3.2 as:

[5]When sampling with replacement, a random observation is drawn from the original data sample, and put back before drawing the next observation.

Algorithm 3.2 The bootstrap procedure

1: Create resample $\mathcal{X}^*$ by sampling with replacement [6] from the original sample $\mathcal{X}$

2: Calculate the bootstrap statistic $\hat{\theta}^*$ of the bootstrap sample $\mathcal{X}^*$

3: Repeat steps 1 and 2 hundreds of times to obtain B bootstrap statistics $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_B$

4: The distribution of the set $\theta^*_1, \theta^*_2 \ldots, \theta^*_B$ is the bootstrap distribution of the sample $\mathcal{X}$

The bootstrap estimate of standard error (BESE) is calculated using Equation [3.16] as:

$$\hat{\sigma}^* = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\theta^*_i - \frac{1}{B}\sum_{i=1}^{B}\theta^*_i)^2} \quad [3.16]$$

The bootstrap estimate of bias (BEB) is calculated using Equation [3.17]as:

$$B^*_{Boot} = \frac{1}{B}\sum_{i=1}^{B}\theta^*_i - \theta \quad [3.17]$$

During the bootstrap procedure observations are drawn from the original sample $\mathcal{X}$ with probability $\frac{1}{n}$ where n is the size of the sample. The observations are drawn with replacement to generate B bootstrap samples to calculate the bootstrap distribution of the original sample. For a large number of bootstrap samples (1000 or 2000) the sample distribution follows the population distribution F (Davison and Hinkley, 1997; Efron and Tibshirani, 1993). Furthermore, for a large number of bootstrap samples non parametric and parametric methods agree in term of performance (Cox and Hinkley, 1974; LePage et al., 1992; DiCiccio and Efron, 1996).

The general bootstrap method to estimate a bootstrap distribution is shown in figure 3.5.

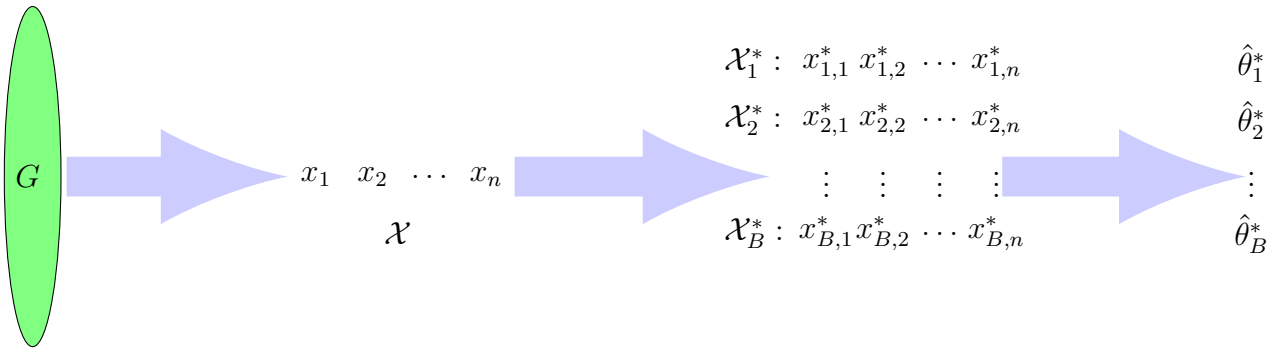


Figure 3.5: General bootstrap procedure. The figure is an illustration of a general bootstrap procedure. G is a population with unknown distribution. $\mathcal{X}$ is a random sample drawn from the population G and is called the original sample with parameter $\hat{\theta}$. $\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_B^*$ are resamples from the original sample and are called the bootstrap samples. $\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*$ are the parameters of the bootstrap samples and represent the bootstrap distribution.

3.7.7.2 Bootstrap method to compare two samples

The bootstrap method described in section 3.7.7.1 can be extended to compare two distinct populations. The bootstrap method in this case is performed by resampling each sample individually. Given two set of observations $\mathcal{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ and $\mathcal{Y} = \{\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}\}$ with their respective estimated parameters $\hat{\theta}_x$ and $\hat{\theta}_y$. The bootstrap method to compare the two samples is performed according to Algorithm 3.3 as:

Algorithm 3.3 The bootstrap procedure to compare two populations

1: Create resample $\mathcal{X}^*$ by sampling with replacement from the original sample $\mathcal{X}$
2: Calculate the bootstrap statistic $\hat{\theta}_x^*$ for the bootstrap sample $\mathcal{X}^*$
3: Create resample $\mathcal{Y}^*$ by sampling with replacement from the original sample $\mathcal{Y}$
4: Calculate the bootstrap statistic $\hat{\theta}_y^*$ for the bootstrap sample $\mathcal{Y}^*$
5: Calculate the bootstrap statistic $\hat{\theta}^*$ that compares $\mathcal{X}^*$ and $\mathcal{Y}^*$
6: Repeat steps 1 to 5 hundreds of time to obtain a set $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$ of statistics
7: The distribution of the set $\hat{\theta}_1^*, \hat{\theta}_2^* \ldots, \hat{\theta}_B^*$ is the estimate of the bootstrap distribution that compare the original samples $\mathcal{X}$ and $\mathcal{Y}$.

In the previous sections, two methods to determine a bootstrap distribution were explained.

Algorithm 3.2 provides the general bootstrap procedure while Algorithm 3.3 derives a method to compare two different populations.

In the following sections, methods to construct the confidence interval (CI) of a bootstrap distribution will be explored. A myriad of methods exist to estimate the CI of bootstrap a distribution and (Carpenter and Bithell, 2000) provides a comprehensive overview of methods to estimate the CI. However, only two essentials approaches will be discussed in the following sections. The two methods are noteworthy for one uses the standard error of the distribution while the second method is distribution-free.

3.7.8 The normal standard confidence interval

Consider a random sample $\mathcal{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ with estimated parameter $\hat{\theta}$ and standard error $\hat{\sigma}$, drawn from a distribution F with parameter θ and unknown variance. For a large sample size, the distribution of $\hat{\theta}$ tends to normal with parameter θ and variance $\hat{\sigma}^2$) and is represented by Equation [3.18]:

$$\hat{\theta} \sim N(\theta, \hat{\sigma}^2) \quad or \quad Z = \frac{\hat{\theta} - \theta}{\hat{\sigma}} \sim N(0, 1) \tag{3.18}$$

The CI of the standard normal distribution is defined by Equation [3.19] as :

$$\left[\hat{\theta} - z^{(1-\alpha)} \cdot \hat{\sigma}, \hat{\theta} - z^{\alpha} \cdot \hat{\sigma}\right] \tag{3.19}$$

where α is the confidence level and $z^{(1-\alpha)}$ and z^{α} are the $100(1-\alpha)^{th}$ and $100\alpha^{th}$ percentiles of the standard normal distribution. As $z^{\alpha} = -z^{(1-\alpha)}$, Equation [3.19] can be rewritten as:

$$\hat{\theta} \pm z^{(1-\alpha)} \cdot \hat{\sigma} \tag{3.20}$$

It can be seen from Equations [3.19] and [3.20] that the evaluation of the standard normal CI requires the standard error $\hat{\sigma}$ and the interval is always symmetric around the parameter $\hat{\theta}$.
A bootstrap equivalent of this CI can be estimated by using the BESE which will provide a more accurate CI. However, this interval will still depend on the standard error. The next section provides a CI that does not require the standard error.

3.7.9 The percentile confidence interval

Another type of CI is the percentile confidence interval. This type of interval is used to mark out a proportion of the values of a distribution within a certain range.
In the sample $x_1, x_2, \ldots, x_n$ it is assumed that the probability of each observation is $P = \frac{1}{n}$. Then the random sample is uniformly distributed with its cumulative distribution function (c.d.f) $F_1(x)$ and probability density functions $f_1(x)$ respectively. $y_1 < y_2 < \ldots, y_n$ is the order statistics of the distribution and it represents $x_1, x_2, \ldots, x_n$ in ascending order. Consider the transformation $w_i = F_1(y_i), \forall i \leq n$. Then $w_1 < w_2 < \ldots, w_n$ is the order statistics of the sample $F_1(x_1), F_1(x_2) \ldots, F_1(x_n)$. the expected value of each element of the order statistic of the transformed sample can be estimated as follows (Hollander and Wolfe, 1973)

$$E\left[w_i\right] = E\left[F_1(y_i)\right] = E\left[\int_{-\infty}^{y_i} f_1(x)dx\right] = \frac{i}{n+1}. \qquad [3.21]$$

From Equation [3.21] $E[w_i]$ can be interpreted as the expected area under $f_1(x)$ on the left of the i^{th} element (y_i) of the order statistics. Then y_i is defined as the $(100p)^{th}$ percentile of the ordered distribution where $p = \frac{i}{n+1}$ and is used to estimate the $(100p)^{th}$ percentile of the population. The position i of the $(100p)^{th}$ percentile is determined according Algorithm 3.4 as:

Algorithm 3.4 Calculation of the $100p^{th}$ percentile

1: Calculate $(n+1)p$ where $0 \leq p \leq 1$
2: **if** $(n+1)p$ is an integer **then**
3: The position of the $(100p)^{th}$ percentile is $(n+1)p$
4: **else**
5: The position of the $(100p)^{th}$ percentile is $\lfloor (n+1)p \rfloor$
6: **end if**

$\lfloor \cdot \rfloor$ denotes the integer part of a given decimal quantity. For example $\lfloor 2.1 \rfloor$ and $\lfloor 2.9 \rfloor$ will both result in 2.

Examples:

For a random sample $x_1, x_2, \ldots, x_{99}$ with 99 elements, to determine 95^{th} percentile of this sample, its order statistics $y_1, y_2, \ldots, y_{99}$ is determined and the $(n+1)p = (99+1)0.95 = 95$. Then the 95^{th} percentile of the sample corresponds to the 95^{th} element of its order statistics.

For a sample of size $b = 13$, the 65^{th} percentile will be calculated as follows:

$(13+1)0.65 = 9.1$. The 65^{th} percentile is the 9^{th} element of the order statistics. Similarly, the 25^{th} percentile is calculated by first determining $(13+1)0.25 = 3.5$. Then the 25^{th} percentile is the 3^{rd} element of the order statistics.

The percentile CI is a distribution free interval for it does not require the standard error of the distribution. Furthermore, the interval is not symmetric and it is shifted toward the skewness of the distribution.

Consider a normal and a skew distributions with the same parameter μ and same standard deviation. The estimation of the CIs of the two distributions shows that The percentile and standard CIs will strongly agree and will not be distinguishable for the normal distribution. However, for the skew distribution, the normal standard CI will remain unchanged while the percentile CI will shift toward the skewness of the distribution (see figure 3.6).

When the distribution is skewed, the standard normal CI will overcover the left side of the distribution and undercover its right side and therefore will cover the distribution in a more balanced way. However, it will still overall misscover the distribution.

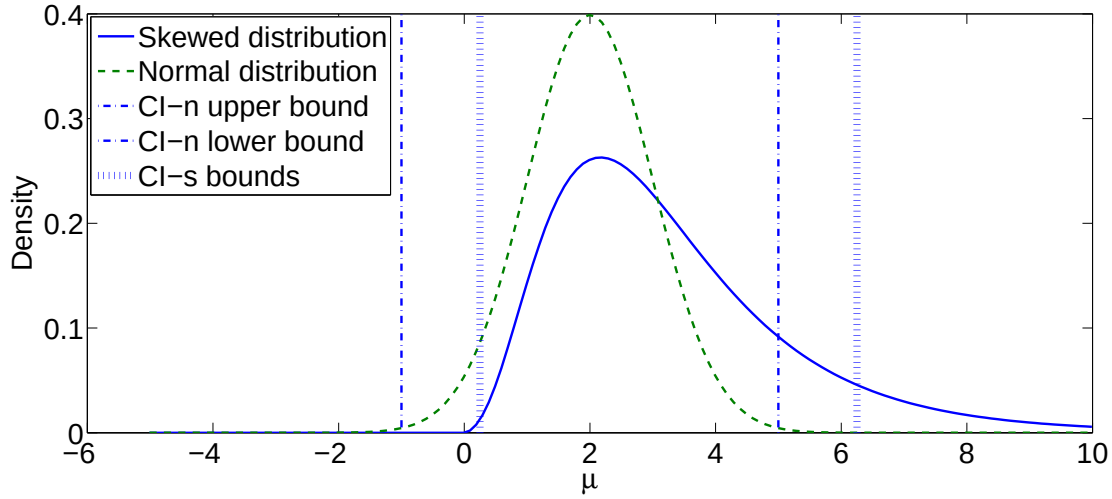


Figure 3.6: Normal (dashed) and normal skew (solid) density distributions with their corresponding CIs (CI-n and CI-s). The figure illustrates how accurate are the CIs estimated by using the normal standard and the percentile methods for different distributions. For the normal standard method, the normal and skewed distributions will have the same CI defined by CN-n. When using the percentile method to evaluate the CIs of the two distributions, CN-n will represent the CI of the normal distribution while CI-s will be the CI of the skewed distribution. The percentile method shifts the CI toward the skewness of the skewed distribution while for the normal standard method the CI remains unchanged. The percentile method is said to be transformation respecting as it follows changes in the shape of the distribution.

The percentile method perform better when the distribution is skewed but still needs some improvement. The goal of the following section is to provide an improved version of the percentile CI.

3.7.10 The Bias corrected and accelerated method

Consider the bootstrap distribution $\hat{\theta}^*_1, \hat{\theta}^*_2 \ldots, \hat{\theta}^*_B$ and its following corresponding order statistics given by Equation [3.22]

$$\tilde{\theta}^*_1, \tilde{\theta}^*_2 \ldots, \tilde{\theta}^*_B, \qquad [3.22]$$

where $\tilde{\theta}_i^* < \tilde{\theta}_j^*, for 1 \leq i \leq j \leq B$.

Algorithm 3.4 can be used to determine the bootstrap CI denoted as $[\tilde{\theta}_{q1}^*, \tilde{\theta}_{q2}^*]$ where $q_1 = \left\lfloor \frac{\alpha}{2}(B+1) \right\rfloor$ and $q_2 = \left\lfloor (1-\frac{\alpha}{2})(B+1) \right\rfloor$

This CI using the percentile method does not require normality and the use of standard error. However, it suffers from coverage error as previously stated.

The Bias corrected and accelerated (BCa) method provides a more accurate estimation of the confidence interval and improves the percentile CI by incorporating an acceleration (a) and a bias correcting (b) factors. The extremities of the BCa and CI are also given as $100p^{th}$ percentiles and are given as $\tilde{\theta}_{Q1}^*$ and $\tilde{\theta}_{Q2}^*$ where Q_1 and Q_2 are function of the constants factors a and b

The bias correction factor b is evaluated as:

$b = \Phi^{-1}(G(\hat{\theta})) = \frac{1}{B}\sum_{i=1}^{B}(\theta_i^* < \hat{\theta})$ where Φ^{-1} is the inverse of the cdf and G is the empirical distribution function and is defined as follows: $G(\hat{\theta}) = \frac{1}{B}\sum_{i=1}^{B} 1_{\hat{\theta}_i^* \leq \hat{\theta}}$.

Various methods are used to evaluate the acceleration factor a. However, the Jacknife estimation of the constant a is used in this section and is calculated as follows(Hall, 1995; Efron, 1987; DiCiccio and Efron, 1996; Hall, 1993).

Let $\hat{\theta}_d^i$ denote the statistic of the sample $\mathcal{X}_d^i$, obtained by deleting the ith observation from the original sample $\mathcal{X} = (x_1, x_2, \ldots, x_n)$ and let $\hat{\theta}_d = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_d^i$ represent the average of the set of the statistics $\hat{\theta}_d^i$. The constant a is defined as : $a = \frac{\sum_{i=1}^{n}(\hat{\theta}_d - \hat{\theta}_d^i)^3}{6(\sum_{i=1}^{n}(\hat{\theta}_d - \hat{\theta}_d^i)^2)^{\frac{3}{2}}}$.

Having calculated the constant a and b the values of Q_1 and Q_2 which represent the positions of the lower and upper bonds of the BCa CI ($\tilde{\theta}_{Q1}^*$ and $\tilde{\theta}_{Q2}^*$) can be evaluated using Equations [3.23] and [3.24] respectively as

$$Q_1 = \left\lfloor (B+1)\Phi(b - \frac{z_{\frac{\alpha}{2}} - b}{1 + a(z_{\frac{\alpha}{2}} - b)}) \right\rfloor, \qquad [3.23]$$

and

$$Q_2 = \left\lfloor (B+1)\Phi(b - \frac{z_{(1-\frac{\alpha}{2})} - b}{1 + a(z_{(1-\frac{\alpha}{2})} - b)}) \right\rfloor . \quad [3.24]$$

Having evaluated the upper and lower bounds of the BCa confidence interval, the CI is defined by Equation [3.25] as

$$\left[\left\lfloor (B+1)\Phi(b - \frac{z_{\frac{\alpha}{2}} - b}{1 + a(z_{\frac{\alpha}{2}} - b)}) \right\rfloor , \left\lfloor (B+1)\Phi(b - \frac{z_{(1-\frac{\alpha}{2})} - b}{1 + a(z_{(1-\frac{\alpha}{2})} - b)}) \right\rfloor \right] . \quad [3.25]$$

In section 3.7 feature extraction which aims to transform the data for more efficient further processing of the data was described. In the next sections, different approaches to detecting changes in EEG data are explored.

3.8 Event detection approaches

Event detection is of great important as it can be used in numerous applications such as system monitoring for fault detection, hand written digit recognition or radar target detection (Carpenter et al., 1996; Dasgupta and Forrest, 1996; Tax and Duin, 1998). In such applications, the systems are required to perform as detectors rather than mere classifiers. The role of the detector is to determine whether the input is part of the data; whilst the classifier is trained to handle something novel. Event detection is such a challenging task that there is a huge range of methods devoted to the task. The methods have different denominations such as novelty detection, outlier detection, change detection, deviation detection, anomaly detection or noise detection according to the application. This thesis focuses on event detection, concerning the detection of a subset of the data with a statistical property that has deviated from that of the original data. The task of the present dissertation is to detect movement initiation in EEG data, and event detection aims to perform the task by analysing and comparing non movement and movement associated EEG time series. There are various methods to address event detection. However, the present dissertation presents the most

important methods that have been mostly applied in biomedical signals, namely, statistical and machine learning approaches.

3.8.1 Statistical methods

Statistical methods are the earlier event detection methods, and generally model the data based on its statistical properties and test whether a test data sample belong to the underlying distribution. In the modelling of the data, event associated data are called abnormal data and the original data are called normal data. In the statistical approach, the normal data and (or) the abnormal data are modelled and the proximity of the models (or the proximity of the normal data model and an sample test abnormal data) is used to decide whether a test sample belong to the normal data or not. The distance measure between normal and abnormal data can be euclidean or some other form of statistical distance (Andrew, 2000). The two main methods used in the statistical approach to model data are parametric and non-parametric methods.

3.8.1.1 Parametric approach

Parametric methods make assumption of the statistical distribution of the data. During this approach, it is assumed that the distribution of the data is known and that the model is based on the data. A selection of parametric methods are presented in the next sections.

3.8.1.2 Gaussian mixtures

In the Gaussian mixtures model (GMM), the data are assumed to originate from a set of Gaussian distributions called kernels. The GMM is trained by choosing a given number of kernels which are fewer than the training data set. An alternative training strategy is based on increasing dynamically the number of Gaussian kernels based on autonomous criterion which determines the number of Gaussian kernels. The GMM, though very powerful, suffers from the curse of dimensionality in the sense that if the dimensionality of the data is high, a very large number of samples are needed

to train the model (Roberts and Tarassenko, 1994; Bishop, 1994). A GMM is shown in Figure 3.7.

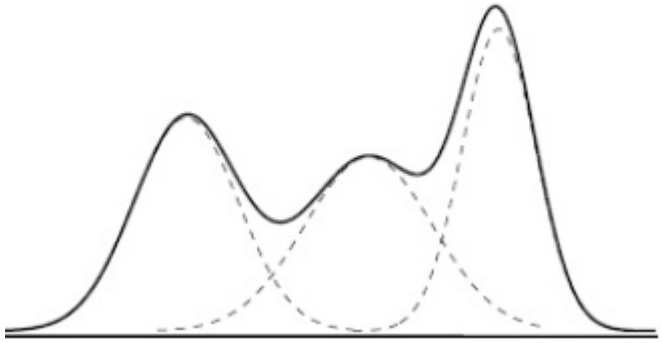


Figure 3.7: A one dimensional GMM composed of three Gaussian distributions. The dashed curves represent the the three underlying Gaussian distributions while the solid curve represents the GMM. [Adapted from http://www.computerrobotvision.org/2010/tutorial_day/GMM_said_crv10_tutorial.pdf]

3.8.1.3 Hidden Markov models

Hidden Markov Models (HMM) are stochastic models used in temporal pattern recognition (Rabiner and Juang, 1986; Rabiner, 1989). The HMM contains a finite number of internal states known as hidden states. The states sequence obey a stochastic process and the transition from one state to another is governed by a transition probability. Each state is associated with observable events with a probability distribution (Rabiner, 1989). The HMM is fully determined by the transition probabilities between all the states, the density distributions of the states and the distribution of the initial state. HMM can be used in different ways, for instance in the BCI field, the states could be defined as right hand movement (M_R) and left hand movement (M_L). The observations could be changes in EEG O_R and O_L associated with M_R and M_L, respectively. The problem to resolve in this case could be stated as follow: Given a sequence of observations $(O_R O_R O_R O_L O_R O_L O_L)$, how can the corresponding state sequence $(M_R M_R M_R M_L M_R M_L M_L)$ that best explains the observations be estimated? A HMM with three states has been used to implement an ERD/ERS based BCI system (Obermaier et al., 2001). The use of the HMM relies on the assignment of some prior probabilities to the initial states in order to determine the transition probabilities, and the choice of such prior

probabilities is not always obvious.

3.8.1.4 Linear Discriminant

The Linear Discriminant (LD) method is a linear transformation that projects each feature vector **v** of a n dimensional vector space into a scalar d such as $d = \mathbf{w}^T\mathbf{v} + w_0$, where **w** is a tunable weight vector and w_0 is termed as the bias or threshold (Bishop et al., 1995).

Another type of LD is the Fisher Linear Discriminant Analysis (LDA). The aim of the LDA is to separate data classes using a hyperplane (Fukunaga, 1990). LDA assumes that the data classes are Gaussians with equal covariance matrices. The separating hyperplane maximises the interclass distance between the means and minimises the within class variance (Fukunaga, 1990). The method is extended to more than two classes by considering several hyperplanes. Due to its simplicity and low cost computation LDA, is a very popular method in the BCI field. However, it results in poor performance on non stationary data (Garcia et al., 2003).

3.8.1.5 Hypothesis testing

A very simple method used in event detection is to test whether a given test sample comes from the same distribution as the training data or not (Markou and Singh, 2003). The method relies strongly on the distribution of the data and is very sensitive to noise in the data.

3.8.1.6 Mahalanlobis distance

The last group of parametric methods to mention are the distance based methods. The metric based method of interest in the present thesis is the MD approach. In general the Mahalanolobis distance Δ^2 is used as a similarity measure between several groups and it assumes that each considered data population has a Gaussian distributions. Given two Gaussian populations P_1 and P_2 with means $\mu_\mathbf{1}$ and $\mu_\mathbf{2}$ and a common covariance matrix $\mathbf{\Sigma}$, the Mahalanolobis squared distance Δ^2 between the

two populations is defined by Equation [3.26] as:

$$\mathbf{\Delta^2} = (\mu_\mathbf{1} - \mu_\mathbf{2})^\mathbf{T} \mathbf{\Sigma^{-1}} (\mu_\mathbf{1} - \mu_\mathbf{2}) \quad [3.26]$$

where the upper-script T is the transposition operator.

In practice, the means $\mu_\mathbf{1}$ and $\mu_\mathbf{2}$ and the common covariance matrix $\mathbf{\Sigma}$ of the populations are unknown. In this case, the MD is estimated from random samples from P_1 and P_2 with bias corrected sample means $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ respectively, and sample covariance matrices $\mathbf{C_1}$ and $\mathbf{C_2}$. The common covariance matrix $\mathbf{\Sigma}$ is estimated by the pooled covariance matrix $\mathbf{C}$ defined by Equation [3.27]

$$\mathbf{C} = \frac{\mathbf{(n1-1)C_1 + (n2-1)C_2}}{\mathbf{N}} \quad [3.27]$$

Where $(N = n_1 + n_2 - 2)$ and n_1 and n_2 are the sample sizes. The sample estimate $\mathbf{D^2}$ of the Mahalanolobis squared distance Δ^2 is given by Equation [3.28]

$$\mathbf{D^2} = (\bar{\mathbf{x}}_\mathbf{1} - \bar{\mathbf{x}}_\mathbf{2})^\mathbf{T} \mathbf{C^{-1}} (\bar{\mathbf{x}}_\mathbf{1} - \bar{\mathbf{x}}_\mathbf{2}) \quad [3.28]$$

The Mahalanolobis distance method is a very powerful and simple method as it calculate the distance from a point to the centroid and takes into account, the correlation of the data. The Mahalanolobis distance is computational expensive compared to the Euclidean distance (Babiloni et al., 2001; Cincotti et al., 2002). Despite its simplicity and good performance for classification of EEG data (Babiloni et al., 2001), the MD method is barely used in the BCI field.

3.8.2 Non-parametric methods

Non-parametric methods do not make any assumptions about the underlying distribution of the data. Two types of classifiers are presented as non-parametric classifiers. The first type consists of assigning a test sample to a class according to its closeness to that particular class. The second type

uses a template that is compared to any test template to evaluate how the test sample matches the original template.

3.8.2.1 Nearest neighbour classifiers

A classifier known as the k Nearest Neighbours (kNN) classifier assigns a test data sample to the dominant class among its k nearest neighbours within the training data set (Duda et al., 2001; Bishop et al., 1995). The Parzen window method is a variant of the kNN method with the difference that the Parzen window uses a smoothing kernel function as class boundaries (Bishop et al., 1995).

Another method based on the nearest neighbour approach is the K-means method. The K-means method clusters the data by initially guessing k centroids. The guessed centroids are then iteratively re-estimated using an Expectation Maximisation (EM) algorithm (Nairac et al., 1999; Bishop et al., 1995). The methods mentioned in this subsection suffer from the curse of dimensionality. Furthermore, the methods are clustering the data in classes and are not suitable for event detection purpose.

3.8.2.2 String matching method

The string matching method uses a training data set to create a template mapping the data using quantisation or binning. This template, which is an encoding of the data, is compared to each test template; their degree of matching is used to classified the test template as an event or not (Dasgupta and Forrest, 1996). The string matching method method requires a large data set to create the template, making the method prone to the curse of dimensionality.

3.9 Neural networks

ANN are the most popular methods used in event detection in the field of BCI. ANN based methods are generally non-parametric. They are model based, can learn the data and can approximate

any complex boundaries to separate classes. Compared to statistical methods, ANN methods are computationally expensive while training, and in some cases they need to be retrained, adding to the computational complexity. ANN can be classified as supervised and unsupervised ANNs. Supervised ANNs require a pre-classified data set used for learning while unsupervised ANNs do not require such pre-classified data set. In the following sections, some of the ANNs that have been used in event detection are presented.

3.9.1 Supervised ANNs

Supervised ANNs are a class of networks that adjust the weights and thresholds of the nodes of the network to ensure the ANN can correctly classify the input. In these topologies of ANNs the output represents the classification and the input data is modelled by distributing each data point across the whole network. The Multi-Layer Perceptron (MLP) is the most widely used supervised ANN. The MLPs have a good performance on data inside the bounds of the training set but perform poorly on unseen data (Bishop et al., 1995). Various training strategies are used leading to various MLPs topologies. The MLPs use hyper-planes to classify the data.

Another type of supervised ANN is the Bayesian Radial Basis Function (RBF) network which in contrast to the MLP approach uses hyper-ellipsoids to classify the data (Brotherton et al., 1998). Various topologies based on the RBF approach have been used in event detection.

3.9.2 Unsupervised ANNs

Unsupervised ANNs contain nodes which compete to represent portions of the data set and require training data set to learn the data. Unsupervised ANNs autonomously cluster the data using node locations to allow the underlying data distribution to be modelled and the classes to be differentiated. Unsupervised ANNs topologically model the data distribution by identifying common features of related vectors.

Self Organising Maps (SOM) are unsupervised ANNs performing vector quantisation and non-

linear mapping to project the data distribution on a lower dimensional grid network where the topology of the grid network is pre-specified by the user. Each node in the grid has an associated weight vector representing each cluster in a k-means system. During the learning process, each input from the training data set is used to find the best matching unit and the winner node's weight is updated to reflect the new match like k-means, clustering the network into regions of similarity (Kohonen, 2001).

Evolutionary networks are ANNs that grow during classification. The Adaptive Resonance Theory (ART) network is a type of ANN which is plastic during learning and stable during classification but can return to plasticity to relearn to adapt to a specific new input, making the ANN ideal for time varying data. When a new input does not match an existing class, a new class is added by creating a new node, making the ANN grow (Carpenter and Grossberg, 1987; Caudell and Newman, 1993). ANN in general requires extensive learning to classify data. Additionally, they are subject to non-linear optimisation, adding to their computational complexity.

3.10 Characteristics of statistical distributions

Statistical distributions are often characterised by the first four moments which are the mean μ, the variance σ, the skewness and the kurtosis β_2 respectively. μ measures the central tendency, σ measures the variability, the skewness is a measure of the deviation from symmetry and β_2 which is the fourth moment is a measure of deviation from normality (Pearson and Please, 1975; Moors, 1986; Ruppert, 1987). The methods assume that the distributions are normal (Hopkins and Weeks, 1990). β_2 is often misinterpreted as a measure of peakedness (Kaplansky, 1945). β_2 as a measure of deviation from normality can be interpreted as a measure of how outlier prone a distribution is compared to the normal distribution. This measure can indirectly be used to compare how outlier prone two non normal distributions are (DeCarlo, 1997). A statistical distribution with a Kurtosis greater than 3 is considered to deviate from normality.

3.11 Variance and distance metric

Both parametric (MD) and non parametric(RBF, Eucludian distance (ED), LDA, LD) methods use distance metrics for classification. However, variance or correlation in the data has an impact on the feasibility of the distance metric used to classify the data. Feature extraction technique such as PCA projects the data along orthogonal directions affecting the covariance matrix in distance calculation of the MD method. The covariance of two random variable X and Y with means μ_X and μ_Y and standard deviations σ_X and σ_Y is defined by Equation [3.29] and [3.30] respectively

$$\sigma_{XY} = E\left[(X - E\left[X\right])(Y - E\left[Y\right])\right] \tag{3.29}$$

$$\rho_{XY} = \frac{E\left[(X - E\left[X\right])(Y - E\left[Y\right])\right]}{\sigma_X \sigma_Y} \tag{3.30}$$

Where $E\left[\cdot\right]$ is the expected value operator. It can be seen from Equations 3.29 and 3.30 that the notions of covariance and correlation are very similar. The covariance with a random variable with itself denoted as σ_X^2 is known as its variance.

3.11.1 Variance covariance matrix

Given a set of N observations with V dimensions which can be represented in a $N \times V$ matrix matrix form as

$$X_{NV} = \begin{pmatrix} X_{11} & \dots & X_{1V} \\ \vdots & \ddots & \vdots \\ X_{N1} & \dots & X_{NV} \end{pmatrix}$$

The covariance matrice is defined as the product of the transpose of the observation data matrix by itself and can be written as

$$Q_{VV} = X_{VN}^T X_{NV} N^{-1} = N^{-1} \begin{pmatrix} X_{11} & \dots & X_{N1} \\ \vdots & \ddots & \vdots \\ X_{1V} & \dots & X_{NV} \end{pmatrix} \begin{pmatrix} X_{11} & \dots & X_{1V} \\ \vdots & \ddots & \vdots \\ X_{N1} & \dots & X_{NV} \end{pmatrix}$$

The product above results in the following matrix

$$Q_{VV} = N^{-1} \begin{pmatrix} \sum_{i=1}^{N} X_{i1}^2 & \dots & \sum_{i=1}^{N} X_{i1} X_{iV} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} X_{i1} X_{iV} & \dots & \sum_{i=1}^{N} X_{iV}^2 \end{pmatrix}$$

When referring to Equation [3.29], the diagonal elements of the matrix Q_{VV}represent the variances while its off diagonal elements are the covariances assuming that the mean value of element in the observation matrix has been removed. Finally the variance covariance matrix can be written as

$$Q_{VV} = N^{-1} \begin{pmatrix} Var(X_1) & \dots & Cov(X_1, X_V) \\ \vdots & \ddots & \vdots \\ Cov(X_1, X_V) & \dots & Var(X_V) \end{pmatrix}$$

The correlation is very important when calculating a distance as it can result into erroneous classification. This often happens when there is variance into specific directions. The MD distance take into account correlation through the covariance matrix. Another way to remove the correlation is to use technique such as PCA to align the data along the few directions with maximum variation. This orthogonal projection of the data on the directions with maximum variance does not improve the classification of the data whereas it yields into a diagonal covariance matrix. Another interpretation of correlation might be that there is some variance in the data that is not aligned with the axes of the coordinate system. The absence of correlation also means that the data are evenly distributed around the centroid. Removing the variance by projecting the data along the axes with

maximum variation using PCA does not achieve anything in term of classification. However, when the data are normalised to achieve unit variance by dividing the variance in each direction by the square root of its variance, the Euclidean distance can be used as a metric to classify the data. The correlation in data has an impact when calculating distances. When removing the variance using PCA, it is advised to normalise the data to unit variance then use the ED distance. When using the MD distance as a metric, it is not advised to use PCA prior to the distance calculation as it remove the covariance in the covariance matrix resulting in the distance calculation in an erroneous ED.

3.12 Aims and objectives

In chapter 2, the neurophysiological background of human brain activity and the mechanisms underlying movement initiation were presented. Chapter 3 explored various signals used in BCI systems, the pre-processing of EEG data and various methods use for event detection in EEG data. In section 3.3 is was reported that BCI systems mostly focused on detecting movements intention during movements execution and that their mental strategies are mainly based on long term training of subjects. In section 3.7, various pre-processing techniques used to enhance both the temporal and spacial resolution of the EEG signal and time-frequency representation used to emphasise features of the EEG signal that are not observable in the raw data were explored. The chapter further discussed the availability of various methods to detect events in EEG data namely, statistical and ANN based approaches in section 3.8. It was reported that generally, statistical methods which are computational less expensive compared to ANN based approaches are rarely used to detect events in EEG data. However, the lack of comparison make it difficult to state which classification method is the most appropriated method and the choice of any method depends on the application.

It was also discussed that the mental strategy to derive EEG signal are mainly focused on training subjects to generate desired EEG patterns through motor imagery using protocol that are constrained by the generation of those EEG patterns. Relieving the rigidity imposed to the protocol design by considering natural movements to generate EEG patterns of interest could contribute to

develop more realistic devices. Section 2.1.4 also reveal that there is a huge application of EEG communication devices and that it is well established that underlying neuronal mechanisms triggering movement initiation takes place prior to movement initiation. However, few of the studies focus on the detection of movement initiation before its actual onset. Detection of movement intention before it take place could contribute to advanced upper limb prosthetics, motor related diseases rehabilitation, effective wheelchair control and decision making.

With the previously mentioned weaknesses in mind, the present thesis aims to address the following issues:

- Use of natural movements to detect self-paced reaching hand movements on single trial scalp EEG data
- Detection of movement initiation before the event takes place
- Explore the time locations of the detection of the movements
- Use of simple statistical detection methods simple with comparable performance with complex non linear detection approaches

Chapter 4

Methodology

4.1 Introduction

The aim of the present research was to explore the possibility to detect initiation of natural movements before their onset. The task was achieved by assessing healthy subjects during the execution of self-paced centre-out natural arm movements in 3D. It was expected during this thesis that patterns generated in EEG data during natural arm movements in 3D could reveal characteristic features to detect movement initiation before movement onset. With the main goal in mind, various strategies were adopted to simultaneously record brain waves and the trajectory of arm movements. In this chapter, signal recording techniques to record EEG during self-paced movements are presented followed by the novel paradigm proposed to record EEG during elf-paced arm movements in 3D. The chapter also presents the recording of arm movements and the synchronisation strategy of EEG and movement recording equipment.

4.2 Subject recruitment

Ethical approval was obtained from the departmental ethical committee at the Department of Biomedical engineering at the University of Strathclyde, Glasgow. Ten healthy subjects (six males

and four females), all right handed and between the ages of 24 to 34 (mean age 25), took part in the study. All participants gave written informed consent with the possibility to withdraw from the study at any time. One participant was excluded from the study because of body movement artifacts and a lack of compliance with the experimental protocol.

4.3 Experimental setup

In this section, the experimental setup for recording EEG and movement data are explained, followed by the description of synchronisation of the systems used to record the data.

4.3.1 EEG recording

The goal of the studies was to investigate MRCPs associated with hand movements. It was shown that MI and supplementary brain regions were responsible for such movements (Kurata and Tanji, 1985; Tanji and Kurata, 1985). The performance of the electrodes placed on the motor area is subject dependent so there is a need to use a dense electrode grid for EEG recording. Furthermore, placing the electrodes over the entire head enables the use of the CAR method to re-reference the electrodes. The CAR method ensures that EEG signal from each electrode is exempt from the influence of the remaining electrodes (Dien, 1998; Osselton, 1965; Offner, 1950). The 10-10 international system was used for a dense recording of EEG (Sharbrough et al., 1990) (see Figure 4.1).

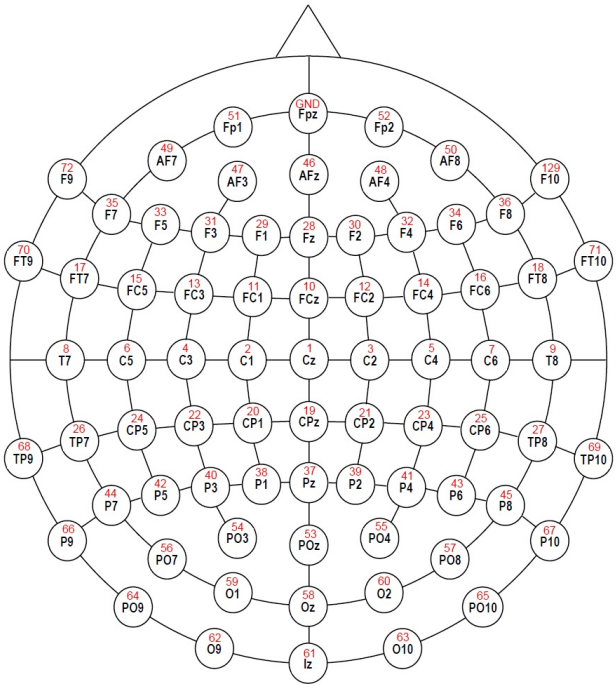


Figure 4.1: The 10-10 standard electrode system used in the present experiment [Modified from EASYCAP GmbH, Germany]

Ear lobes reference method was adopted in this experiment for it exhibits little muscular activity. The EEG was recorded using a Compumedics Neuroscan $Synmaps^2$®. The system has four amplifiers, and each amplifier has 64 monopolar electrode input connections for EEG or Electrooculogram (EOG) recording, one analogue input and four bipolar inputs. The Neuroscan system can receive an external analogue voltage of maximum $5\,\mathrm{V}$ through its analogue input. The amplifiers are connected to a desktop computer (Neuroscan PC) dedicated to EEG recording in the Neurophysiology laboratory in the Department of Bioengineering at the University of Strathclyde. The Neuroscan PC runs SCAN4.3 software which is used to perform EEG and EOG recording. During the present experiment EEG was recorded using a single amplifier where 61 and 3 electrodes were used for EEG and EOG recordings respectively. The amplifier was connected to an electrode cap with Ag/AgCl sintered ring electrodes with a ground electrode (FP_z) at the forehead position (see Figure 4.1). Eye movements were recorded using three EOG electrodes. The vertical and horizontal movements of the eyeball are caused by the individual contraction of four muscles

attached to the surface of the eyeball, namely the Inferior Rectus (I.R.), Superior Rectus (S.R.), Medial Rectus (M.R.), Lateral Rectus (L.R.) muscles, respectively. The different directions of the right eye associated with each contraction of the four recti muscles are summarised in Table 4.1. EOG electrodes positioned below (I.R.), above (S.R.) and lateral (L.R.) the right eye of each participant were used to record the downward, upward and (left or right) eye movements, respectively (see Figure 4.2).

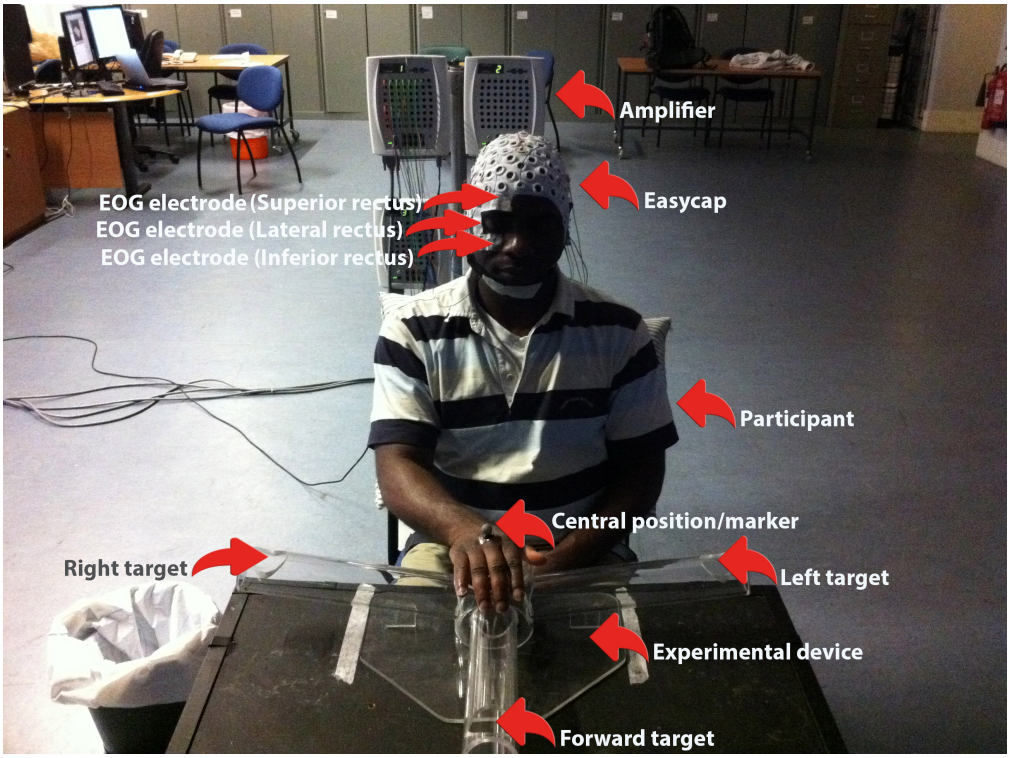


Figure 4.2: The participants were instructed to execute self-paced reaching hand movements from the central position toward three (right, left and forward) target positions. An Easycap fitted with 64 active electrodes positioned according to the 10-10 standard electrodes position system was connected to a Neuroscan amplifier to record EEG data. Hand movement data was recorded using a fluorescent marker positioned on the second carpometacarpal joint of the participants' hand. Three EOG electrodes were used to record the participants' upward (Superior rectus), downward (Inferior rectus) and lateral (Lateral rectus) right eye movements.

