warwick.ac.uk/lib-publications

Connectivity Analysis from EEG Phase Synchronisation in Emotional BCI

by

Lorena Santamaria Covarrubias

A thesis submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy in Engineering

University of Warwick

Warwick Manufacturing Group

March, 2017

Contents

List of Figures

List of Tables

Nomenclature

LDA	Linear Discriminant Analysis
3-nn	Nearest algorithm k=3
ADHD	Attention Deficit Hyperactivity Disorder
ALS	Amyotrophic Lateral Sclerosys
BCI	Brain Computer Interfaces
BOLD	Blood Oxigenation Level Dependence
CBF	Cerebral Blood Flow
CPL	Characteristic Path Length
CSF	Cerebro-Spinal Fluid
CSP	Common Spatial Pattern
CWT	Continuous Wavelet Transform
DCM	Dynamic Causal Modelling
ECoG	Electrocorticography
EEG	Electroencephalogram
EP	Evoked Potential
ERD	Event Related De-synchronization
ERS	Event Related Synchronization
FDR	Fisher Discriminant Ratio
fMRI	functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
GCM	Granger Causal Modelling
GE	Global Efficiency
iEEG	intracranial electroencephalography
MEG	Magnetoencephalography
MI	Motor Imaginery
MI-BCI	Motor Imagery based Brain Computer Interface

MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MVAR models	Multivariate Autoregressive models
NIRS	Near Infrared Spectroscopy
NMR	Nuclear Magnetic Resonance
PET	Position Emission Tomography
PLV	Phase Locking Value
Q	Modularity
QDA	Quadratic Discriminant Analysis
RF	Radio Frequency
RID-Rihaczek	Reduced Interference Rihaczek
ROC	Receiver Operating Curve
SCPs	Slow Cortical Potentials
SPECT	Single Photon Emission Computed Tomorgaphy
SSEP	Steady-State Evoked Potentials
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
SVM	Support Vector Machine
T	Transitivity
Thinking L	Thinking in moving your left hand (MI task)
Thinking R	Thinking in moving your right hand (MI task)
TNr	True Negative Rate
TP	True Positive
TPr	True Positive Rate
VAR models	Vector Autoregresive models
WHO	World Health Organization

Acknowledgements

I would like to thank my supervisor Professor Christopher James for giving to me the opportunity of undertake my PhD and for supporting me throughout this period. Also, many thanks to James Amor for his help and friendly suggestions.

I would also like to take the opportunity to acknowledge my colleagues in my research groups, both of them, for their cooperation, help and suggestions; but overall for their great company, friendship and for the good moments in the office and outside. Thanks for helping me to make this PhD a reality.

I would like to thanks to all these people I have meet on my way here, that in one or another way, have made me the person who I am, supporting and encouraging me to pursue my dreams. Many people, many places and many countries, to all of them thank you for always being there for me.

Finally, I would like to express to my parents and all my family an enormous gratitude for always believing in me and for supporting me in all my crazy adventures. Finally, to Mike, who has always been there helping me in so many ways that I cannot express how grateful I am.

<div align="right">Thank you!</div>

To my partner, my family, my friends, my supervisor... and especially to Amanda

Declaration of authorship

I, Lorena Santamaria Covarrubias, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

 I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Parts of this thesis have been published by the author: [1], [2], [3], [4], [5].

Signed: . . . .

Date: . . . .

Abstract

A Brain Computer Interface (BCI) is a device that uses the brain activity of the user as an input to the system to select the desired output on a computer, giving the person a different pathway to establish communications with the surrounding environment. There are many types and uses of BCIs. They can be defined by which technique is used to record the brain activity of the user and which variety of stimuli is used to trigger a consistent response from the user, following the signal processing methodology selected to produce a response on the computer. Each one of the selected choices will determine the reliability and efficiency of the BCI system. However, even with this flexibility, the performance of BCI systems used for assistive technology or rehabilitation processes still remains behind other systems and the percentage of people unable to use one of these systems remains too high.

The main objective of this thesis is to improve the classification performance and reliability of the current electroencephalogram (EEG) based BCI systems. Firstly, a novel paradigm based on emotional faces is used with the aim of enhancing a stronger response from the user, therefore a higher amplitude of brain activity. Two types of emotional faces have been used during this work. Initially, emotional schematic faces or emoticons were used. Posteriorly, human emotional faces were introduced into the experiments. Additionally, the evolution of the phase synchronisation over time is studied to achieve a deeper understanding of the latent communication mechanisms of the different parts of the human brain. Wavelet families and their ability to retain temporal and frequency information simultaneously have been used to study the phase relationships between the EEG signals when a specific task is being performed. This study has led to the identification of a reduced number of discrete states with a quasi-stable phase synchronisation of the order of milliseconds, named synchrostates. Those synchrostates present switching patterns over time, clearly distinctive for each one of the tasks performed by the user. In order to establish a classification protocol the temporal stability of each task-specific synchrostate was studied by means of the synchronisation index and posteriorly translated into connectivity network maps based on graph theory. From this connectivity network, a series of connectivity metrics was obtained and used to feed a variety of classification algorithms. This process led to accuracies of 83% for a two-tasks classification problem and rose to a 93% averaged accuracy for a four- tasks problem.

Chapter 1

Introduction

1.1 Motivation

The human brain plays a central role in the control of the human body. It is the centre of the emotions, senses and is responsible for the principal biological and physical functions. However, damage to the brain, either through a physical trauma such as a driving accident or a disease (e.g. cardiovascular disorders, Amyotrophic Lateral Sclerosis (ALS) or brainstem stroke) can cause a wide range of physical, cognitive, and behavioural/emotional impairments that may be either temporary or permanent [6]. World wide statistics regarding brain injury are disheartening. According to the World Health Organization (WHO), brain injury is the leading cause of death and disability in children and young adults around the world. Brain injury is a public health concern that demands ongoing study, increased efforts to prevent occurrence, research to advance medical options and therapeutic interventions.

There are several areas where research can assist in the improvement of medical treatments and rehabilitation therapies. It can be from a psychological approach [7, 8], rehabilitation [9] or providing the brain with a new, non-muscular, communication and control channel [10, 11]. The science and technology of devices and systems responding to neural processes in the brain that generate motor movements and cognitive processes that modify the motor movements are called Brain Computer Interfaces (BCI) [12]. All other assistive technology methods depend on the brain's natural output pathways of peripheral nerves and muscles and take some of the outputs that the person still retains, using them to replace the missing functions. In contrast, BCIs provide the brain with completely new output routes, using the brain activity of the patient as an input to the system to select a desired output on the

computer [13]. The principal target population is those patients displaying a severe impairment in their motor activity, unable to communicate through normal muscle-bases pathways giving them the opportunity to control a robotic prosthetic, form complete sentences so as to communicate with the external world or move a wheelchair. In a nutshell, increasing their quality of life and offering them a limited but vital new form of communication.

BCI research has made considerable progress recently, particularly in the last decade, when BCI research has been extended to different arenas such as entertainment and marketing [14]. However, this methodology still presents some limitations. BCI studies generally take place in a highly controlled environment, either a laboratory or a hospital, where the BCI user assumes a specific position, in a place free of distractions, for a short amount of time and with a reduced number of devices interfering with the recorded signals. In addition, BCI displays an ineradicable variability due to the patient's tendency to perform better one day than others or even between one trial to the next [15]. Finally, BCI can function as an excellent communication and control instrument for many patients, but not for all users. Actually, 20% of BCI users are not proficient with a typical BCI system [16]. The phenomenon has been called "BCI Illiteracy" for some research groups [17, 18, 19]. This is due to a user's natural behaviour. While the functional divisions of the brain are common for all, there are individual variations in brain structure. Furthermore, it can be a consequence of the low amplitude brain signals that are undetectable to a particular neuroimaging methodology, as in an electroencephalogram (EEG) or because the participant's muscular artifacts are too numerous.

Extensive efforts have been made to overcome this problem through various mechanisms, improving technology for recording devices, new methodologies to refine signal processing, increase the training period for the BCI user or modifying the instructions that patients have to follow. Some of these options have functioned for some previously illiterate users. By contrast, some people still remain unable to manage a BCI system [20]. Our aim is find a combination of novel paradigms and new processing algorithms to develop more accurate BCI systems and also increase the range of people that can use them without extensive and extenuated training periods.

Over the last decade the number of studies has increased that are focused on the understanding of connectivity patterns, how the exchange of information happens, for normal and disrupted brain behaviour [21]. The study of connections within the brain has resulted in the conclusion that brain organisation is non-random [22]. For example, the absence of some connection patterns or

the disproportionate presence of them between specific brain regions can be a manifestation of some disorders such as autism or attention hyperactivity disorder (ADHD) [23]. Brain network science has also been employed in the study of maturation, epilepsy or schizophrenia among others [24, 25]. Connectivity can be inferred from a variety of techniques as functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG) among others. fMRI provides a high spatial resolution compared with EEG or MEG. By contrast, EEG and MEG offer a larger temporal resolution which is required to quantify the temporal evolution of the relationships between different areas of the brain [21].

The study of connectivity needs a mathematical framework; graph theory is the methodology that has been increasingly utilised in recent years [26]. Graph theory provides a way to not only capture the topology of the network under study, but also allows the researcher to quantify the multivariate relationships among these brain regions [21, 22].

Deeper understanding about how the different regions of the brain interact under different situations will lead to the researchers being successful in more personalised brain mapping, potentially leading to a more efficient BCI systems and lower BCI illiteracy rates.

1.2 Aims and objectives

The main aim of this thesis is to improve the performance of the current EEG-based BCI systems, primarily through the increment of the classification accuracy rates. Motivated by the fact that more practical and efficient systems can alleviate the situation of isolation of many patients with any type of brain injury, different components of BCI systems will be examined and consequently, novel variations developed to pursue this end.

The primary components of BCI systems are: the brain activity recording system (EEG in this particular case), the stimuli used to induce a specific response in the user, then a series of processing steps to finally reveal a small set of features that will be used to control an external device (e.g. a prosthetic arm, wheelchair or PC cursor). This study introduces novelties in two of these components: the stimuli used that is responsible for generating a specific response in the user named evoked potential and the signal processing component.

The focus of this work is finding quantitative metrics from the temporal

dynamics of the brain connectivity that can be used as features to control a wheelchair, speller or similar devices. In order to achieve this objective the specific stimuli and the adequate processing to calculate the connectivity dynamics are a fundamental prerequisite in order to achieve higher accuracy rates, hence more efficient BCI systems.

As a novel stimulus, the use of emotional faces is proposed for the first time as a base for a motor imaginary BCI system. Emotional faces have been widely used in other research. For instance, psychological experiments [27, 28], to classify Parkinson's disease from healthy control patients [29] or Autism Spectrum Disorder patients from typically developed children [30]. However, emotional faces have never been used as stimuli in a motor imagery BCI. Subsequently, a set of complex signal processing algorithms has been employed to accurately determine and characterise the connectivity between different regions of the brain giving further insight into the integration process within the brain and when a specific stimulus is utilised. The algorithm, involving wavelet transformation, pattern recognition techniques and graph theory, will be detailed in the following chapters.

1.3   Challenges

The choice of technique to record the brain activity has a series of associated difficulties. Specifically, in this particular study, the election of the EEG as the recording method has an associated lack of spatial accuracy, but is compensated by its temporal resolution at the millisecond level. Furthermore, the low amplitude of brain signals due to the attenuation occurring in reaching the scalp, the presence of artifacts or the volume conduction problem force the scientist to interpret these signals very carefully.

Another complication added to the study is the complexity of the human brain and its interconnections; these, in conjunction with our restricted knowledge of the physiology behind the brain, makes it even more challenging to map its connectivity.

A major hinderance to this project is the need to finding volunteers to participate in the designed experiments aiming to achieve enough EEG registers to be a statistically significant study.

1.4   Contributions

The main contributions of this research work are listed below:

- Validate the use of emotional faces, both schematic and human, as stimuli for a motor imagery EEG based BCI system.

- Establish the presence of quasi-stable states of the order of milliseconds, named synchrostates, during the execution of motor imagery tasks. These states are based on the instantaneous phase difference between EEG electrodes for a specific band of interest.

- Show that these states are consistent for both the averaged population and for each individual belonging to the cohort.