Table 4.1: The directions of the right eye associated with the contractions of the I.R., S.O., M.R. and L.R., respectively.

Muscles	Right eye directions
Inferior rectus	Downward
Superior rectus	Upward
Lateral rectus	Laterally
Medial rectus	Medially

Most of the EEG signature of interest falls between the delta and gamma frequency range with an upper bond of around $75\,\mathrm{Hz}$. Hence theoretically, a minimum Nyquist sampling frequency of $150\,\mathrm{Hz}$ should be sufficient to capture the highest frequency components in the EEG. In the present studies EEG was band-pass filtered between $0.05\,\mathrm{Hz}$ and $500\,\mathrm{Hz}$ by the Neuroscan system and sampled at $2\,\mathrm{kHz}$ for a higher resolution.

The Neuroscan system has both Direct Current (DC) and Alternative Current (AC) recording modes. The DC recording mode is most appropriate to record slow varying potentials and the system provides possibility to refilter the data with different high pass values. This is not the case for recording in the AC mode where the high pass value cannot be changed. However, the DC mode recording is sensitive to environmental factors (such as sweat) that produce transient and unpredictable results. Despite its sensitivity to noise, the DC mode was then chosen as the most convenient recording method, as it provides the possibility to record slow EEG activity and it provides the flexibility to refilter the recorded data after recording.

4.3.2 Movement recording

Arm movement trajectory was recorded using a VICON MX Ultranet HD system. A fluorescent marker was placed on the dorsum of the right hand just below the head of the second metacarpal (see Figure 4.3). The Vicon system is equipped with 12 cameras, and each of the 12 cameras takes a picture of the fluorescent marker attached to the arm and reconstructs its 3D coordinates using

images of the marker from at least two cameras. During the recording of arm movement, it was ensured that the marker was not occluded.

Figure 4.3: Marker position on the right hand. The pictures of the fluorescent marker taken by twelve cameras is used to reconstruct the trajectory of the hand during the self-paced arm movements.

A mechanical device was used by the participants to perform the arm movements. The device is made of three hollow transparent plastic tubes of equal length (28 centimetres) soldered on a vertical tube of 10 centimetres diameter glued to a horizontal basal plate. The three tubes are 45 degrees apart and are inclined with a 10 degree angle, with respect to the horizontal plate, with a hole of 5 centimetres in diameter drilled at the extremity of each tube (Figure 4.4). The hole at the centre of the device is referred to as the central position while the holes at the end of the tubes are referred to as target positions. The mechanical device provides a means for each subject to perform centre out reaching hand movements from the central position to the three target positions, namely,

in the left, right and forward directions. The coordinate system of the experimental laboratory is oriented such that the forward, left and right positions point toward the negative X, positive Y directions, respectively, while the Z axis is directed upward (Figure 4.4).

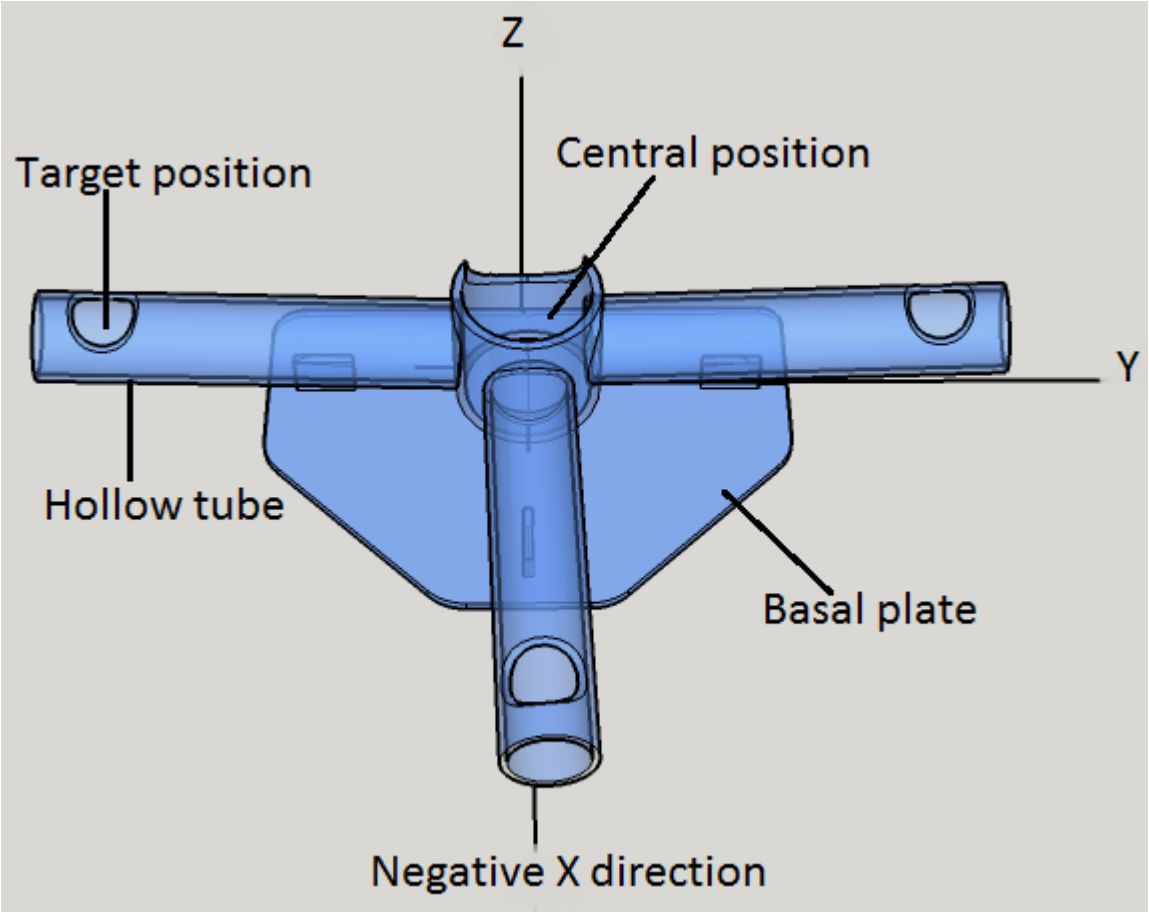


Figure 4.4: The device used during the experiments to perform voluntary reaching hand movement execution in three directions. The coordinate system attached to the picture is seen from the negative orientation of the x axis of the coordinate system of the laboratory

4.3.3 Synchronisation

EEG and movement data were recorded by two different systems which started at different times. The two recording systems were synchronised by using a mechanical switch. The switch was used to send a $1.5\,\mathrm{V}$ DC current both to the Neuroscan system (through the high voltage input) and the Vicon system simultaneously. As the movement data and the EEG data were sampled at $100\,\mathrm{Hz}$

and $2\,\mathrm{kHz}$, respectively, the movement data were upsampled at $2\,\mathrm{kHz}$ to be able to compare the two signals, like for like. The resampled movement and EEG data were synchronised by shifting one of the signals to align the DC pulse positions of the two signals. The diagram of the mechanical switch and its connections with the recording systems are shown in Figure A.3.

4.4 Experimental task

Various experimental tasks have been explored for MRCPs, mainly in the field of BCI. However, as stated in section 3.12, most of the protocols are controlled in order to generate expected patterns during the execution of the protocol. In the present study, self-paced natural reaching hand movements were proposed. The only instruction given to the subject was to start each session when ready. Before the experiment, the subjects performed a test trial to familiarise themselves with having a minimum duration between two consecutive trials, which was set to approximatively ten seconds. No further instruction was given to the participants apart from giving a go sign at the beginning of each session. As the inter-trial duration was solely controlled by the subject, its duration was variable; hence, there was no risk of habituation during the execution of the protocol. Using this type of spontaneous movement does not allow the introduction of an idle state, that could be used to evaluate the event detection methods. Most self-paced BCI design protocols introduce a 'no-control' (idle) state which represents a time period where the subject has no intention to control the BCI, and the 'no-control' state is incorporated by means of visual or auditory cues (K.Ward, 2007; Mason and Birch, 2000; del R. Millán and Mouriño, 2003; Valsan, 2007), making the protocols less self-paced in nature. The present studies propose a protocol without such a no intention state, hence it is exempt from cues and the reaching hand movements from the protocol are free and natural.

4.5 Experimental protocol

The experiment took place in a quiet environment in the Biomechanics laboratory in the Bioengineering Unit at the University of Strathclyde. The Neuroscan PC was set up in the Biomechanics laboratory, which is equipped with a Vicon motion recording system used during the present experiment. The participants sat comfortably in a chair and were instructed to perform centre out self-paced arm movements from the central position toward the target positions in the left, right and forward directions. The subjects had to move back to the central position before initiating another movement.

The subjects were instructed to fixate their eyes on a grey board situated at three metres distance to minimise eye-blink during movement execution. The experiments were divided into twelve sessions and each session began when the subject placed his or her hand in the central position and was asked to relax his arm in order to avoid muscular activity contamination of the EEG. The subject started to execute movements when the synchronisation switch was switched on and when he or she was ready to do so. The subject was waiting randomly during a period of time which allowed at least ten seconds between consecutive movement initiations. At the beginning of each experiment, a calibration session was performed by the subjects to familiarise themselves with having ten seconds duration between consecutive movement initiations. Each trial began when the subject placed his or her hand in the central position. The experimental set up is shown in figure 4.2.

A trial was defined as an arm movement from the central position to one of the target position and back at the central position. The participants were instructed to avoid swallowing and blinking their eyes during the execution of arm movement. The experiment started by switching on the synchronization device. The subject then put his or her arm in the central position and freely released the arm from the central position to one of the target positions and back to the central position. The participant was instructed to rest at least for 10 seconds (or longer if necessary) before attempting another arm movement. Libet provided evidence that neuronal activities preceding the onset of self-paced voluntary movement are activated several hundreds of milliseconds (about $800\,\mathrm{ms}$) and

they persist about $500\,\mathrm{ms}$ or more after movement onset (Libet et al., 1983b). Furthermore, there is evidence that the duration of neuronal activity differs between self-paced and cued movements. It was reported that self-paced movements exhibit longer lasting activity in the SMA compared to cued movements those activities originate from the primary premotor areas (PMA) (Lu et al., 2012; Thut et al., 2000).

The resting period between two consecutive movements was introduced to prevent any superposition of neural activities induced by consecutive movements. A trial will consist of moving the arm from the central position to a target position, and from the target position back to the central position. The time-line of the experimental protocol is shown in Figure 4.5.

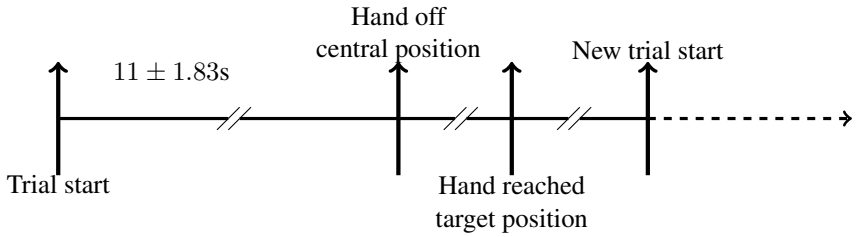


Figure 4.5: The time-line of the experimental protocol. Each experimental session starts when the subject places his or her arm in the central position. The subject then freely moves his or her arm from the central position to one of the target positions at random, and back to the central position. Each trial starts when the subject places his or her hand to the central position. The subject waits at least ten seconds before movement initiation for each trial.

For each participant, the experiment was divided into twelve sessions of approximately three minutes duration each, with a resting period between two consecutive sessions. 61 EEG electrodes, 3 EOG electrodes and 3D motion data were recorded simultaneously. There were on average 156 trials recorded per subject. The average time to complete a full set of trials was approximately three hours including the preparation of the subject [1]. The data recorded by the Neuroscan system (EEG, EOG and DC) were sampled at $2\,\mathrm{kHz}$. While the data recorded by the Vicon system (movement and DC) were sampled at $100\,\mathrm{Hz}$ to limit the amount of data generated by the Vicon system [2].

[1]The preparation of the subject included abrading the electrode recording sites with a gel, and the application of conductive gel. The calibration of the laboratory and the setting of the electrodes on the Easycap was done before the participant arrived in the laboratory

[2]The Vicon recording system is used in motion capture analysis. The system captures trajectory of movements and

This sampling frequency was judged sufficient to capture all the important features in the slow arm movements.

additionally calculates the forces on the force plates in the laboratory, generating a huge amount of data which can make the system crash when the recording period is long as in this set of experiments. One way to avoid this was to record data over a short period of time and reduce the sampling frequency as much as possible

Chapter 5

Data analysis

The aims of this study was to explore methods to detect the initiation of self-paced reaching hand movements in single trials in scalp EEG data. To enable this, various assumptions were made regarding the underlying neuronal correlates of the data prior to movements onset. First, it was assumed that neuronal correlates related to movements onset were modulated by gradual changes in the EEG data. The second assumption stipulated that movements onset induces a change in the statistical distribution of the EEG data. In this chapter, methods based on the assumption above to detect movements intention are explored. After providing the various strategies to filter, map the data in the time-frequency domain to improve the signal to noise ratio and to enhance characteristics features in the data. The chapter presents methods to detect movements intention based on the first and second assumptions respectively (see sections 5.4.1, 5.7 and 5.8). The chapter finally explores techniques to evaluate the performance of the methods (see section 5.10) and to analyse their running time (see section 5.11).

5.1 Overview of the detection of movements onset

The detection of the reaching hand movements onset was performed using the following steps. The movement data were used to determine initiation time points of the movements which were used

to time stamp the raw EEG data. The EEG data were epoch in windows of three seconds (two seconds before movement initiation and one second after movement initiation). The epoched data were then used to detect the initiation of the movements. Three methods explained in depth in sections 5.6, 5.7 and 5.8 were used to predict hand movement onset, as shown in the block diagram of the data processing (Figure 5.1).

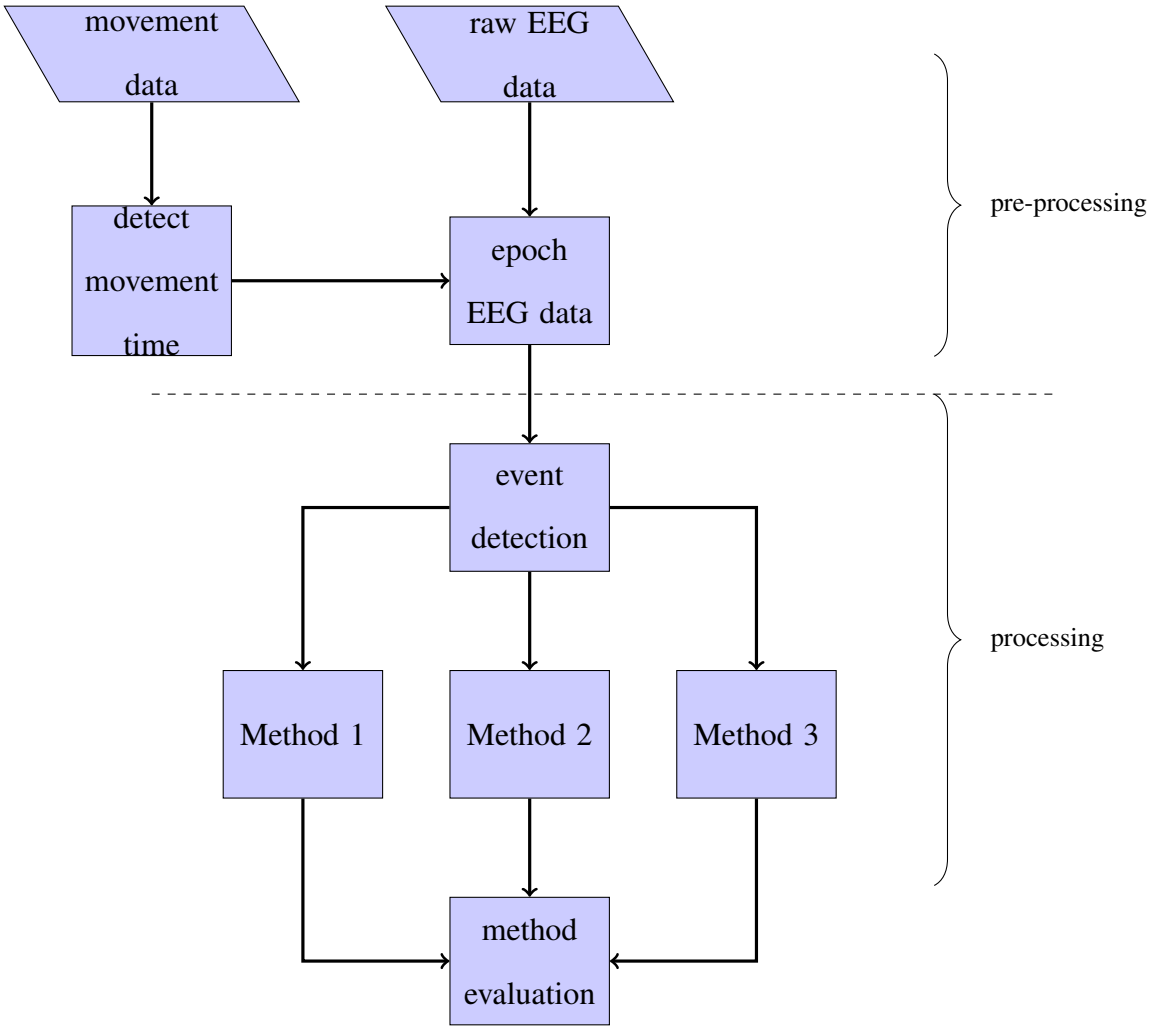


Figure 5.1: A schematic block diagram of the data processing.

5.2 Detecting movement initiation

The movement data profile of the arm during the execution of the free self-paced arm movements in the different target positions was recorded by a 12 cameras motion analysis system (Vicon). The trajectory was calculated as the distance from the origin $(0, 0, 0)$ of the coordinate system in the laboratory to the coordinates (x, y, z) of a marker on the hand of the subject at each time point in the 3D space. The movement profile is analytically defined by Equation [5.1]:

$$s_t = \sqrt{x_t^2 + y_t^2 + z_t^2} \qquad [5.1]$$

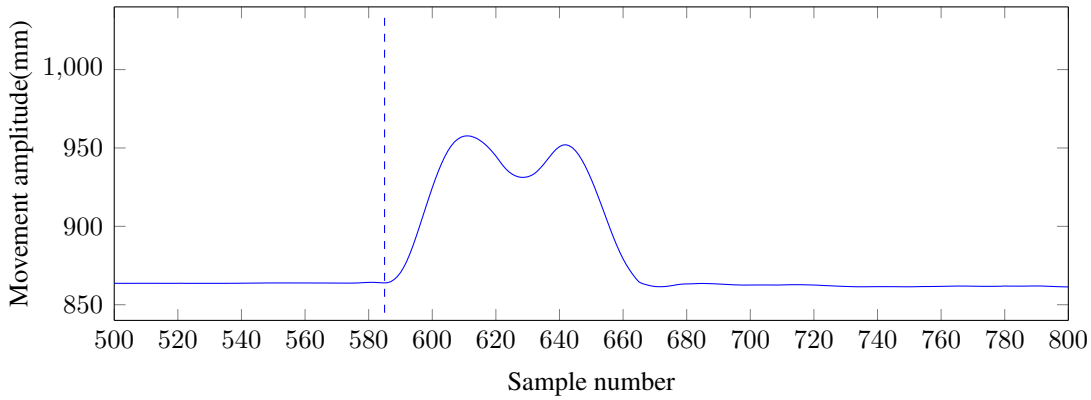
The movement profile consists of two bell shaped curves corresponding to movement from the central position toward the target position (reaching movement) and movement from the target back to the central position. The movement initiation was determined by calculating the velocity profile defined as the first derivative of Equation [5.1]. The first derivative of the movement profile was numerically calculated by Equation [5.2]:

$$\frac{ds}{dt} \approx \frac{\Delta s}{\Delta t} = \frac{s_{i+1} - s_{i-1}}{t_{i+1} - t_{i-1}} \qquad [5.2]$$

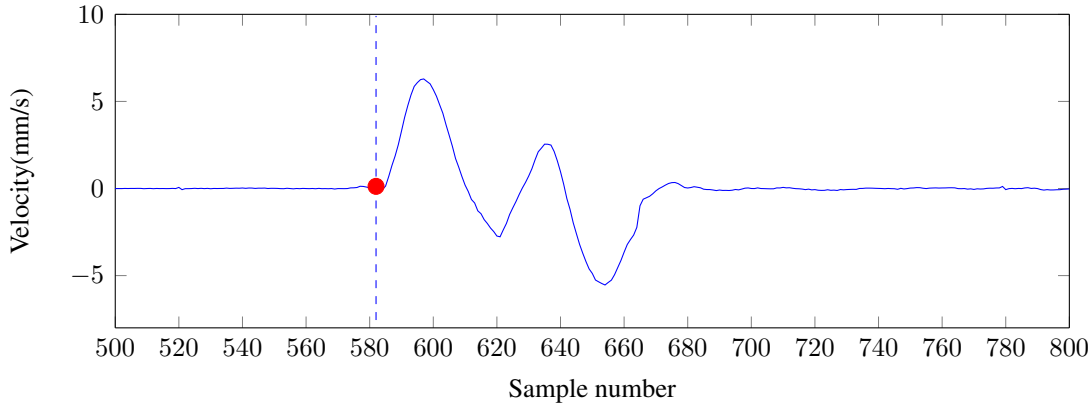
Figure 5.2(a) shows the movement profile of a reaching hand movement from the central position to a target position and back to the central position. The movement profile consists of two parts. The first part corresponding to the first peak represents hand movement from the central position to the target and the second peak represents hand movement from the target position back to the central position. Figure 5.2(b) shows the corresponding velocity profile which exhibits two opposite bell shaped curves. The velocity profile of each part of the movement profile shows a gradual increase of the velocity up to a maximum value and down to a minimum value. The velocity profile of the reaching movement is the opposite of that of the movement from the target back to the central position.

The initiation time of the movement from the central position to the target position was then defined

as the nearest local minimum to the maximum velocity during arm reaching as shown in Figure 5.2(b). The movement initiation time points were stored as an event file for further processing.



(a) Movement profile



(b) Velocity profile

Figure 5.2: Movement (a) and velocity (b) profiles. The dot on the vertical lines represent movement initiation position.

5.3 Time stamping of EEG data

The event files generated using movement detection time points were used to time stamp the EEG data. Data segments centred at each event initiation time (two seconds prior and one second post movement initiation) were considered to perform the event detection task. The EEG data were visually inspected to identify tagged EEG segments which were corrupted by ocular, unexpected

movements and swallowing artifacts. The process of time stamping resulted in marked EEG data at each movement initiation position. The spectrum of quality of movement initiation positions was defined in such a way that an EEG segment associated with a given position is either accepted (artifact free) or rejected (corrupted) otherwise. Table 5.1 shows the total number of trials recorded during the experiment and the number of artifact free trials. During further processing of the time stamped EEG data, only the artifact free segments were considered and all other segments mere excluded from the analysis.

Table 5.1: The statistics of the data recorded for each subject. The table shows the total number of recording sessions, the total numbers of epochs recorded during the experiment and the number of artifact free epochs

Subjects	sessions	Total epochs	artifact free epochs
S1	12	147	80
S2	12	192	72
S3	13	203	83
S4	12	125	57
S5	12	135	64
S6	12	127	53
S7	12	129	55
S8	12	121	76
S9	12	148	51

5.4 Feature extraction

The objective of feature extraction is to enhance information specific to movement in order to differentiate between movement and non movement data. The quality of the data can be improve by using spatial filters as discussed in section ref and the spatial filter that will be used in this study is the CAR method. The use of the CAR method is justified as a dense electrode configuration

covering the full head is used (Dien, 1998; Osselton, 1965; Offner, 1950).

Various feature extraction were presented in section 3.7. In the present study, the PCA technique was used to reduce the dimensionality of the EEG data and to remove any variance in the data. The CWT was chosen to represent the EEG data in the time-frequency domain. The method was preferred as it provides a way to represent the data in different frequency bands.

5.4.1 Wavelet transform

Various methods are available for time frequency representation, namely the STFT and the WT. The Time-frequency representation was selected as it provides not only the frequency components of the signal but also the time at which the frequency component occurred. The WT was preferred than the STFT as the WT provides a multi-resolution representation of the analysed signal in contrary of the STFT method which provides a constant resolution time-frequency representation. The CWT was particularly used as it provides a continuous time resolution. The WT was first introduced by Grossman (Mallat, 1989a). A given complex valued function $\psi(x)$ is said to be a wavelet if and only if it satisfies Equation [5.3]:

$$\int_{-\infty}^{+\infty} \Psi(x)dx = 0. \qquad [5.3]$$

Let $\psi_s(x) = \frac{1}{s}\psi(\frac{x}{s})$ be the dilation of $\Psi(x)$ by the scale factor s. The continuous wavelet transform of a function $f(x)$ is defined by Equation [5.4] as

$$Wf(s,x) = f(x) * \Psi_s(x), \qquad [5.4]$$

where $f(x) * \Psi_s(x)$ represents the convolution between $f(x)$ and $\Psi_s(x)$ and $Wf(s,x)$ represents the wavelet coefficients at scale s. The coefficient value at a given point in the scale - space (s,x)

plane is defined by Equation [5.5] as

$$Wf(s_0, x_0) = Wf_{Re}(s, x) + JWf_{Im}(s, x), \quad [5.5]$$

where $Wf_{Re}(s, x)$ and $Wf_{Im}(s, x)$ respectively, represent the real and imaginary parts of the wavelet coefficients for a complex wavelet. The imaginary part is zero for a real value Wavelet function. In this thesis a complex Morlet Wavelet was used and is defined by equation 5.6

$$\Psi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2+j2\pi x}}. \quad [5.6]$$

The CWT is a mathematical transformation which maps a time domain signal $f(x)$ in a time-frequency (s, x) plane where s represents the scales (pseudo frequencies) and x represents the time. In the time-frequency representation, higher scaling factors correspond to low frequencies while low scale values represent high frequencies. The CWT function from MATLAB was used in the present study.

5.4.2 Principal component analysis

The PCA technique is mainly used to reduce the dimension of the data. In multi-channel EEG recording, PCA can for example be used to reduce the number of channels by projecting the data on the orthogonal directions with maximum variance. PCA method is used to transform correlated data into fewer uncorrelated data variables called principal components. The first principal component accounts for as much of the of the variability as possible in the data and each successive component accounts for as much of the remaining variability as possible. PCA is a linear transformation that rotate the set of points to align with the first principal components. This projection results in a data set those components are unidentifiable. As it was vital to consider the performance of individual electrode in this study, the method was used not to reduce the number of EEG channels but to select frequency components with maximum variation within each EEG channel; Hence, prevailing the

integrity of the data from individual electrode. This was achieved by splitting the data from a single channel into different pseudo frequency bands corresponding to very slow variations (SCP), Mu and Beta bands using appropriate wavelet decomposition scales. The frequency bands used are slow cortical potential ($0.5\,\mathrm{Hz}$ - $1.5\,\mathrm{Hz}$), Mu ($8\,\mathrm{Hz}$ - $13\,\mathrm{Hz}$) and Beta ($12\,\mathrm{Hz}$ - $30\,\mathrm{Hz}$) rhythms.

5.4.2.1 Number of principal components

Various strategies such as the Kaiser's stopping rule, the Scree plot and the percentage of cumulative variance methods are used to select the number of principal components for dimensionality reduction (Gorsuch, 1983; Bryant and Yarnold, 1995). The Kaiser's stopping rule stipulates that only eigen values equal or greater than one should be selected. The Scree plot is a visual representation of the eigen values and the number of variables. The Scree plot selects the number of components to be included in the analysis by selecting where the eigen values plot stops decreasing precipitously and levels out (Bryant and Yarnold, 1995). The percentage of cumulative variance is related to the Kaiser's stopping rule. In this method the use of all the eigen values accounts 100% of the variance of the data. An intuitive threshold value is used to set the desired cumulative percentage of variance to be used during the analysis of the data. In the present dissertation, PCA is used to select a reference population which entails a use of a representative percentage of data. With this in mind the cumulative percentage of variance appears to be the method of choice for it provides a mean to determine the percentage of data to be used during the analysis.

In sections 5.4.1 and 5.4.2, two feature extraction methods were presented. The first method represents the data into three frequency bands using the CWT. The second feature extraction method uses PCA to select data from each frequency bands with maximum variation. In section 5.4.2.1 methods to select the cardinality of the dimension of the transformed data were presented. The two sets of features are evaluated with different event detection algorithms which are presented in the following sections.

5.5 Event detection methods

This section is devoted to transform the features derived in the previous sections in order to predict the initiation of self-paced reaching hand movements. The task is termed as event detection and aims at identifying the time location of a group of EEG signal samples which underwent changes due to the initiation of free arm movements. Three different approaches are used in this research for event detection. A method based on singularity detection in time - frequency representation is proposed in section 5.6. In section 5.7, a method using the statistical difference between movement and non movement data is presented. Finally, the third method uses a bootstrap of the statistic difference metric between movement and non movement data to construct a confidence interval. The significance of the difference between movement and non movement data is assessed using the constructed confidence interval and the Bootstrap method is presented in section 5.8.

5.6 Event detection using Wavelet modulus maxima (WMM)

This section presents event detection method based of the evaluation of EEG signals singularities. The signal is first represented in the time frequency representation using the CWT. In section 5.4.1, the Wavelet coefficients were calculated using Equation [5.4] as $Wf(s,x)$ in the scale-space (s,x) plane. A maximum in the (s,x) plane is defined by Equation [5.7]:

$$\frac{\partial Wf(s_0,x)}{\partial x} = 0 \qquad [5.7]$$

A modulus maximum is defined such that for any point (s_0,x_0) in the (s,x) plane, $|Wf(s,x)|$ is a local maximum at $x = x_0$. The connected curve between WMM is called the maxima line. It was shown that such local maxima contain most of the useful information of the signal and that for each modulus maximum there exists a positive constant A and a real number α such that Equation [5.8]

is satisfied (Mallat, 1989b; Mallat and Hwang, 1992).

$$|Wf(s,x)| \leq As^{\alpha} \quad [5.8]$$

The number α is known as the Lipschitz exponent (LE) and represents an upper bound of the rate of change of the first derivative along the maxima line. LE is also often used to investigate the singularity of a function. Equation [5.8] can be approximated as:

$$log_2 |Wf(s,x)| \leq log_2 A + \alpha log_2 s \quad [5.9]$$

Various methods have been used to determine LE using Equation [5.9] (Venkatakrishnan et al., 2012; Struzik, 2001; Peng et al., 2002). In the present study, a simplified solution of Equation [5.9] is proposed using Equation [5.10] (Hong et al., 2002):

$$LP(x) = \frac{\sum_{s=1}^{m} (log_2 |W(s,x)| \times log_2 s) - \frac{\left(\sum_{s=1}^{m} log_2 |W(s,x)|\right)\left(\sum_{s=1}^{m} log_2 s\right)}{m}}{\sum_{s=1}^{m} (log_2 s)^2 - \frac{\left(\sum_{s=1}^{m} log_2 s\right)^2}{m}} \quad [5.10]$$

Equation [5.10] has been used previously with good results in machine vibration monitoring (Hong et al., 2002; Robertson et al., 2003). The equation evaluates the Lipschitz function LP(x) which represents the value of LE across the scales as a function of time. In the investigation of LP(x), it is assumed that changes in the EEG will result in significant variability in LP(x). The Lipschitz function is evaluated by calculating the LE along an EEG data window of $3\,\mathrm{s}$. The window consists of the baseline (the initial $500\,\mathrm{ms}$ of the analysis window), an unconscious activity window (from $500\,\mathrm{ms}$ to $250\,\mathrm{ms}$ before movement onset) and a conscious activity window (from $500\,\mathrm{ms}$ before movement onset to $1\,\mathrm{s}$ after movement onset). The unconscious and conscious data windows refer to the periods during which the subject is unaware and aware to execute self-paced hand movement. The computation the Lipschitz function is shown in Figure 5.3.

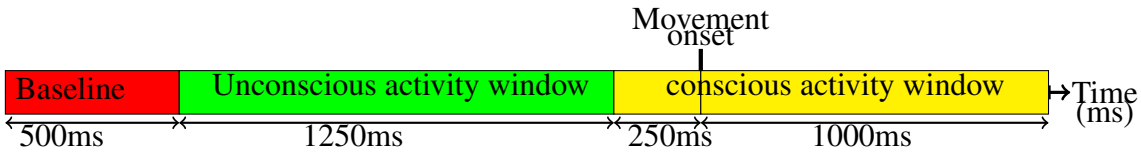


Figure 5.3: A schematic representation of the 3 s *EEG data window used to evaluate the Lipschitz function.*

The steps to calculate the Lipschitz function LP(x) are given in Algorithm 5.1.

Algorithm 5.1 Evaluation of the Lipschitz function

1: Let $|Wf(s,x)|$ represents the Wavelet maxima in a three seconds window as described in section 5.3
2: **for** $x = 1$ to L **do**
3: **for** $s = 1$ to m **do**
4: Calculate LP(x) using equation 5.10
5: **end for**
6: **end for**

5.6.1 Evaluation of the event index

Having calculated LP(x), movement initiation time is calculated by defining an event index which is calculated using the Kurtosis of the LE distribution over different time intervals. The kurtosis of a population distribution with mean μ is defined by Equation 5.11

$$\beta_2 = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} \qquad [5.11]$$

where E is the expectation operator. For a sample of the population the sample counterpart of β_2 is given by Equation [5.12]

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^4}{\frac{1}{n}\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right]^2} \qquad [5.12]$$

where $\overline{X}$ is the sample mean and n is the number of observations. Equation [5.12] shows that b_2 is calculated using a biased variance as n is used in the denominator instead of (n-1). This in turn leads to a biased sample kurtosis. The unbiased estimator of the sample kurtosis is given by the Fisher statistic g (Fisher, 1992) and is defined by Equation 5.13

$$g_2 = \frac{n(n+1)\sum_{i=1}^{n}(X_i - \overline{X})^4}{(n-1)(n-2)(n-3)\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^2} - \frac{3(n-1)^2}{(n-2(n-3)} \qquad [5.13]$$

The Kurtosis of a statistical distribution is a measure of how outlier prone the distribution is. A normal distribution has a Kurtosis of three. Hence, a distribution with a kurtosis greater than three is more outlier prone compared to the normal distribution as mentioned in section 3.10. In the present dissertation, it is assumed that the baseline which consists of EEG period with no activity is less outlier prone compared to activity EEG period. Furthermore, it is assumed that an event is characterised by lasting outliers. Figure 5.4 shows a schematic segmentation of an LP(x) data analysis window into a baseline and subsequent windows of the remaining data.

To evaluate the event index, the Kurtosis between the baseline ($500\,\mathrm{ms}$) and 99% overlapping windows along the unconscious and conscious activity data are computed. The first activity window is taken from the middle of the baseline window to ensure that the first time value (t_1 as shown in Figure 5.4) of the event index time line corresponds to the beginning of the activity window. The last time point (t_q as illustrated in Figure 5.4) corresponding to the Kurtosis between the last window and the baseline is located at $250\,\mathrm{ms}$ before the the end of the $3\,\mathrm{s}$ window. Activity windows of $500\,\mathrm{ms}$ were used, leading to a time resolution of $0.5\,\mathrm{ms}$ for the timeline of the Kurtosis.

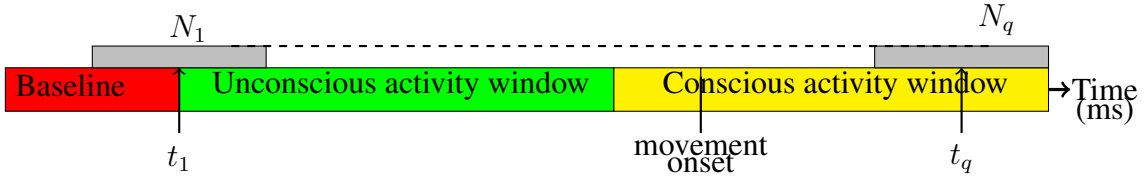


Figure 5.4: Graphical illustration of The computation of the event index as function of time.

The event index $I(i)$ related to a given window $\mathbf{w}_i$ calculated using equation 5.13 is defined as the ratio of the kurtosis of $\mathbf{w}_i$ and that of the baseline window $\mathbf{w}_b$ of the LP data (see figure 5.4). The event index is given by equation 5.14

$$I(i) = \frac{g_{2_{\mathbf{w}_i}}}{g_{2_{\mathbf{w}_b}}} \qquad [5.14]$$

where $g_{2_{\mathbf{w}_b}}$ and $g_{2_{\mathbf{w}_i}}$ are the kurtosis of the baseline and a subsequent window of the remaining LE data. Based on the assumption above, movement initiation detection is characterised by n successive event indices greater than a constant value above one. The steps to determine movement initiation time are given by algorithm 5.2.

Algorithm 5.2 Event detection algorithm using WMM

1: Let λ represents the event counter
2: Let BW represents the latency of the event
3: Let δ represents the event index threshold
4: Let W represents a 3 seconds LE data vector
5: Calculate the kurtosis K_{base} of the baseline using the first 500 milliseconds of W
6: Segment the remaining LE data into L non overlapping windows
7: $\lambda \leftarrow 0$
8: **for** $i = 1$ to L **do**
9: Calculate the kurtosis K_i of the i^{th} window
10: Calculate the event index $I(i) = \frac{K_i}{K_{base}}$
11: **if** $I(i) > \delta$ **then**
12: $\lambda \leftarrow \lambda + 1$
13: **if** $\lambda \geq BW$ **then**
14: Determine the time location t_d of movement
15: **return** t_d
16: **else**
17: $\lambda \leftarrow 0$
18: GO TO step 8
19: **end if**
20: **end if**
21: **end for**

5.7 The statistic distance method

This section presents a method to detect events by comparing EEG where there is no movement activity, known as baseline EEG, with EEG where there is activity, known as activity EEG. The

statistic difference between a baseline and activity EEG data is evaluated and an event is detected if the difference between the baseline and activity distributions exceeds a given threshold value.

5.7.1 The Mahalanobis distance method

The Mahalanobis assumes normality and is used in pattern recognition, discriminant analysis or in classification to discriminate between several groups (McLachlan, 1999). Given a normal distributed population $N(\mu, \sigma)$ with unknown mean μ and unknown variance σ, the method is used to test whether a given observation differs from the population by calculating the statistical distance between the population and the observation. The parameters of the population are often unknown so a sample of the population is often considered and the distance is calculated based on the sample of the population.

The MD approach was chosen to avoid the intensive training involving the use of machine learning methods. Furthermore, the method was preferred compared to other distance metrics as the MD takes into account any correlation in the data when evaluating the distance.

The MD method was performed by considering EEG epochs of three seconds(two seconds pre and one second post movement) from a single trial as shown in Figure 5.3. The CWT coefficients in the SCP, Mu and Beta bands were considered separately. These frequency bands were chosen for they have been shown in previous literature to cover frequency bands with interesting feature related to movements (Babiloni et al., 2001; Müller-Gerking et al., 1999; Babiloni et al., 2001). The data were represented by a $n \times p$ matrix where the columns (number of frequency components) represent the number of variables and the rows (coefficient values over time) are the observations. The data matrix is further divided into a baseline and activity data. The baseline is represented by the first half second and the remaining data are the activity data. The PCA spatial filtering was not performed on the data as it would project orthogonally the data on components with maximum variation. Under this condition, the covariance matrix would be diagonal and the use of the MD would be justified as the ED would be more appropriate metric (see section 3.11.1).

Given a $n \times m$ baseline (reference population) $\boldsymbol{X}_{B_{n\times m}}$ and a $q \times m$ observation data $\boldsymbol{X}_{O_{q\times m}}$ with respected sample means $\bar{\boldsymbol{X}}_B$ and $\bar{\boldsymbol{X}}_O$, the squared Mahalanobis distance $\boldsymbol{d}^2_{1\times m}$ between the observation data and the baseline population is calculated using Equation [5.15]:

$$\boldsymbol{d}^2 = (\bar{\boldsymbol{X}}_B - \bar{\boldsymbol{X}}_O)^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{X}}_B - \bar{\boldsymbol{X}}_O) \qquad [5.15]$$

Where $\boldsymbol{\Sigma}$ is the pooled covariance matrix of the two sample populations given by Equation [3.27] (see section 3.8.1.6).

To evaluate the Mahalanolobis distance as a function of time, the MD between the baseline ($500\,\mathrm{ms}$) and 99% overlapping windows of $50\,\mathrm{ms}$ along the unconscious and conscious activity data is computed leading to a time resolution of $0.5\,\mathrm{ms}$. The first activity window is taken from the end of the baseline window so that the first time value (t_1) of the MD time line corresponds to $1.5\,\mathrm{s}$ before movement initiation. The last time point (t_q) corresponding to the MD between the last activity window and the baseline is located at $50\,\mathrm{ms}$ before the end of the $3\,\mathrm{s}$ data window. The computation of the MD distance is illustrated by Figure 5.5.

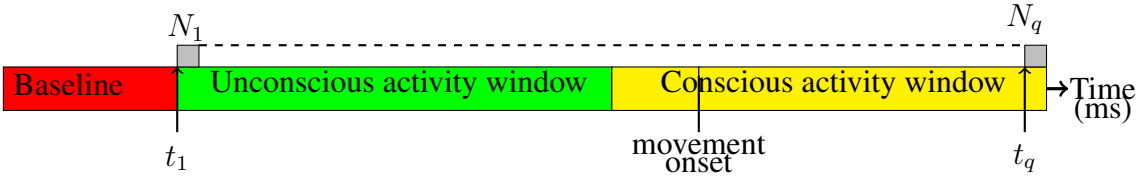


Figure 5.5: Graphical illustration of the computation of the MD as a function of time.

By using an appropriate threshold value, the Mahalanobis distance can be used to investigate outliers from the reference population. An outlier is defined in this report as the median value of the MD vector $\boldsymbol{d}$ falling outside a given threshold interval. The threshold interval is calculated by first evaluating each median value of the MD between the baseline and 99% overlapping windows of $50\,\mathrm{ms}$ across the baseline. The upper and lower bounds of the threshold interval were then calculated as the two standard deviations above and below the mean of the median values of the calculated MD of observation data from the baseline. The threshold interval is calculated from the

set of the median values $\left\{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_L\right\}$ by Equation [5.16]

$$[T_H, T_L] = \frac{1}{L}\sum_{i=1}^{L}\tilde{d}_i \pm 2\sqrt{\sum_{i=1}^{L}\frac{\left(\tilde{d}_i - \frac{1}{L}\sum\limits_{i=1}^{L}\tilde{d}_i\right)^2}{L-1}}, \qquad [5.16]$$

where $\tilde{d}_i$ is the median value of the MD vector $\boldsymbol{d}$ calculated using Equation [5.15] the baseline and observation data from the baseline and L is the number of the median values. The steps involved in the calculation of the baseline interval are shown in algorithm 5.3

Algorithm 5.3 Calculation of the baseline interval

1: Consider a baseline window $\boldsymbol{X}_B$ length L_B
2: Take $\log(CWT)$
3: Select frequency band of interest
4: Consider an observation window $\boldsymbol{X}_O$ length L_O
5: **for** $i = 1$ to $L_B - L_O$ **do**
6: Calculate $\boldsymbol{d}_i$ using Equation [5.15]
7: Calculate the median value $\tilde{d}_i$ of $\boldsymbol{d}_i$
8: **end for**
9: Calculate the threshold interval from $\left\{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_L\right\}$ using Equation [5.16]

Event detection is performed by assessing whether the median value $\tilde{d}$ of the MD of each observation data from the activity window falls outside the threshold interval as defined in Equation [5.16]. An event is detected when n consecutive $\tilde{d}$ values fall outside the threshold interval. Data from 100 observations were used resulting in windows of $50\,\mathrm{ms}$ for the data are sampled at $2\,\mathrm{kHz}$. The different steps to perform event detection based of the Mahalanoobis distance are shown in algorithm 5.4

Algorithm 5.4 Event detection using the Mahalanobis distance

1: Consider CWT of EEG window $(W_L)$from a single EEG channel of length L as described in Figure 5.3.
2: Take $\log(CWT)$
3: Set event latency BW
4: Let λ represent the event counter
5: Select frequency band of interest
6: construct the base line population $\boldsymbol{X}_B$
7: Calculate the threshold interval $[T_H, T_L]$ using algorithm 5.3
8: $\lambda \leftarrow 0$ {initialise event counter}
9: Consider the activity window of length L_a from $500\,\mathrm{ms}$ to $3\,\mathrm{s}$
10: **for** $i = 1$ to L_a **do**
11: Calculate MD $\boldsymbol{d}$ using equation 5.15
12: Calculate the median value $\tilde{d}$ of $\boldsymbol{d}$
13: **if** $\tilde{d} \notin [T_H, T_L]$ **then**
14: $\lambda \leftarrow \lambda + 1$
15: **if** $\lambda \geq BW$ **then**
16: Determine the time location t_d
17: **return** t_d
18: **else**
19: $\lambda \leftarrow 0$
20: Slide the window by one sample to the right
21: GO TO step 10
22: **end if**
23: **else**
24: Slide the window by one sample to the right
25: GO TO step 10
26: **end if**
27: **end for**

5.8 The bootstrap method

In this section an event detection method based on the calculation of the difference between two statistical distributions is proposed. The method detects events by calculating the distance between a population and an observation and assign the observation to the population if the distance between the two populations is significant. The method does not rely of an empirical threshold value to investigate and observation.

5.8.1 Hypothesis testing

Given an EEG data window of $3\,\mathrm{s}$ where the first $500\,\mathrm{ms}$ are non activity data known as baseline and the remaining data are activity data (see Figure 5.3). Let define W_B be the baseline window and W_{A_i} a sliding window along the activity data. If baseline and activity data are assumed to come from statistical empirical distributions F_B and F_A, respectively, the distance d_i between W_B and W_{A_i} can be calculated as the difference between the underlying distributions and its significance can be used to assess the occurrence of an event. The significance of $d_i = D\left(F_A - F_B\right)$ can be assessed by the null hypothesis (H_0) formulated by Equation [5.17]

$$H_0 : F_A = F_B. \qquad [5.17]$$

H_0 can be investigated by using a Bootstrap estimate of the distance. B bootstrap estimates $\left\{\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_B\right\}$ are calculated (see section 3.7.7.2, Algorithm 3.3) and a confidence interval $[0, d_\alpha]$ is constructed from the Bootstrap estimate. H_0 is rejected if the distance d falls in the interval $]d_\alpha, \infty[$ which is known as the critical region (see section 3.7.9, section 3.7.8 and section 3.7.10). In the present event detection method, the Mahalanolobis distance is used. The rational of the MD is that it takes into account correlation in the data. The different steps of the event detection method based on the Bootstrap distribution of the MD is shown in Algorithm 5.5.

Algorithm 5.5 Event detection algorithm using bootstrap method

1: Let λ represent the event counter
2: Let δ represent the event counter's threshold
3: Construct a EEG signal window W_L from a single channel of length L
4: Calculate CWT of W_L at increasing scales $s_0, s_1, s_2, \ldots, s_q$ within a given frequency band
5: Construct baseline $\boldsymbol{X}_B$
6: Resample $\boldsymbol{X}_B$ to create B bootstrap baselines $\boldsymbol{X}^*_{B1}, \ldots, X^*_{BB}$
7: $\lambda \leftarrow 0$ {initialise the event width to 0}
8: **for** $j = 1$ to L **do**
9: Construct observation data $\boldsymbol{X}$
10: Resample $\boldsymbol{X}$ to create create B bootstrap observations $\boldsymbol{X}^*_1, \ldots, \boldsymbol{X}^*_B$
11: calculate Bootstrap distance distribution $\boldsymbol{d}^*_1, \boldsymbol{d}^*_2, \ldots, \boldsymbol{d}^*_B$ and their medians $\tilde{d}^*_1, \tilde{d}^*_2, \ldots, \tilde{d}^*_B$
12: Construct the critical region $]d_\alpha, \infty[$
13: Compute distance $\boldsymbol{d}_i$ and its median $\tilde{d}_i$
14: **if** $\tilde{d}_i \in]d_\alpha, \infty[$ **then**
15: $\lambda \leftarrow \lambda + 1$
16: **if** $\lambda > \delta$ **then**
17: An event is detected
18: **else**
19: GO TO step 8
20: **end if**
21: **else**
22: GO TO step 8
23: **end if**
24: **end for**

5.9 Event detection criteria

In sections 5.6, 5.7 and 5.8, methods were derived to predict the initiation of hand movements. The prediction was performed by considering a data window of $3\,\mathrm{s}$ (see Figure 5.3). The following assumptions were made for the purpose of the prediction of hand movements in single trial EEG data:

1. During the analysis window, the subject did not veto the movement
2. Significant changes in EEG data underlying the occurrence of an event should last a certain period of time
3. The location of any event is between the end of the baseline period and precede the onset of the movement

Assumption 1 confers events a voluntary nature as the subject voluntarily decided to move (Soon et al., 2008). Assumption 2 makes a difference between an event with a sustained lasting period of time and outliers with short latencies. The duration of EEG changes due to movements was estimated to $250\,\mathrm{ms}$. In the last assumption, the time location of an event spans a $[-1500, -250]$ milliseconds time interval with respect to movement onset. This time interval was denoted as activity window and is schematically shown in Figure 5.3. The choice of the activity interval was motivated by previous findings that brain activity was taking place before movement onset (Deecke et al., 1969). A further justification of event location was based on previous evidence that awareness of the intention to move precede movement initiation (Libet et al., 1983a; Libet, 1993; Libet et al., 1999) (see Figure 2.4).

5.10 Evaluation of the performance of the event detection methods

An event detection method can be evaluated either according to its accuracy and detection power or according to the amount of information that can be transferred per unit of time, a measure also known as ITR. In the next sessions, the detection power and the ITR are explored respectively.

5.10.1 Detection power

The power of detection and accuracy of an event detection method can be measured by the sensitivity (also known as the true positive rate), the specificity and the accuracy. To calculate these measures, consider P trials during which there is an event and N trials during which there is no event. The P and N trials are called positive and negative trials, respectively. The False Negative (FN) can be defined as the number of positive trials undetected, the True Negative (TN) as the number of negative trials effectively detected as negative, the False Positive (FP) as the number of negative trials detected as positive, and the True Positive (TP) as the number of positive trials effectively detected as positive. The sensitivity is the proportion of positive trials effectively detected as positive and the specificity is the proportion of negative trials effectively detected as negatives. The performance measures (sensitivity, specificity and accuracy) can be calculated (Hudson and Cohen, 2000) from TP, N,FP and FN using Equations [5.18], [5.19], and [5.20].

$$sensitivity = \frac{TP}{TP + FN} \quad [5.18]$$

$$specificity = \frac{TN}{N} \quad [5.19]$$

$$accuracy = \frac{TP + TN}{P + N} \quad [5.20]$$

5.10.2 Information transfer rate

A detection method can also be evaluated based on the amount of information that can be transmitted during a unit of time (Dornhege, 2006). The ITR combines both the speed and the accuracy of the detection method. The ITR is commonly used to assess the performance of a BCI (McFarland and Krusienski, 2012) and is often calculated based on Shannon's information theory (Shannon, n.d.; Wolpaw et al., 1998, 2002a) using Equation [5.21]

$$BTR = log_2 n + \nu log_2 \nu + (1-\nu)\, log_2 \frac{1-\nu}{n-1}, \quad [5.21]$$

where *n* is the number of mental tasks and ν is the accuracy and BTR is the bit rate in (bits/symbol). The number of mental tasks is defined in the present dissertation as the number of directions in which movements are executed. *n* corresponds to three and it is further assumed that there is an equal probability to execute a movement in each of the three directions. In practice, the bit rate in (bits/min) is used to indicate the ITR and is calculated according to equation 5.22:

$$BTR_t = \frac{60}{\tau} BTR, \quad [5.22]$$

where τ in (seconds/symbol) is the time required to transmit a symbol. In the present thesis, τ corresponds to the time between two consecutive movements.

5.11 Evaluation of the running times of the event detection methods

It can be very tempting to define the actual number of milliseconds it takes a program to be executed as a mean to determine its running time performance. However, such measure is severely biased as this measure depends on various factors such as the machine running the program, the programming language and the compiler to name a few. A more objective measure of the running time of a

program that does not depend on the factors mentioned previously is the complexity analysis of that program. The complexity analysis is one of the most desirable attribute of a given algorithm and is very important to estimate such an attribute in order to discuss the performance of an algorithm. The complexity is defined as a numerical function $T(n)$ representing the running time as function of the input size n. $T(n)$ will measure the numbers of elementary steps required by a given program to be executed provided each step is executed in constant time.The term constant time means that the computer CPU performs each elementary steps as one instruction. Complexity analysis is a technique used to evaluate the the variation of the execution time as a function the the input data and to objectively compare algorithms. This technique can also be used to predict the running time behaviour of the algorithm when the input data size change. In the next sections, a method to derive the time function $T(n)$ is provided and the the function is used to analyse the complexity of the algorithm developed in the present thesis.

5.11.1 Determination of the running Time function

The running time function is evaluated by counting the elementary steps in the algorithm to be analysed. During the counting of the elementary steps, it is assumed that the following operation are performed in constant time:

- Addition and multiplication
- Incrementing a value
- Comparing values
- Assigning a variable
- Looking up a value in an array
- storing a value

The counting of the elementary steps obeys to a certain number of rules. The first states that when tasks are executed in sequence their results are added while when tasks are executed within a loop, their results are multiplied and third rule is illustrated by figures 5.6 and 5.7.

Algorithm 5.6 Sigle loop counting program

1: Initialise X
2: **for** $i = 1$ to n **do**
3: X=X+1
4: **end for**

In figure 5.6, the following steps are required to execute the algorithm. First the number of counts outside the loop and the number of counts inside the loop:

1. initialise *X*
2. Initialise the loop counter *i*

and the number of counts inside the loop:

1. Increment *X*
2. Store the value of *X*
3. Increment the loop counter
4. Store the value of the loop counter
5. Compare the loop counter to *n*

Which gives $T(n) = 5n + 2$ steps.

Algorithm 5.7 nested loop counting program

```
1: Initialise X
2: for i = 1 to n do
3:     for i = j to n do
4:         X=X+1
5:     end for
6: end for
```

In the nested loop in figure 5.7, the inner loop is executed and the outer loop is executed. The count of the algorithm is as follows:

1. Initialise X
2. Initialise the outer loop i
3. initialise the loop counter j
4. Execute the inner loop
5. increment outer loop counter
6. store the value of the loop counter
7. Compare the loop counter to n

The outer loop is executed by initialising and executing the inner loop in $5n + 1$ steps then incrementing the outer loop counter, storing its value and comparing it with n, resulting into a total of $5n+1+3 = 5n+4$ steps for each run of the outer loop. The total number of steps is $n(5n+4)$ and the initialisation of the outer loop and the initialisation of x, which yields into a total of $5n^2+4n+2$ steps. The time function of the program in figure 5.7 is $T(n) = 5n^2 + 4n + 2$.

In some cases and more often, the time function does not solely depend on n but on the input, and in such a case the count of the instruction is made based on the "Worse case analysis" which

is to analysis when the algorithm executes the most instructions to complete as illustrated in fifure 5.8.

Algorithm 5.8 Single loop counting program

```
1: Initialise X where X is a n dimensional vector
2: M = max(X)
3: for i = 1 to n do
4:     if X[i] ≥ M then
5:         X[i]=X[i]+1
6:     end if
7: end for
```

In the algorithm in figure 5.8, depending on the body of the if statement, the for loop may be completely or partially executed. For example if the fist element of the vector **X** is the maximum value of the vector, the loop will only be executed once.In the case the maximum value of the vector is its last element, the loop will be executed exactly n times. The execution of the body of the if statement in the present case depends on the vector **X**. In this case it is assumed that the body of the if statement is executed n times, in accordance with the "Worst case analysis". The analysis of the algorithm in Figure 5.8 is done as follow:

1. Initialise **X**
2. initialise M
3. Look up the maximum element of **X**
4. Store the value of M
5. Initialise the loop counter
6. Execute the body of the if statement
7. Increment the loop counter

8. Store the value of the loop counter

9. Compare the loop counter with n

The if statement is executed by first looking up the i^{th} element of the vector then incrementing and saving the value which is 3 steps. According to the "Worst case analysis", the loop is executed in 6 steps. The initialisation of the loop, the vector and looking up and storing the maximum value of the vector results into 5 steps giving a total number of $6n + 5$ steps. The time function of the algorithm in Figure 5.8 is $T(n) = 6n + 5$.

5.11.2 Asymptotic running time

In the analysis of complexity of an algorithm, it is of interest to investigate the behaviour of the algorithm when n is sufficiently large , e.g. when $n \to \infty$, leading to the asymptotic or the 'Big O' notation. The 'Big O' notation is used to find an function that grows at least as fast as the function $T(n)$ or an upper bound. Given a function g(n), the 'Big O' notation $T(n) = O(g(n))$ means that the function $g(n)$ grows at least as fast as the function $T(n)$. Given the time function $T(n) = 6n+5$, there is a constant term and a term depending on the input size n. For this function, 6n grows faster than 5 so the constant term can be dropped in the analysis. Furthermore, n grows at least as fast as 6n according to the asymptotic method. Hence $6n + 5 = O(n)$ and the function is said to grow linearly. Following the same argument, $5n^2 + 4n + 2 = O(n^2)$ and the corresponding algorithm is called a quadratic algorithm. Complexity analysis then provides a framework to objectively assess or compare algorithms. In the previous sections, Algorithm 5.8 is linear while Algorithm 5.7 is quadratic. As a quadratic function is growing faster than a linear function, it can be concluded that Algorithm 5.8 is more efficient than Algorithm 5.7. Table 5.2 provides a summary of the common names of the most common upper bound functions.

Table 5.2: Common names of complexity functions.

Function	Common name
1	constant
$\log n$	logarithmic
$\sqrt{n}$	root n
n	linear
$n \log n$	quasi-linear
n^2	quadratic
n^3	cubic
$n^d, d > 3$	polynomial
2^n	exponential
$n!$	factorial

The rules to to determine complexity of algorithms can be summarised as follow:

(a) Always assume the worst possible case

(b) Add complexities for independent modules

(c) Multiply complexities for nested loops.

(d) If an algorithm consists of independent modules, the most complex module determines the complexity of the whole algorithm.

An in-depth discussion of complexity analysis is provided in (Cormen, Leiserson and Rivest, 2001; Dasgupta et al., 2006; Nievergelt and Hinrichs, 1999).

5.11.3 Complexity analysis of the different algorithms

Another evaluation of a detection method is how long the method takes to detect an event and this is provided by the complexity analysis of the method as described in section 5.11. In the present section, the evaluation of the detection algorithms developed in the present chapter is discussed. The section will only discuss custom built algorithms and not inbuilt functions of the Matlab software. Most of the operations involved in the algorithms are matrix operations that are computational demanding. For such operations, the number of steps required to perform the key matrix operations are provided in Appendix A in (Higham, 2008). Given two real $n \times n$ general non symmetric matrices **A** and **B**, the multiplication $(\mathbf{AB})$of the two matrices and the computation of the inverse of such matrix requires $2n^3$ basics operations.

The WMM method requires the computation of the Lipschitz function while the MD and the Bootstrap methods rely on the computation of the Mahalanolobis distance. The following sections provide the count of the basic steps required to execute the functions previously mentioned.

Complexity analysis of the Lipschitz function

The functions involved in the WMM method are the LE computation and the Lipschitz function evaluation, respectively. The evaluation of the Lipschitz coefficients is defined by Equation [5.10] which is a sequential execution of four additions requiring a linear running time $4n$. The computation of the Lipschitz function using Algorithm 5.1 calls Equation [5.10] in a nested loop. Algorithm 5.1 as quadratic asymptotic running time $T(n) = O(n^2)$. The WMM uses Algorithm 5.2 , calling the Lipschitz function in a single loop, resulting in a $(T(n) = O((n^2)n))$ that is a cubic $(T(n) = O(n^3))$ running time. The WMM algorithm has a cubic running time.

Complexity analysis of the Mahalanolobis distance function

The modules involved in the MD and Bootstrap methods are the the evaluation of the MD baseline, the Bootstrap distance distribution and MD function.

The Mahalanolobis is computed using Equation [5.15] involving one matrix inversion and two matrix products(one left and one right products of the inverse matrix) resulting in $3(2n^3)$. The MD function has a cubic time function ($T(n) = O(n^3)$.

The Bootstrap method calls the MD function to compute the bootstrap distribution ($3(2n^3) * n$), resulting in a $T(n) = O(n^4)$. The running times of the different functions and algorithms are summarised in Table 5.3.

Table 5.3: Computation complexity of the event detection algorithms of the different methods.

Asymptotic running times of the event detection algorithms		
WMM method	Computation of the Lipschitz exponent	$O(n)$
	Computation of the Lipschitz function	$O(n^2)$
	Computation of the Kurtosis	$O(n)$
	Computation of the event index	$O(n)$
	Event detection using the WMM algorithm	$O(n^3)$
MD method	Computation of the MD distance	$O(n^3)$
	Computation of the threshold interval	$O(n)$
	Computation of the baseline	$O(n^3)$
	The whole MD algorithm	$O(n^3)$
Bootstrap method	Computation of the Bootstrap distance distribution	$O(n^3)$
	Computation of the MD distance	$O(n^3)$
	The whole Bootstrap algorithm	$O(n^4)$

5.12 Summary

In this chapter three methods were developed to detect events in EEG data. The first method uses LEs which are the rate of change of the derivative of wavelet maxima across scales. The method uses the ratio between the baseline and activity Kurtosis as an empirical threshold. The second

models the EEG signal as a Gaussian distribution and calculates the MD between the baseline and activity EEG data to mine the data for events. A two standard deviation interval is used in this method to assess event occurrence. The third method uses a Bootstrap of the MD and assesses event occurrence based on the significance of the difference between the baseline and activity EEG data. Finally, methods were presented to evaluate the different methods in terms of their accuracy and information transfer. In the following chapter, the methods will be applied and evaluated on EEG data.

Chapter 6

Results

This chapter evaluates the algorithms described in chapter 5, by applying each algorithm to EEG data. Three of the algorithms were used to predict movement initiation, the results are described in sections 6.2, 6.3 and 6.5 respectively. Finally, the overall performance of each algorithm is described in sections 6.6 and 6.7.

6.1 Kinematics

In the present section, the peak velocities and trial latencies of reaching arm movements were examined, in an attempt to compare how each subject executed the protocol. In this section, the reaching velocity, which is the maximum velocity achieved during a reaching movement is considered. In Figure 6.1, the box-plots of the peak velocities for an individual subject is shown. A statistical examination of the mean value of the peak velocities of the nine participants using a Kruskal-Wallis test demonstrated that there was a significant difference in peak velocity between participants $(p = 0)$ at a 5% level of significance (Gibbons and Chakraborti, 2003; Hollander and Wolfe, 1973). The horizontal lines on the box-plots show the median values of the peak velocity. Figure 6.1 shows that the average peak velocity varies significantly and ranges from approximately $200\,\mathrm{mm\,s^{-1}}$ (for subject 6) to approximately $900\,\mathrm{mm\,s^{-1}}$ (for subject 4). Furthermore, peak velo-

cities tend to exhibit greater variability in participants with high peak velocities while the variability is smaller for subjects with low peak velocities. This difference in peak velocities might suggest that individual subjects employ different approaches to the experimental protocol.

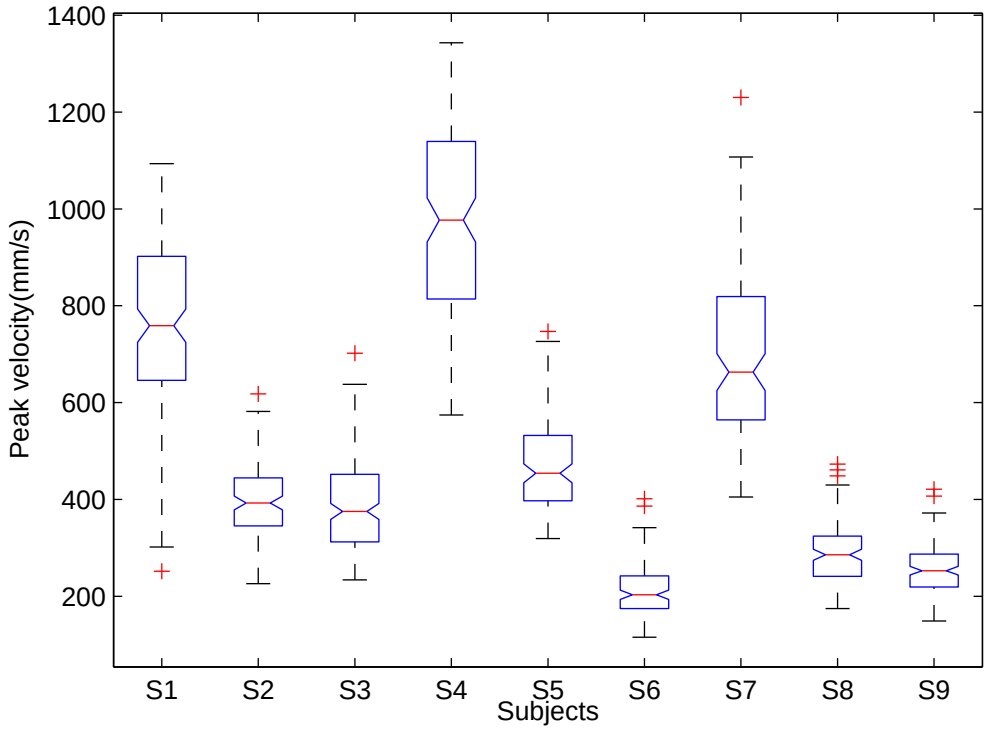


Figure 6.1: Box-plots of the peak velocity for individual subjects. The horizontal lines show the median peak velocity for each subject. The crosses indicate outlier peak velocity values.

Figure 6.2 shows the average trial duration for each subject. A Kruskal-Wallis test demonstrated that there was a significant difference in the duration of trials between subjects $(p = 0)$ at a 5% level of significance. This difference in trial latency between the subjects is an indication that there was no consistency between subjects in the way the movements were executed.

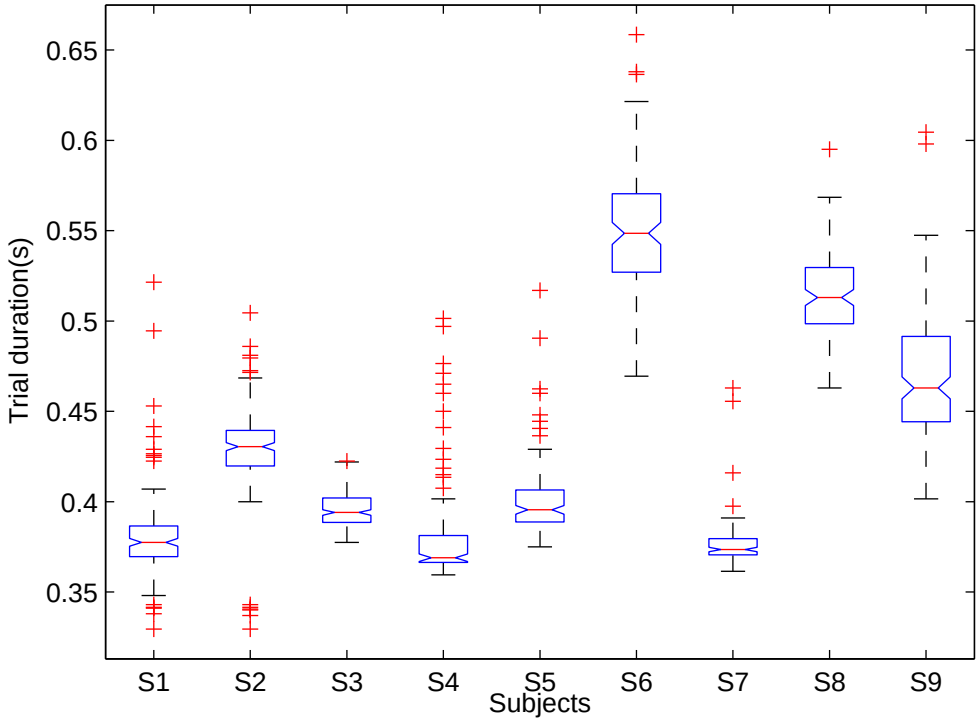


Figure 6.2: Box-plots of the duration of reaching hand movement execution, defined as the execution of hand movement from the central position to one of the target positions as defined in Figure 4.4.

6.2 Predicting the initiation of reaching hand movements using the WMM

In this section, the WMM method used to predict the initiation of reaching hand movements is presented. The detection rate of both a single electrode and the pooled electrodes are presented to investigate the inter subject variability. Finally, in this section, the detection times of the movement initiation between $1.5\,\mathrm{s}$ and $250\,\mathrm{ms}$ before movement onset are provided. The raw EEG data from each recording channel was resolved in 71 frequency components in the Beta, Mu and SCP frequency bands using the CWT. PCA technique was used to reduce the dimension of the time-frequency data in the Beta, SCP and Mu frequency bands respectively.

6.2.1 PCA sub-spaces

Figure 6.3 shows the grand average percentage of cumulative variance of the 71 eigen values in the Beta, Mu and SCP frequency bands respectively. All the 71 eigen values represent 100% of the variance of the data. The Figure shows that the cumulative variance does not increase significantly from 90% of the cumulative value in the Beta and Mu frequency bands, corresponding to 8 and 21 grand average principal components in the respective frequency bands. In conclusion, 8 and 21 PCA components were used on average to predict the initiation of reaching hand movements in the Beta and Mu frequency bands. In the SCP band, the percentage of cumulative value of 90% corresponds to 51 grand average PCA components. 51 PCA components on average were used to predict the initiation of reaching hand movements in the SCP frequency band. The selection of the number of PCA components is described in section 5.4.2.1.

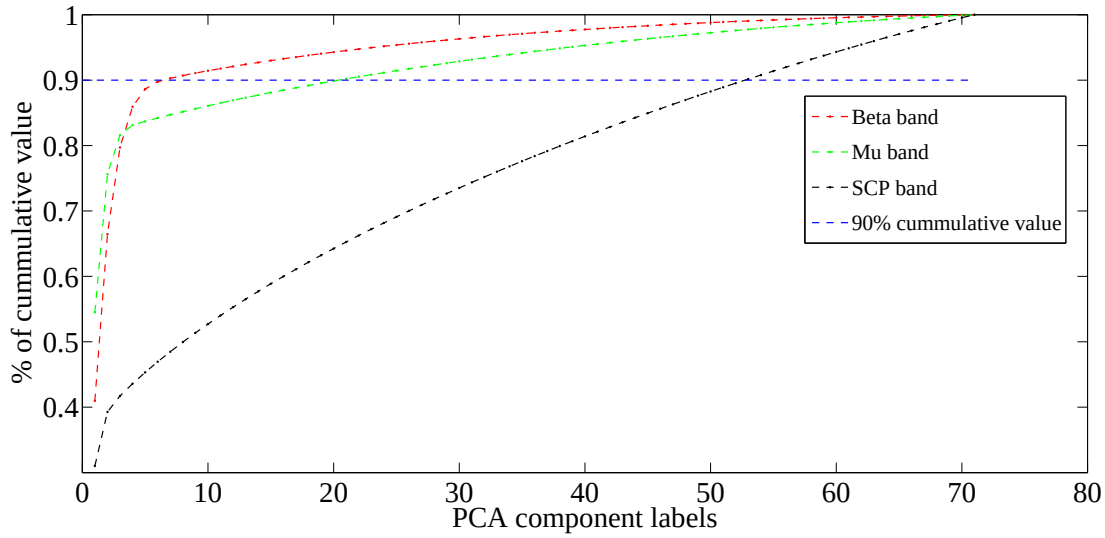


Figure 6.3: The grand average percentage of cumulative variance plots for all 71 singular values in the Beta, Mu and SCP frequency bands. The intersections of the 90% cumulative plot (horizontal dotted line) and the cumulative variance plots of the Beta (red), Mu (green) and SCP (black) frequency bands show the grand average number of singular values of $(8, 21, 51)$ that correspond to 90% of the data in the respective frequency bands.

6.2.2 The event index

The algorithm described in section 5.6 was used to detect the initiation of reaching movements (see Algorithm 5.2). A data window of three seconds was investigated as illustrated in Figure 5.3. To illustrate how the prediction of reaching hand movement initiation was performed on a single trial, the time course of the event index, computed as described in Figure 5.4, is depicted in Figure 6.4. The Figure shows that the event index is above a predefined threshold value of one during at least $250\,\mathrm{ms}$ indicating the occurrence of an event (see section 5.9). The prediction time of the initiation of the reaching movement corresponds to the vertical dotted line in Figure 6.4.

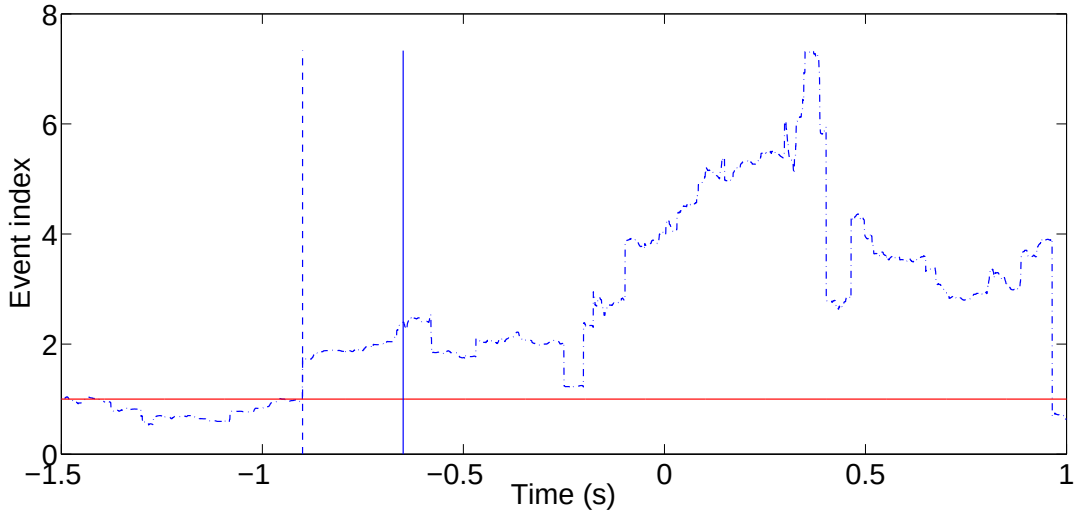


Figure 6.4: Event index plot as a function of time. The event index is above three during at least $250\,\mathrm{ms}$ *represented by the time interval between the solid and dotted lines in the figure. The solid line indicates the occurrence of an event about* $900\,\mathrm{ms}$ *before movement initiation which corresponds to time point zero.*

6.2.3 Grand Average Detection Rate of pooled subjects

To investigate the performance of an individual electrode in each frequency band (SCP ($0.5\,\mathrm{Hz}$ - $1.5\,\mathrm{Hz}$), Mu ($8\,\mathrm{Hz}$ - $13\,\mathrm{Hz}$) and Beta ($12\,\mathrm{Hz}$ - $30\,\mathrm{Hz}$)), the detection rate of pooled subjects was estimated for the WMM method. Figure 6.5 shows the grand average movement intention detection in the SCP, Mu and Beta frequency bands, respectively, for each electrode for all the subjects. The results show that the WMM method exhibits a high grand average detection rate in the Mu

frequency band for electrodes Fc4, Fc6, Cp5, Cp3, Cp1, Cpz, Cp2, Cp4 and Cp6. For the remaining electrodes, the results show that there is no electrode with a consistent high performance in the SCP, Mu and Beta bands, respectively. The average detection rate of individual subjects shows that the maximum average detection rate using the WMM method is about 80% in the SCP, Mu and Beta bands (see Figures B.1(a), B.3(a), B.5(a), B.6(a), B.8(a) in Appendix B) and that there is no electrode with a consistent high detection rate.

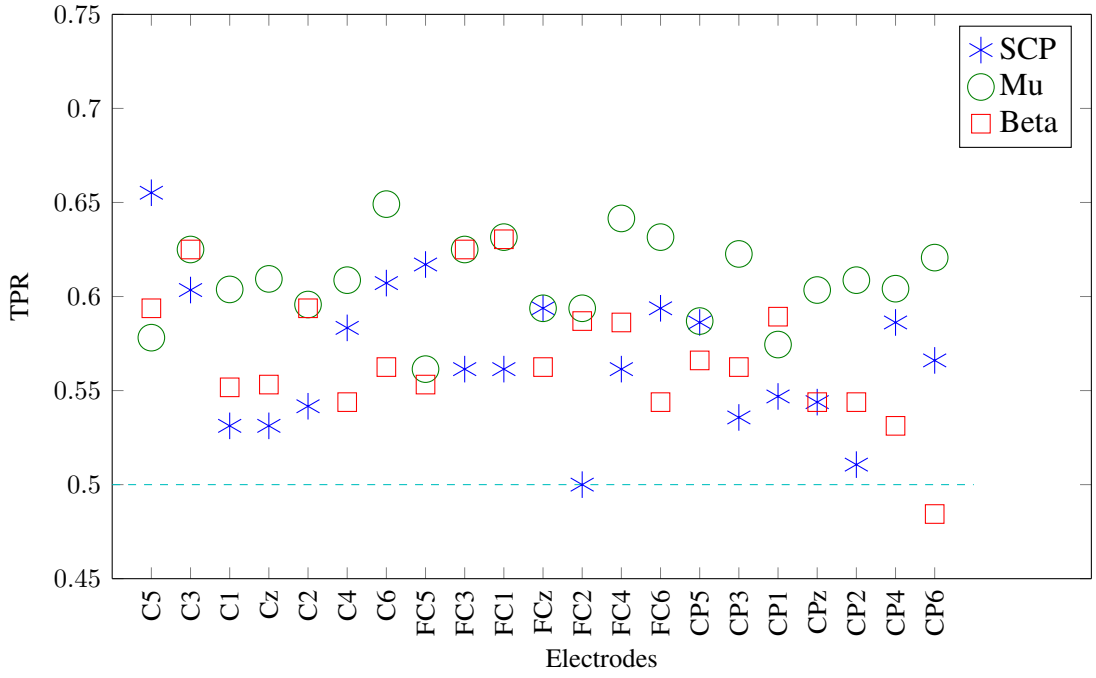


Figure 6.5: Grand average TPR using the WMM in the SCP, Mu and Beta frequency bands. The horizontal line represents the chance level defined as 50% *detection rate.*

6.2.4 Topographic plots

To investigate the prediction of reaching hand movement onset according to their spatial location on the scalp, the topological plots of the grand average prediction rates in the SCP, Mu and Beta bands are presented in sections 6.2.4.1, 6.2.4.2 and 6.2.4.3. Each plot represents a topographic map of the grand average prediction rate at each electrode position in a 2-D view using a spline interpolation on a Cartesian grid.

6.2.4.1 Grand average TPR topographic plot in the SCP (0.5 Hz - 1.5 Hz) band

The topographic plot is presented to investigate the performance of each electrode according to its position on the scalp. Figure 6.6 shows the topographic plot in the SCP band for the WMM method. The results show that on average, the contralateral electrodes, particularly electrodes C5 and FC5 exhibit higher prediction rates. The topographic plots of individual subjects show that the majority of the subjects (8 out of 9) show a good detection rate for contra lateralelectrodes (see Figures C.1(a), C.2(a),C.4(a),. . . , C.9(a) in Appendix C), five subjects show a good detection rate for the ipsilateral electrodes (see Figures C.3(a), C.4(a),C.6(a),. . . , C.8(a) in Appendix C) and four subjects exhibit a good detection rate for both contralateral and ipsilateral electrodes (see Figures C.4(a), C.6(a),. . . , C.8(a) in Appendix C). The results from the topographic plots of individuals subjects confirm the results from the grand average topographic plot that the contralateral electrodes have the higher detection rate (see Figure 6.6). Furthermore, the results from the topographic plots of individual subjects show that there is no electrode with a consistent high performance.

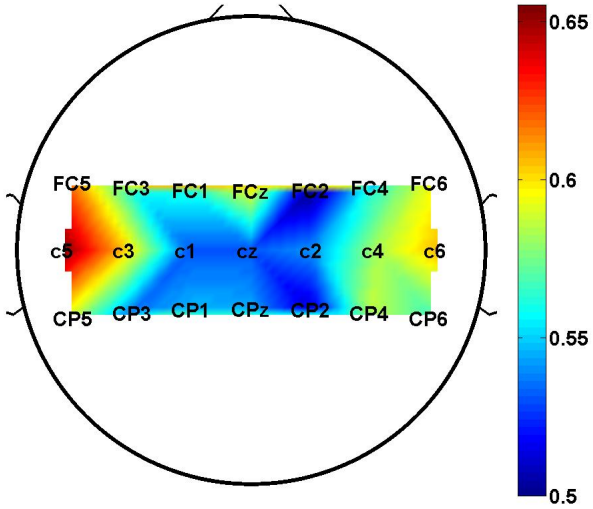


Figure 6.6: Grand average TPR topographic plot in the SCP (0.5 Hz - 1.5 Hz) frequency band for the WMM method. The figure shows that C5 and FC5 are the best performing electrodes.

6.2.4.2 Grand average TPR topographic plot in the Mu (8 Hz - 13 Hz) band

Figure 6.7 shows the topological plot in the Mu band for the WMM method. It can be seen in the figure that on average electrodes from both contralateral and ipsilateral sides have a good detection rate in the Mu band with the ipsilateral electrodes showing a higher detection rate. The topographic plots for individual subjects show that, out of the nine subjects, eight subjects show a good detection rate for ipsilateral elctrodes (see Figures C.1(b), ..., C.5(b), C.7(b),..., C.9(b) in Appendix C), and six subjects show a good detection rate for both ipsilateral and contralateral electrodes (see Figures C.2(b), ..., C.5(b), C.7(b),..., C.9(b) in Appendix C). The results from the topographic plots of individuals subjects show that both contralateral and ipsilateral electrodes have a higher detection rate, in agreement with the results from the grand average detection rate (see Figure 6.7), and that there is no electrode with a consistent high performance.

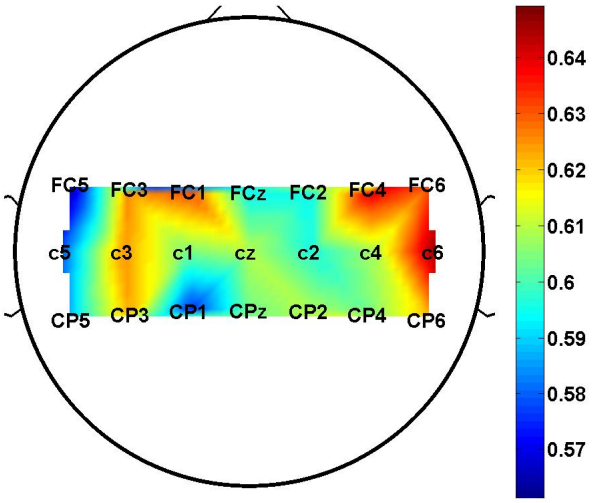


Figure 6.7: Grand average TPR topographic plot in the Mu frequency band for the WMM method. The figure shows that C6, FC4, FC6, CP6, C3, FC3, FC1 and CP3 are the best performing electrodes.

6.2.4.3 Grand average TPR topographic plot in the Beta (12 Hz - 30 Hz) band

The topological plot in the Beta band for the WMM method is depicted in Figure 6.8 in an attempt to evaluate the detection rate of each electrode according to its location on the scalp. Fig-

ure 6.8 shows that contralateral electrodes in general show a good detection rate and particularly electrodes C3, FC3 and FC1. The topographic plots of individual subjects show that, out of the nine subjects, all the subjects show a good detection rate for the contralateral electrodes (see Figures C.1(c),. . . , C.9(c) in Appendix C), and five subjects show a good detection rate for both the ipsilateral and contralateral electrodes (see Figures C.2(c), C.3(c),. . . , C.5(c), C.7(c) in Appendix C). The results from the topographic plots of individuals subjects are in line with the results from the grand average topographic plot that the contralateral electrodes have the higher detection rate (see Figue 6.8) and that there is no electrode with a consistent high performance.

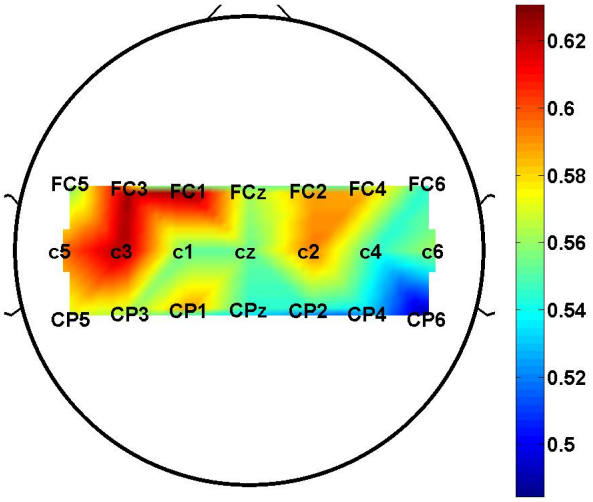


Figure 6.8: Grand average TPR topographic plot in the Beta frequency band for the WMM. C3, FC3 and FC1 are the best performing electrodes.

6.2.5 Average detection rate of individual subjects

The average performance of all the electrodes is investigated in order to assess the variability of event detection between individual subjects. Figures 6.9, 6.10 and 6.11 show the box-plots of the average TPR of the pooled electrodes for each subject. A Kruskal-Wallis test showed that there is a significant difference between the performance of each subject in the SCP $(p = 0.0003)$, Mu $(p = 0.0041)$ and the Beta $(p = 0.0029)$ frequency bands at a 0.05 significance level.