- Show that the temporal switching transition of these synchrostates is different for each task, reporting that the synchrostate phenomenon is task-specific. Therefore, they can be used for classification purposes.

- Validate the use of the graph theory metrics from the synchrostates as features for classification between different motor imagery tasks with high classification performance.

1.5   "Newton fund" programme

During the research project period I was awarded a Newton Fund programme scholarship sponsored by the British Council. The aims of this programme are among others, to build stronger links between UK and China research centres, develop individual capacity through an international training and create a deeper understanding of both cultures.

The awarded scholarship had a duration of four months and the host institution was the Department at Bioengineering of Tsinghua University Medical School. The collaborative project between both universities has, as the main purpose, the study of new classification algorithms of different tasks from high density intracranial EEG (iEEG). Further details regarding this project are explained in Appendix C.

1.6   Outline

Following this introductory chapter, a brief overview of the literature review

performed is carried out in chapter 2. It includes biological aspects of brain functioning, a summary of the main recording techniques of brain activity and an overview of BCI systems and their classification. It also reviews the current state of the art of EEG signal processing techniques and their disadvantages that motivated the introduction of the proposed methodology. Furthermore, it explains the motivation behind the use of emotional faces as stimuli. Chapter 3 covers the proposed algorithm to study the phase synchronisation in the EEG signals which leads to the phenomena called synchrostates. Their stability over time is studied in chapter 4 using a synchronisation index as a measure. Furthermore this chapter also explains how connectivity metrics can be derived from the synchrostates obtaining quantitative measures of brain networks. Chapter 5 deals with the details of the classification process between different motor imagery tasks using the complex network metrics derived from synchrostates. Chapter 6 presents a validation of the method described in chapters 3 to 5 for a more complex system using human emotional faces as stimuli and four different tasks to classify. Finally, the conclusions drawn from this thesis, how well the objectives were met and future work are detailed in chapter 7. Following the main chapters the relevant appendices and references are listed.

1.7   Publications

Accepted papers for publication:

1. L. Santamaria, C. James, *"Use of graph metrics to classify motor imagery based BCI"*, The International Conference for Students on Applied Engineering (ICSAE), Newcastle, United Kingdom, October 2016, pp 469-474. *(Awarded with the best poster prize in the category of bioengineering)*.

2. L. Santamaria, C. James, *" Classification in emotional BCI using phase information from the EEG"*, 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) , Florida, USA, pp. 371 - 374, August 2016.

3. L. Santamaria, C. James, *"Classification in Emotional EEG- based BCI Using Connectivity Measures"*, WMG Doctoral Research and Innovation Conference, Coventry, United Kingdom, pp. 14, June 2016.

4. L. Santamaria, C. James, "*Single Phase Locking Value Classification Algorithm for Motor Imagery based BCI*", 2nd WIN Conference, Coventry, United Kingdom, pp. 57, January 2016.

5. L. Santamaria, C. James, "*Electroencephalogram analysis in emotional BCI*", 8th IEEE EMBS UK and RI Postgraduate Conference in Biomedical Engineering and Medical Physics (PGBiomed), pp. 45-46, July 2014.

Papers under preparation or under review:

1. L. Santamaria and C. James, "*Existence of Synchrostates during Motor Imagery Tasks.*" **(in preparation)**

2. L. Santamaria and C. James, "*Using brain connectivity metrics from synchrostates to perform motor imagery discrimination in EEG based BCI systems.*" **(in preparation)**

Chapter 2

Literature Review

This chapter introduces a theoretical background regarding the origins and physiology behind the electroencephalography (EEG) signals. This introduction to the EEG covers the basic components of the brain and its functions, including details of the main techniques currently used to record brain activity. A discussion of the state of the art of brain connectivity techniques is also included. Among the existing mechanisms to register brain activity, EEG is one of the most accessible and affordable options, which makes it perfectly suitable for the purpose of this work in classifying mental states and tasks. The study and analysis of EEG signals can be time or frequency orientated. Some commonly utilised methodologies that are used in both domains are described within this chapter.

This chapter focuses on the use of phase synchrony to study the interaction of different parts of the brain; the currently available methods used to study these interconnections and also highlights the fact that these techniques are unable to preserve the information of the temporal evolution of phase. This gap of knowledge inspired this research on phase differences along brain areas across the time domain.

The structure of the chapter is as follows: Section 2.1 is a brief introduction to the brain and its components followed by a list of the principal techniques to measure brain activity in Section 2.2. Section 2.3 focuses on EEG as the technique used during this research to record brain activity. Section 2.4 includes information about different BCI approaches emphasising MI-based BCI. Also, in this section, a novel paradigm is described, based on schematic emotional faces and is used during this work as a stimulus to enhance a stronger response in participants than established paradigms already used for this purpose. Section 2.5 includes information regarding brain connectivity performing

a recount of some recent findings and justifies our approach of using source level connectivity analysis. Finally, Section 2.6 briefly introduces some theoretical concepts regarding complex brain networks based on graph theory as a concept related to the research conducted on phase synchrony.

2.1   The brain: introduction

The human brain, located in the head and protected by the skull, is the main organ of the human nervous system. This approximately three-pound weight organ is the centre of emotions, interpreter of the senses and controller of the main biological and physical functions [31, 32, 33]. The brain and the rest of the human nervous system are composed of many different types of cells. However, the primary working functional unit is the neuron. Neurons are electrically excitable cells that transmit information by electro-chemical signalling. Each neuron can be connected to up to 10,000 of its neighbours transmitting information to each other via the estimated 100 trillion synaptic connections existing in our brain [34].