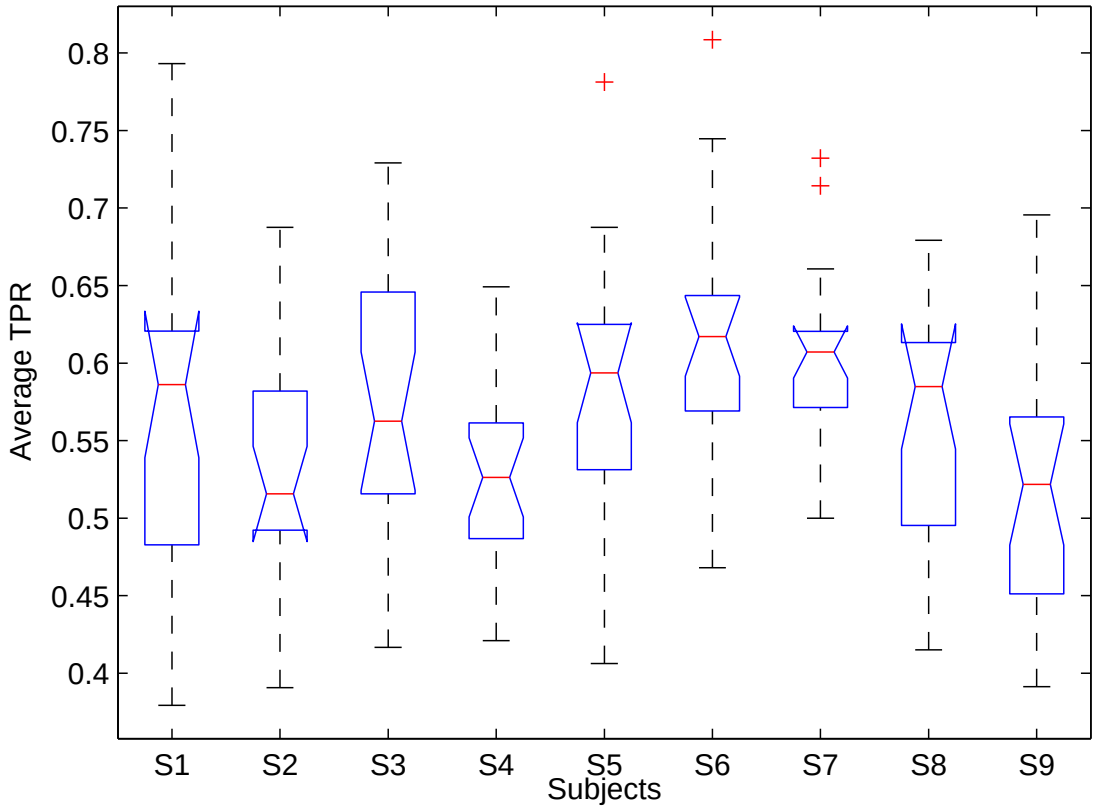


Figure 6.9: The average detection rate of the WMM method in the SCP band for each subject.

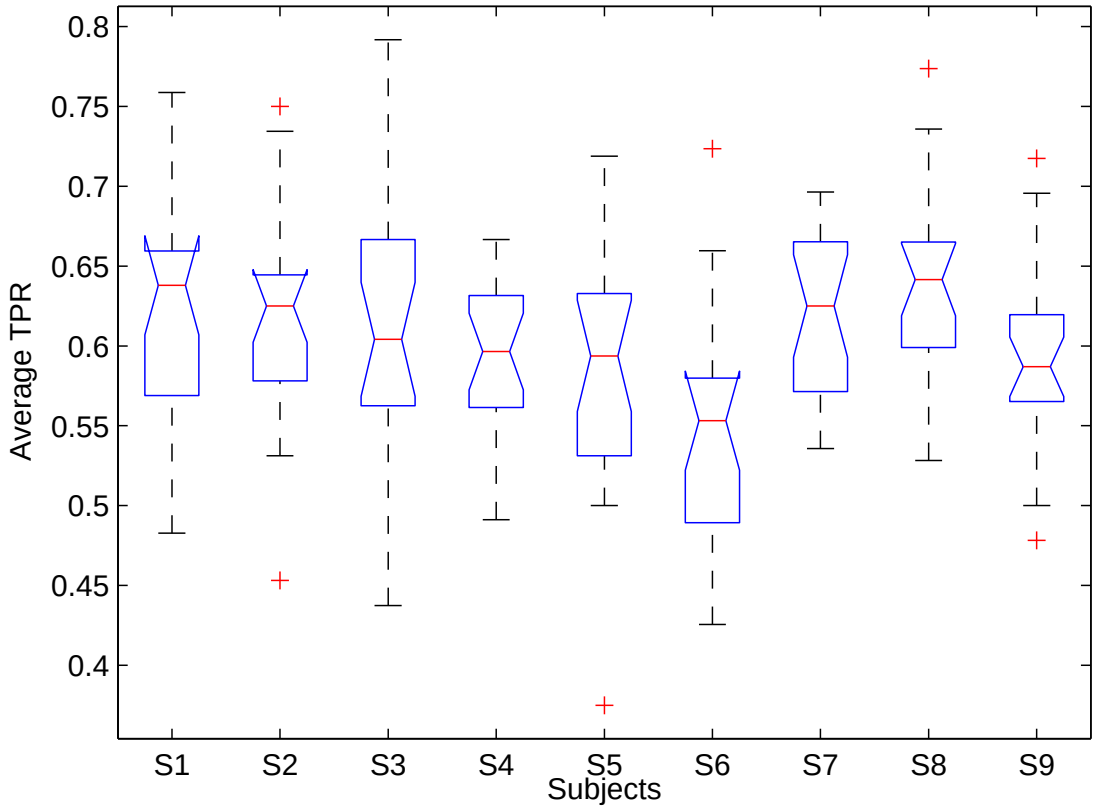


Figure 6.10: The average detection rate of the WMM method in the Mu band for each subject.

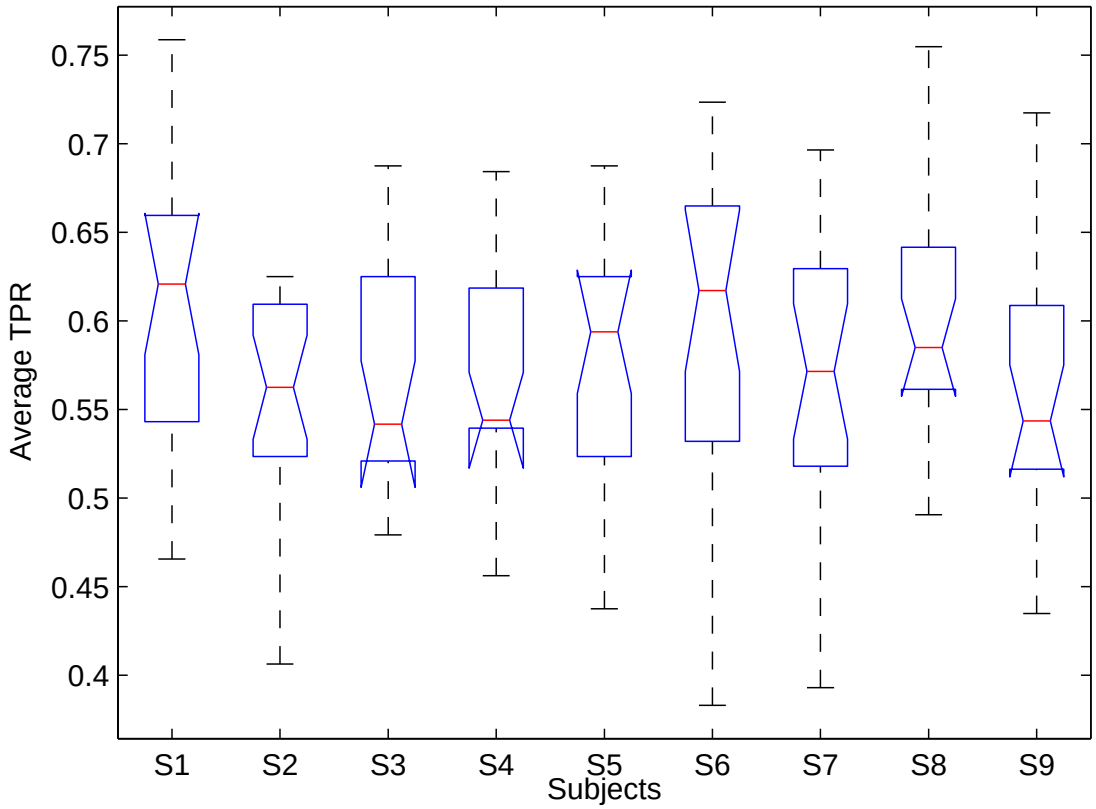


Figure 6.11: The average detection rate of the WMM method in the Beta band for each subject.

6.2.6 Detection time

The box-plots of movement initiation detection time grand average for each subjects are depicted in Figures 6.12, 6.13 and 6.14 in an attempt to investigate the time of movement initiation in the time interval from $-1500\,\mathrm{ms}$ to $-750\,\mathrm{ms}$ with respect to movement onset. A Kruskal-Wallis test demonstrated that there is no significant difference in the event detection time between the subjects in the SCP ($p = 0.7013$), mu band $p = 0.129$ and beta band $p = 0.1027$) at a 0.05 significance level. It can be seen in the figures that the mean detection times in the three different frequency bands are around $1\,\mathrm{s}$ prior movement initiation.

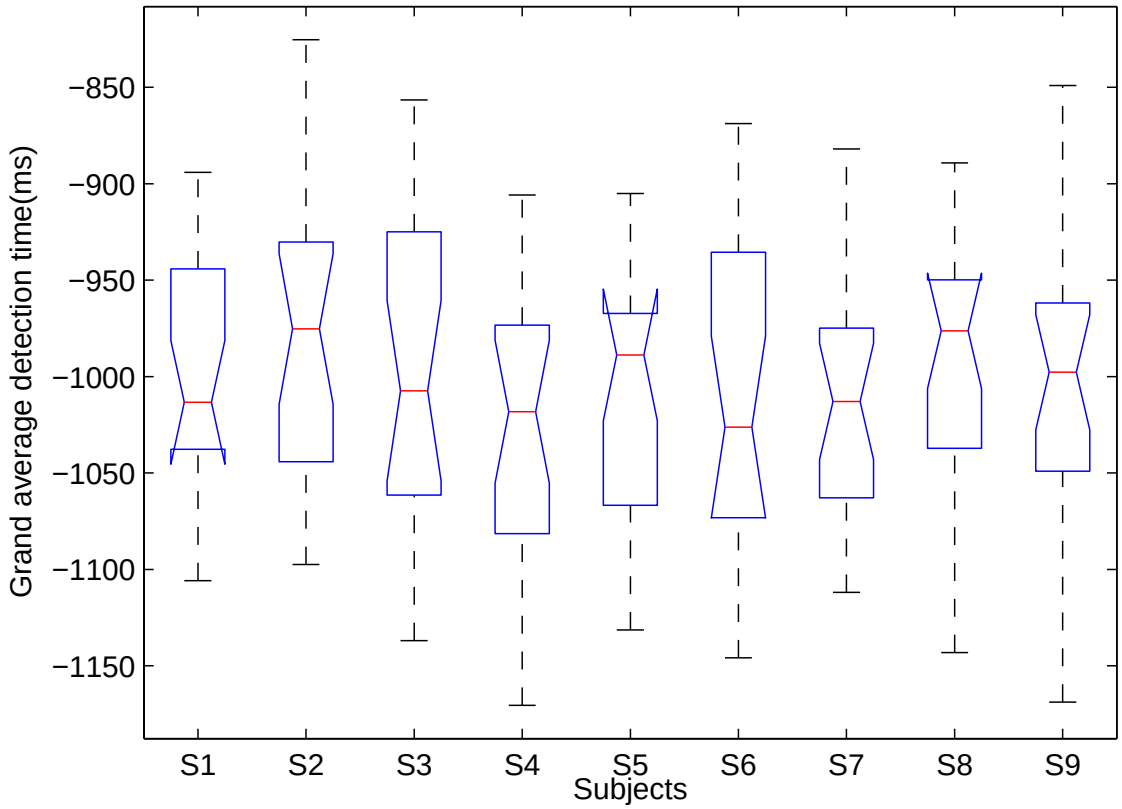


Figure 6.12: The grand average detection time for all the electrodes in the SCP band for each subject. The horizontal lines show the median values of the detection times. Time point zero represents the movement initiation time.

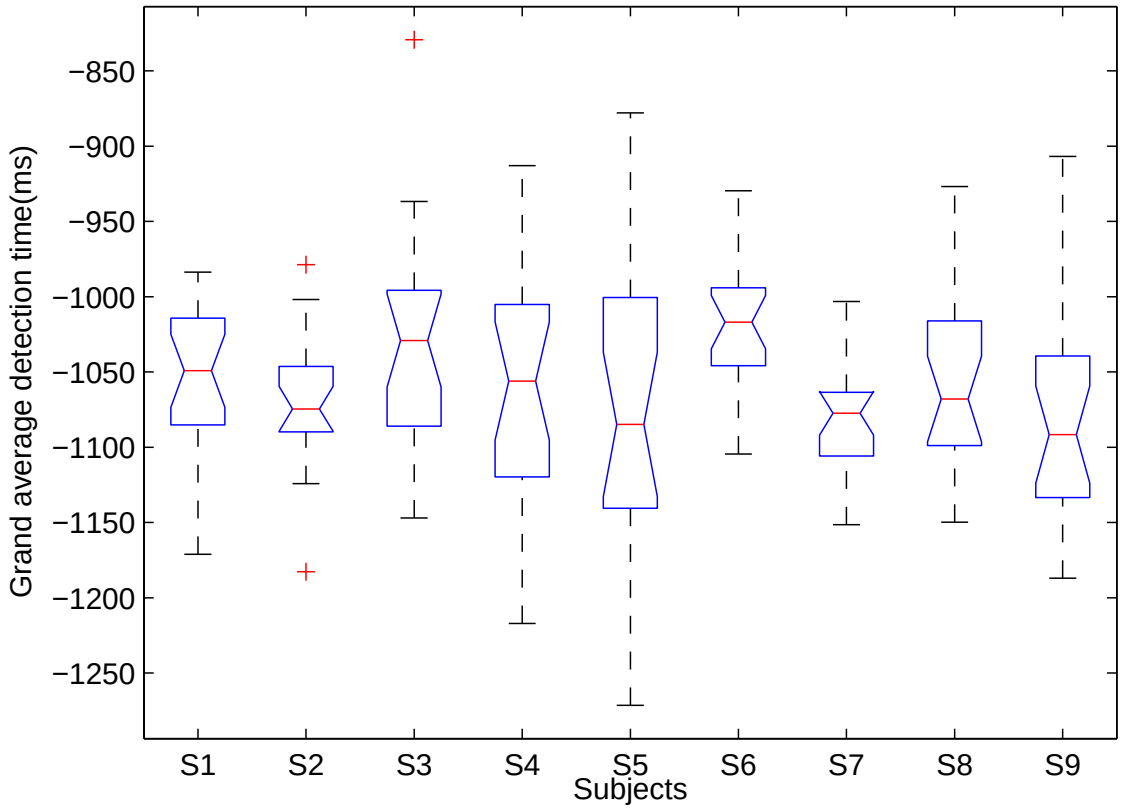


Figure 6.13: The grand average detection time for all the electrodes in the Mu band for each subject. The horizontal lines show the median values of the detection times. Time point zero represents the movement initiation time.

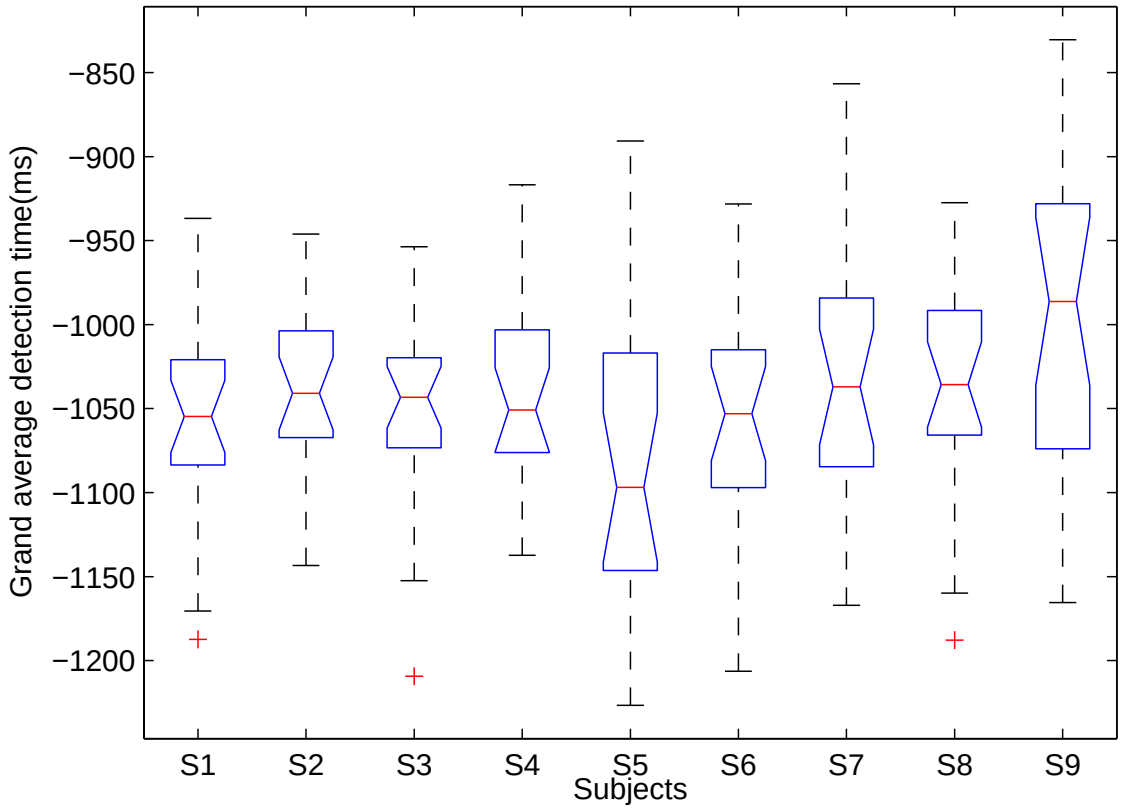


Figure 6.14: The grand average detection time for all the electrodes in the Beta band for each subject. The horizontal lines show the median values of the detection times. Time point zero represents the movement initiation time.

6.3 Predicting the initiation of reaching hand movements using the MD

This section illustrates how the MD was used to detect reaching hand movement onset. The detection rate of each electrode for the pooled subjects (grand average detection rate) and the average detection rate for the pooled electrodes for each subject (average detection rate) are presented to investigate the performance of each electrode and the variability of the performance between the subjects. The movement initiation detection times are also presented in this section.

6.3.1 The timeline of the MD

The algorithm described in section 5.7 is used to predict reaching hand movements (see Algorithm 5.4). A time window of three second was used (see Figure 5.3). This method uses the Mahalanolobis distance which was evaluated according to Figure 5.5. The timeline of the MD is shown in Figure 6.15 to illustrate how it was used to predict the onset of reaching hand movements. It can be seen in Figure 6.15 that the MD consistently exceeds two standard deviations of the mean baseline distance (see Algorithm 5.3) for at least $250\,\mathrm{ms}$ indicating the initiation of a movement (See the vertical solid line in Figure 6.15).

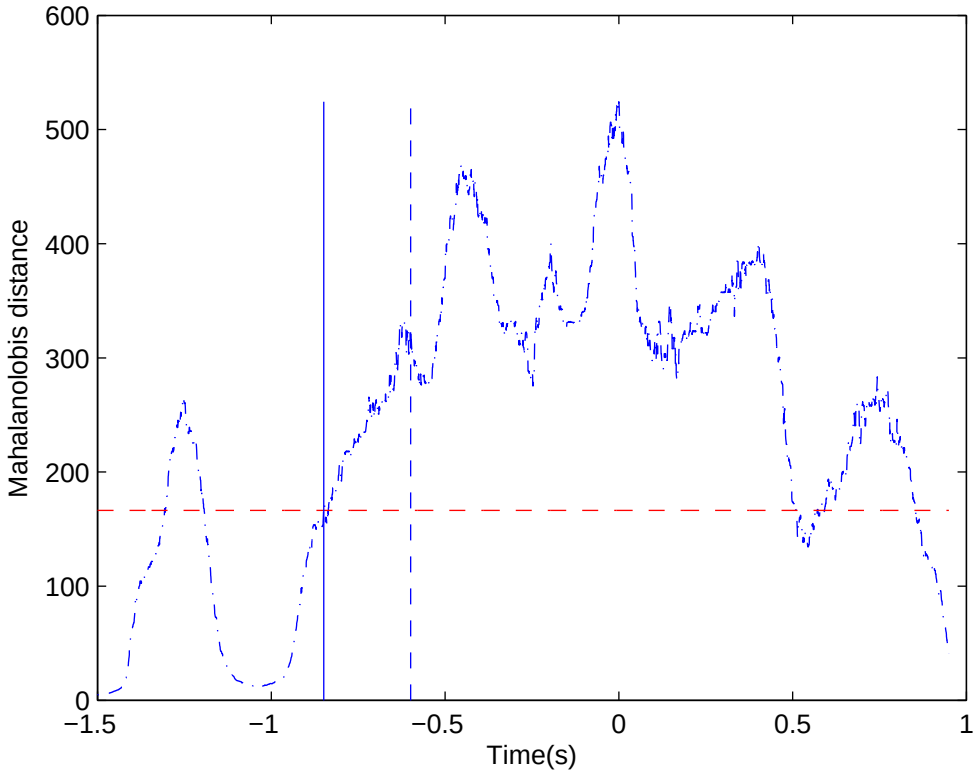


Figure 6.15: The Mahalanobis distances between the baseline ($500\,\mathrm{ms}$*) and the activity windows (*$50\,\mathrm{ms}$*). The figure shows that the MD consistently above two standard deviations of the mean baseline distance (horizontal line) during* $250\,\mathrm{ms}$ *from about* $1.25\,\mathrm{s}$ *which is the detection time.*

6.3.2 Grand average detection rate using the MD

The grand average detection rate of each electrode is shown in Figure 6.16 to investigate the performance of individual electrodes during the detection of reaching hand movements. In the present section only the detection rate in the SCP band is reported as the detection rates in the Mu and Beta bands were consistently below the chance level of 50%. The grand average detection rate shows that the detection rate of the electrodes in each electrode group is symmetric around their central electrodes (cz, FCz and CPz). Furthermore, the results show no particular performance affinity between ipsilateral and contralateral electrodes. The performances of individual subjects show that the highest detection rate using the MD is around 70% for particular subjects. See Figures B.2(b), B.3(b), B.6(b), B.7(b), B.8(b) in Appendix B.

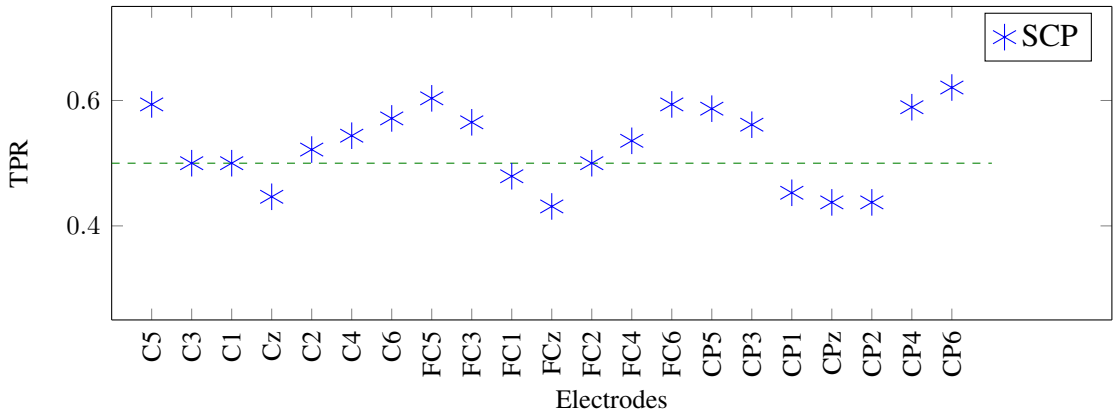


Figure 6.16: The grand average detection rate of the pooled subjects for each electrode in the SCP (0.5 Hz - 1.5 Hz*) frequency band using the MD.*

6.3.3 Grand average TPR topographic plot for the MD method

The topological plot in the SCP band for the MD method is shown in Figure 6.17 in an attempt to compare the detection rate of each electrode according to its location on the scalp. It can be seen in Figure 6.17 that both contralateral and ipsilateral electrodes exhibit a good detection rate, particularly electrodes Fc5, C5, Cp5, Fc6, C6, Cp6 and Cp4. The topographic plots of individual subjects show that, out of the nine subjects, eight subjects show a good detection rate for the contralateral

electrodes (see Figures C.1(d),. . . , C.7(d), C.9(d) in Appendix C), seven subjects show a good detection rate for the ipsilateral electrodes (see Figures C.1(d), C.3(d),. . . , C.8(d) in Appendix C), and six subjects show a good detection rate for both the ipsilateral and contralateral electrodes (see Figures C.1(d), C.3(d),. . . , C.7(d) in Appendix C). The results from the topographic plots of individuals subjects are in line with the results from the grand average topographic plot that both the contralateral and the ipsilateral electrodes have the higher detection rate (see Figure 6.17) and that there is no electrode with a consistent high performance.

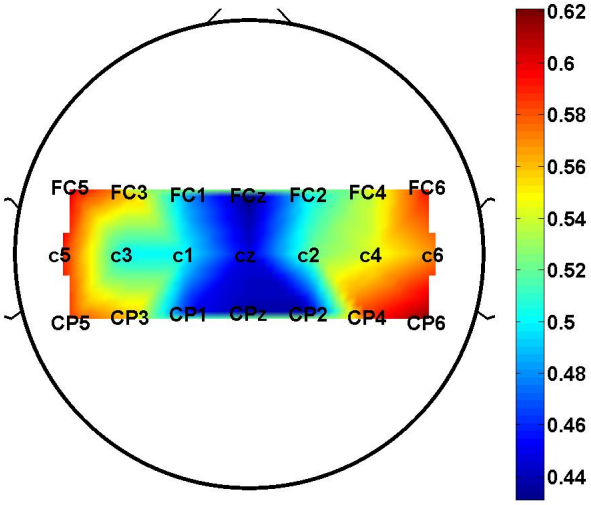


Figure 6.17: Grand average TPR topographic plot in the SCP frequency band for the MD method. The figure shows that C6, FC6, CP4, CP6, C5, FC5 and CP5 are the best performing electrodes.

6.3.4 Average detection rate for individual subject using the MD

The average detection rate of individual subjects is shown in Figure 6.18 to gain an insight in to the consistency of the performance of each subject. A statistic analysis (Kruskal-Wallis test) demonstrated that there was a significant difference in performance between subjects $(p = 0.0002)$ at a 0.05 significance level.

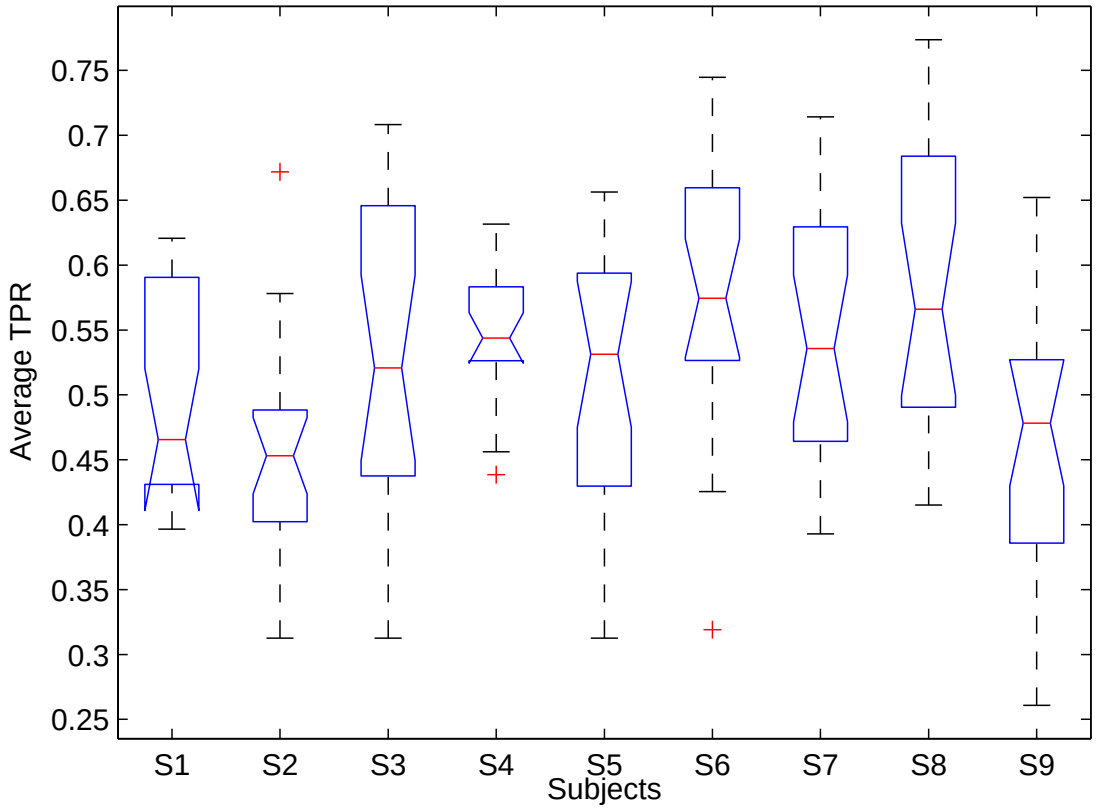


Figure 6.18: The average detection rate of all the electrodes for each subject in the SCP ($0.5\,\mathrm{Hz}$ - $1.5\,\mathrm{Hz}$*) frequency band using the MD.*

6.4 Detection time of the MD method

The detection of reaching movement onset times were investigated to see the consistency of the movement onset time (see Figure 6.19). Figure 6.19 shows the average detection time of all the electrodes for each subject. A Kruskal-Wallis test demonstrated that there was a significant difference ($p = 0.0029$) in the detection time between the subjects at a 0.05 significance level. The results show that the mean detection time is around $1\,\mathrm{s}$.

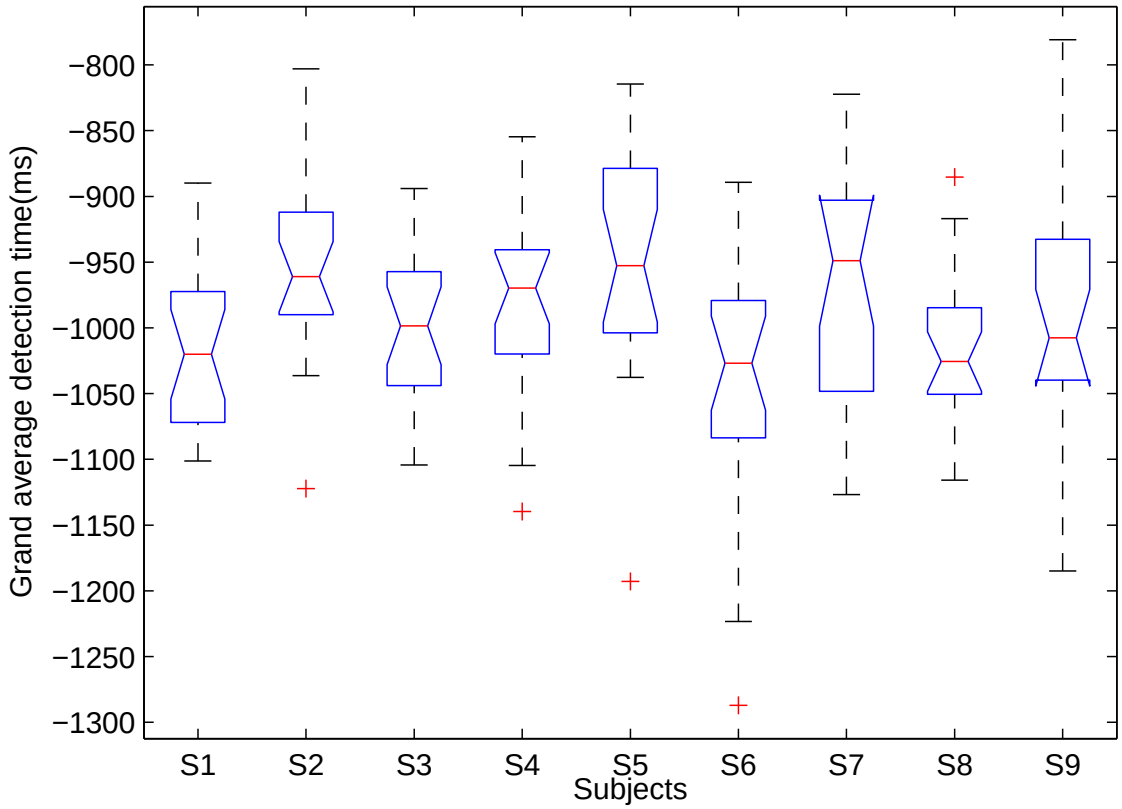


Figure 6.19: The average detection times with respect to movement initiation of all the electrodes for each subject in the SCP band using the MD.

6.5 Detecting the initiation of reaching hand movements using the bootstrap method

The MD method presented in section 5.7 depends on a tuning parameter which was set empirically. The bootstrap method, based on Algorithm 5.5 (see section 5.8), is another potential method to detect events, and does not use a tuning parameter like the MD method. The present method is closely related to the MD method. However, in contrast to the MD method, a bootstrap distribution of the Mahalanolobis distance is used. The prediction of reaching hand movements was performed

by testing that the MD distance falls in the corresponding confidence interval of its bootstrap distribution. In the present section, a 95% percentile confidence interval (see Algorithm 3.4) of the MD bootstrap distribution was considered as this confidence interval does not require the data to be normally distributed and the algorithm was more computationally efficient.

6.5.1 Grand average detection rate using the bootstrap method

The detection rate of individual electrodes was investigated by evaluating the average detection rate of each electrode for the pooled subjects (see Figure 6.20). The figure shows that the performance between electrodes is inconsistent. The performance of individual electrodes for each subject shows a maximum average detection rate of around 70% (See Figures B.1(c), B.2(c), B.4(c), B.5(c), B.7(c), B.9(c) in Appendix B).

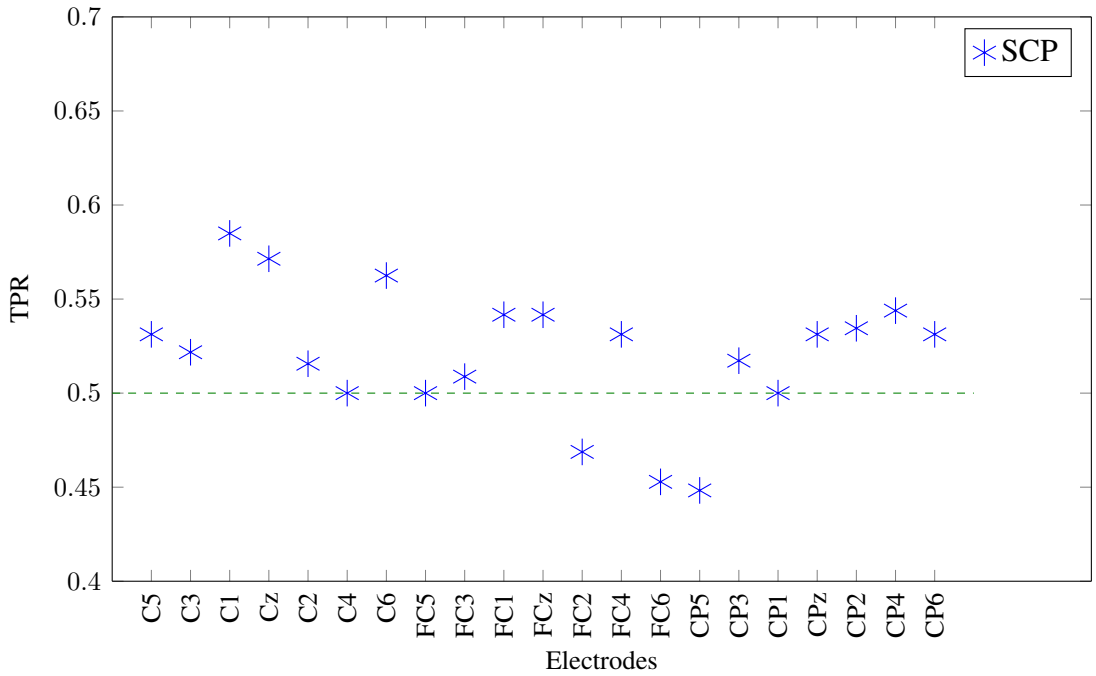


Figure 6.20: The grand average detection rate for individual electrode in the SCP ($0.5\,\text{Hz}$ - $1.5\,\text{Hz}$*) band using the bootstrap method. The detection rate represents the average detection rate of the pooled subjects for each electrode using the bootstrap method. The horizontal line shows chance level.*

6.5.2 Average TPR topographic plot using the bootstrap method

Figure 6.21 shows the topographic plot in the SCP band for the bootstrap method. It can be seen in Figure 6.21 that the electrodes at the centre of the scalp exhibit a higher detection rate. The topographic plots of individual subjects show that six subjects exhibit a good detection rate for the electrodes covering the central cortical region (see Figures C.2(d), C.5(d), ..., C.8(d) in Appendix C) and the remaining subjects have a good detection rate for both the contralateral and the ipsilateral electrodes. The average detection rate of the topographic plots of individual subjects show that the method did not indicate any dependency on the topographic positions of the electrodes.

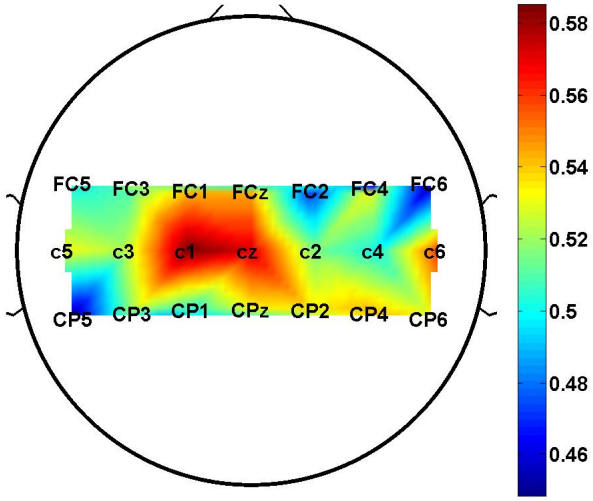


Figure 6.21: The topographic plot in the SCP (0.5 Hz - 1.5 Hz) frequency band using the bootstrap method. The figure shows that C1 and CZ are the best performing electrodes.

6.5.3 Average detection rate using the bootstrap method

The variability of the detection of the initiation of reaching hand movements between the subjects was investigated by evaluating the detection rate of all the electrodes for each subject (see Figure 6.22). A Kruskal-Wallis test demonstrated that there was a significant difference in the performance between subjects ($p = 3.75489 \times 10^{-12}$) at a 0.05 significance level.

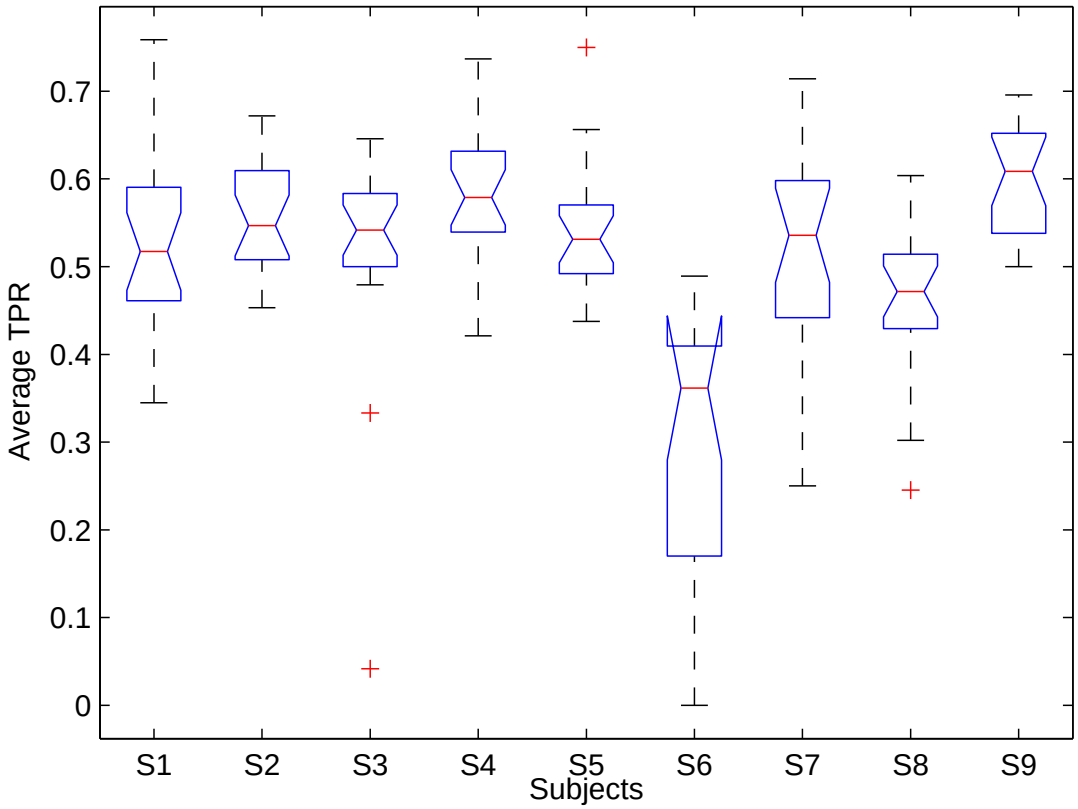


Figure 6.22: The average detection rate of all the electrodes for each subject using the Boot strap method.

6.5.4 Detection time using the bootstrap method

The temporal onset of reaching hand movements was investigated (see Figure 6.23) to gain an apprehension of the average detection time of reaching hand movements for each subject. A Kruskal-Wallis test demonstrated that the temporal onset of reaching hand movements between subjects was not significantly different ($p = 0.4294$) at a 0.05 significance level. It can be seen in the figure that the mean detection time is around $900\,\mathrm{ms}$ before movement initiation.

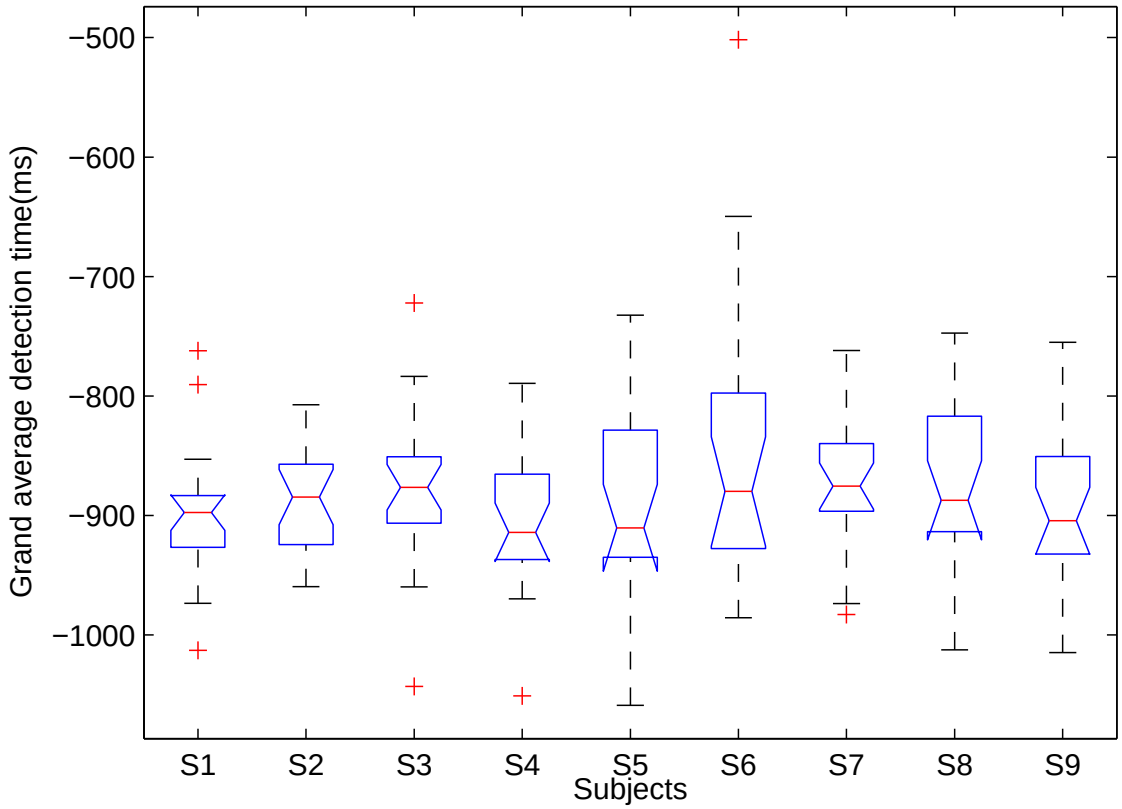


*Figure 6.23: The grand average detection time all the electrodes for each subject in the SCP (*0.5 Hz - 1.5 Hz*) band using the bootstrap method.*

6.6 Information transfer rate

In the present section, the movement initiation detection methods are evaluated by calculating the ITR which is a measure of the amount of information transferred per unit of time (minute). During the present analysis, the true positive rate, which is defined as the number of trials which are detected as movement during the movement intention period over the total number of trial per subject, is assumed to represent the accuracy. The rational behind this assumption is that there is always a symbol generated during the experimental paradigm and there is no idle trial during the protocol. Tables 6.3, 6.2 and 6.1 show the ITRs for the WMM method in the SCP, Mu and Beta frequency

bands, respectively. The results show a maximum ITR of approximately 4 Bits per minute in the Mu and SCP frequency ranges and a maximum ITR of 3 Bits per second in the Beta frequency range.

Tables 6.5 and 6.4 show the ITRs for the bootstrap and Mahalanobis detection methods, respectively. The tables show that the bootstral and the MD methods have a maximum ITR of approximately 3 Bits per minute. In the present study, there are three mental tasks (N=3) and the time to transfer a single symbol is about $10\,\mathrm{s}$, resulting in a generation of six trials per minute. Under these conditions the theoretical ITR (Bits per minute when six trials are generated per minute) is approximately 9 bits per minute as shown in Figure 6.24.

Table 6.1: The ITR (Bits/Min) in the Beta band for all the subjects using the WMM method when six trials are generated per minute. The highlighted values in each column show the maximum ITR values for each subject.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
C5	0.87	1.37	0.64	1.25	1.23	1.92	2.37	0.98	0.65
C3	0.41	1.53	1.53	1.60	0.51	0.72	1.53	2.15	2.66
C1	0.87	1.53	2.00	0.81	0.34	2.75	0.75	1.71	0.80
Cz	1.01	1.53	0.79	2.22	1.53	0.88	0.11	0.69	0.65
C2	0.62	1.37	0.79	1.25	1.23	1.68	0.88	0.83	1.37
C4	1.32	0.51	1.53	0.81	0.71	0.58	1.18	1.14	0.19
C6	1.86	0.95	0.51	0.81	0.71	2.45	1.53	0.69	0.97
FC5	1.86	0.83	0.64	2.00	1.53	0.88	0.07	1.71	0.80
FC3	2.07	0.71	1.76	1.42	1.53	0.88	1.73	1.93	0.97
FC1	2.29	1.23	0.39	1.25	1.88	1.92	1.03	3.22	1.59
FCz	1.67	0.95	0.79	0.68	0.95	1.92	0.62	1.14	0.65
FC2	1.49	0.10	0.64	0.56	1.23	1.45	1.93	1.14	1.16
FC4	1.16	0.95	2.26	1.80	2.26	0.88	0.75	0.45	1.59
FC6	0.32	0.71	1.53	0.81	0.71	2.17	1.93	0.98	0.51
CP5	1.49	0.34	0.79	0.46	1.53	2.17	2.15	0.98	0.51
CP3	2.29	0.95	0.79	1.60	1.53	1.68	0.62	0.57	0.39
CP1	3.28	1.37	0.79	0.94	1.88	0.72	1.18	0.98	1.37
CPz	1.49	0.71	1.13	0.81	0.20	0.58	1.03	2.15	0.80
CP2	2.51	1.23	0.51	0.81	0.34	0.45	0.62	1.51	0.97
CP4	0.51	0.51	2.00	0.28	0.71	2.75	0.32	1.14	1.37
CP6	0.51	0.42	0.64	0.36	0.20	0.05	1.18	1.32	0.39

Table 6.2: The ITR (Bits/Min) in the Mu band for all the subjects using the WMM method when six trials are generated per minute. The highlighted values in each column show the maximum ITR values for each subject.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
C5	0.87	1.08	1.13	0.94	0.95	1.05	2.37	1.51	1.37
C3	2.07	3.14	1.53	0.46	0.95	0.16	2.15	0.83	2.66
C1	1.01	1.70	3.83	1.80	2.68	0.88	1.03	1.32	0.65
Cz	2.07	1.37	0.79	1.09	1.53	2.75	1.35	1.51	0.51
C2	1.86	1.08	2.26	0.81	0.51	1.24	2.15	3.52	0.39
C4	2.51	0.71	1.76	0.46	2.68	0.72	1.03	1.93	1.37
C6	1.86	1.70	3.14	1.80	0.51	1.05	1.53	1.93	1.83
FC5	1.49	0.27	0.95	0.94	0.51	0.45	0.75	1.14	0.97
FC3	1.01	1.53	2.00	1.60	2.26	0.45	1.93	0.69	1.16
FC1	0.41	1.08	0.95	1.60	1.53	1.92	1.73	1.71	1.83
FCz	1.67	1.23	0.95	1.25	1.53	0.34	1.35	1.14	0.80
FC2	0.74	0.95	2.26	1.42	1.23	0.34	1.73	0.69	2.36
FC4	3.28	1.88	1.32	2.00	1.88	0.58	1.18	1.71	1.16
FC6	3.01	1.08	1.32	1.60	1.53	0.88	2.15	2.15	1.83
CP5	1.67	2.91	0.20	0.81	2.26	0.45	0.88	1.93	1.16
CP3	1.67	1.53	2.00	1.25	0.95	0.58	2.37	1.51	1.16
CP1	1.16	2.26	0.95	1.42	0.71	1.05	1.03	1.93	0.97
CPz	1.32	1.53	1.13	1.25	0.71	1.92	1.93	2.93	0.97
CP2	1.86	2.46	1.53	1.09	0.03	1.05	0.75	2.15	1.37
CP4	0.62	1.37	1.32	1.60	0.95	0.88	1.73	1.71	0.97
CP6	1.49	1.70	1.76	2.00	1.23	1.24	1.35	2.40	1.37

Table 6.3: The ITR (Bits/Min) in the SCP band for all the subjects using the WMM method when six trials are generated per minute. The highlighted values in each column show the maximum ITR values for each subject.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
C5	1.86	2.26	2.00	0.94	2.26	1.68	1.35	1.14	2.36
C3	1.32	0.61	0.13	0.21	2.26	1.24	2.62	2.15	2.09
C1	0.32	0.71	0.39	0.28	0.95	1.24	1.03	1.32	0.19
Cz	0.41	1.08	0.64	1.09	0.71	0.58	1.35	0.57	0.97
C2	1.01	0.34	0.79	0.81	0.34	1.45	2.87	0.19	0.65
C4	1.49	0.51	1.13	1.80	0.95	1.05	1.18	1.71	0.19
C6	1.16	1.23	2.83	0.68	2.26	1.68	1.35	1.71	0.80
FC5	3.86	0.51	1.76	0.68	1.23	1.45	1.93	1.71	1.37
FC3	1.49	0.61	0.51	0.94	3.65	2.45	1.35	0.27	0.39
FC1	0.32	0.71	1.13	0.94	1.53	1.92	1.93	0.19	0.28
FCz	0.04	0.51	2.00	0.56	1.23	1.68	1.35	1.32	0.65
FC2	0.08	0.51	1.76	0.56	0.51	1.68	0.51	0.69	0.51
FC4	0.41	0.95	1.76	0.94	0.71	1.92	0.51	1.14	0.51
FC6	1.16	1.23	1.76	0.36	0.71	3.06	1.93	1.93	0.97
CP5	1.16	1.37	0.95	1.60	0.95	4.13	1.18	0.98	0.80
CP3	1.49	1.37	0.51	0.46	0.51	0.34	0.75	1.32	0.80
CP1	1.49	0.83	0.13	0.14	1.53	1.68	0.62	0.13	1.59
CPz	1.32	0.27	0.79	0.81	1.23	0.88	1.03	0.19	0.19
CP2	1.67	0.34	0.95	0.28	0.10	0.58	1.18	1.14	0.06
CP4	1.16	0.06	1.53	0.46	1.23	1.24	1.35	1.14	0.19
CP6	1.32	0.34	0.64	1.09	1.53	0.88	1.35	0.98	0.80

Table 6.4: The ITR (Bits/Min) in the SCP band for all the subjects using the MD method when six trials are generated per minute. The highlighted values in each column show the maximum ITR values for each subject.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
C5	0.18	0.51	0.64	1.25	1.23	0.16	1.73	1.32	1.83
C3	0.51	0.42	1.76	0.28	0.71	0.72	0.75	0.35	0.12
C1	0.18	0.06	0.51	1.42	0.20	0.88	0.51	0.83	0.00
Cz	0.32	0.10	0.03	0.81	0.10	0.34	0.24	0.35	0.03
C2	0.08	0.42	0.20	0.81	0.95	0.88	0.07	1.51	0.65
C4	0.32	0.27	1.13	0.81	0.71	1.92	1.53	2.65	0.12
C6	1.32	0.34	1.76	1.09	0.51	0.88	1.03	2.40	0.06
FC5	1.32	2.06	2.53	1.25	0.71	0.00	0.88	1.71	1.83
FC3	1.49	0.71	2.00	0.68	1.23	0.24	1.35	0.45	0.97
FC1	0.08	0.06	0.39	1.42	0.03	2.17	0.51	0.45	0.00
FCz	0.18	0.00	0.00	0.46	0.51	0.58	0.07	0.45	0.00
FC2	1.16	0.00	0.01	0.21	1.88	2.45	0.51	0.98	0.39
FC4	0.51	0.34	0.29	0.81	1.53	1.24	0.75	0.45	0.80
FC6	0.62	0.95	2.00	0.68	1.23	1.45	2.62	3.52	0.65
CP5	0.41	1.08	1.76	0.68	0.95	1.92	1.73	2.15	1.16
CP3	0.25	0.15	0.95	0.94	0.10	1.68	2.15	0.98	0.51
CP1	0.25	0.10	0.79	0.56	0.00	1.05	0.32	0.27	0.03
CPz	0.08	0.34	0.20	0.68	0.20	0.88	0.17	0.45	0.06
CP2	0.32	0.03	0.20	0.68	0.03	1.05	0.32	0.13	0.06
CP4	1.32	0.27	0.20	1.60	0.95	1.92	1.18	2.65	0.51
CP6	1.49	0.27	1.13	0.94	1.53	3.06	1.93	2.65	0.65

Table 6.5: The ITR (Bits/Min) in the SCP band for all the subjects using the bootstrap method when six trials are generated per minute. The highlighted values in each column show the maximum ITR values for each subject.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**
C5	0.62	1.37	0.79	1.60	0.71	0.00	0.24	0.19	0.97
C3	0.62	2.06	0.95	0.81	0.34	0.00	0.88	0.45	0.65
C1	0.62	1.23	1.13	1.25	0.20	0.16	2.62	1.14	1.37
Cz	1.49	1.53	1.76	1.60	0.71	0.45	1.03	0.57	0.97
C2	1.16	0.61	0.95	0.21	0.20	0.05	1.53	0.19	1.37
C4	0.41	1.88	1.32	0.28	1.53	0.00	0.51	0.13	2.09
C6	0.08	1.23	1.13	1.60	0.95	0.00	0.00	0.00	1.83
FC5	0.51	1.23	0.51	2.94	1.88	0.00	0.32	0.04	1.16
FC3	0.74	1.37	0.51	0.56	0.20	0.00	1.18	0.35	0.65
FC1	1.32	0.27	0.79	0.94	1.23	0.05	0.75	0.69	1.83
FCz	0.51	1.70	0.79	0.94	0.34	0.24	1.03	0.19	1.37
FC2	1.67	0.34	0.95	0.14	0.51	0.02	0.32	0.19	1.59
FC4	0.32	0.71	0.95	0.68	0.71	0.00	0.17	0.83	0.80
FC6	0.00	0.83	0.00	1.09	0.95	0.00	0.00	0.27	1.59
CP5	0.25	0.61	0.00	2.69	0.51	0.00	0.07	0.01	2.36
CP3	0.62	0.42	0.39	1.09	0.51	0.00	0.88	0.83	0.97
CP1	0.25	0.83	0.39	1.42	0.71	0.09	1.53	0.35	0.51
CPz	0.25	0.34	0.79	1.60	0.71	0.34	1.73	1.32	0.65
CP2	0.74	0.34	1.13	1.09	0.95	0.34	1.93	0.35	0.65
CP4	3.28	0.61	1.53	0.81	1.53	0.02	0.17	0.35	2.09
CP6	1.32	0.71	0.64	2.45	3.14	0.05	0.41	0.00	2.09

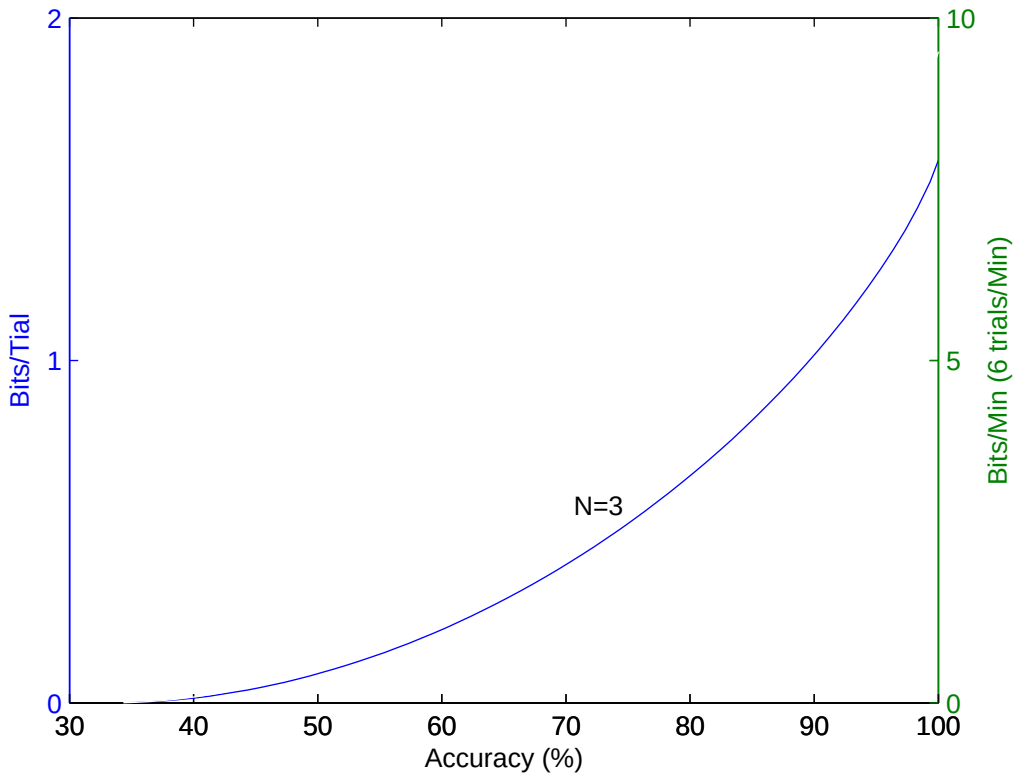


Figure 6.24: The ITR in Bits/trial and in Bits/min (for 6 trials/min) when the number of possible selections N is 3. The bit rate is shown only for accuracy greater than 33% *(i.e accuracy* $> 1/N$*)*

6.7 Running time of the different algorithms

The running time of the various algorithms in the present report were elaborated in section 5.11.3 and summarised in Table 5.3. Table 5.3 shows that the WMM, MD algorithms have a cubic ($O(n^3)$) running time and the bootstrap algorithm has a polynomial ($O(n^4)$) asymptotic running time.

In the present study, the performance of the algorithms were evaluated on an EEG data window from $-1.5\,\mathrm{s}$ to $1\,\mathrm{s}$ with respect to movement onset, resulting in 3000 data points as the data was sampled at $2\,\mathrm{kHz}$. Table 6.6 shows the asymptotic running times for quadratic, cubic and polynomial functions as a function of the problem size of 3000. It can be seen in the figure that for a problem size of 3000, the asymptotic running time has a order of magnitude of microseconds, milliseconds and seconds for quadratic, cubic and polynomial functions respectively.

Table 6.6: Examples of asymptotic running time for quadratic, cubic and polynomial functions for different input size. The time is computed by assuming that a basic instruction is executed in one nanosecond ($1\,\mathrm{ns}$*).*

	Asymptotic running times			
Function	$n = 500$	$n = 750$	$n = 1000$	$n = 3000$
n^2	$250\,\mu\mathrm{s}$	$562.5\,\mu\mathrm{s}$	$1\,\mathrm{ms}$	$9\,\mathrm{ms}$
n^3	$125\,\mathrm{ms}$	$421.9\,\mathrm{ms}$	$1\,\mathrm{s}$	$27\,\mathrm{s}$
n^4	$1.04\,\mathrm{min}$	$5.3\,\mathrm{min}$	$16.7\,\mathrm{min}$	$22.5\,\mathrm{min}$

Improving the asymptotic running time of an algorithm often tremendously reduces its performance much more than any smaller technical optimisation such as using a faster programming language.

The WMM, MD and the Bootstrap algorithms were executed in Matlab R2013a on a dual core Dell Computer with an Intel(R) Xeon(R) CPU E5440 @2.83GHz (2 processors). The computer runs Microsoft Widows 7 64-bit and has 16 GB memory. Table 6.7 shows the running time of the three methods on the desktop PC and the asymptotic running time when the programs are supposed to run on a machine where a basic instruction is executed in one nano second. The detection interval is $[^{-}1.5{,}0]$ seconds with respect to movement initiation, which results in the an input size of 3000 for a sampling rate of $2\,\mathrm{kHz}$. In a rehabilitation settings, a realistic detection time in the order of milliseconds is required: Table 6.6 shows that a quadratic algorithm is sufficient to solve such a problem.

As the detection time require to be in the magnitude of milliseconds to be useful in rehabilitation, table 6.6 shows that a quadratic algorithm is sufficient to solve such a problem.

Table 6.7: The figure shows the measured and the corresponding asymptotic running time using a cubic (WMM, MD) and a polynomial bootstrap) algorithms for a problem size of 3000. In the evaluation of the asymptotic running time, it is assumed that a single instruction is executed in $1\,\mathrm{ns}$ *while the measured running time is estimated on a dual core PC of 2.83GHz speed.*

	Running time($n = 3000$)		
	WMM	MD	Bootstrap
Measured	$4.47\,\mathrm{s}$	$9.23\,\mathrm{s}$	$20.2\,\mathrm{min}$
Asymptotic	$27\,\mathrm{s}$	$27\,\mathrm{s}$	$22.5\,\mathrm{min}$

6.8 Summary

In this chapter, the prediction of self-paced reaching hand movements was presented. It was shown that it is possible to predict self-paced reaching hand movements before the actual movement onset with a TPR greater than chance. An investigation of movement peak velocity revealed that individual participants approached the experimental protocol differently and that there was no consistent method employed by participants to execute movements. The prediction of reaching hand movements was performed in three frequency bands (SCP, Mu and Beta) using three different methods. When using the WMM method, it was possible to predict movement onset in the three frequency bands and it was shown that the highest detection rate was achieved in the Mu band. The movements prediction using the MD and bootstrap methods was only possible in the SCP band, and did not demonstrate consistent results at each electrode. The three methods showed that there was inter-subject variability in the prediction of the movements but the timing of the predicted movement onset was consistent. The methods exhibited an ITR of 3 to 4 bits/minute compared to a maximum theoretical ITR of 9 bits/minute when the experimental protocol consists of three mental tasks. The complexity analysis of the different algorithms used in the study (cubic and polynomial) showed that the time taken to detect movement initiation ranges from seconds to minutes in magnitude; making these algorithms unsuitable for real time application. The next chapter provides a discussion and an interpretation of the results, and compares the present results to previous studies.

Chapter 7

Discussion

The aim of this study was to explore methods of detecting self-paced centre-out reaching hand movement initiation in EEG signals on a single trial basis. Slow reaching hand movements in three directions (right, left and forward) were investigated for this purpose. Three different methods were used to detect reaching hand movement initiations and their timing in relation to movement onset. In this chapter, the findings concerning the prediction of reaching hand movements before the movement occurs and the timing of these movement initiations are highlighted and their implication in rehabilitation and BCI implementation are discussed.

7.1 Detection of self-paced reaching hand movement initiation

In this study, it was demonstrated that it is possible to predict self-paced centre-out reaching hand movement initiation in single trial EEG data approximately $1\,\mathrm{s}$ before movement onset. The predictions were made based on an activity time window spanning $-1500\,\mathrm{ms}$ to $-250\,\mathrm{ms}$ with respect to movement onset. The activity window was determined from early onset of RP (Haggard et al., 2002) and from the evidence that voluntary movement might be unconscious $250\,\mathrm{ms}$ before movement onset (Libet et al., 1983b). The detection time of about $1\,\mathrm{s}$ prior to movement onset is close to the early MRCP onset, suggesting that most of the subjects were merely planning their movements

(Libet et al., 1983b).

To the best of the knowledge of the author, the only study which has explored the detection of self-paced reaching hand movements before the onset of the actual movements was conducted by Lew et al. (2012). In their experimental paradigm, they considered a more accurate activity time interval from $^{-}750\,\mathrm{ms}$ to $^{-}250\,\mathrm{ms}$ with respect to movement onset and achieved an average detection time of about $0.5\,\mathrm{s}$ before movement onset.

Previous studies on self-paced movements have been conducted. However, the studies focused on movements with no reaching goal. Some of the movements were self-paced typewriting and key strokes (both left and right hand) on a computer keyboard (Blankertz et al., 2003; Bai et al., 2007) whilst other focused on wrist movements (Bai et al., 2011). The neuronal correlates of such movements are different from those of reaching movements as neuronal correlates associated with movements in general are modulated by the complexity, speed, sequence and precision of the movement (Lang, 2003). In their studies, (Bai et al., 2011) investigated wrist movements with no reaching goal in a time interval from $^{-}1.5\,\mathrm{ms}$ to $0\,\mathrm{ms}$ with respect to movement onset and achieved an average detection time of about $0.6\,\mathrm{s}$ before movement onset, in line with the results reported in Lew et al. (2012).

7.2 Methodology

EEG and movement data were recorded using Neuroscan and a Vicon motion tracking device, respectively. The two recording systems were synchronised using a switch that send a simultaneous DC pulse to the two recording systems. As an analog signal was used to synchronise the recording devices, the synchronisation method might be prone to precision error due to the rising time of the DC pulse. The synchronisation method was dictated by the lack of an appropriate method of synchronising the Vicon motion tracking system with any other recording device. This synchronisation error might impact the movement initiation marker positions. During the study, self-paced reaching tasks were considered and ten healthy subjects, were recruited. One subject was excluded

as the subject could not comply the the execution of the protocol due to the generation of unwanted movement artefacts. The present study proposed a protocol without any cue signal. Hence, the movements from the present protocol are natural and unconstrained reaching hand movements. Lew et al. (2012) used a similar protocol to detect self-paced reaching hand movements. The movements were centre-out movements in four directions (right, left, forward and backward) (Lew et al., 2012). The protocol by Lew et al. (2012) introduced an auditory cue in order to inform the subjects in which direction to move. During their protocol, the participants moved on average $5\,\mathrm{s}$ after the auditory cue. Their experimental protocol also defined an idle (no action) period immediately before the auditory cue. The experimental device used during their experiment was a $48\,\mathrm{cm}\times47.5\,\mathrm{cm}$ horizontal plane with micro switch buttons (dics) with a diameter of $27\,\mathrm{cm}$. Two experiments were conducted during the study in Lew et al. (2012). Eight healthy subjects (seven of them were right-handed) took part in their first experiment and four subjects (two healthy and two stroke patients) took part in their second experiment. During their first experiment, the subjects touched four target buttons in four directions (up, down, right and left) from the central position where the distance from the central position to the target position was approximately $20\,\mathrm{cm}$. In their second experiment, the subjects were instructed to move a cursor on a computer screen, from the central position to one of the four target positions using a manipulandum.

The difference between the experimental paradigm in the present study and the protocol by Lew et al. (2012) is the introduction of the auditory cue instructing the subjects in which direction to move. Though the participants moved when they wanted to after waiting for a minimum of $2\,\mathrm{s}$ following the audio cue, the auditory cue might have made the subjects think about the direction of the movement. Furthermore, the average mean time for the subjects moving after the cue was approximately $5\,\mathrm{s}$, and this time might not be sufficient to prevent the superposition of any eventual neuronal correlates due to the auditory cue with neuronal correlates due to movement execution. The instructions that consist of moving randomly toward a target and moving toward a specific target might be associated with different neuronal correlates as the later instruction contains extra information that adds to the complexity of the movement (Lang, 2003).

Another difference between the two experiments is that in part of the experiment carried out by Lew et al. (2012), the participants had to look at the target position to be able to place the computer cursor in the appropriate position and this not only induced some eye movement artifacts, but required more precision during movement execution. Another difference between the present study and the study performed by Lew et al. (2012) is the use of stroke patients. This could impact the outcome of the study in terms of the region of the motor cortex to be investigated as stroke might induce brain plasticity. Stroke patients whose brains have undergone brain plasticity must have brain regions reacting differently compared to able body participants. Furthermore, the performance of stroke patients might be impacted by their movement speed and precision which are directly linked to the neuronal correlates of movement onset (Lang, 2003).

Another difference between the two experiments is the choice of the baseline period. In the present study, the baseline was chosen as a time interval from $-2\,\mathrm{s}$ to $-1.5\,\mathrm{s}$ with respect to movement onset while in the study by Lew et al. (2012), the baseline was selected as the immediate $500\,\mathrm{ms}$ before the auditory cue. The later baseline selection might result in a major rejection of trials as there is no consistency in the concentration level of the participants before and after the auditory cue.

Another major difference between the two studies is that in this study, the EEG signals were represented in the time-frequency domain using the CWT enabling the signals to be resolved in three frequency bands, namely, the SCP, Mu and Beta bands while the alternative study uses a IIR filter to isolate only the SCP frequency band. It is not clear whether the filter was a hardware or software filter. The advantage of working in the time-frequency domain is that there is no need for averaging while working in the time domain requires averaging multiple epochs to cancel out random noise in order to enhance the SCP features. The time-frequency representation is therefore more appropriate to predict movement initiation on a single trial basis.

7.3 Data analysis

Various computational methods were considered to predict the intention preceding the actual onset of self-paced hand movements from single trial EEG data. EEG data was recorded from 61 monopolar electrodes. Single trial EEG data was explored offline by following a preprocessing (visual inspection, spatial filtering, temporal filtering, features selection) and postprocessing (pattern classification) analysis routine.

7.3.1 Preprocessing

The data was visually inspected to remove each trial contaminated with muscular, eye and movement artefacts. The CAR method was used to enhance the spatial resolution of the EEG data following a suggestion from previous study stating that the method was more efficient when a large recording site is considered (Pizzagalli, 2007; Pivik et al., 1993). The PCA technique was used to reduce the dimensionality of the data by projecting the data to orthogonal directions that maximised the variation. The PCA method has been extensively used to reduce the dimensionality of EEG data in ERPs (Carretié et al., 2004; Dien et al., 2005) and in BCI applications (Vallabhaneni and He, 2004). PCA selects the eigenvectors corresponding to the higher eigenvalues to explain most of the variance of the data. However, there might be a possibility that eigenvectors corresponding to small eigenvalues might contain important information to classify the data pattern (Dien et al., 2005). The CWT was used as a temporal filter to represent the EEG signal in a multi-resolution representation in the time-frequency domain. During the temporal filtering, the CWT was performed on the full dataset where each session had an average duration of $2.5\,\mathrm{mins}$. Using such a long window allows the resolution of very low frequency components such as the SCP. To process the data online, short windows can be considered and this is suitable for high frequency components such as the Mu and Beta frequency bands. However, a minimum window of $2\,\mathrm{s}$ is needed to process data with a resolution of $0.5\,\mathrm{Hz}$ which is the lower bound of the SCP in the present study. For instance, an online implementation of the algorithms in the present study in the SCP band would require a minimum

data window of $2\,\mathrm{s}$ for both the baseline and activity data, resulting in considerable data sets to be processed online.

7.3.2 Processing

Three methods (MD, WMM and Boot strap) were used to classify the data. The WMM method used PCA to reduce the dimensionality of the data. The method compares the Gaussian distribution of the baseline and overlapping activity windows. This method was able to predict the initiation of reaching hand movements in the SCP, Mu and Beta frequency bands and yielded a maximum TPRs of 80% in the respective frequency bands.

The MD method used a pooled covariance matrix to compare the baseline population with a sample population in the activity window. The rational of comparing the baseline with a sample of activity was that a single observation in the activity window might not be representative of the underlying neuronal correlates of the activity. During the prediction of movement initiation, the pooled covariance matrix was evaluated every $500\,\mathrm{ms}$ in order to compare the baseline with a sample active population, making the algorithm slow. The rational of comparing the baseline with a sample population of the activity window was made on the assumption that the baseline and the activity EEG data were different. An alternative comparison might be to test the mean vector of the sample activity window with the baseline, resulting in the computation of only the covariance matrix of the baseline. If the calculation of the MD only relied on the covariance matrix of the baseline, the covariance matrix could be computed only once for the examination of an entire trial, increasing the speed of the algorithm. A method using such a covariance matrix was used to classify movements directions (left and right) using a MD classifier (Babiloni et al., 2001).

The bootstrap method used in the present study increases the complexity of the MD algorithm by re-evaluating the MD a certain number of times where there is a trade off between the number of repetitions of the bootstrap and the statistical properties of the bootstrap distribution. A very efficient evaluation of the MD can significantly improve the performance of the bootstrap method. The

bootstrap method can also be improved by reducing the number of repetition during the evaluation of the bootstrap distribution and using a confidence interval that does not require the data to be normally distributed.

The results in the present study show that neuronal correlates to predict the onset of reaching hand movements were modulated by the SCP as the three methods used to predict movements intention performed equally in the SCP frequency band. However, when considering the rate of change of power across the frequency in a given frequency band, such changes seemed to be also modulated by neuronal correlates in the Mu and Beta bands.

Movement intention has also been investigated in previous studies using signals other than scalp EEG and it was shown that when using Local Field Potential (LFP), a higher detection rate of up to 95% can be achieved (Loukas and Brown, 2004).

7.4 Information transfer rate

The information transfer rates were calculated for the three different classifiers for three selections ($N = 3$) where the number of selections represents the reaching targets. The studies by Lew et al. (2012) did not report an ITR in contrast to the present study. It was challenging to define an 'idle' state in the present study to evaluate the specificity of the classifiers and the TPR was defined as the accuracy in the calculation of the ITR. The intertrial latency was defined in the present study as the time duration between two consecutive trials ($10\,\mathrm{s}$). This duration was different from the intertrial duration from previous studies, considering the length of the processing window, centred around movement onset, as the intertrial duration (Valsan, 2007). The ITRs of 3 and 4 bits per minute as reported in the present study are close to half the highest theoretical ITR which is around 9 bits per minute. Even though encouraging, it is worth noting that the evaluation of the ITR is more relevant for real time applications as it indicates how accurate the detection is with respect to the frequency of the movements.

7.5 Evaluation of the algorithms

The running time of the algorithms can be optimised in various ways. The most intuitive way is to run the algorithm on a faster machine or implement the algorithm in a faster programming language such as low level machine language. This is a rather technical optimisation with the following limitations. The speed of a program depends in general on its implementation on the idea level rather than the architecture and speed of the CPU. A program which has dependent modules will not run four times faster on a quad core processor as the processor will wait for each individual module to terminate before executing the next dependent module. It is obvious that an inefficient algorithm running slowly in high level programming languages can relatively run faster when translated in low level languages. Despite such apparent improvement, the inherent attribute of the program remains unchanged (the algorithm is still a badly implemented algorithm) as the efficiency of an algorithm depends on its implementation rather than its speed.

Another method used to optimised the running time of an algorithm is to reduce the problem size. This is one of the most challenging tasks in a project as this method requires expert knowledge of the features of the data and is termed as feature selection. Feature selection can be a very useful technique to reduce the problem size. However, there is a trade off between the quality of the input and feature selection and on the other hand, there is simply a limit in the reduction of the size of the problem. The last optimisation of the running time to be mentioned in the present study is the analysis of the asymptotic behaviour of the algorithm also known as complexity analysis. Complexity analysis provides a means to predict the running time of an algorithm as a function of the problem size. Improving the asymptotic running time of an algorithm seems to be an ideal optimisation technique, as the method makes no abstraction of the programming language the algorithm is implemented in. However, the result of this optimisation is only guaranteed for very large problem sizes. For example, an asymptotically optimal algorithm will outperform a non-optimal algorithm given a large enough input. However, the asymptotically optimal algorithm might not necessarily outperform a non-optimal algorithm for a small problem size. Furthermore, the determination of

the asymptotic running time function can be a laborious task (Ribeiro and Guerreiro, 2009; Cormen, Leiserson, Rivest and Stein, 2001). From the discussion above, it is clear that there is no gold standard method to optimise the running time of a given algorithm, and all the optimisation methods should be carefully combined to achieve the expected result.

7.6 Contributions and interpretation of results

The discussions of the results from section 7.3.2, showed that self-paced reaching hand movement intentions can be predicted with a good accuracy using simple linear classification methods without training. This is in agreement with the research hypothesis which was to predict the initiation of self-paced reaching hand movement onset. In the next sections, the contributions of the present research to the prediction of self-paced reaching hand movement onset and the contribution of the study to the understanding of natural movements and their impact to develop real time BCI systems are presented.

7.6.1 Contributions to event detection

The accuracy of a classification method depends on what kind of signals were used for the classification. For example very different performances can be achieved when different mental strategies (motor imagery or actual movements) are used (Neuper et al., 2005; Pfurtscheller et al., 1996). Most BCIs intend to detect movements during movement execution whilst the present study focuses on movement detection during movement planning. According to previous studies, detection of movements during movement execution is much higher (up to 80%) than detection of movements during the planning phase (up to 50%) (Müller-Gerking et al., 2000; Blankertz et al., 2003). In some BCI strategies such as sustained motor imagery, learning the generation of specific EEG patterns can cause long term fatigue, putting a mental load on the participant that might not be appropriate for people with disability such as ALS who are potential target users of BCI systems (Birbaumer, 2006).

It might not be appropriate to directly compare general BCI studies based on various mental strategies (Pfurtscheller et al., 1996; Gerwin Schalk, 2008; Gert Pfurtscheller and Neuper, 2000) with the present study as the main goals are different. The present study focuses on movement detection during the planning of movement and proposes natural movements that do not impose an additional mental load on the participants. From the observations above, the results provided in the present dissertation can be regarded as unique by providing a protocol for the first time, to the best of the author's knowledge, totally free from any cue signal.

Previous studies detected movement intention by modelling a baseline period using training techniques requiring a huge training data set (Cincotti et al., 2002; Babiloni et al., 2001) and the modelled baseline is then compared to the activity data to detect movement intention. In the present study, the detection of movement intention is performed by directly comparing the baseline of each single trial to its corresponding activity data (see sections 5.7 and section 5.8). The later methods were used to detect movement intention in the SCP frequency band and yielded an average maximum detection rate of 70%, below the performance of previous studies which reported maximum average TPRs of respectively 81% (Lew et al., 2012) and 85% (Blankertz et al., 2003) exploring the same frequency band.

In section 5.6, the gradual change along the maxima of the moduli of the wavelet coefficients across the scales in a given frequency band was proposed as a method to detect movement intention. The method can detect movement onset with a maximum detection rate of 80% in the SCP band, in line with previous methods which detected movement intention with a maximum average TPR of 81% (Lew et al., 2012) and below the performance of another method with an average maximum detection rate of 85% (Blankertz et al., 2003). The method can also detect movement intention with a maximum average TPR of 80% in the Mu and Beta bands, below the detection rates of previous studies exploring ERD with average maximum detection rates of 85% (Wang et al., 2004) and 90% (Bai et al., 2011) respectively. To the best of the author's knowledge, the method using the rate of the change of the wavelet maxima is novel in the detection of movement intention in humans.

7.6.2 Self-paced reaching movements and implementation of real time BCI systems

In addition to the feasibility of predicting the initiation of self-paced reaching hand movement as mentioned above, general understanding of self-paced reaching hand movement and using such movements to develop real time BCI systems are of interest. Reaching movements might be slow, hence more difficult to generate brain patterns to detect reaching movement initiation compared to brisk movements. In the present study, CWT technique was used to represent the EEG signals in the time frequency domain. The lower frequency band of interest was the SCP frequency band with a lower bound of $0.5\,\mathrm{Hz}$ requiring a minimum window of $2\,\mathrm{s}$ to resolve such low frequency components. Using such a long processing window in the implementation of a BCI system will impact the response time of the BCI system and mainly will result in a considerable computational load which is unsuitable for real time BCI implementation. The window length could be shortened by using EEG data in the time domain as in Lew et al. (2012). However, time domain data are far more prone to noise compared to data in the time-frequency representation. The random noise in the time domain data is suppressed by averaging several repeats of the data. Averaging technique is not appropriate to predict movement intention in a single trial which is the aim of the present research.

Self-paced movements are not well understood as the outcome of a self-paced movement intention can be difficult to be interpreted due to the possibility that the subject can change his mind any time from his intention to move up to the time he is conscious to move (Libet et al., 1983b; Kane, 2011). Such a flexibility in self-paced movement is a severe limitation in real time BCI implementation and evidence was provided that such a flexibility in self-paced movements makes self-paced movements less decodable compared to cued movements (Salvaris and Haggard, 2014).

The lack of comprehension of self-paced movements is one of the main reason of introducing cue in BCI protocols. For instance an external cue could be used to ensure that the subject did not change his mind during the execution of a self-paced movement and this in turn can lead to a

non natural movement. The need of cues is also indispensable to develop real time BCI systems. Cues in BCI paradigms are often used in an attempt to detect the control tasks such as movement execution, imagination or movement intention (Koo et al., 2015; Salvaris and Haggard, 2014), to evaluate the sensitivity of the BCI system (Lew et al., 2012) or to merely monitor the mental state of the subject (Bai et al., 2011; Alonso-Valerdi et al., 2015). However, the introduction of such cues violates the self-paced nature of the control tasks. In addition, the evidence that the decoding accuracy of cue delay response tasks depends on the perception and the cognition of the cue in use (Alonso-Valerdi et al., 2015), casts a shadow on the plausibility of BCI studies including cues and claiming to be self-paced BCI studies such as the hybrid BCI (Alonso-Valerdi et al., 2015) and the studies decoding free choice hand tapping movements (Salvaris and Haggard, 2014).

The reaching tasks used in the present study are limited to single limb, limiting the topological distribution of the brain area involved in the neuronal correlates underlying the movements.

In the light of the complexity inherent to self-paced movements, it is tempting to state that despite the possibility of successfully detecting self-paced reaching hand movements, using such movements to implement real time BCI systems remain a considerable challenge.

7.6.3 Self-paced vs cued studies

In the present section, previous cue studies will be compared to the present study which is cue free (self-paced). Previous study (Salvaris and Haggard, 2014) showed evidence that during pre cue delay tasks; the decoding of free choice decisions is merely triggered by the cue. From this assumption, previous study (Lew et al., 2012) and (Bai et al., 2011) can be considered as cue paradigms. Previous study (Bai et al., 2011) used a visual cue to detect the control signal while previous study (Lew et al., 2012) used an audio cue to instruct the subjects in which direction to move. The present study, study (Lew et al., 2012) and study (Bai et al., 2011) use a prediction interval of $[-1500, -250]$, $[-750, -250]$ and $[-1500,0]$ milliseconds with respect to movement onset respectively. The cued studies exhibit a higher detection rate compared to the self-paced method

as shown in Table 7.1. Although the detection methods and the features extraction are different to make a like for like comparison of the methods, the major difference in the detection rates may be linked to previous findings that it is more difficult to decode neuronal correlates associated with self-paced movements than neuronal correlates associated with cued movements due to the low discriminative strength of self-paced neuronal correlates (Salvaris and Haggard, 2014; Fleming et al., 2009). Previous studies showed that free decisions such as self-paced movements exhibit low attention level compared to instructed decisions (cued movements) as free decisions are subject to change (Fleming et al., 2009; Kane, 2011; Filevich and Haggard, 2013).

	Strathclyde	Lew et al	Bai et al
TPR	0.7, 0.7 and 0.8	0.92	0.90
time	-1s	-0.5s	-0.62s
Window	-1.5s to -250ms	-750ms to -250ms	-1.5s to 0
Task	Reaching	Reaching	Wrist flexion
Cue	None	Audio	Visual

Table 7.1: Detection rates and times of movement prediction before movement onset of the present study and previous studies (Lew et al., 2012) and (Bai et al., 2011). The three TPRs of the Strathclyde methods correspond to the detection rates of the MD, Bootstrap and WMM methods respectively.

Chapter 8

Conclusion

The present study has explored the possibility of detecting self-paced reaching hand movements in single trial scalp EEG. Nine subjects were recruited to take part in the present study were the task was to execute self-paced centre-out reaching hand movements in three directions (left, right, and forward). An appropriate time interval of $[-1500, -250]$ milliseconds with respect to movement onset was selected to analyse the EEG data.

The protocol in the present study was designed to prevent a mental load on the participant, in contrast to BCI protocols using motor imagery or training the participant to generate specific brain patterns. The protocol was designed to generate voluntary self-paced movements. The movement is not triggered by any external cue. The lack of a cue means that the sensitivity of the detection methods can not be investigated as it is not possible to define idle trials.

The quality of the recorded data was assessed by recording EOG data to remove ocular artefacts. EEG data associated with high amplitude EOG was excluded.

The protocol provided subjects with the flexibility to avoid fatigue and minimise artefacts such as eye blinks, yawning and swallowing by allowing the subject to wait until they are ready for movement execution. Previous studies reported that habituation can lead to the attenuation of the neuronal correlate during voluntary movements, especially repeated arm movements, which result in less pronounced readiness potential preceding movements. To avoid habituation, the experiment

was divided into small sessions of up to three minutes. Furthermore, the time between consecutive sessions was dictated by the subject, reducing the possibility of automatic movement execution.
The EEG signals recorded were transformed in the time-frequency domain using the CWT in the SCP, Mu and Beta frequency bands.
Analysis of the movement kinematics emphasised that there is a significant difference in movement peak velocity between the subjects during the reaching phase of the movement. Simple linear classifiers comparable with computationally intense methods were used to predict the onset of reaching movements. The results demonstrated that self-paced movement initiation can be detected before the execution of the movement and that there is a statistically significant difference in performance between subjects. The results also demonstrated that it was possible to predict reaching hand movement approximately $1\,\mathrm{s}$ before movement onset and that there was no significant difference in the time prediction of movement between subjects.