A typical neuron needs both electrical and chemical stimulation for the excitation processes due to the voltage gradient characterising the neuron membranes. In normal conditions the neuron remains in the resting potential state, meaning that the interior of the neuron contains a greater number of negatively charged ions than the area outside of the cell does. When the neuron is triggered by an electrical signal, the membrane of the neuron opens its gates allowing the positive ions to pass through it, generating a change in voltage and becoming temporarily positively charged. This generated impulse is called an *action potential* [35]. This pulse travels along the neuron's axon by creating similar voltage changes from segment to segment within the axon. The neural signals not only travel via electrical changes along the neuron, but they can also communicate across neurons via chemical transmission. This is possible thanks to a special connection called a *synapse* (see Figure 2.1). This connection is a gap between the end of one neuron axon and the dendrite of one neuron nearby [36]. The main function of this gap is to allow neurotransmitters to jump to the next neuron, hence spreading the information from neuron to neuron [35]. A neurotransmitter is a special chemical that relays signals across the synapses between neurons. These chemicals travel across the synaptic space between the end of an axon to the dendrites of the receiving neuron starting the process all over again in the receiving neuron. Figure 2.1 exemplifies the movements of neurotransmitters between different neurons during the synapse. It is notice-

able that different neurons release different neurotransmitters and in the same manner, each dendrite is particularly sensitive to a neurotransmitter [35].
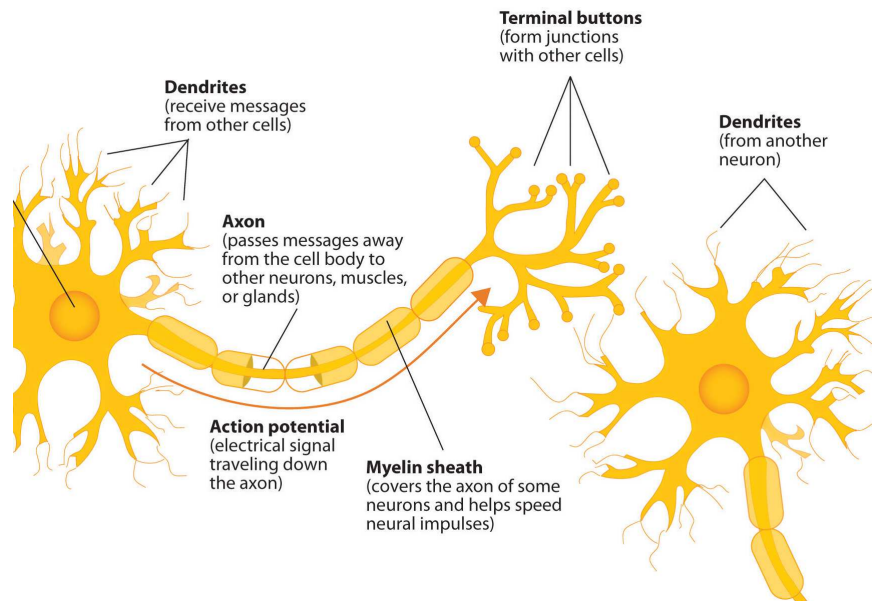
(a) Components of a neuron and how the electrical signal or action potential travel along the neuron cells.
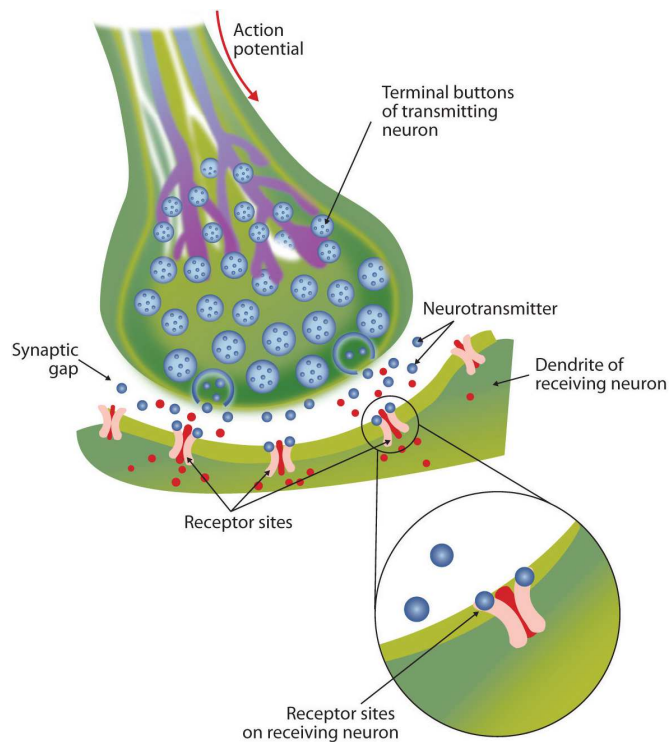
(b) Detail of the synapse showing how the neurotransmitters pass across two neuron cells by the named synaptic gap.

Figure 2.1: Neuron structure and functioning of the action potential and synapse. Pictures taken from [35].

The transmission of information between neurons in the nervous system is based on changes in the membrane potential. The exchange of ions in the neuron membrane is governed by the action potential that produces electrical activity. This electrical activity can be registered, measured and displayed and is known as *brain activity* [35].

An electroencephalogram is the time series of data of those electrical signals generated by intercommunication across neurons and their connections recorded by special electrodes [37]. Those electrodes can be placed internally, directly on the brain surface, or externally when placed on the scalp. All the possible electrode locations share the same objective, measuring the potentials produced by synapse connection between neurons during the excitation. The array of electrodes placed on the scalp measure the potential difference between dipoles formed by axons and dendrites and a selected reference [38].

At the same time that the electrical currents are initiated by the action potential of one neuron an associated electromagnetic field is generated. Similar to the electrical current, the orientation of the magnetic field pattern reflects the direction of the action potential [39]. This magnetic field can be also captured by magnetoencephalography (MEG) and can also be used to study brain activity. Attending their different sensitivities to source orientation and location it can be said that both techniques, EEG and MEG, complement each other. These techniques to measure brain activity are explained in the following section.

In addition to these electrical and magnetic currents there is another series of physiological and functional effects related to the increase or decrease in neural activity. These phenomena can also be registered and they are known as indirect measures of the brain activity. They are based on the study of changes in blood flow instead of directly measuring the electrical or magnetic current across neurons when information is transmitted. In particular, functional magnetic resonance imaging (fMRI) is a technique based on the level of blood oxygenation [40]. Specifically, it studies the magnetic properties of the haemoglobin.

The increase in blood flow due to neuronal activity is also accompanied by a gain in oxy-haemoglobin concentration in a particular activated region of the brain. However, although there is an increase in oxygen consumption, there is an excess in oxygen supply causing the ratio between oxy/deoxy-haemoglobin tissue concentration to rise which can be registered, measured and translated into 3D images [41]. What remains vague is the understanding of how tissue oxygenation is related to neuronal activity.