8.1 Functional significance

The present study has demonstrated the possibility of predicting the intention of a human to move approximately $1\,\mathrm{s}$ before the execution of the actual movement. It raises enormous possibilities to implement effective control and assistive devices, design of advanced prosthetics and improvements to BCI capabilities. Predicting movement intention prior to movement onset could contribute to develop devices that can anticipate desired actions. Furthermore, the ability of predicting movement initiation before movement onset could contribute to prevent the execution of badly planned actions.

8.2 Limitation of the studies

The experimental design involved simultaneous recording of EEG and movement data and the synchronisation of both recordings. A significant amount of time was spent during the research

to develop a device to track arm kinematics in three dimensions but the effort was unsuccessful. Finally, a Vicon movement tracking system was used to record arm kinematics. During the experiment, a marker was placed on the hand of the subject and the trajectory of the marker was used as an estimation of the arm kinematics. The drawback of using such marker is that the skin of the subject can move during arm movement, and this can affect the motion of the marker, introducing a relative motion in the kinematics of the arm. However, it was assumed that the marker vibration was minimal as the movement were very smooth. Another limitation of the movement recording system was its lack of synchronisation with another recording device. To overcome this limitation, a mechanical switch was used to synchronise the movement tracking and EEG recording systems. The use of such a device could introduce inaccuracy in the determination of movement execution time. To compensate for this source of error, movement initiation time was corrected by subtracting the rising time of the mechanical switch from the evaluated movement initiation time. Moreover, the impact of such an error is minimised as the data was analysed two seconds prior to the movement initiation time.

Introducing an idle trial is a very challenging task in the attempt to design a true self paced protocol. Several previous studies attempted to introduce such an idle trial. However, as the idle trial is mostly associated with a cue, this makes the protocol less self-paced. Without the use of a cue, it was difficult to include an idle state which could improve the ITR analysis. Furthermore, the analysis of the ITR is more meaningful in online detection rather than off line detection.

The algorithms in the present dissertation use a dynamic baseline, meaning that during event detection, each activity window is referred to its own baseline which is the first $0.5\,\mathrm{s}$ of the trial. The rational of this type of baseline was to ensure a direct correlation between activity and baseline period for each trial as the subject could be in a different state of mind from trial to trial. Another reason to dynamically select the baseline was to avoid training. However, choosing to dynamically select the baseline makes the application of the algorithms for online applications very challenging. This also increase the complexity of the algorithms as the baseline is recalculated for each trial in contrast to machine learning methods where the baseline is calculated once during a training pro-

cess.

The weakness of the protocol is that such a long intertrial duration results in the generation of very few trials compared to other protocols. This lack of sufficient trials creates a trade-off between sufficient number of trials and the duration of the experiment. As trials need to be rejected if corrupted, there is a need to perform a longer experiment which in turn might affect the quality of recorded data due to subject fatigue. The protocol design does not provide an easy way to introduce an idle trial which could be used to assess the specificity of the event detection algorithms.

8.3 Future work

The feasibility of developing an algorithm that does not require training raises the prospect of developing BCI systems that can operate in real life communication and the development of advanced prosthetics. Such algorithms will be more generalised as they will not be subject-dependent such as retraining subjects due to subject fatigue. BCIs will then become out of the box systems and their users will not require BCI literacy. Protocol design will be democratised by removing the stringent constraint imposed by BCI literacy during protocol design leading to the development of real life communication BCI systems.

8.3.1 Automatic artefact removal

In the present study, data epochs were cleaned visually and corrupted data epoch were discarded resulting in a huge loss of data. Development of automatic artefacts rejection and a means to recover corrupted data using blind source separation is needed (Akhtar et al., 2012). The availability of such techniques will increase the prospect of online event detection and save substantial amounts of data.

8.3.2 Baseline selection

Most algorithms compare baseline with activity EEG data. Methods based on machine learning techniques model baseline using huge training data sets. This approach is suitable for online the algorithms as there is no need to calculate a new baseline for each activity epoch. The present thesis uses a dynamic baseline which is tailored to each individual activity data epoch. The use of such dynamic allocation of baseline might be more realistic as it expresses the comparison between a given activity EEG epoch to its ultimate corresponding activity free epoch. Such comparison is crucial for it can cope with both endogenous and exogenous temporal EEG dynamics. Although it is possible to perform off line event detection, techniques to develop dynamic baseline are needed to obtain a more realistic comparison between baseline and activity EEG epochs during online event detection.

8.3.3 Planned vs. unplanned movements

There is evidence that planned movements are associated with early RPs. Previous studies reported and average movement prediction time of about $0.5\,\mathrm{s}$ before movement onset which correspond to the late RP associate with spontaneous movements, hence corroborating the hypothesis of unplanned movement. The present study reports a movement prediction between $1\,\mathrm{s}$ and $800\,\mathrm{ms}$ and the prediction does not take into account, the latency of an event of $250\,\mathrm{ms}$. A realistic detection will include the latency of the event, leading approximately to a similar average detection time with previous studies. The results seem to support that most of self-paced movement are not planned and there is a need to further investigate this question.

8.3.4 Subjects' fatigue

Subject fatigue can be a serious issue during movement execution. This can impact the quality of the EEG data and subsequently the detection accuracy. This issue regarding long term stability needs to be investigated.

8.3.5 Best performing electrodes

The event detection methods show that there is no consistent electrode with a high detection rate. One way to address this problem is either to use a dense electrode grid or identify in advance the electrodes that perform best. A need to develop a method to identify the best set of electrodes in advance could be beneficial. The development of such method will not only reduce the problem size, then subsequently improve the performance of event detection methods, but will also improve their accuracy.

Appendices

Appendix A

Technical details of the synchronisation box

Figure A.1: Top view of the synchronization box

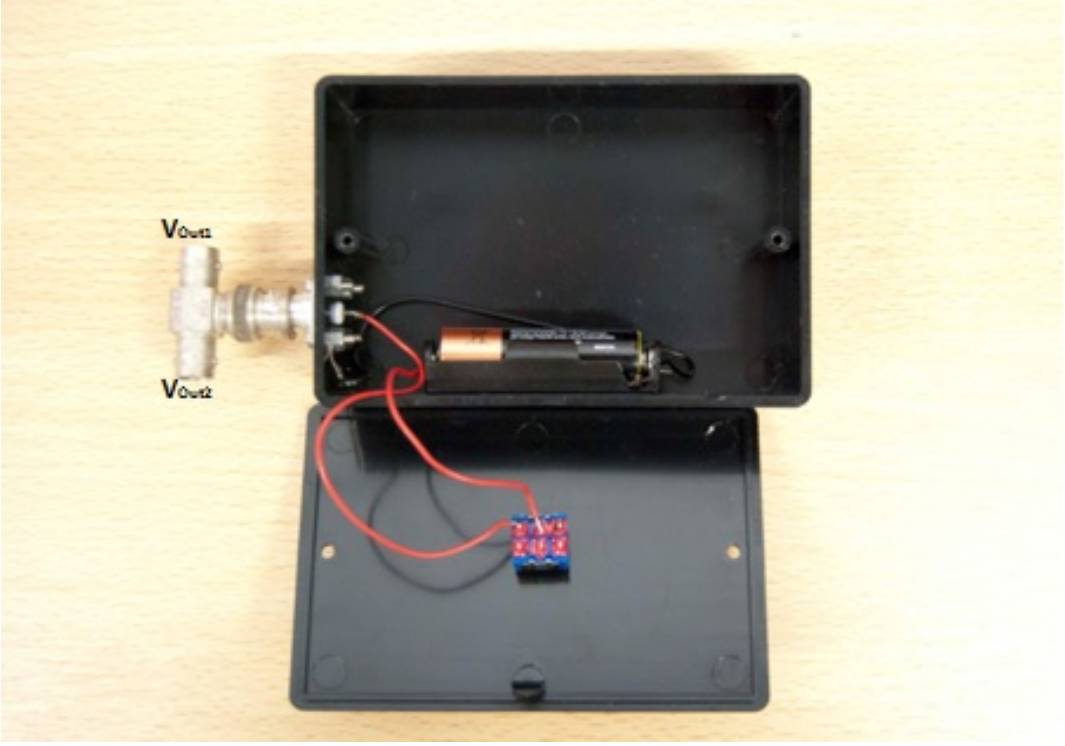


Figure A.2: Inside view of the synchronization box

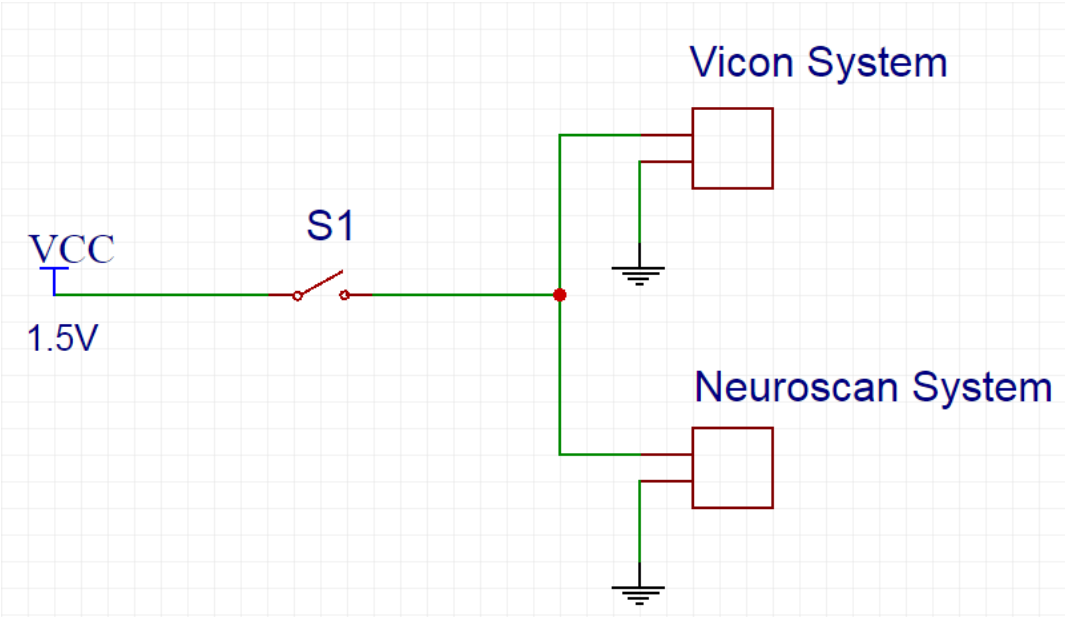


Figure A.3: Circuit diagram of the synchronization box

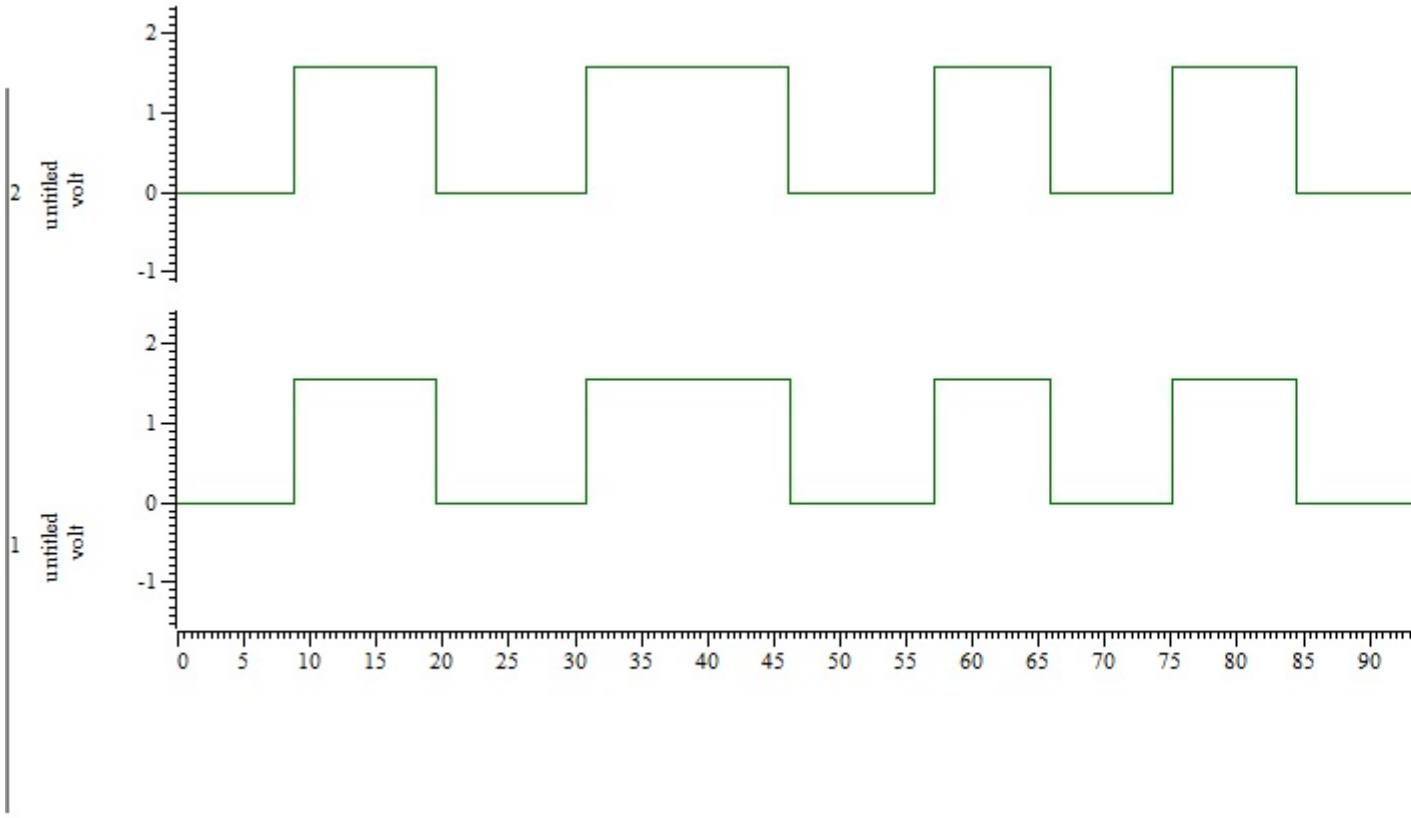


Figure A.4: Wave forms (Vout1 and Vout2) from the switch.

Table A.1: First five pulses of output channel 1 of the switch

Pulse 1	Pulse 2	Pulse 3	Pulse 4	Pulse 5
Amplitude (Volt)	Amplitude (Volt)	Amplitude (Volt)	Amplitude (Volt)	Amplitude (Volt)
0.0012	0.0000	0.0012	0.0000	0.0012
0.0012	0.0000	0.0012	0.0000	0.0012
0.0000	0.0000	-0.0000	0.0000	0.0012
0.0000	0.0000	0.0000	0.0000	0.0024
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
0.0012	0.0000	0.0012	0.0000	0.0024
0.7922	0.7910	0.7910	0.7910	0.7922
1.5820	1.5820	1.5820	1.5820	1.5820
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
1.5820	1.5820	1.5820	1.5833	1.5820
1.5820	1.5820	1.5820	1.5833	1.5820
1.5820	1.5820	1.5820	1.5820	1.5820

Table A.2: First five pulses of output channel 2 of the switch

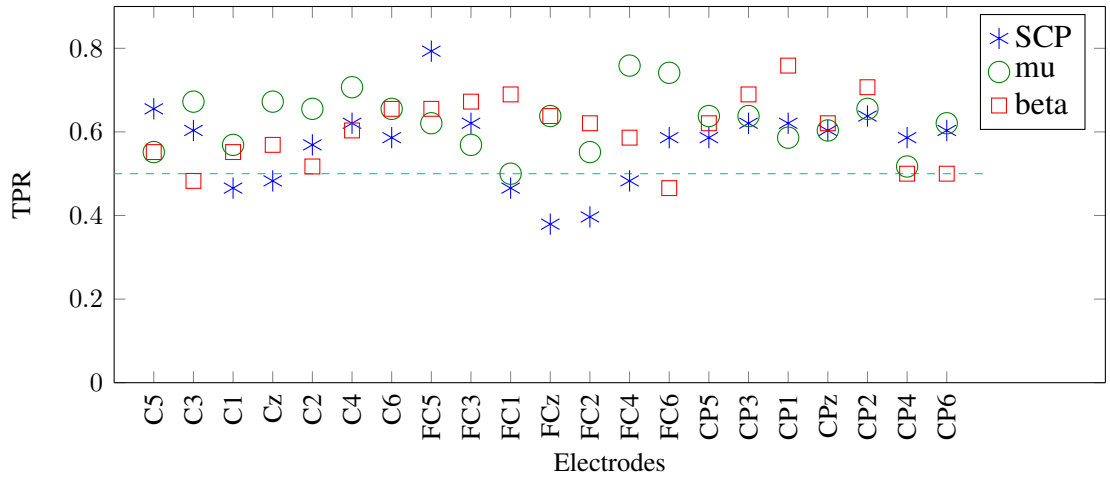
Pulse 1	Pulse 2	Pulse 3	Pulse 4	Pulse 5
Time (ms)	Time (ms)	Time (ms)	Time (ms)	Time (ms)
0.0012	0.0000	0.0012	0.0012	0.0012
0.0012	0.0000	0.0012	0.0000	0.0012
0.0012	0.0012	0.0012	0.0000	0.0012
0.0000	0.0012	0.0012	0.0000	0.0012
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
0.0000	0.0012	0.0012	0.0000	0.0012
0.7910	0.7922	0.7922	0.7129	0.7922
1.5820	1.5820	1.5820	1.5039	1.5820
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
1.5820	1.5820	1.5820	1.5820	1.5820
1.5820	1.5820	1.5820	1.5820	1.5833
1.5820	1.5820	1.5820	1.5820	1.5833

Tables A.2 and A.2 show the output voltages from the two channels of the synchronisation box when it is switched on. The voltage signals are sampled at a frequency of 100Hz. The dashed entries in the tables mark the time it takes to a single pulse to rise from zero to its peak value (1.5V) and this time interval is termed the rising time. The figures show that the raising time is 20 millisecond for the voltage to switch from zero to 1.5 Volt and this time is termed as the rising time of the switch. This time delay it takes the voltage from zero to its maximum (i.e 1.5 volt) is defined as the rising time of the switch. This time delay can be critical f the switch is used as an event marker.

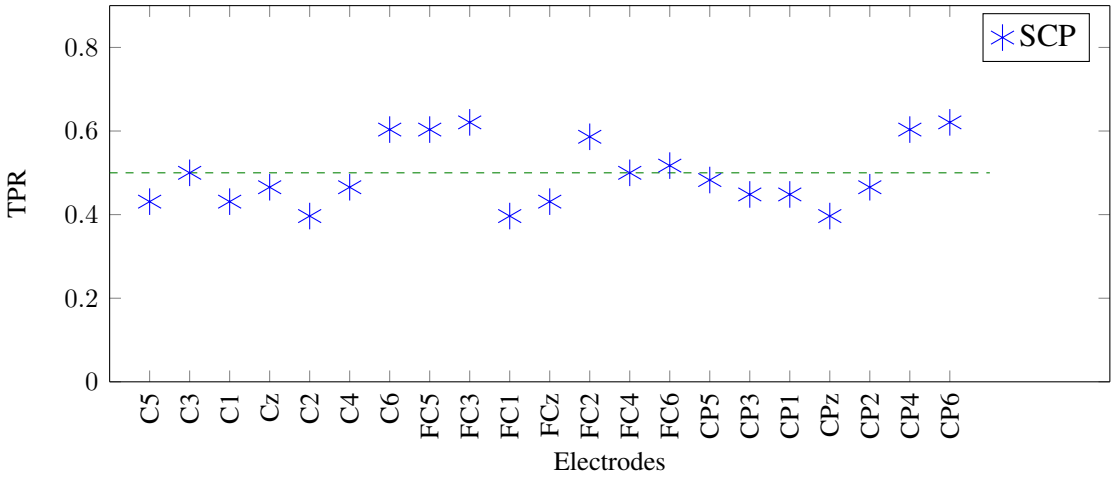
Appendix B

Average detection rate of each subject

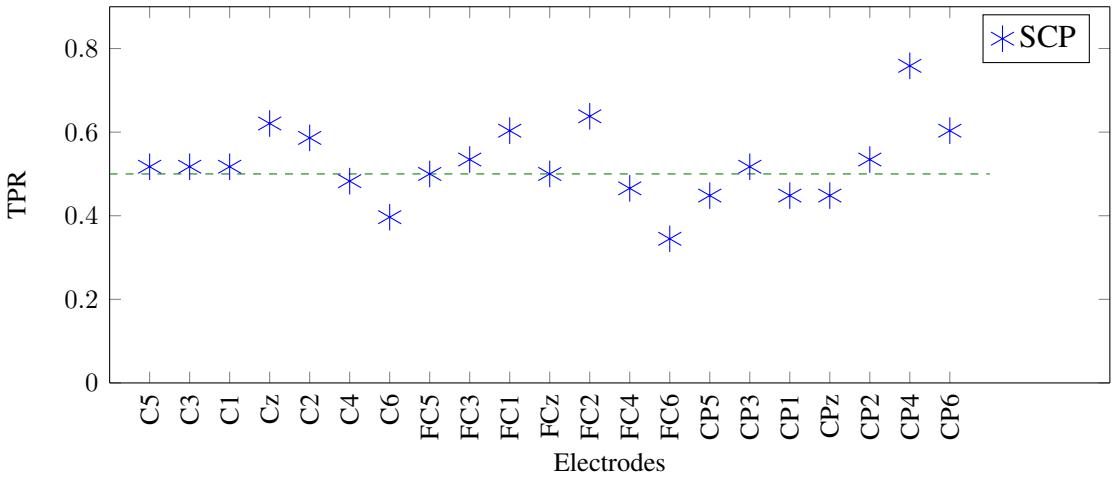
This appendix shows the average detection rate of individual electrode for each subject.



(a) Subject 1 (WMM)

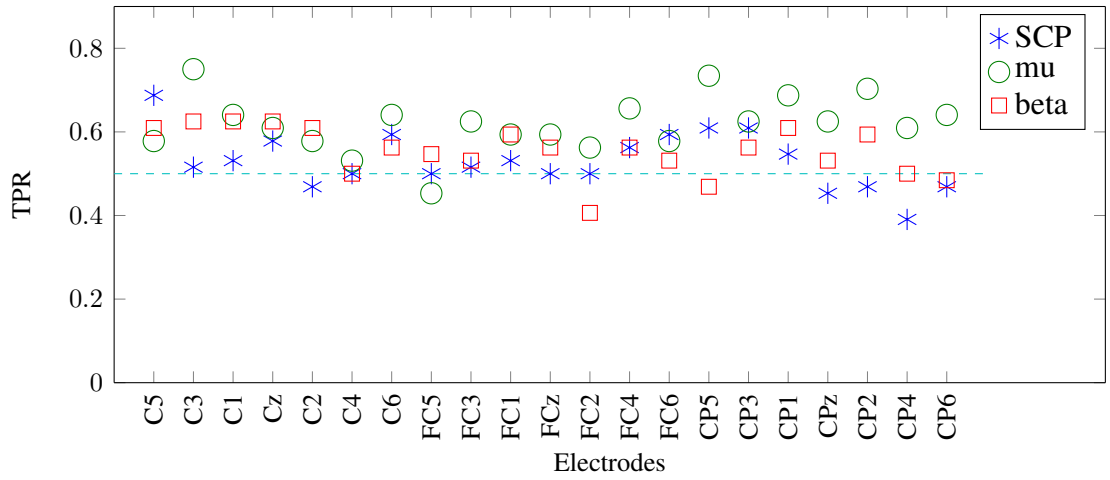


(b) Subject 1 (MAhalanolobis)

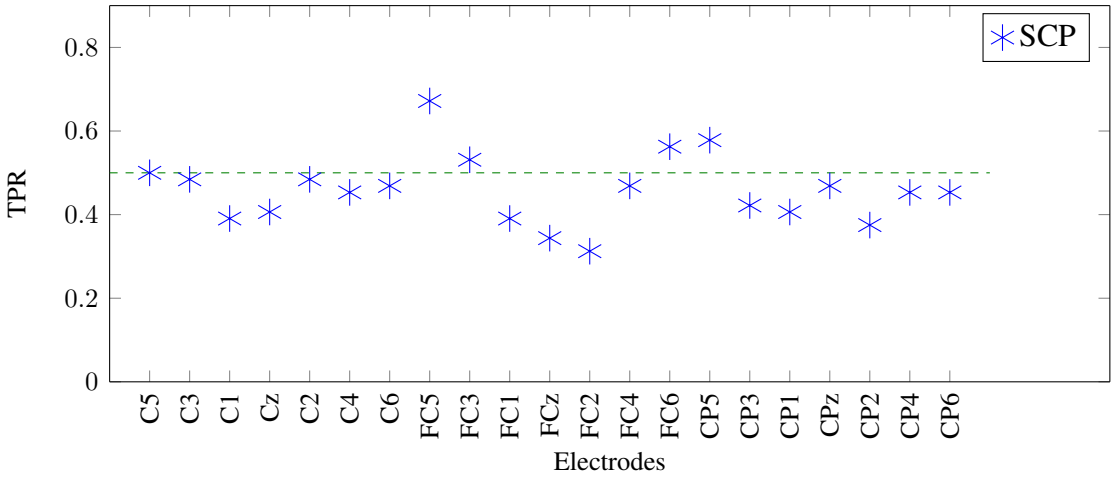


(c) Subject 1 (Bootstrap)

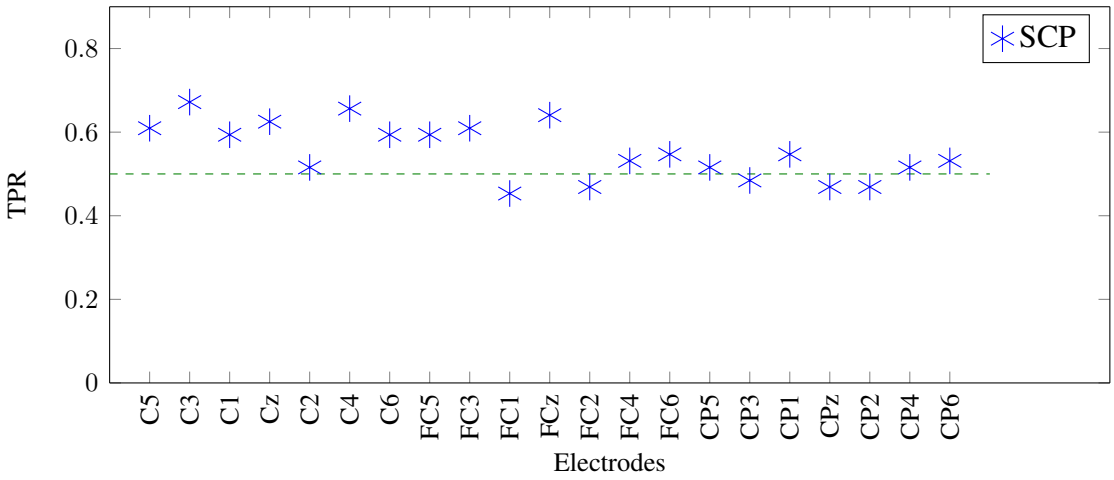
Figure B.1: The average detection rate for subject S1 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 2 (WMM)

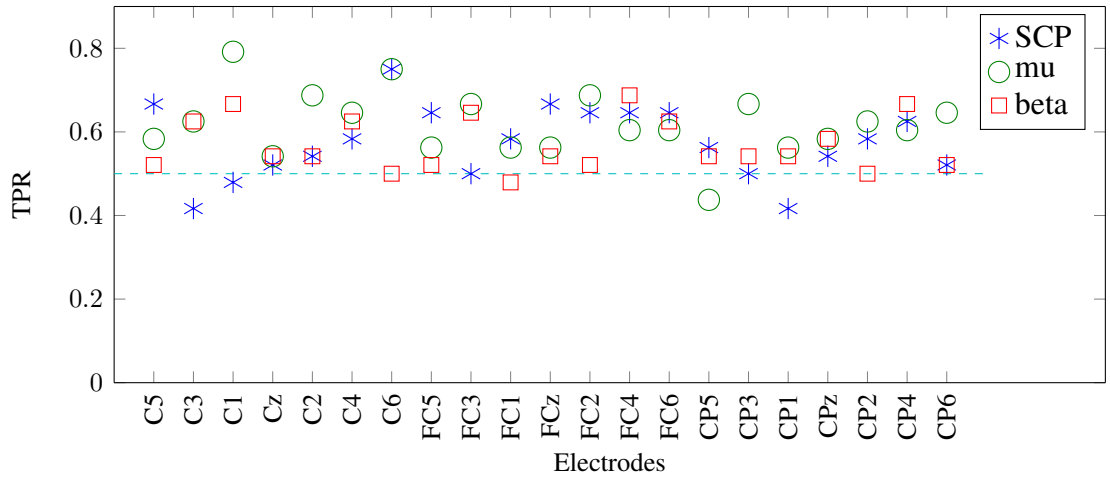


(b) Subject 2 (MAhalanolobis)

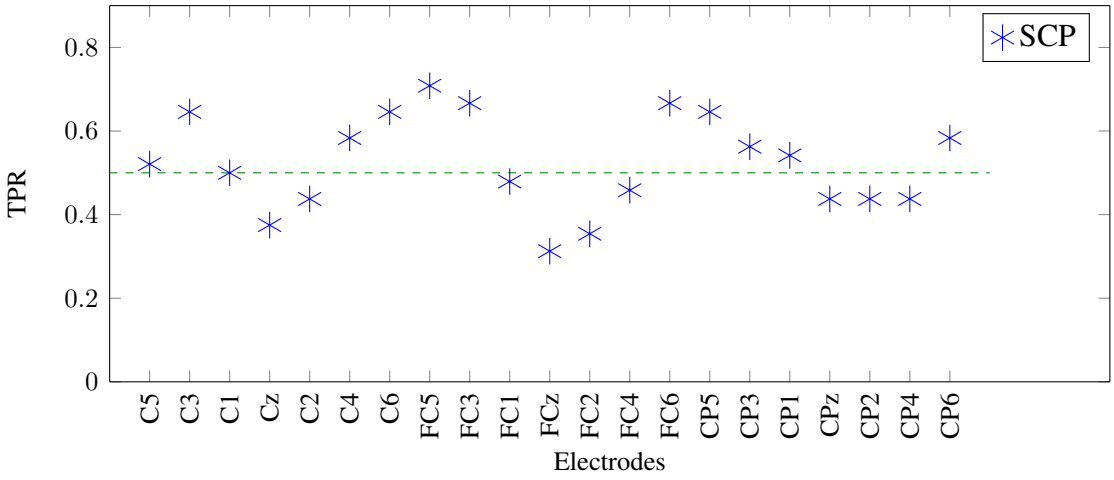


(c) Subject 2 (Bootstrap)

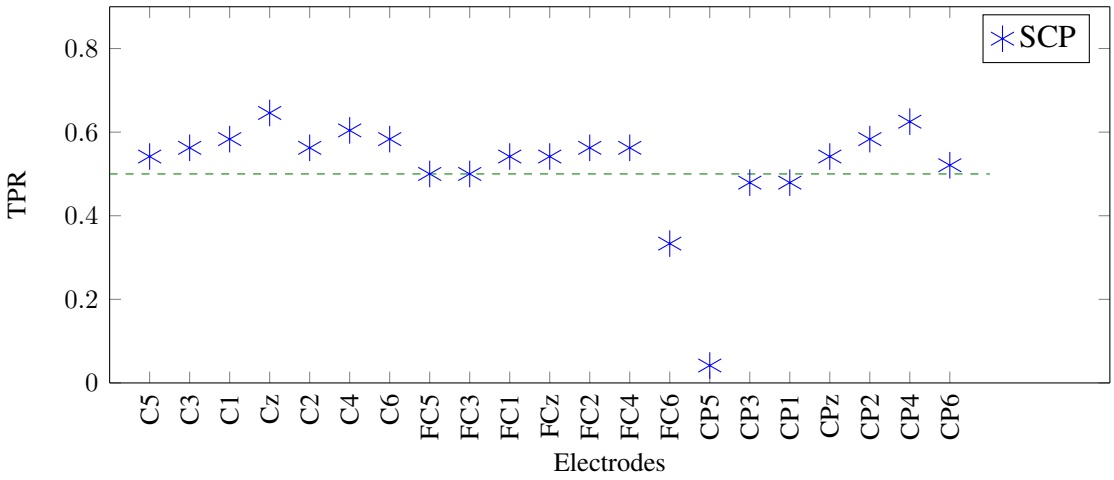
Figure B.2: The average detection rate for subject S2 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 3 (WMM)

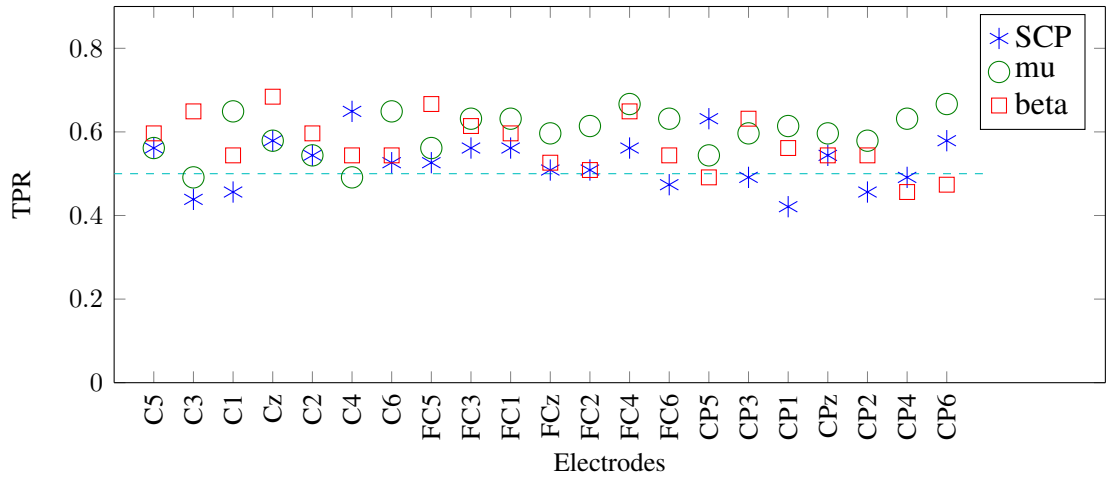


(b) Subject 3 (MAhalanolobis)

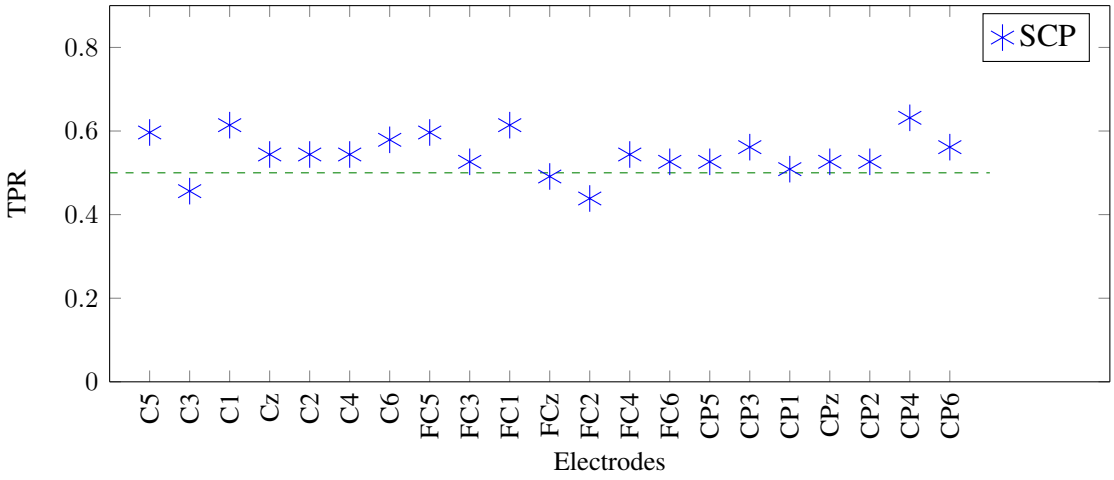


(c) Subject 3 (Bootstrap)

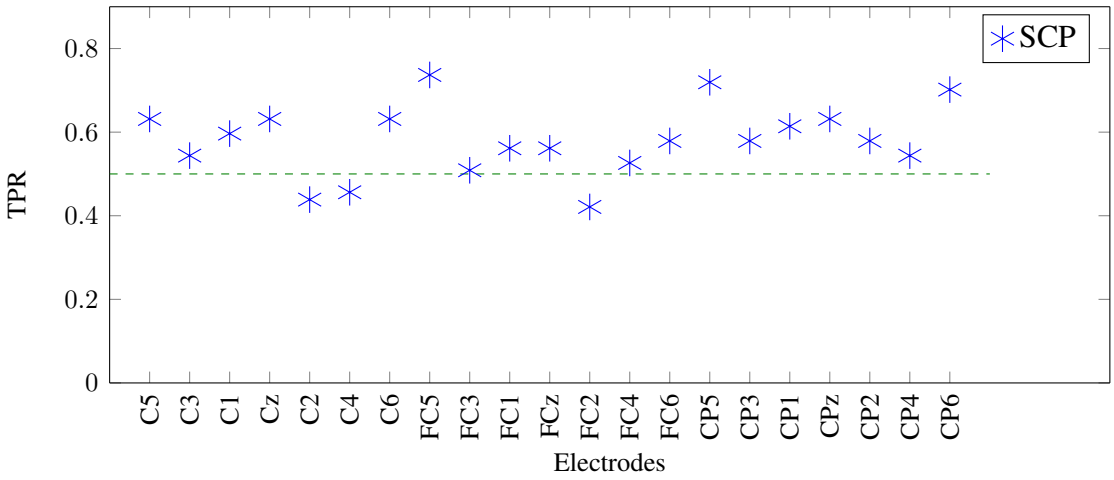
Figure B.3: The average detection rate for subject S3 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 4 (WMM)

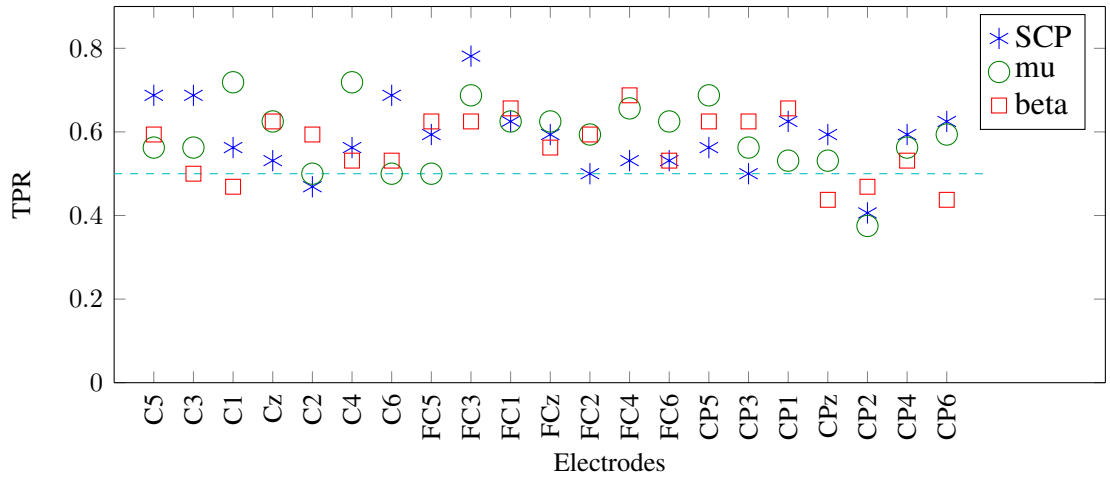


(b) Subject 4 (MAhalanolobis)

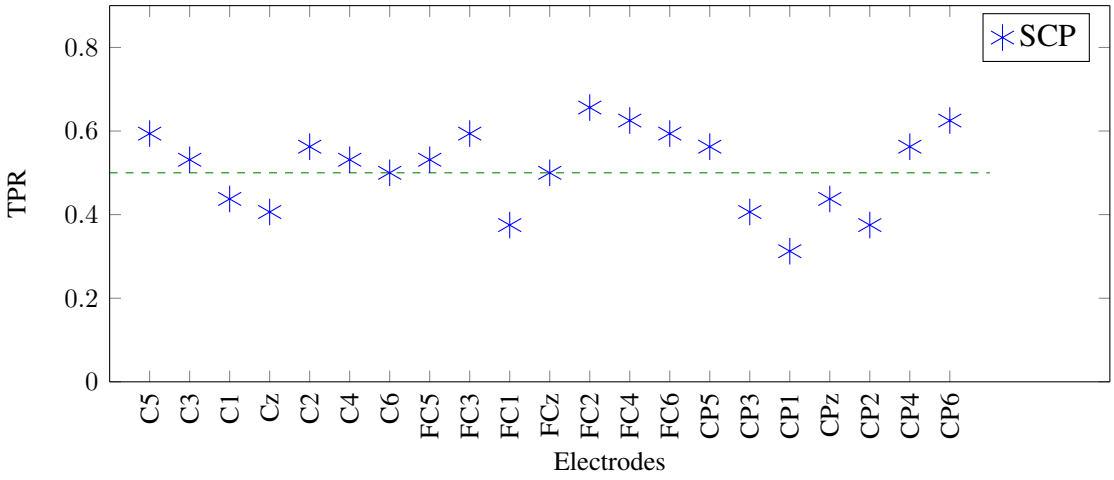


(c) Subject 4 (Bootstrap)

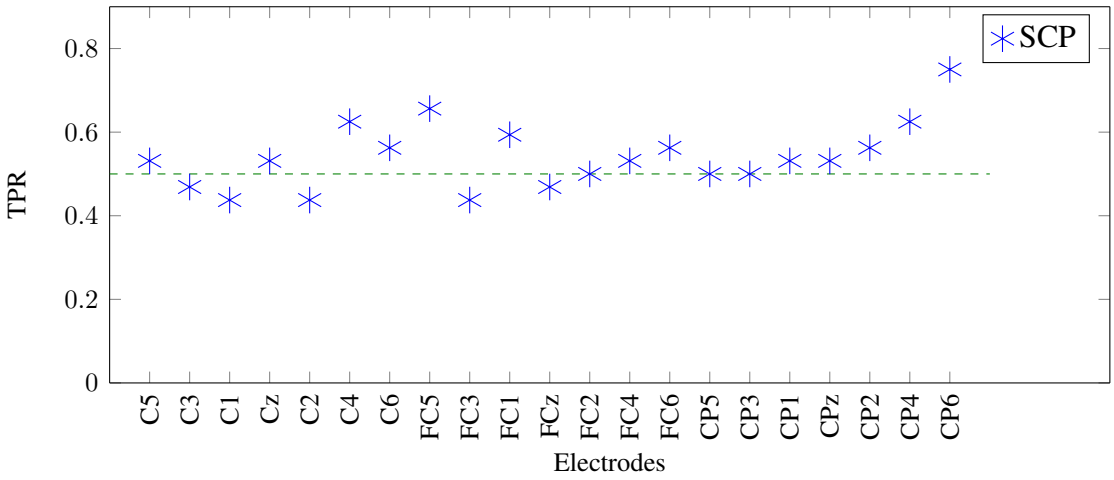
Figure B.4: The average detection rate for subject S4 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 5 (WMM)

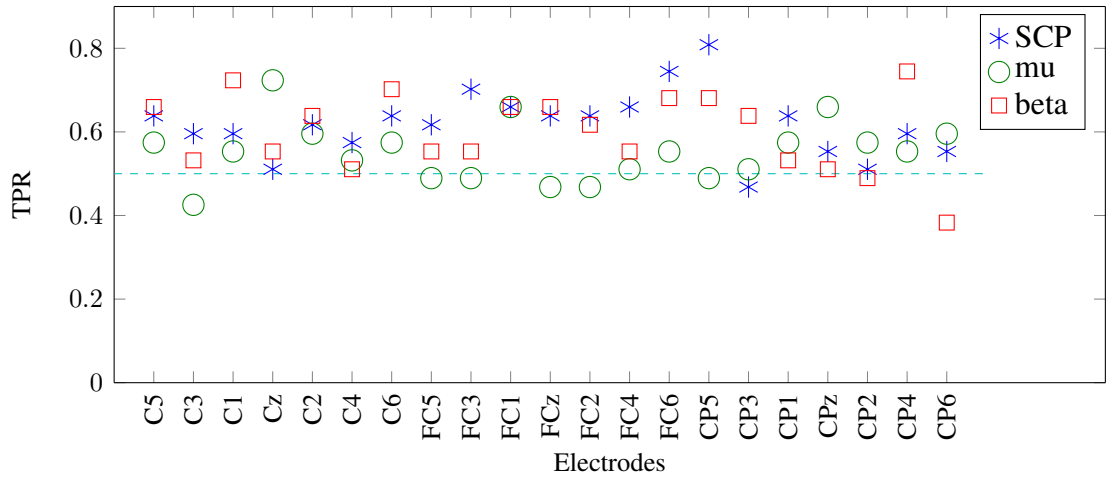


(b) Subject 5 (MAhalanolobis)

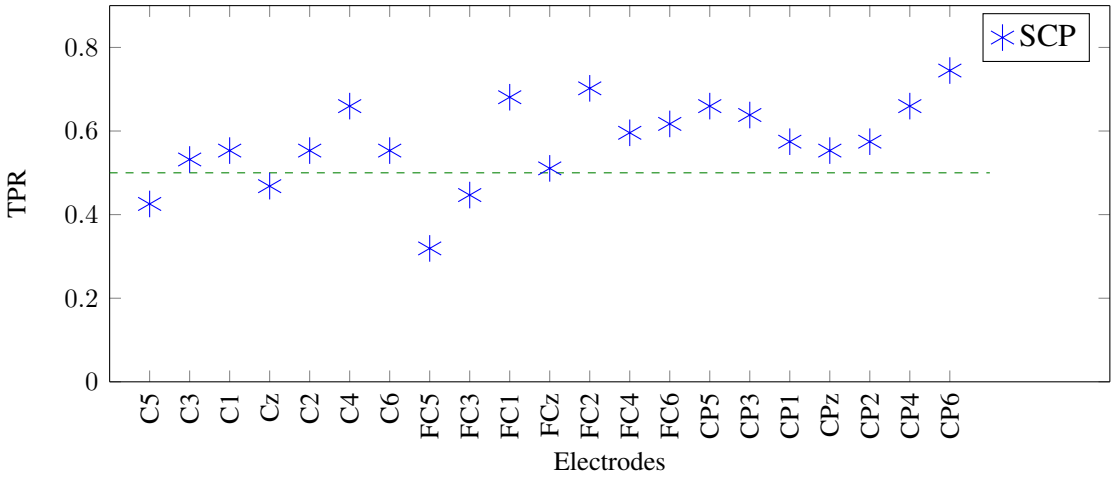


(c) Subject 5 (Bootstrap)

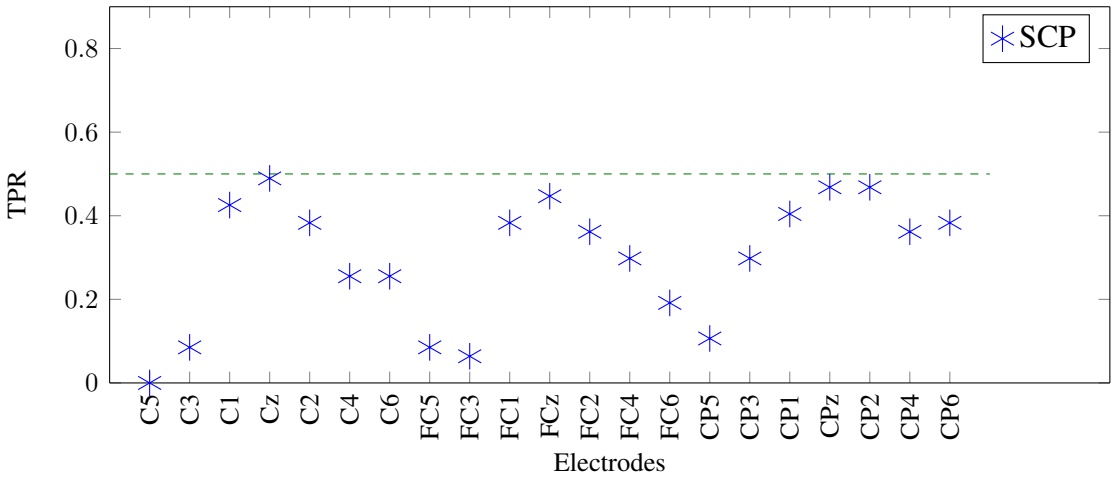
Figure B.5: The average detection rate for subject S5 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 6 (WMM)

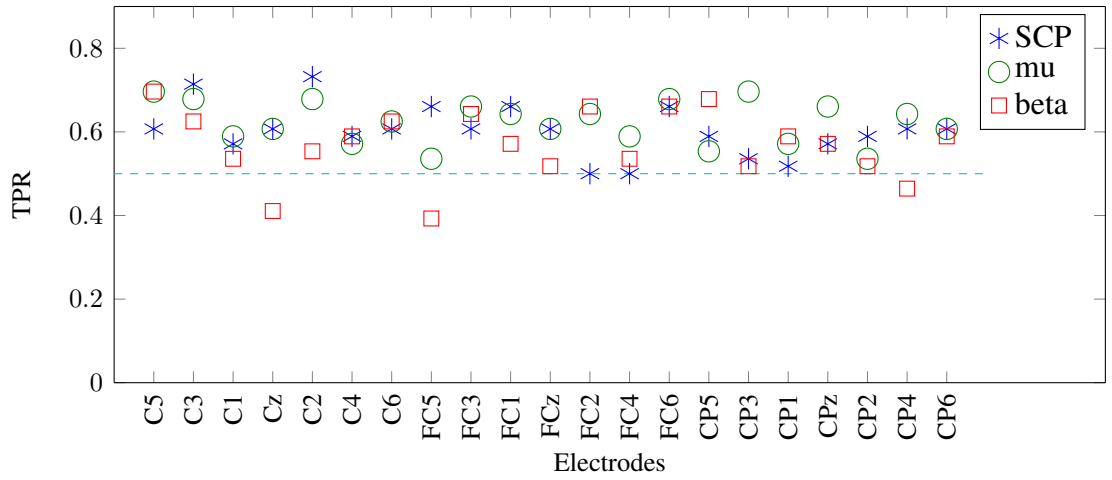


(b) Subject 6 (MAhalanolobis)

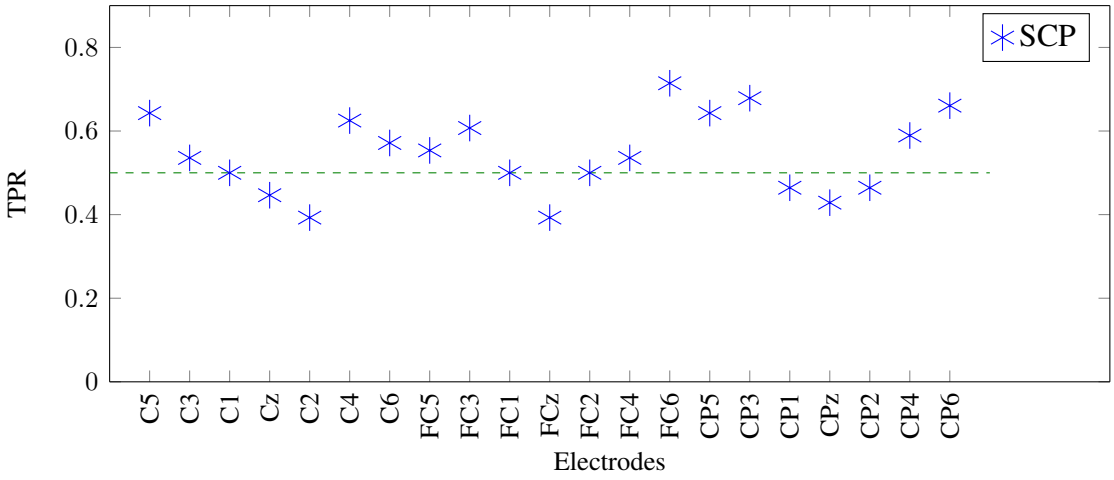


(c) Subject 6 (Bootstrap)

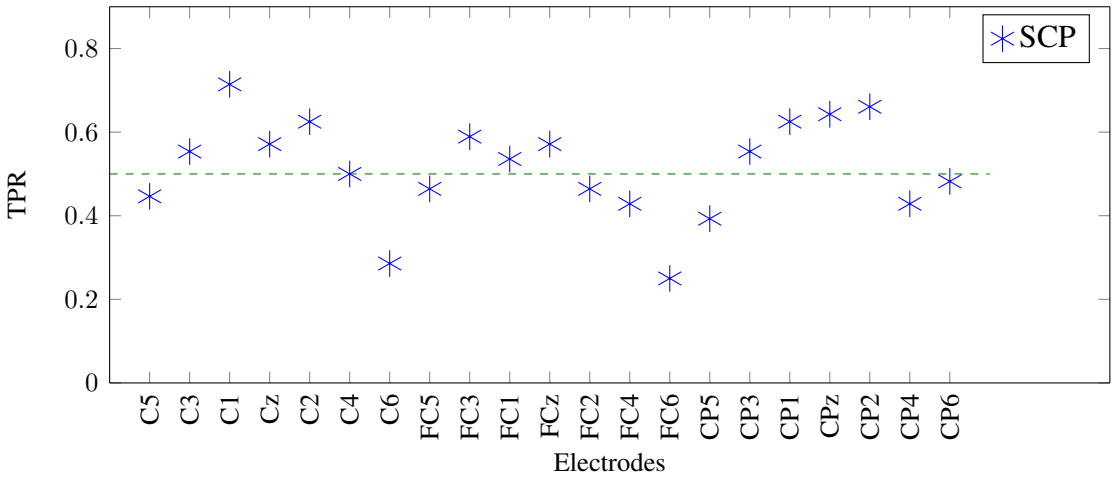
Figure B.6: The average detection rate for subject S6 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 7 (WMM)

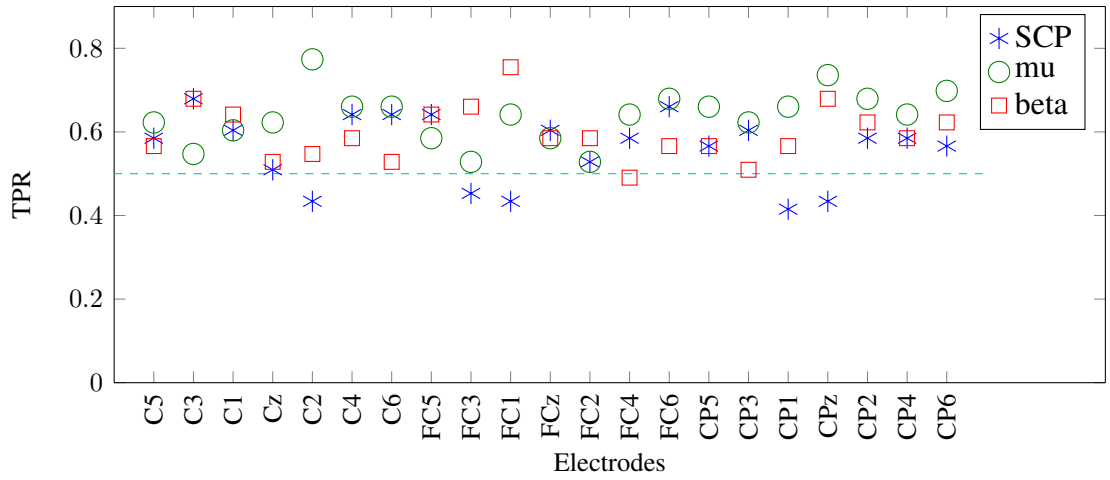


(b) Subject 7 (MAhalanolobis)

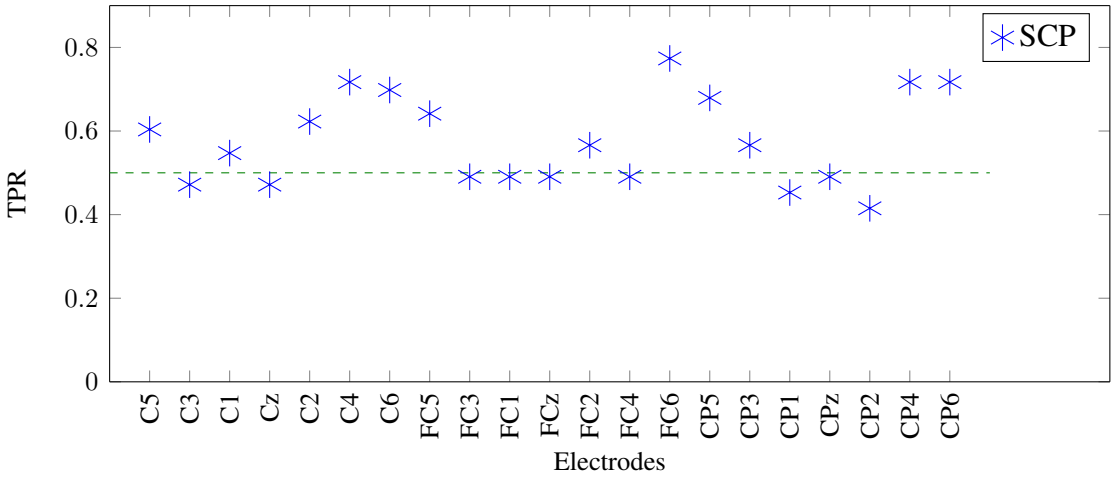


(c) Subject 7 (Bootstrap)

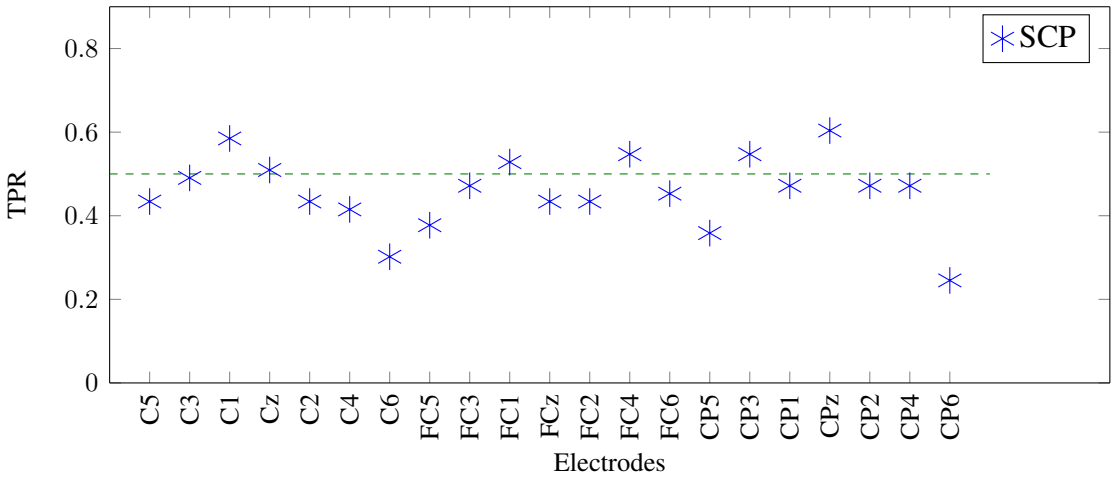
Figure B.7: The average detection rate for subject S7 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 8 (WMM)

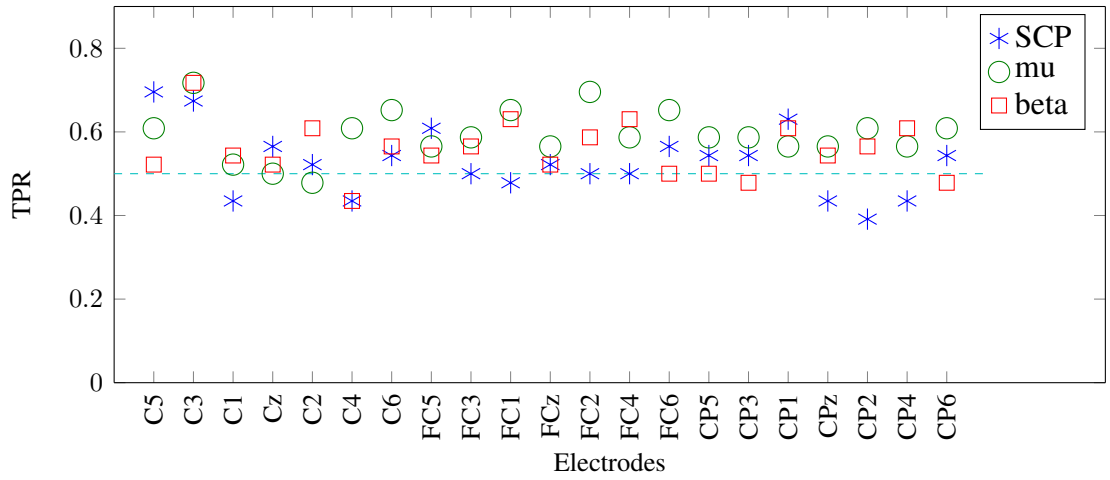


(b) Subject 8 (MAhalanolobis)

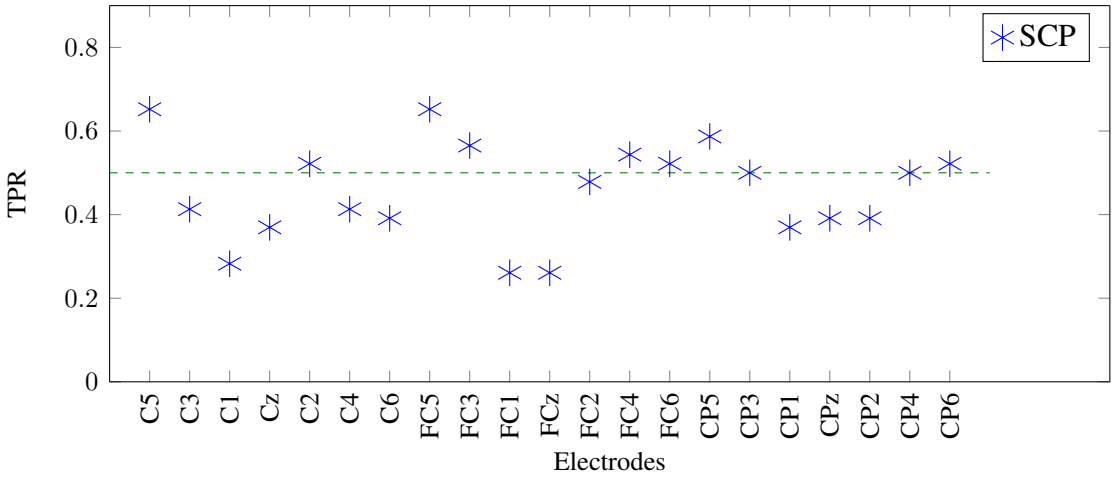


(c) Subject 8 (Bootstrap)

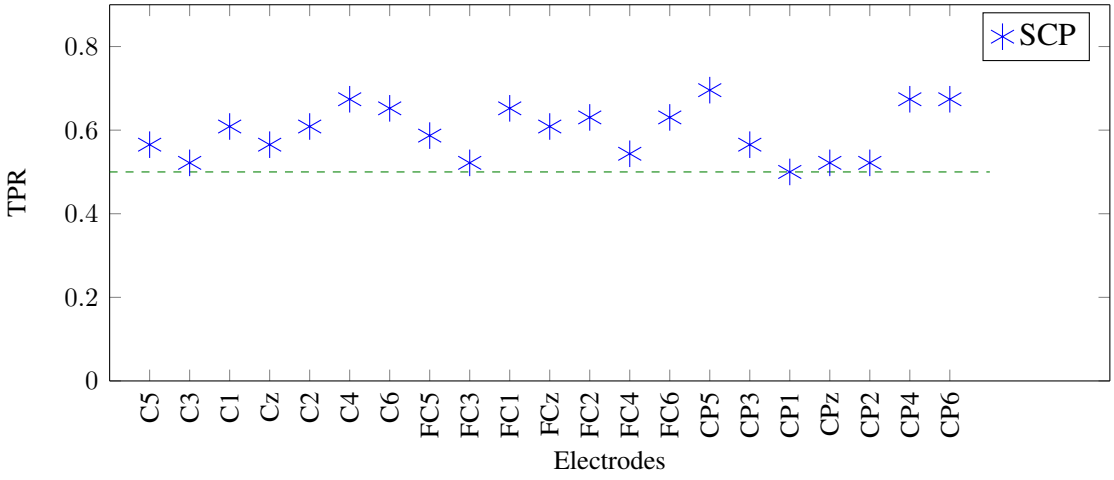
Figure B.8: The average detection rate for subject S8 using the WMM, Mahalanolobis and Bootstrap methods



(a) Subject 9 (WMM)



(b) Subject 9 (MAhalanolobis)



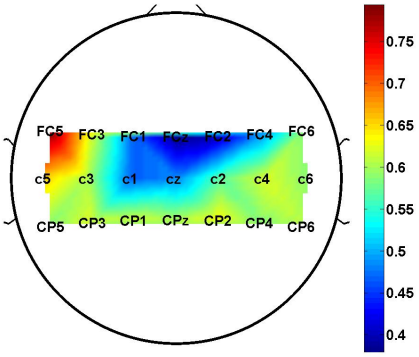
(c) Subject 9 (Bootstrap)

Figure B.9: The average detection rate for subject S9 using the WMM, Mahalanolobis and Bootstrap methods

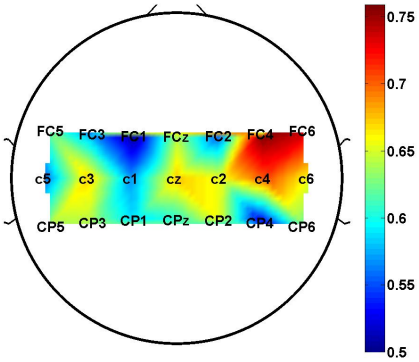
Appendix C

Topographic plots

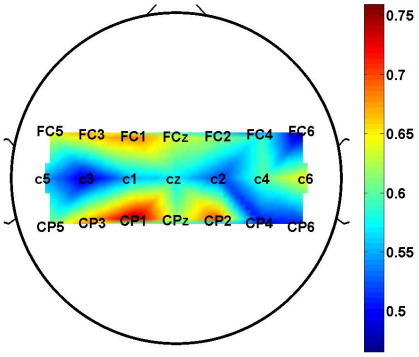
This appendix summarises the topographic plots for the WMM in the SCP, Mu and Beta frequency bands and the topographic plots for the MD and bootstrap methods in the SCP frequency band for each subject.



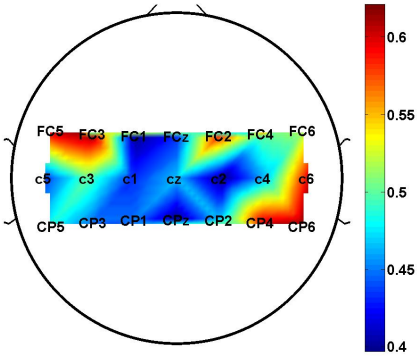
(a) Topographic plot fro S1 in the SCP band (WMM)



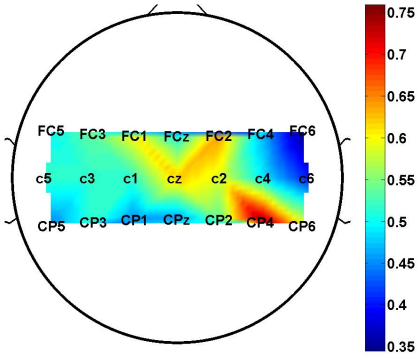
(b) Topographic plot for S1 in the Mu band (WMM)



(c) Topographic plot for S1 in the Beta band (WMM)

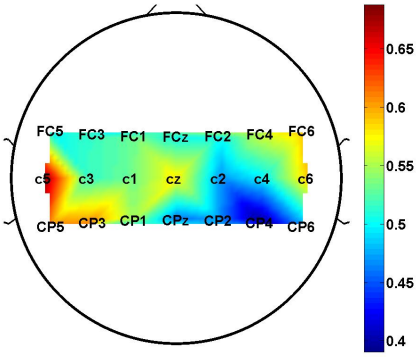


(d) Topographic plot for S1 in the SCP band (MD)

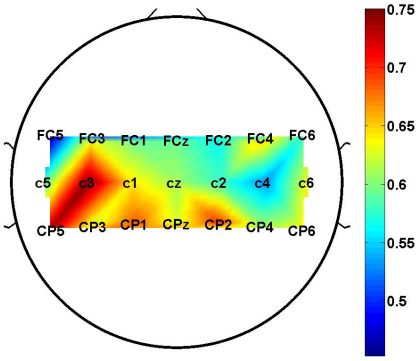


(e) Topographic plot for S1 in the SCP band (boot)

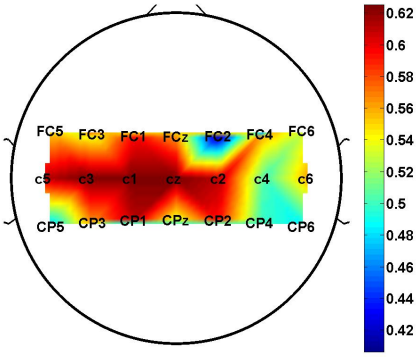
Figure C.1: Average TPR topographic plots for subject s1 for the WMM method in the SCP (C.1(a)), Mu (C.1(b)) and Beta (C.1(c)) bands and for the MD (C.1(d)) and bootstrap (C.1(e)) methods in the SCP band.



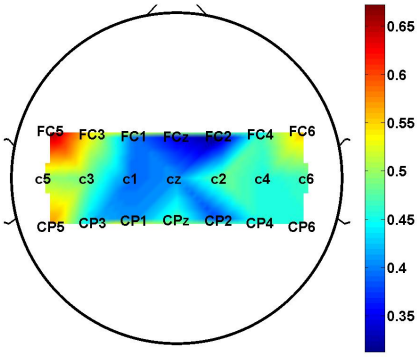
(a) Topographic plot fro s2 in the SCP band (WMM)



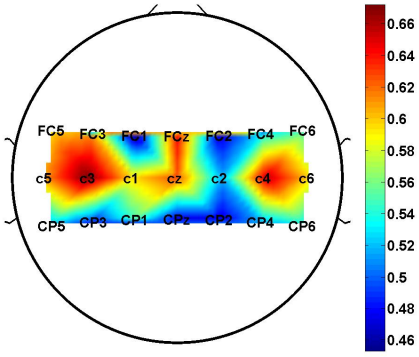
(b) Topographic plot for s2 in the Mu band (WMM)



(c) Topographic plot for s2 in the Beta band (WMM)

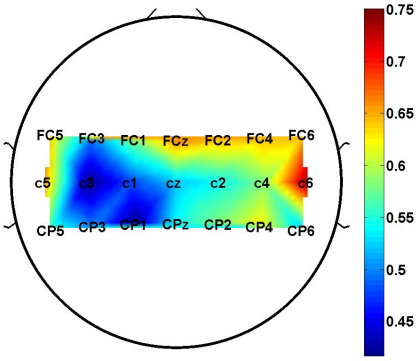


(d) Topographic plot for s2 in the SCP band (MD)

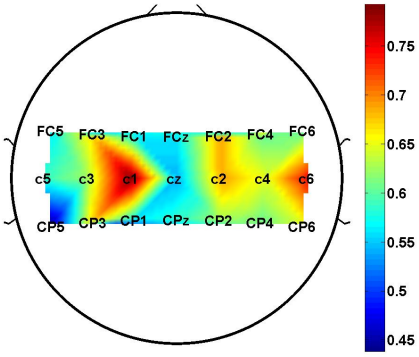


(e) Topographic plot for S2 in the SCP band (boot)

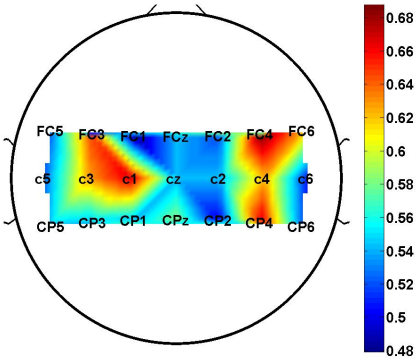
Figure C.2: The topographic plots for the average detection rate for the WMM method in the SCP (C.2(a)), Mu (C.2(b)) and Beta (C.2(c)) bands and for the MD (C.2(d)) and bootstrap (C.2(e)) methods in the SCP band.



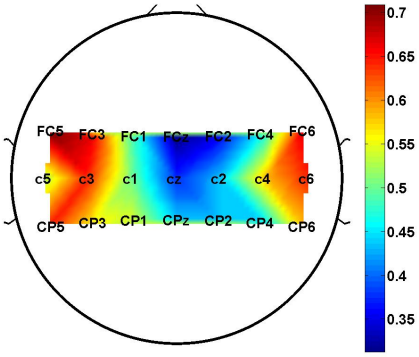
(a) Topographic plot fro s3 in the SCP band (WMM)



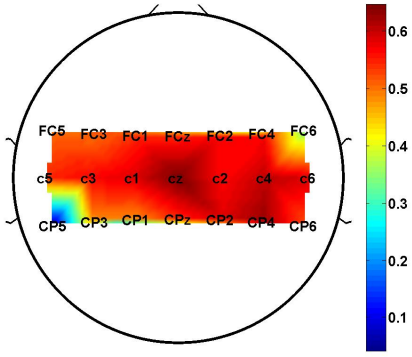
(b) Topographic plot for s3 in the Mu band (WMM)



(c) Topographic plot for s3 in the Beta band (WMM)

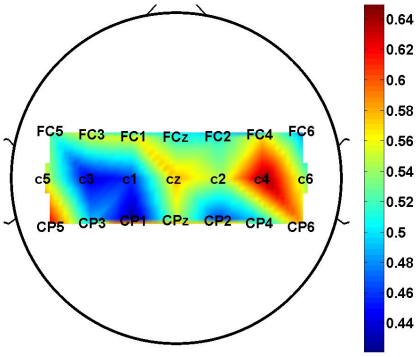


(d) Topographic plot for s3 in the SCP band (MD)

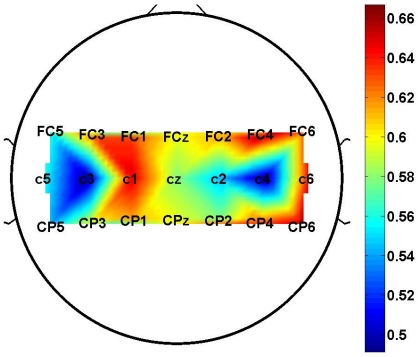


(e) Topographic plot for s3 in the SCP band (boot)

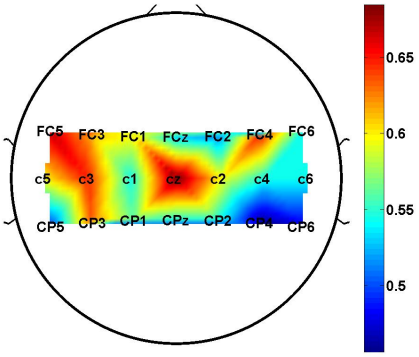
Figure C.3: The topographic plots for the average detection rate for the WMM method in the SCP (C.3(a)), Mu (C.3(b)) and Beta (C.3(c)) bands and for the MD (C.3(d)) and bootstrap (C.3(e)) methods in the SCP band.



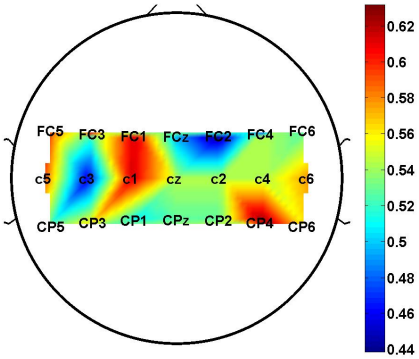
(a) Topographic plot fro s4 in the SCP band (WMM)



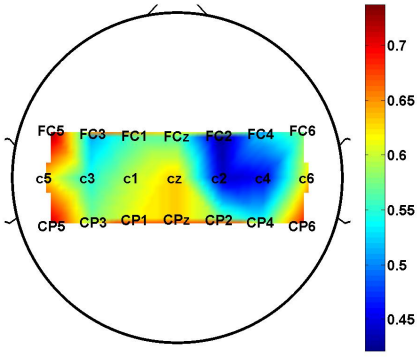
(b) Topographic plot for s4 in the Mu band (WMM)



(c) Topographic plot for s4 in the Beta band (WMM)

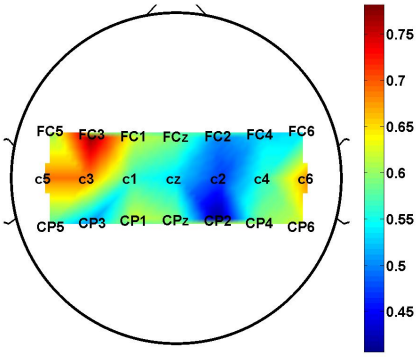


(d) Topographic plot for s4 in the SCP band (MD)

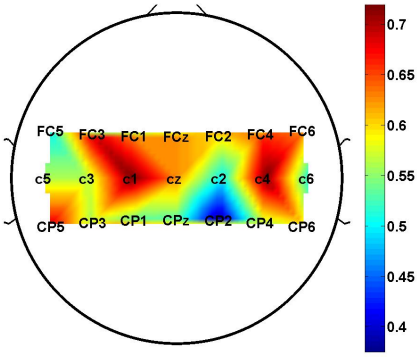


(e) Topographic plot for s4 in the SCP band (boot)

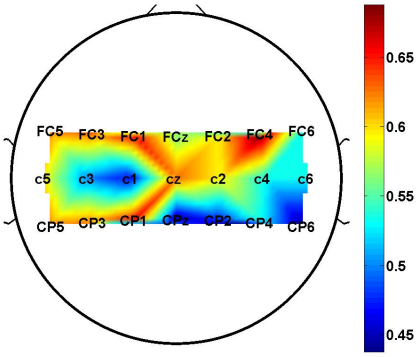
Figure C.4: The topographic plots for the average detection rate for the WMM method in the SCP (C.4(a)), Mu (C.4(b)) and Beta (C.4(c)) bands and for the MD (C.4(d)) and bootstrap (C.4(e)) methods in the SCP band.



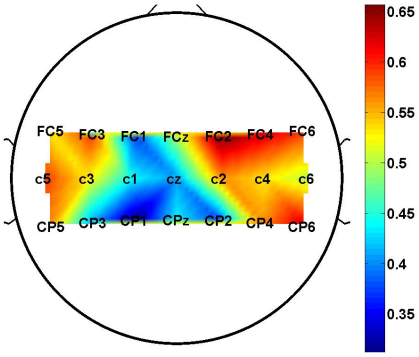
(a) Topographic plot fro s5 in the SCP band (WMM)



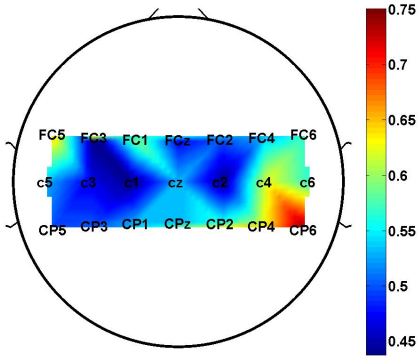
(b) Topographic plot for s5 in the Mu band (WMM)



(c) Topographic plot for s5 in the Beta band (WMM)

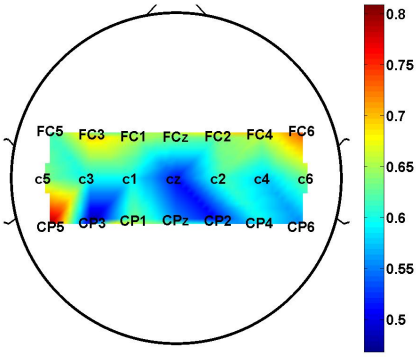


(d) Topographic plot for s5 in the SCP band (MD)

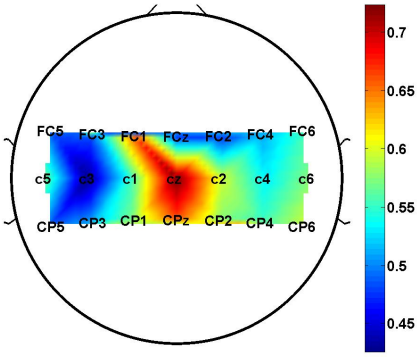


(e) Topographic plot for s5 in the SCP band (boot)

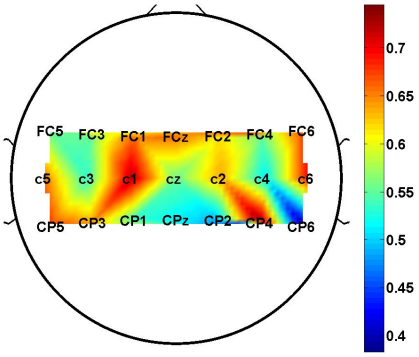
Figure C.5: The topographic plots for the average detection rate for the WMM method in the SCP (C.5(a)), Mu (C.5(b)) and Beta (C.5(c)) bands and for the MD (C.5(d)) and bootstrap (C.5(e)) methods in the SCP band.



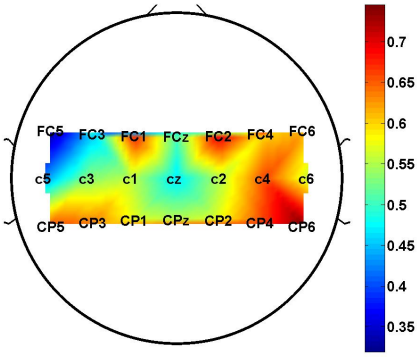
(a) Topographic plot fro s6 in the SCP band (WMM)



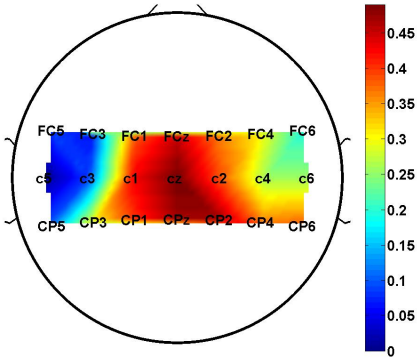
(b) Topographic plot for s6 in the Mu band (WMM)



(c) Topographic plot for s6 in the Beta band (WMM)

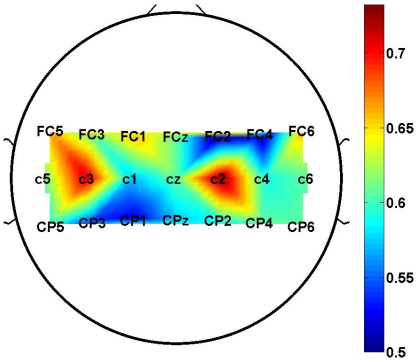


(d) Topographic plot for s6 in the SCP band (MD)

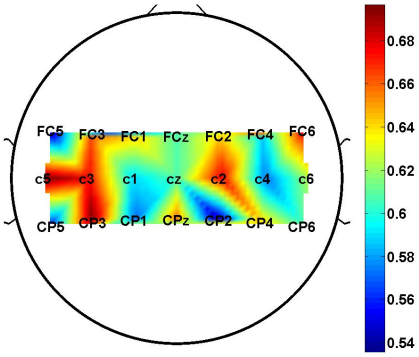


(e) Topographic plot for s6 in the SCP band (boot)

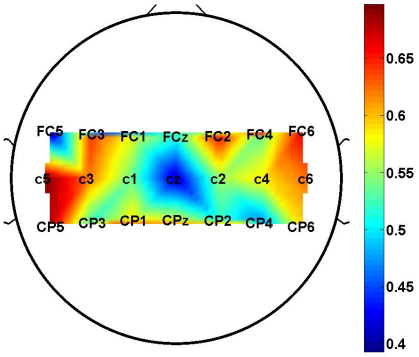
Figure C.6: The topographic plots for the average detection rate for the WMM method in the SCP (C.6(a)), Mu (C.6(b)) and Beta (C.6(c)) bands and for the MD (C.6(d)) and bootstrap (C.6(e)) methods in the SCP band.



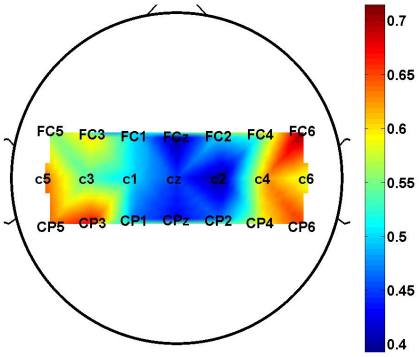
(a) Topographic plot fro s7 in the SCP band (WMM)



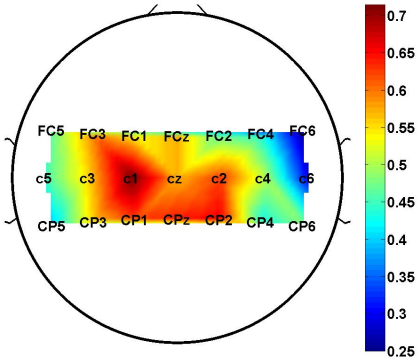
(b) Topographic plot for s7 in the Mu band (WMM)



(c) Topographic plot for s7 in the Beta band (WMM)

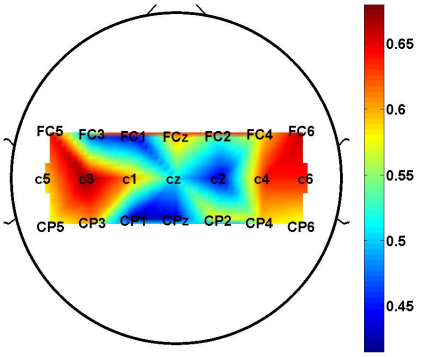


(d) Topographic plot for s7 in the SCP band (MD)

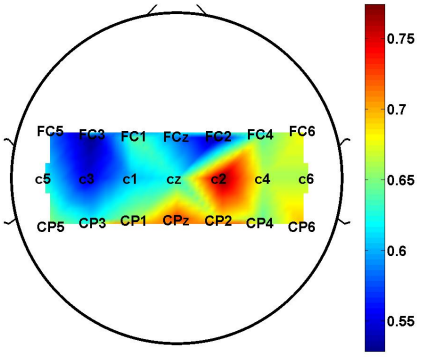


(e) Topographic plot for s7 in the SCP band (boot)

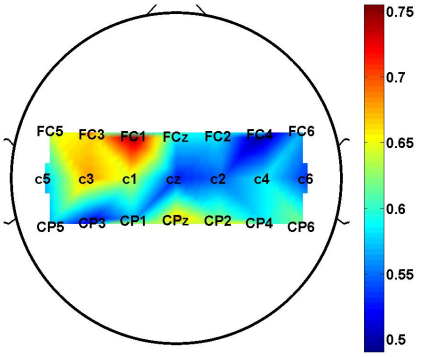
Figure C.7: Average TPR topographic plots for subject s7 for the WMM method in the SCP (C.7(a)), Mu (C.7(b)) and Beta (C.7(c)) bands and for the MD (C.7(d)) and bootstrap (C.7(e)) methods in the SCP band.