2.2   Measuring Brain Activity

Being able to understand brain functions is one of the prominent challenges in neuroscience. One of the aspects of this arduous research is exploring how the brain and nervous system control the enormous quantity of processes involved with daily functions such as cognition, perception and motor actions. Modern technology has provided the scientific community with several techniques to measure brain activity; each one exploiting different principles and recording various classes of processes. For this reason, the selection of a technique that surpasses the rest will depend on several criteria:

- The application or aim of the study.

- Spatial resolution needed.

- Temporal resolution required.

- Budget.

The variety of mechanisms with which to measure brain activity can be divided into two major groups. The first of these is those directly measuring the electrical activity of the brain and associated with the variations of the post-synaptic potentials as mentioned in the previous section. The main methodologies within this group are EEG and MEG. The main advantage of this group is its temporal resolution, in the region of milliseconds. Spatial resolution, conversely, is rather poor as only a few hundred simultaneous data positions can be acquired.

The second group of techniques includes all methods for indirectly recording neuronal activity. Some examples of these procedures are fMRI, position emission tomography (PET) or near infrared spectroscopy (NIRS). Its main feature and antagonistic with the first group, is the high spatial resolution. Attempts to use conventional fMRI techniques for high temporal resolution needs fail overall due to the technical limitations of the scanner. In addition there is a limitation in the haemodynamic response to prolonged episodes of stimulation [42].

Table 2.1 shows a comparison of the different methods with which to measure neural activity of the brain. The following section describes briefly the basic principles of each one of the techniques mentioned.

Table 2.1: List of the principal techniques of measuring brain activity. Including their main features, advantages and disadvantages [16].

Technique	Physical Property	Direct / Indirect measure	Spatial resolution	Temporal resolution	Advantages	Disadvantages
PET	Blood flow	Indirect	good	low	signal not affected by small movements of the subject	expensive, need of inject tracers
SPECT	Blood flow	Indirect	good	low	slightly less expensive than PET	expensive need of tracers
fMRI	Blood flow	Indirect	excellent (~1mm to 1cm)	relatively good (in the order of seconds)	widely available, predominant neuroimaging technique	expensive, need of an expert to manage the equipment
MEG	Magnetic Potential	Direct	relatively low (~ 1cm)	good (in the order of milliseconds)	enables much deeper imaging and is more sensitive than EEG as skull is almost transparent to magnetic waves	expensive equipment due to need of superconductivity
EEG	Electrical Potential	Direct	low (in the order of centimetres)	high (tens to hundreds of milliseconds),	wearable, affordable, easy to use, widely available	requires careful placement of electrodes directly on scalp to avoid noise

2.2.1   Indirect measures of neural activity

The use of neuroimaging techniques to gain unobtrusive access to the brain are relatively recent and has continued to thrive from technical and methodological standpoints. Lately, as a consequence of the advancements within the area and a reduction of associated costs, the number of functional magnetic resonance imaging (fMRI) scanners in hospitals around the world has increased enormously; as a result it has become the more popular modality to approach the brain in action. fMRI is the neuroimaging technique more commonly used, but not the only one. Others, such as positron emission tomography (PET), single photon emission computed tomography (SPECT) or the predecessor of fMRI, magnetic resonance imaging (MRI) are widely used as medical and research tools.

PET

PET is a nuclear imaging technique that uses brain stimulation which leads to an increase in cerebral blood flow (CBF) which in turn reflects the activation of a population of neurons. PET studies blood flow and metabolic activity in the brain helping to visualise biological changes taking place. The concept of radioactive tracers was introduced in the late 50s by David E. Kuhl, Luke Chapman and Roy Edwards and this technique was first applied in humans in 1963 [43].
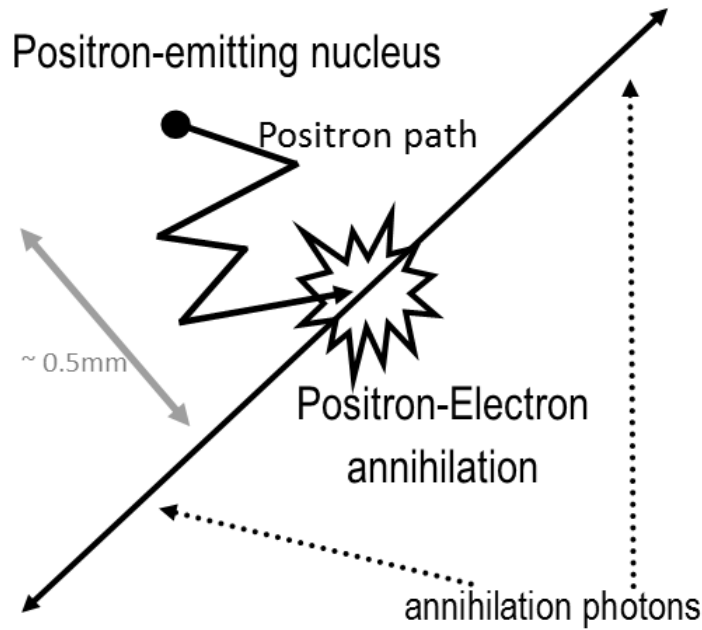
Figure 2.2: Basic principle of PET: collision between a positron and electrode where both particles annihilate each other. The result of the collision is the emission of the two high-energy protons that will be detected by the PET scan sensors. Adapted figure from [44].

In a PET scan, a small amount of radioactive tracer is injected into a vein, arriving in the brain a few seconds later. The tracer is usually a substance, such as a type of sugar like glucose, that can be metabolised by cells in the body and is labelled with a radioactive isotope. This particle is highly unstable and after being injected into the bloodstream starts to decay after a few seconds. This means the isotope becomes less radioactive over time. During this process it emits positrons from the nucleus that annihilate on contact with electrons after travelling a short distance within the body. The final result of this collision is the emission of two high-energy photons travelling in nearly opposite directions (Figure 2.2).

A PET scan consists of a set of detectors that surround the object to be imaged. When the two high-energy photons or gamma rays leave the subject's body they are sensed by two detectors positioned 180 degrees from each other in the scanner. The scan is able to convert these rays into an electrical signal than can be fed to subsequent electronics. Finally, using standard tomographic techniques, the computer output is a three dimensional image of a volume from the brain [45, 46].