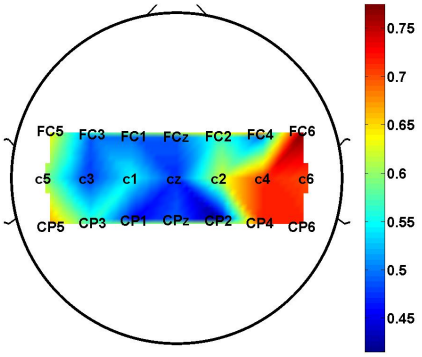
(a) Topographic plot fro s8 in the SCP band (WMM)



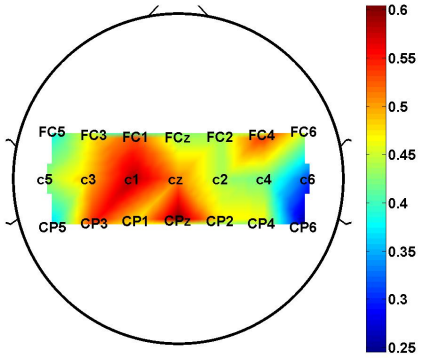
(b) Topographic plot for s8 in the Mu band (WMM)



(c) Topographic plot for s8 in the Beta band (WMM)

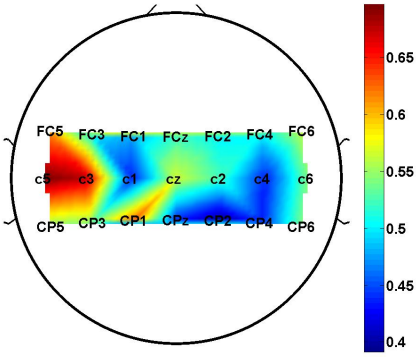


(d) Topographic plot for s8 in the SCP band (MD)

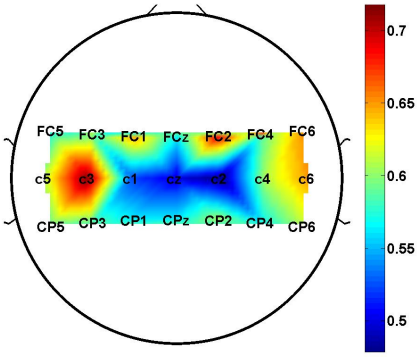


(e) Topographic plot for s8 in the SCP band (boot)

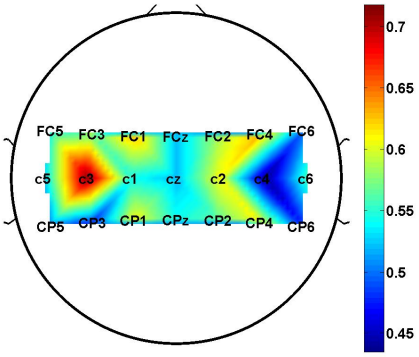
Figure C.8: Average TPR topographic plots for subject s8 for the WMM method in the SCP (C.8(a)), Mu (C.8(b)) and Beta (C.8(c)) bands and for the MD (C.8(d)) and bootstrap (C.8(e)) methods in the SCP band.



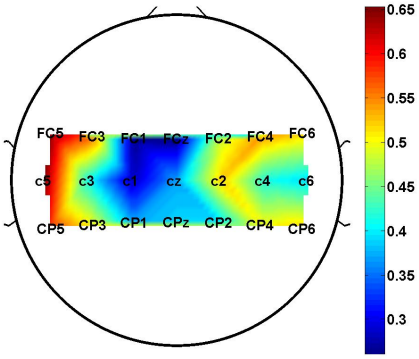
(a) Topographic plot fro s9 in the SCP band (WMM)



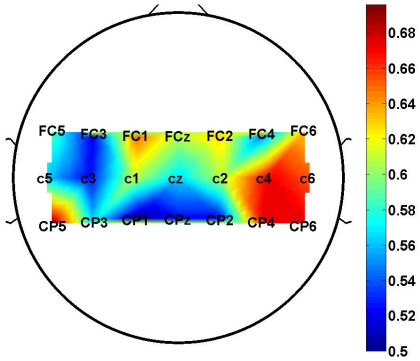
(b) Topographic plot for s9 in the Mu band (WMM)



(c) Topographic plot for s9 in the Beta band (WMM)



(d) Topographic plot for s9 in the SCP band (MD)



(e) Topographic plot for s9 in the SCP band (boot)

Figure C.9: Average TPR topographic plots for subject s9 for the WMM method in the SCP (C.9(a)), Mu (C.9(b)) and Beta (C.9(c)) bands and for the MD (C.9(d)) and bootstrap (C.9(e)) methods in the SCP band.

Appendix D

Participants information sheet

Participant Information Sheet

Name of department:
Bioengineering Unit

Title of the study:
Event detection in EEG data

Main investigator
Mr. Ange Guillaume Tano, Doctoral student, Bioengineering Unit, University of Strathclyde, Glasgow, UK, ange.tano@strath.ac.uk

Associate investigators
Dr H Lakany, lecturer, Bioengineering Unit, University of Strathclyde, Glasgow, UK, heba.lakany@strath.ac.uk
Prof. B.A. Conway, Professor, Bioengineering Unit, University of Strathclyde, Glasgow, UK, b.a.conway@strath.ac.uk
Prof. John Soraghan, Professor, Department of Electrical Engineering, University of Strathclyde, Glasgow, UK, john@strath.ac.uk

Facility
The study will take place in the biomechanical Lab of the Bioengineering Unit at the University of Strathclyde, Wolfson centre
Glasgow, G4 0NW

Sponsor
This project is funded by the EPSRC and is coordinated by the Doctoral Training Centre in Medical Devices of the University of Strathclyde.

The purpose of this investigation
The long term objective of the project is to develop assistive devices intended to help people with partial or total paralysis to communicate with their environment.
The aim of this study is to detect events in brain waves associated with initiation of voluntary movement. Detecting such event will enable us to develop algorithms that map events in brain signals into commands that can be used to control devices.

Do you have to take part?

There is no obligation to take part in this investigation. The experimental procedure will be clearly explained to the participant by the investigator and participants have a right to withdraw without detriment.

Background

Previous studies have reported that initiation of voluntary movement is preceded by a decrease in the energy level of brain waves followed by an increase of the energy level after the execution of movement. The purpose of this study is to detect the time of movement execution from changes in brain waves.

Experimental protocol

a. Equipment:

A VICON MX Ultranet HD system with 12 cameras will be used to record 3 dimensions coordinates of 14 markers which will be attached to the subject. The markers will be attached to the participant's right arm, right shoulder, chest, and back.

A Synsamps^2 (Neuroscan) amplifier will be used to record the electroencephalogram (EEG) and eye movements (EOG) of the participant. EEG will be recorded from cintered Ag/AgCl electrodes placed according the extended 10-20 standard electrode placement. The participant will wear a 128 electrodes cap (EasyCap).

b. Subjects:

10 participants will take part in the study. We are looking for participants in the age range from 20 up to 60 years with a general good health condition. You cannot take part in this study if you satisfy one of the following criteria:

- *You have a motor neuron disease*
- *You have any form of paralysis*
- *You have a mental disease*
- *You have an implanted electronic medical device in your brain*
- *You have a known brain damage condition*
- *You have known or suspected skin allergy to adhesives used in skin dressing e.g. ECG & EEG abrasive skin Prepping Gel(Nuprep) or Electro-Gel*
- *You cannot understand the protocol or instruction in English*

c. Duration:

The experiment will take approximately two and half hours for each participant including the preparation time (of the participant) e.g. the placement of the electrodes and conductive gel for EEG recording and the calibration of the VICON system and attachment of the markers to the participant for movement recording. The experiment will take place in the Biomechanical laboratory at the Bioengineering Unit at the University of Strathclyde.

d. Description of experiment, data collection and analysis:
The experimental set up will consist in a device comprising three hollow tubes soldered on a plate and forming a T cross of equal arms. There is a cup at the end of each tube and a cup at the crossing of the tubes. There is a communication canal from each cup at the end of each tube to the central cup. The tubes are soldered on a horizontal plate to the central cup making an angle of 15 degrees between the plan of each tube and the plan of the horizontal plate. The position of the central cup is called the central position and those of the cups paced at the end of the tubes are called target positions. The device is design in such a way that when a spherical ball is placed in a cup at the end of a tube, the ball will slide in the central cup through the tube and remain at this position. During the experiment the device will be placed on a board such that the central cap will be placed at the centre of the table and the cups at the end of the tubes are placed at 0, 90 and 180 degree from the central cup, corresponding to the right, forward and left direction of the participant. The experiment will be divided into two parts. During the part, the participant will randomly perform voluntary arm movement from the central position to one of the target position. The participant will be instructed to fix the central position with its eyes and avoid swallowing and blinking during arm movement from the central position to the target position. The participant will return its arm to the central position and rest no more than 8 seconds before repeating the movement. A trial will consist of moving the arm from the central position to a target position. The experiment will be divided into four sessions of 5 minutes each with a maximum rest of 5 minutes between 2 consecutive sessions.
During the second part of the experiment the participant will grasp and move a ball from the central position to a target position randomly, all other things being equal. A trial will consist in grasping the ball from the central position and place it randomly at one of the three target positions.

EEG and EOG data will be recorded using Neuroscan. Three dimensions coordinate data will be recorded using the VICON system. Data from VICON and Neuroscan will be synchronised by sending a constant 1.5V DC voltage to both recording systems.

EOG recording will be used to remove artefact e.g. eye blink from EEG data. Position data will be used to evaluate movement trajectory velocity and to group EEG data according to target locations and movement speed. Movement data will also be used to evaluate the time of movement execution. EEG signals in each target location will be averaged and used to calculate time of movement execution. The time of movement execution derived from movement data will be used as a control to validate that from EEG data.

What are the potential risks to you in taking part?

This project is classified as a medium risk. There is no possible hazard as the EEG recording equipment is properly grounded and isolated. All procedures are non invasive and the only possible source of hazard is any allergy to the gel used to establish a good conductivity between the electrodes and the participant's scalp. There will also be a slight discomfort for very hairy participants while removing the double sides tape used to attach the markers for position data recording.

What happens to the information in the project?
Data recorded during this study will be confidential and will be securely stored within the Bioengineering Unit research facilities. Personal data will not be accessible to unauthorised people and will be used exclusively for the purpose of the present study.

Dissemination of findings
Outcomes of the study will be published in peer-reviewed journals and conference proceedings. The identity of participants will be confidential and anonymous during publication.

The University of Strathclyde is registered with the Information Commissioner's Office who implements the Data Protection Act 1998. All personal data on participants will be processed in accordance with the provisions of the Data Protection Act 1998.

Thank you for reading this information – please ask any questions if you are unsure about what is written here.

What happens next?
You should carefully read this information sheet and if you would like to take part in the study you should sign a consent form before doing so. You also have the right to withdraw from the study at any time without detriment.

This investigation was granted ethical approval by the University of Strathclyde ethics committee.

If you have any questions/concerns, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought from, please contact:
Mr. David Smith
Secretary to the Departmental Ethics Committee
University of Strathclyde
Bioengineering Unit
106 Rottenrow
Glasgow
G4 0NW
Telephone: 0141 548 3142

4

Last updated: March 2010

Email: d.j.a.smith@strath.ac.uk

Researcher Contact Details:
Mr Ange Guillaume Tano, Bioengineering Unit, University of Strathclyde 106 Rottenrow, Glasgow, G4 0NW, ange.tano@strath.ac.uk

Chief Investigator Details:
Dr Heba Lakany, lecturer, Bioengineering Unit, University of Strathclyde, 106 Rottenrow, Glasgow, G4 0NW, Tel: +141 548 3487, Fax: +141 552 6098

5

Last updated: March 2010

Consent Form

Name of department:
Bioengineering Unit

Title of study:
Event detection in EEG data

University of Strathclyde Engineering

Researcher's details:
Name: Ange Guillaume Tano
Contact details: ange.tano@strath.ac.uk

Participant's details
Name:
Contact details:

- I confirm that I have read and understood the information sheet for the above project and the researcher has answered any queries to my satisfaction.
- I understand that my participation is voluntary and that I am free to withdraw from the project at any time, without having to give a reason and without any consequences.
- I understand that I can withdraw my data from the study at any time.
- I understand that any information recorded in the investigation will remain confidential and no information that identifies me will be made publicly available.
- I consent to being a participant in the project

I (PRINT NAME)	Hereby agree to take part in the above project
Signature of Participant:	Date

6

Last updated: March 2010

Bibliography

Akhtar, M. T., Mitsuhashi, W. and James, C. J. (2012). Employing spatially constrained ica and wavelet denoising, for automatic removal of artifacts from multichannel eeg data, *Signal Processing* **92**(2): 401–416.

Allison, B. Z., Wolpaw, E. W. and Wolpaw, J. R. (2007). Brain-computer interface systems: progress and prospects.

Alonso-Valerdi, L. M., Sepulveda, F. and Ramírez-Mendoza, R. A. (2015). Perception and cognition of cues used in synchronous brain–computer interfaces modify electroencephalographic patterns of control tasks, *Frontiers in human neuroscience* **9**.

Andrew, A. M. (2000). Statistical pattern recognition, by andrew webb, arnold, london (cambridge university press, new york, for usa), 1999, xviii+ 454 pp., isbn 0-340-74164-3 (pbk, 29.99).

Babiloni, F., Bianchi, L., Semeraro, F., Millan, D. R., Mouriño, J., Cattini, A., Salinari, S., Marciani, M. G., Cincotti, F. et al. (2001). Mahalanobis distance-based classifiers are able to recognize eeg patterns by using few eeg electrodes, *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, Vol. 1, IEEE, pp. 651–654.

Bai, O., Lin, P., Vorbach, S., Li, J., Furlani, S. and Hallett, M. (2007). Exploration of computational methods for classification of movement intention during human voluntary movement from single trial eeg, *Clinical Neurophysiology* **118**(12): 2637–2655.

Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D.-Y., Schneider, L., Houdayer, E., Chen, X. and Hallett, M. (2011). Prediction of human voluntary movement before it occurs, *Clinical Neurophysiology* **122**(2): 364–372.

Bertrand, O., Perrin, F. and Pernier, J. (1985). A theoretical justification of the average reference in topographic evoked potential studies., *Electroencephalography and clinical neurophysiology* **62**(6): 462–464.

Birbaumer, N. (2006). Brain–computer-interface research: coming of age, *Clinical Neurophysiology* **117**(3): 479–483.

Bishop, C. M. (1994). Novelty detection and neural network validation, *Vision, Image and Signal Processing, IEE Proceedings-*, Vol. 141, IET, pp. 217–222.

Bishop, C. M. et al. (1995). Neural networks for pattern recognition.

Blankertz, B., Dornhege, G., Schäfer, C., Krepki, R., Kohlmorgen, J., Müller, K.-R., Kunzmann, V., Losch, F. and Curio, G. (2003). Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial eeg analysis, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* **11**(2): 127–131.

Bloomfield, P. (2004). *Fourier Analysis of Time Series: An Introduction*, Wiley series in probability and statistics: Applied probability and statistics, John Wiley & Sons.
URL: *http://books.google.co.uk/books?id=zQsupRg5rrAC*

Bokil, H., Purpura, K., Schoffelen, J.-M., Thomson, D. and Mitra, P. (2007). Comparing spectra and coherences for groups of unequal size., *Journal of Neuroscience Methods* **159**(2): 337–345.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/16945422*

Bradshaw, L. and Wikswo Jr, J. (2001). Spatial filter approach for evaluation of the surface laplacian of the electroencephalogram and magnetoencephalogram, *Annals of biomedical engineering* **29**(3): 202–213.

Brotherton, T., Johnson, T. and Chadderdon, G. (1998). Classification and novelty detection using linear models and a class dependent-elliptical basis function neural network, *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, Vol. 2, IEEE, pp. 876–879.

Bryant, F. B. and Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis.

Carpenter, G. A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine, *Computer vision, graphics, and image processing* **37**(1): 54–115.

Carpenter, G. A., Rubin, M. A. and Streilein, W. W. (1996). Artmap-fd: familiarity discrimination applied to radar target recognition, *Technical report*, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Statistics in Medicine* **19**: 1141–1164.

Carretié, L., Tapia, M., Mercado, F., Albert, J., López-Martín, S. and Juan, M. (2004). Voltage-based versus factor score-based source localization analyses of electrophysiological brain activity: a comparison, *Brain topography* **17**(2): 109–115.

Carvalhaes, C. and de Barros, J. A. (2015). The surface laplacian technique in eeg: Theory and methods, *International Journal of Psychophysiology* **97**(3): 174 – 188.

Caudell, T. and Newman, D. (1993). An adaptive resonance architecture to define normality and detect novelties in time series and databases, *IEEE World Congress on Neural Networks, Portland, Oregon*, pp. 166–176.

Cincotti, F., Mattia, D., Babiloni, C., Carducci, F., Bianchi, L., Millán, J. d. R., Mourino, J., Salin-

ari, S., Marciani, M., Babiloni, F. et al. (2002). Classification of eeg mental patterns by using two scalp electrodes and mahalanobis distance-based classifiers, *Methods Inf Med* **41**(4): 337–341.

Cormen, T. H., Leiserson, C. E. and Rivest, R. L. (2001). C. stein introduction to algorithms, *MIT Press* **5**(3): 55.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. and Stein, C. (2001). Introduction to algorithms second edition.

Cox, D. R. D. R. and Hinkley, D. V., j. a. (1974). *Theoretical statistics / [by] D. R. Cox [and] D. V. Hinkley*, London : Chapman and Hall.

da Silva, F. L. and Van Rotterdam, A. (1999). Biophysical aspects of eeg and magnetoencephalogram generation, *Electroencephalography: Basic principles, clinical applications and related fields* pp. 107–126.

Dasgupta, D. and Forrest, S. (1996). Novelty detection in time series data using ideas from immunology, *Proceedings of the international conference on intelligent systems*, pp. 82–87.

Dasgupta, S., Papadimitriou, C. H. and Vazirani, U. (2006). *Algorithms*, McGraw-Hill, Inc.

Davis, L. E. (2005). *Fundamentals of neurologic disease*, Demos Medical Publishing.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis., *Psychological methods* **2**(3): 292.

Deecke, L., Scheid, P. and Kornhuber, H. H. (1969). Distribution of readiness potential, premotion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements, *Experimental Brain Research* **7**(2): 158–168.

del R. Millán, J. and Mouriño, J. (2003). Asynchronous bci and local neural classifiers: An overview of the adaptive brain interface project, *IEEE Transaction on Neuronal Systems and Rehabilitation Engineering* **11, NO. 2**: 159–161.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals, *Statistical Science* **11**(3): pp. 189–212.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations, *Behavior Research Methods, Instruments, & Computers* **30**(1): 34–43.

Dien, J., Beal, D. J. and Berg, P. (2005). Optimizing principal components analysis of event-related potentials: matrix type, factor loading weighting, extraction, and rotations, *Clinical neurophysiology* **116**(8): 1808–1825.

Donchin, E., Spencer, K. M. and Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a p300-based brain-computer interface, *Rehabilitation Engineering, IEEE Transactions on* **8**(2): 174–179.

Dornhege, G. (2006). *Increasing information transfer rates for brain-computer interfacing*, PhD thesis, Universitätsbibliothek.

Duda, R. O., Hart, P. E. and Stork, D. G. (2001). Pattern classification. 2nd, *Edition. New York* .

Efron, B. (1987). *The Jackknife, the Bootstrap, and Other Resampling Plans (CBMS-NSF Regional Conference Series in Applied Mathematics)*, Society for Industrial Mathematics.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Monographs on statistics and applied probability, Chapman & Hall.

Evarts, E. V. (1968). Relation of pyramidal tract activity to force exerted during voluntary movement, *J Neurophysiol* **31**(1): 14–27.

Evarts, E. V. and Tanji, J. (1976). Reflex and intended responses in motor cortex pyramidal tract neurons of monkey, *J Neurophysiol* **39**(5): 1069–1080.

Faloutsos, C., Korn, F., Labrinidis, A., Kotidis, Y., Kaplunovich, A. and Perkovic, D. (1997). Quantifiable data mining using principal component analysis.

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

Fenwick, P., Mitchie, P., Dollimore, J. and Fenton, G. (1969). Application of the autoregressive model to eeg analysis., *Agressologie: revue internationale de physio-biologie et de pharmacologie appliquees aux effets de l'agression* **10**: Suppl–553.

Filevich, E. and Haggard, P. (2013). Persistence of internal representations of alternative voluntary actions.

Filk, T. and Römer, H. (2011). Generalized quantum theory: Overview and latest developments, *Axiomathes* **21**(2): 211–220.

Fisch, B. J. and Spehlmann, R. (1999). *Fisch and Spehlmann's EEG primer: basic principles of digital and analog EEG*, Elsevier Health Sciences.

Fisher, R. (1992). Statistical methods for research workers, *Breakthroughs in Statistics*, Springer, pp. 66–70.

Fleming, S. M., Mars, R. B., Gladwin, T. E. and Haggard, P. (2009). When the brain changes its mind: flexibility of action selection in instructed and free choices, *Cerebral Cortex* **19**(10): 2352–2360.

Fries, P., Nikolić, D. and Singer, W. (2007). The gamma cycle, *Trends in neurosciences* **30**(7): 309–316.

Fukunaga, R. (1990). *Statistical pattern recognition*, Academic Press.

Garcia, G., Ebrahimi, T. and Vesin, J. (2003). Support vector eeg classification in the fourier and time-frequency correlation, *Proceedings of the IEEE-EMBS First International Conference on Neural Engineering 2003*, number LTS-CONF-2003-030, IEEE, pp. 591–594.

Gert Pfurtscheller, B. G. and Neuper, C. (2000). Eeg-based brain-computer interface system, *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING* **47, NO. 10**: 1297–1307.

Gerwin Schalk, Eric C. Leuthardt, P. B. J. G. O. L. A. G. J. R. W. (2008). Continuous eeg classification during motor imagery-simulation of an asynchronous bci, *NeuroImage* **43**: 245–249.

Gibbons, J. D. and Chakraborti, S. (2003). Nonparametric statistical inference marcel dekker, *Inc. New York* .

Gorsuch, R. L. e. (1983). *Factor analysis*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Guger, C., Allison, B. and Edlinger, G. (2013). State of the art in bci research: Bci award 2011, *Brain-Computer Interface Research*, Springer, pp. 1–5.

Haas, L. (2003). Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography, *Journal of Neurology, Neurosurgery & Psychiatry* **74**(1): 9–9.

Haggard, P., Clark, S. and Kalogeras, J. (2002). Voluntary action and conscious awareness, *Nature neuroscience* **5**(4): 382–385.

Hall, P. (1993). On edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation, *Journal of the Royal Statistical Society. Series B (Methodological)* **55**(1): pp. 291–304.

Hall, P. (1995). *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics, Springer.

Halliday, D., Rosenberg, J., Amjad, A., Breeze, P., Conway, B. and Farmer, S. (n.d.). A framework for the analysis of mixed time series/point process data. theory and application to the study of physiological tremor, single motor unit discharges and electromyograms.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A. and Epstein, R. (2005). Bootstrap methods and permutation tests, *Introduction to the Practice of Statistics* **5**: 1–70.

Higham, N. J. (2008). *Functions of matrices: theory and computation*, Siam.

Hinterberger, T., Schmidt, S., Neumann, N., Mellinger, J., Blankertz, B., Curio, G. and Birbaumer, N. (2004). Brain-computer communication and slow cortical potentials, *Biomedical Engineering, IEEE Transactions on* **51**(6): 1011–1018.

Hjorth, B. (1975). An on-line transformation of eeg scalp potentials into orthogonal source derivations, *Electroencephalography and clinical neurophysiology* **39**(5): 526–530.

Hollander, M. and Wolfe, D. (1973). *Nonparametric statistical methods*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section, Wiley.

Hong, J., Kim, Y., Lee, H. and Lee, Y. (2002). Damage detection using the lipschitz exponent estimated by the wavelet transform: applications to vibration modes of a beam, *International Journal of Solids and Structures* **39**(7): 1803–1816.
URL: *http://www.ingentaconnect.com/content/els/00207683/2002/00000039/00000007/art00279*

Hopkins, K. D. and Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting, *Educational and Psychological Measurement* **50**(4): 717–729.

Hudson, D. L. and Cohen, M. E. (2000). Neural networks and artificial intelligence for biomedical engineering, Institute of Electrical and Electronics Engineers.

Hyvärinen, A., Karhunen, J. and Oja, E. (2004). *Independent component analysis*, Vol. 46, John Wiley & Sons.

James, C. J. and Gibson, O. J. (2003). Temporally constrained ica: an application to artifact rejection in electromagnetic brain signal analysis, *Biomedical Engineering, IEEE Transactions on* **50**(9): 1108–1116.

James, C. J. and Hesse, C. W. (2005). Independent component analysis for biomedical signals, *Physiological measurement* **26**(1): R15.

Jansen, B. H., Bourne, J. R. and Ward, J. W. (1981). Autoregressive estimation of short segment spectra for computerized eeg analysis, *Biomedical Engineering, IEEE Transactions on* (9): 630–638.

Jansen, B. H., Hasman, A. and Lenten, R. (1981). Piecewise analysis of eegs using ar-modeling and clustering, *Computers and Biomedical Research* **14**(2): 168–178.

Jenkins, G. M. and Watts, D. G. (n.d.). Spectral analysis and its applications 1968. San Francisco: Holden-Day.

Jurcak, V., Tsuzuki, D. and Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems, *Neuroimage* **34**(4): 1600–1611.

Kandel, E. R., Schwartz, J. H., Jessell, T. M. et al. (2000). *Principles of neural science*, Vol. 4, McGraw-Hill New York.

Kane, R. (2011). *The Oxford handbook of free will*, Oxford University Press, USA.

Kaplansky, I. (1945). A common error concerning kurtosis, *Journal of the American Statistical Association* **40**(230): 259–259.

Kennedy, P. R. and Bakay, R. A. (1998). Restoration of neural output from a paralyzed patient by a direct brain connection, *Neuroreport* **9**(8): 1707–1711.

Kohonen, T. (2001). *Self-organizing maps*, Vol. 30, Springer Science & Business Media.

Koo, B., Lee, H.-G., Nam, Y., Kang, H., Koh, C. S., Shin, H.-C. and Choi, S. (2015). A hybrid nirs-eeg system for self-paced brain computer interface with online motor imagery, *Journal of neuroscience methods* **244**: 26–32.

Kozelka, J. W. and Pedley, T. A. (1990). Beta and mu rhythms., *Journal of Clinical Neurophysiology* **7**(2): 191–208.

Krakauer, J. and Ghez, C. (2000). Voluntary movement, *Principles of neural science* **4**: 756–781.

Kübler, A., Kotchoubey, B., Hinterberger, T., Ghanayim, N., Perelmouter, J., Schauer, M., Fritsch, C., Taub, E. and Birbaumer, N. (1999). The thought translation device: a neurophysiological approach to communication in total motor paralysis, *Experimental Brain Research* **124**(2): 223–232.

Kurata, K. and Tanji, J. (1985). Contrasting neuronal activity in supplementary and precentral motor cortex of monkeys. ii. responses to movement triggering vs. nontriggering sensory signals, *J Neurophysiol* **53**(1): 142–152.

K.Ward, M. F. . G. E. B. . R. (2007). A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms, *J Comput Neurosci* **23**: 21–37.

Lang, W. (2003). Surface recordings of the bereitschaftspotential in normals, *The Bereitschaftspotential*, Springer, pp. 19–34.

Lebedev, M. A. and Nicolelis, M. A. (2006). Brain–machine interfaces: past, present and future, *TRENDS in Neurosciences* **29**(9): 536–546.

Lehtonen, J. and Koivikko, M. (1971). The use of a non-cephalic reference electrode in recording cerebral evoked potentials in man, *Electroencephalography and clinical neurophysiology* **31**(2): 154–156.

LePage, R., Billard, L. and of Mathematical Statistics, I. (1992). *Exploring the Limits of Bootstrap*, Wiley Series in Probability and Statistics, Wiley.

Lew, E., Chavarriaga, R., Silvoni, S. and Mill?n, J. d. R. (2012). Detection of self-paced reaching movement intention from eeg signals, *Frontiers in Neuroengineering* **5**(13).

Libet, B. (1993). Unconscious cerebral initiative and the role of conscious will in voluntary action, *Neurophysiology of consciousness*, Springer, pp. 269–306.

Libet, B., Freeman, A. and Sutherland, J. (1999). Editors introduction, *Journal of consciousness studies* **6**(8-9): 8–9.

Libet, B., Gleason, C. A., Wright, E. W. and Pearl, D. K. (1983a). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act, *Brain* **106**(3): 623–642.

Libet, B., Gleason, C. A., Wright, E. W. and Pearl, D. K. (1983b). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act, *Brain* **106**(3): 623–642.

Loukas, C. and Brown, P. (2004). Online prediction of self-paced hand-movements from subthalamic activity using neural networks in parkinsons disease, *Journal of neuroscience methods* **137**(2): 193–205.

Lu, M.-K., Arai, N., Tsai, C.-H. and Ziemann, U. (2012). Movement related cortical potentials of cued versus self-initiated movements: Double dissociated modulation by dorsal premotor cortex versus supplementary motor area rtms, *Human brain mapping* **33**(4): 824–839.

Mallat, S. (1989a). A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11(7)**: 674–693.

Mallat, S. (1989b). A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(7): 674–693.

Mallat, S. and Hwang, W. L. (1992). Singularity detection and processing with wavelets, *IEEE Trans. Inf. Th* **38**: 617–643.

Markou, M. and Singh, S. (2003). Novelty detection: a reviewpart 1: statistical approaches, *Signal processing* **83**(12): 2481–2497.

Mason, S. and Birch, G. (2000). A brain-controlled switch for asynchronous control applications, *Biomedical Engineering, IEEE Transactions on Biomedical Engineering* **47**(10): 1297–1307.

McFarland, D. J., Krusienski, D. J. and Wolpaw, J. R. (2006). Brain–computer interface signal processing at the wadsworth center: mu and sensorimotor beta rhythms, *Progress in brain research* **159**: 411–419.

McFarland, D. J., McCane, L. M., David, S. V. and Wolpaw, J. R. (1997). Spatial filter selection for eeg-based communication, *Electroencephalography and clinical Neurophysiology* **103**(3): 386–394.

McFarland, D. and Krusienski, D. (2012). Bci signal processing: feature translation, *Brain-computer interfaces: principles and practice. Oxford University Press, Oxford* pp. 147–164.

McLachlan, G. (1999). Mahalanobis distance, *Resonance* **4**(6): 20–26.

Mitra, P. and Pesaran, B. (1999). Analysis of dynamic brain imaging data, *Biophysical Journal* **76**(2): 691 – 708.
URL: *http://www.sciencedirect.com/science/article/pii/S000634959977236X*

Moore, I. C. and Cada, M. (2004). Prolate spheroidal wave functions, anintroduction to the slepian series and its properties, *Applied and Computational Harmonic Analysis* **16**(3): 208 – 230.
URL: *http://www.sciencedirect.com/science/article/pii/S106352030400017X*

Moors, J. (1986). The meaning of kurtosis: Darlington reexamined, *The American Statistician* **40**(4): 283–284.

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

Müller-Gerking, J., Pfurtscheller, G. and Flyvbjerg, H. (2000). Classification of movement-related eeg in a memorized delay task experiment, *Clinical Neurophysiology* **111**(8): 1353–1365.

Nairac, A., Townsend, N., Carr, R., King, S., Cowley, P. and Tarassenko, L. (1999). A system for the analysis of jet engine vibration data, *Integrated Computer-Aided Engineering* **6**(1): 53–66.

Neidermeyer, E. (1999). The normal eeg of the waking adult, *Electroencephalography: basic principles, clinical applications and related fields, 4th ed. Baltimore, MD: Williams and Wilkins* pp. 149–173.

Neuper, C., Scherer, R., Reiner, M. and Pfurtscheller, G. (2005). Imagery of motor actions: Differential effects of kinesthetic and visual–motor mode of imagery in single-trial eeg, *Cognitive Brain Research* **25**(3): 668–677.

Nicolelis, M. A. (2001). Actions from thoughts, *Nature* **409**(6818): 403–407.

Niedermeyer, E. and da Silva, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*, Lippincott Williams & Wilkins.

Niels, B. et al. (2000). The thought translation device (ttd) for completely paralyzed patients, *IEEE Transactions on Rehabilitation Engineering* **8**(2).

Nievergelt, J. and Hinrichs, K. H. (1999). *Algorithms & data structures*, vdf Hochschulverl. an der ETH.

Nunez, P., Silberstein, R., Cadusch, P., Wijesinghe, R., Westdorp, A. and Srinivasan, R. (1994). A theoretical and experimental study of high resolution eeg based on surface laplacians and cortical imaging, *Electroencephalography and clinical neurophysiology* **90**(1): 40–57.

Obermaier, B., Guger, C., Neuper, C. and Pfurtscheller, G. (2001). Hidden markov models for online classification of single trial eeg data, *Pattern recognition letters* **22**(12): 1299–1309.

Offner, F. F. (1950). The eeg as potential mapping: the value of the average monopolar reference, *Electroencephalography and clinical neurophysiology* **2**(1): 213–214.

Olejniczak, P. (2006). Neurophysiologic basis of eeg, *Journal of clinical neurophysiology* **23**(3): 186–189.

Osselton, J. (1965). Acquisition of eeg data by bipolar unipolar and average reference methods: a theoretical comparison, *Electroencephalography and clinical neurophysiology* **19**(5): 527–528.

Pais-Vieira, M., Lebedev, M., Kunicki, C., Wang, J. and Nicolelis, M. A. (2013). A brain-to-brain interface for real-time sharing of sensorimotor information, *Scientific reports* **3**.

Parra, L., Deco, G. and Miesbach, S. (1996). Statistical independence and novelty detection with information preserving nonlinear maps, *Neural Computation* **8**(2): 260–269.

Pearson, E. and Please, N. (1975). Relation between the shape of population distribution and the robustness of four simple test statistics, *Biometrika* **62**(2): 223–241.

Penfield, W. and Rasmussen, T. (n.d.). The cerebral cortex of man, a clinical study of localization of function, 1950, *McMillan, New York* .

Penfield, W. and Welch, K. (1951). The supplementary motor area of the cerebral cortex: a clinical and experimental study, *AMA Archives of Neurology & Psychiatry* **66**(3): 289–317.

Peng, Z., He, Y., Chen, Z. and Chu, F. (2002). Identification of the shaft orbit for rotating machines using wavelet modulus maxima, *Mechanical Systems and Signal Processing* **16**(4): 623–635.

Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications : multitaper and conventional univariate techniques / Donald B. Percival and Andrew T. Walden*, Cambridge University Press, Cambridge ; New York, N.Y., U.S.A. :.
URL: *http://www.loc.gov/catdir/toc/cam021/92045862.html*

Pfurtscheller, G. (2004). *Brain-computer Interface-State of the Art and Future Prospects*, na.

Pfurtscheller, G. (2006). The cortical activation model (cam), *Progress in brain research* **159**: 19–27.

Pfurtscheller, G., Brunner, C., Schlögl, A. and Da Silva, F. L. (2006). Mu rhythm (de) synchronization and eeg single-trial classification of different motor imagery tasks, *Neuroimage* **31**(1): 153–159.

Pfurtscheller, G., Kalcher, J., Neuper, C., Flotzinger, D. and Pregenzer, M. (1996). On-line eeg classification during externally-paced hand movements using a neural network-based classifier, *Electroencephalography and clinical Neurophysiology* **99**(5): 416–425.

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N. and Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts, *Psychophysiology* **30**(6).

Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization, *Handbook of psychophysiology* **3**: 56–84.

Prediction of human voluntary movement before it occurs (2011). *Clinical Neurophysiology* **122**(2): 364 – 372.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3 edn, Cambridge University Press.

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

Rabiner, L. and Juang, B.-H. (1986). An introduction to hidden markov models, *ASSP Magazine, IEEE* **3**(1): 4–16.

Rao, R. P., Stocco, A., Bryan, M., Sarma, D., Youngquist, T. M., Wu, J. and Prat, C. S. (2014). A direct brain-to-brain interface in humans, *PloS one* **9**(11): e111332.

Regan, D. (1977). Steady-state evoked potentials, *JOSA* **67**(11): 1475–1489.

Ribeiro, P. and Guerreiro, P. (2009). Improving the automatic evaluation of problem solutions in programming contests, *Olympiads in Informatics* **3**: 132–143.

Roberts, S. and Tarassenko, L. (1994). A probabilistic resource allocating network for novelty detection, *Neural Computation* **6**(2): 270–284.

Robertson, A. N., Farrar, C. R. and Sohn, H. (2003). Singularity detection for structural health monitoring using holder exponents, *Smart Structures and Materials*, International Society for Optics and Photonics, pp. 569–581.

Ruppert, D. (1987). What is kurtosis? an influence function approach, *The American Statistician* **41**(1): 1–5.

Salvaris, M. and Haggard, P. (2014). Decoding intention at sensorimotor timescales, *PloS one* **9**(2): e85100.

Shanechi, M. M., Hu, R. C. and Williams, Z. M. (2014). A cortical–spinal prosthesis for targeted limb movement in paralysed primate avatars, *Nature communications* **5**.

Shannon, C. (n.d.). ve weaver, w.(1964). the mathematical theory of communication, *Urbana: University of Illinois Press* .

Sharbrough, F., Chatrian, G., Lesser, R., Lüders, H., Nuwer, M. and Picton, T. (1990). Guidelines for standard electrode position nomenclatureamerican eeg society.

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case, *AT T Technical Journal* **57**: 1371–1430.

Soon, C. S., Brass, M., Heinze, H.-J. and Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain, *Nature neuroscience* **11**(5): 543–545.

Stanley, G., Li, F. and Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus., *The Journal of neuroscience: the official journal of the Society for Neuroscience* **19**(18): 8036.

Struzik, Z. R. (2001). Wavelet methods in (financial) time-series processing, *Physica A: Statistical Mechanics and its Applications* **296**(1): 307–319.

Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses, *Journal of Microcomputer Applications* **15**(1): 31–45.

Tanji, J. and Kurata, K. (1985). Contrasting neuronal activity in supplementary and precentral motor cortex of monkeys. i. responses to instructions determining motor responses to forthcoming signals of different modalities, *J Neurophysiol* **53**(1): 129–141.

Tax, D. M. and Duin, R. P. (1998). Outlier detection using classifier instability, *Advances in Pattern Recognition*, Springer, pp. 593–601.

Teplan, M. (2002). Fundamentals of eeg measurement, *Measurement science review* **2**(2): 1–11.

Thut, G., Hauert, C.-A., Viviani, P., Morand, S., Spinelli, L., Blanke, O., Landis, T. and Michel, C. (2000). Internally driven vs. externally cued movement selection: a study on the timing of brain activity, *Cognitive Brain Research* **9**(3): 261–269.

Tressoldi, P. E., Pederzoli, L., Caini, P., Fedele, P., Ferrini, A. and Melloni, S. (2014). Brain-to-brain (mind-to-mind) interaction at distance: A pilot study, *Available at SSRN 2423852* .

Vallabhaneni, A. and He, B. (2004). Motor imagery task classification for brain computer interface applications using spatiotemporal principle component analysis, *Neurological research* **26**(3): 282–287.

Valsan, G. (2007). Brain computer interface using detection of movement intention, *PhD thesis* .

Vaughan, T. M., Heetderks, W. J., Trejo, L. J., Rymer, W. Z., Weinrich, M., Moore, M. M., Kübler, A., Dobkin, B. H., Birbaumer, N., Donchin, E. et al. (2003). Brain-computer interface technology: a review of the second international meeting., *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* **11**(2): 94–109.

Vaughan, T. M., Wolpaw, J. R. and Donchin, E. (1996). Eeg-based communication: prospects and problems, *Rehabilitation Engineering, IEEE Transactions on* **4**(4): 425–430.

Venkatakrishnan, P., Sangeetha, S. and Sundar, M. (2012). Measurement of lipschitz exponent (le) using wavelet transform modulus maxima (wtmm), *International Journal of Scientific & Engineering Research* **3**(6).

Wang, Y., Zhang, Z., Li, Y., Gao, X., Gao, S. and Yang, F. (2004). Bci competition 2003-data set iv: an algorithm based on cssd and fda for classifying single-trial eeg, *Biomedical Engineering, IEEE Transactions on* **51**(6): 1081–1086.

Wennberg, A. and Isaksson, A. (1976). Simulation of nonstationary eeg signals as a means of objective clinical interpretation of eeg, *Quantitative Analytical Studies in Epilepsy* pp. 493–509.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. and Vaughan, T. M. (2002a). Brain–computer interfaces for communication and control, *Clinical neurophysiology* **113**(6): 767–791.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. and Vaughan, T. M. (2002b). Brain-computer interfaces for communication and control, *Clinical Neurophysiology* **113**(6): 767 – 791.

Wolpaw, J. R., Loeb, G. E., Allison, B. Z., Donchin, E., do Nascimento, O. F., Heetderks, W. J., Nijboer, F., Shain, W. G. and Turner, J. N. (2006). Bci meeting 2005-workshop on signals

and recording methods, *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* **14**(2): 138–141.

Wolpaw, J. R. and McFarland, D. J. (1994). Multichannel eeg-based brain-computer communication, *Electroencephalography and clinical Neurophysiology* **90**(6): 444–449.

Wolpaw, J. R. and McFarland, D. J. (2004). 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.

Wolpaw, J. R., Ramoser, H., McFarland, D. J. and Pfurtscheller, G. (1998). Eeg-based communication: improved accuracy by response verification, *Rehabilitation Engineering, IEEE Transactions on* **6**(3): 326–333.

Woolsey, R. M. (2005). Fundamentals of neurologic disease, *The journal of spinal cord medicine* **28**(4): 348.

Yoo, S.-S., Kim, H., Filandrianos, E., Taghados, S. J. and Park, S. (2013). Non-invasive brain-to-brain interface (bbi): establishing functional links between two brains, *PloS one* **8**(4): e60410.

Zetterberg, L. H. (1969). Estimation of parameters for a linear difference equation with application to eeg analysis, *Mathematical Biosciences* **5**(3): 227–275.

Zhou, P.-b. (2012). *Numerical analysis of electromagnetic fields*, Springer Science & Business Media.