The concept is that blood is more concentrated in activated brain areas than in the inactivated ones, meaning that the scanner will detect more gamma rays coming from those parts that are working more. This is translated into the

image as a range of colours depending on the activity shown in these regions. Reddish colours indicate high brain activity and by contrast, bluish colours means little to no brain activity as can be seen in figure 2.3.
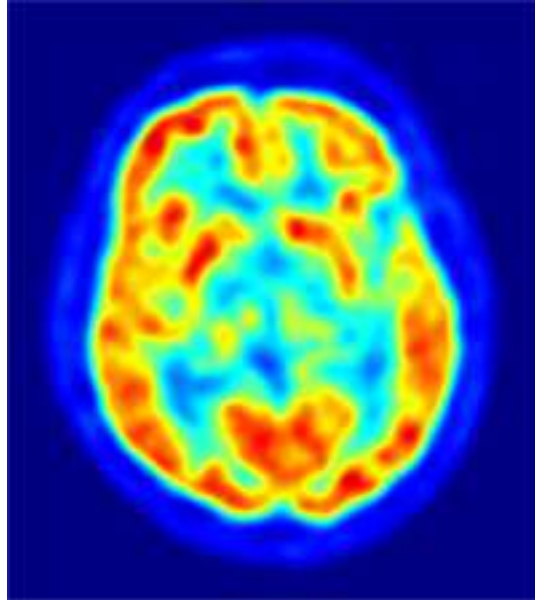
Figure 2.3: Example of a PET scan image of a human brain. The reddish colours mean high brain activity and the bluish colours are associated with those parts of the brain with little to no brain activity. (Source: public domain).

One of the advantages of PET, unlike other imaging tests, is its ability to detect irregularities in body function caused by diseases which often occur before anatomical changes can be observed. In addition, the quality of the PET scan is not affected by small movements of the subject. Consequently, the subject does not need to remain as still as they would for a MRI or EEG recording. On the other hand, its limited spatial resolution can result in images that are not very clear. Furthermore the use of radioactive tracers always involves some risk.

SPECT

SPECT is based on the same principles as PET to produce its images. However, important differences in instrumentation and radiochemistry are dictated by the physics of photon emission. In a SPECT scan the photons emitted from the radio-tracer in the body are detected as independent events. Detections of photons are performed by a special camera, namely a gamma camera; comprising one or multiple detector heads. In front of the camera crystal a collimator is placed, featuring a large number of holes allowing photons to enter only in the direction parallel to the holes. Photons arriving at the collimator in other

directions have a high probability of being absorbed by the material of the collimator [47]. Those photons that finally reach the crystal will interact with it resulting in an ionisation of the crystal that can be translated to an electric pulse detectable by the electronics of the gamma camera.
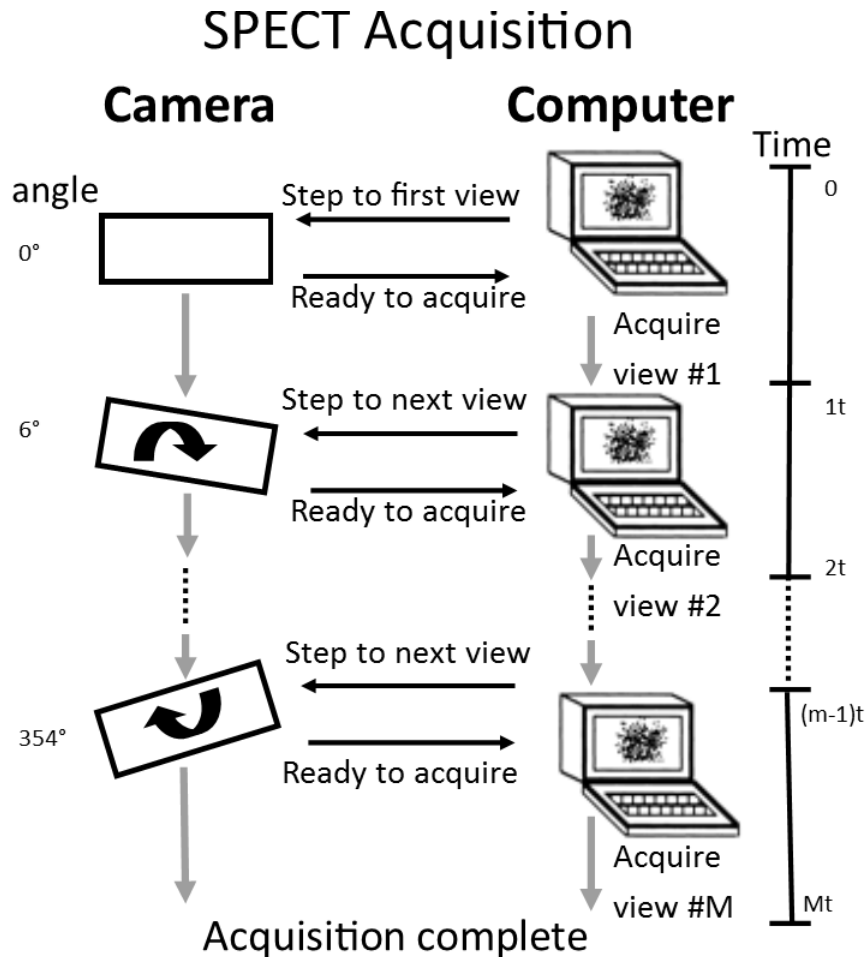
Figure 2.4: Schematic diagram of SPECT scanner data acquisition. For each projection the camera acquires an image at that angle and at that specific time. Afterwards, all the collected images are used to reconstruct the three-dimensional object. Image adapted from [48].

A SPECT scanner performs a series of planar projections acquired by rotating the gamma camera at different angles around the patient. At each of the projection angles one static image is acquired for that angle for a specific time as shown in figure 2.4. Afterwards, all of the images stored in a matrix format are used by the computer to reconstruct the 3D image of the activity distribution within the brain. This is done by the utilisation of mathematical image reconstruction algorithms [48].

SPECT has inferior image quality than PET as the spatial resolution is

limited by the scan technology. By contrast, SPECT produces a 3D image with
an improved image contrast than with PET scanners. In addition, SPECT
radioisotopes have longer lives (from hours to days), opening the possibility of
investigating relatively longer-lasting tasks as walking [49].

MRI

In MRI the primary element used to generate almost all clinical images
comes from the nuclei of the tissue's hydrogen atoms. A traditional MRI
scanner contains a very strong electromagnet that can generate a powerful
magnetic field inside the scanner.

Hydrogen nuclei consist of an odd number of protons that carry a positive
electrical charge. The presence of charged particles spinning around produces
a small magnetic moment. These proton moments are generally random in
orientation. However, when they are placed inside an external magnetic field,
as in an MRI scanner, many of the free hydrogen nuclei align themselves in
the same (denominated parallel) or opposite (antiparallel) direction than the
external source. In order to flip over, the protons have to absorb some energy
from the radio waves coming from the MRI scan. This particular movement
of the group of protons is called precession (Figure 2.5). The speed of the
precession, or the number of times the protons precess per second is named
the frequency of Lamor and it is proportional to the external applied magnetic
field strength, represented in Figure 2.5 as ω_0 and B_0 respectively.

In an MRI study the patient essentially becomes a magnet with a magnetic
vector aligned at a specific angle to the external magnetic field. The participant
is placed in a cylindrical coil that surrounds the head. Following the application
of a strong static magnetic field, the brain's hydrogen nuclei consequently align
with the magnetic field, resulting in a magnetic moment parallel to the applied
field. The magnetic force of the patient cannot be measured as it is in the
same direction as the external field. For this reason, a radio-frequency (RF)
pulse with a very specific frequency and duration of milliseconds is applied
perpendicular to the external field.

As soon as the external RF pulse is switched off, the protons start to realign
themselves, returning a lower energy state and the protons relax. Protons con-
tinue this process until they come back to their original state, releasing the
energy they have absorbed during the precession process. This generated en-
ergy is named a MR (magnetic resonance) signal and can be measured by the
electromagnetic detectors of the MRI scan. In addition the protons start to
precess out of frequency. As a result of this process the longitudinal magnetisa-

tion gradually increases (named *T1 recovery*) and the transverse magnetisation does the opposite, by gradually decreasing (a process called *T2 decay*). The rate at which *T1* and *T2* occur vary depending on the type of tissue. The duration, repetition, timing and amplitude over which the RF pulse sequences are varied to produce a signal which can be analysed in different ways in order to "weight" the image. The resulting image has contrast if it contains areas of high signal intensity (white), low signal intensity (black) and intermediate signal intensity (grey). The intensity of the signal generated by a tissue depends on its water and fat content, the proton density and the presence of any blood flow [50].
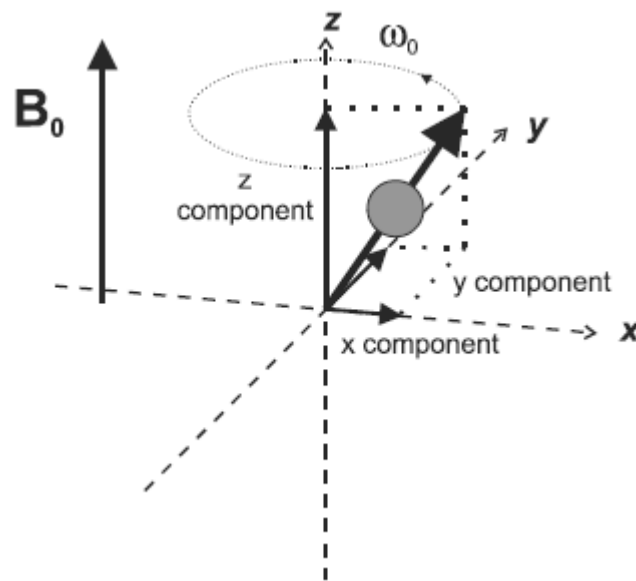
Figure 2.5: MRI principles: effect of an external magnetic field B_o. Inside a magnetic field a proton precesses in a parallel direction to the field B_o with a precession speed ω_0.

The magnetic resonance signal generated from the sum of the magnetisation vector is termed *free induction decay*. Generally this signal is not measured in MRI; instead it is more common practice to generate the magnetic resonance signal in the form of echoes. In order to produce a 3D image the echoes must be recorded for each dimension using three separate magnetic field gradients: a slice-selection gradient, a phase-encoding gradient and a frequency-encoding gradient. Data collected from the three axis are reconverted into a volume image using mathematical techniques such as a 2D Fourier Transform and spatial encoding. The majority of standard MRI examinations take 20-30 minutes to complete, with each sequence of echoes lasting around 5 minutes each. Additionally an extra 10-15 minutes are needed in order to re-construct the 3D image [50].

The advantages of MRI compared to PET or SPECT are clear: lower cost, higher spatial resolution and no need for radioactive isotopes. There are however some disadvantages. For instance, it is not suitable for patients with metal implants in their body or for those suffering from claustrophobia. Additionally, the patient has to remain as still as possible during the recording process in contrast to recordings taken by PET scanners.

fMRI

fMRI is a widely used technique to study brain function. This methodology began with the discovery of nuclear magnetic resonance (NMR) followed by MRI [40]. However, it was not until the early 1990s when its potential for neuroimaging was discovered. Essentially, MRI and fMRI differ from each other in that MRI is used to produce structural images of subject's brains useful for anatomical and morphometric studies while a fMRI views the metabolic function. The functional methodology calculates the levels of oxygen in the blood in the brain. By contrast, MRI studies water molecule's hydrogen nuclei. Due to these differences, the features of both techniques are distinctive. For instance, MRI views the difference between tissue types at high resolution with respect to space. On the other hand, fMRI views those differences with respect to time. In other words, MRI has a high spatial resolution whereas fMRI has better temporal resolution.

The fMRI method exploits magnetic differences between oxygenated and de-oxygenated blood called blood oxygenation level dependence (BOLD). Basically, haemoglobin in the blood becomes strongly paramagnetic in its de-oxygenated state, therefore it can be used as a natural contrast agent, eradicating the need to inject a tracer. In this way, highly oxygenated brain regions - activated areas - produce a larger magnetic resonance signal than lower oxygen concentration regions - low to no activity - [51].

The temporal evolution of the BOLD under the presence of a brief stimulus is a dynamic process that can be represented by mathematical models. After the stimulus onset the BOLD signal presents an initial dip linked to an increase in deoxy-haemoglobin concentration. Afterwards, the BOLD signal rises considerably as the ratio of oxy/deoxy-haemoglobin increases. The increment is directly proportional to the neural activity performed by the subject [41]. If the stimulus lasts for long enough, the BOLD signal will reach a plateau otherwise the signal will return to the original baseline when the stimulus is eliminated.

Consequently, the BOLD effect can be used to detect the increasing neural

activity at the moment that a subject is performing a particular task. A typical use of fMRI for the detection of a task is having a person inside the scanner performing a series of cognitive tasks (paradigms). At the same time, a set of images of the brain are collected, normally every 2 or 3 seconds which accumulates an enormous quantity of brain volumes by the end of the scan. The signal intensity of each pixel can be associated with a specific task using statistical analysis [41]. Several approaches have been proposed for the analysis of fMRI data, with the aim of producing an image able to identify those regions, showing a significant change in response to a specific task. One of the simplest methods for a two state fMRI experiment, is a simple subtraction of the averaged images acquired for the task, minus the averaged images for when the task is 'off'. Using this method, the null hypothesis cannot be tested, so instead of a subtraction, it is more common to use a Student's t-test. This method weights the difference in means, giving larger t-test scores to large differences [52]. Other more complex methods used in fMRI analysis are based on correlation techniques or the widely used general linear model [53]. Figure 2.6 represents a scheme of a typical fMRI experiment and the associated BOLD signal evolution.
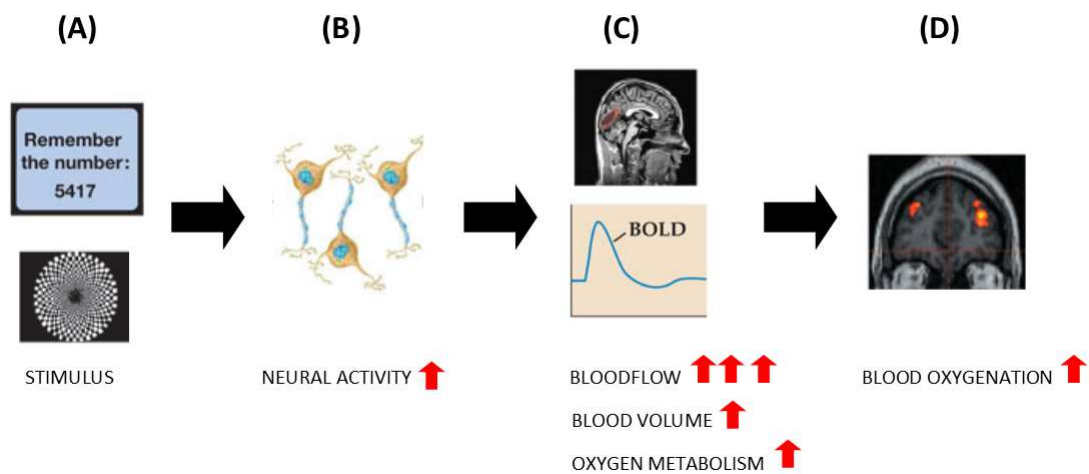
Figure 2.6: Change in a BOLD signal during a fMRI experiment. (A) A stimulus is presented to the participant, triggering the neural activity. (B) The increment of the neural activity is followed by a rise in blood flow, blood volume and oxygen. (C) The combination of these events alters the level of deoxyhemoglobina, which affects the MR signal (D). [41, 54].

fMRI has become a predominant technique in the field of neuroimaging research as it provides an excellent spatial resolution (average resolution of 3-4 mms) and relatively good temporal resolution (in the order of seconds) [55]. The fact that it is non-invasive and the use of haemoglobin as a natural tracer are also factors that have contributed to its standardisation.

2.2.2   Direct measures of neural activity

Direct measures of brain activity such as EEG or MEG present, in general, lower spatial resolution (in the order of cms) than the techniques explained in the previous section. On the other hand, they allow a higher temporal resolution (in the order of tens of ms) which directly measures the brain function and generally speaking, they are easier to use. The technique to select depends of the needs and the resources of the research as all of them have advantages and disadvantages over each other. For the present study, both temporal and spatial resolution would be highly favourable. Furthermore the system should be economically viable and with the possibility of being portable for real-time applications. For this reason EEG was the technology chosen to perform the brain activity recording across this study. A further explanation of this procedure is developed in the next sections of this chapter.

MEG

In the human body there are several electric currents of a diverse nature but all of them induce electromagnetic fields. Some of those electrophysiological currents are well-known as they are strong and easy to measure. For example, the ones induced by the heart or the muscles. On the other hand, other organs from the human body also produce ionic current flows of less intensity [56]. This is true in the case of the brain, where the neurons are stronger generators of ionic flows as mentioned at section 2.1.

When the neurons are activated synchronously as a consequence of a stimulus, electrical currents are generated inducing magnetic fields (see section 2.1 for more detailed information about the origin of this magnetic field). Those magnetic fields generated by the exchange of information between neurons pass almost unaffected through brain tissue and the skull. This is due to the permeability of biological tissues which is almost equal to empty space [57]. The magnetic field, however, diminishes with distance, resulting in an extremely small signal reaching the scalp. This handicap has been solved with technology using sensors based on superconductivity [58]. These superconducting devices are placed as close as possible to the subject's head and are normally housed in a helmet-shaped container for ease of use. Magnetic field changes are detected by these special sensors and transformed into voltage changes.

The use of these remarkably efficient detectors requires an extremely cold environment; cooling at -269°C, which is achieved using liquid helium. About 70 litres of liquid helium are necessary on a weekly basis to keep the system at

optimum performance. Thermal isolation is obviously a challenge in terms of
security and makes the system immobile. In addition, the extreme sensibility
of the sensors makes the device very sensitive to any other magnetic fields
nearby. Consequently, MEG scanners are typically placed in magnetically
shielded rooms.

There is an added problem of how to compute what area of the brain is
active. By analysing the spatial distributions of magnetic fields, it is possible to
estimate the intracranial localisation of the generator source and superimpose
it on an MRI scan. The steps to achieve a MEG image are schematically
described in figure 2.7.
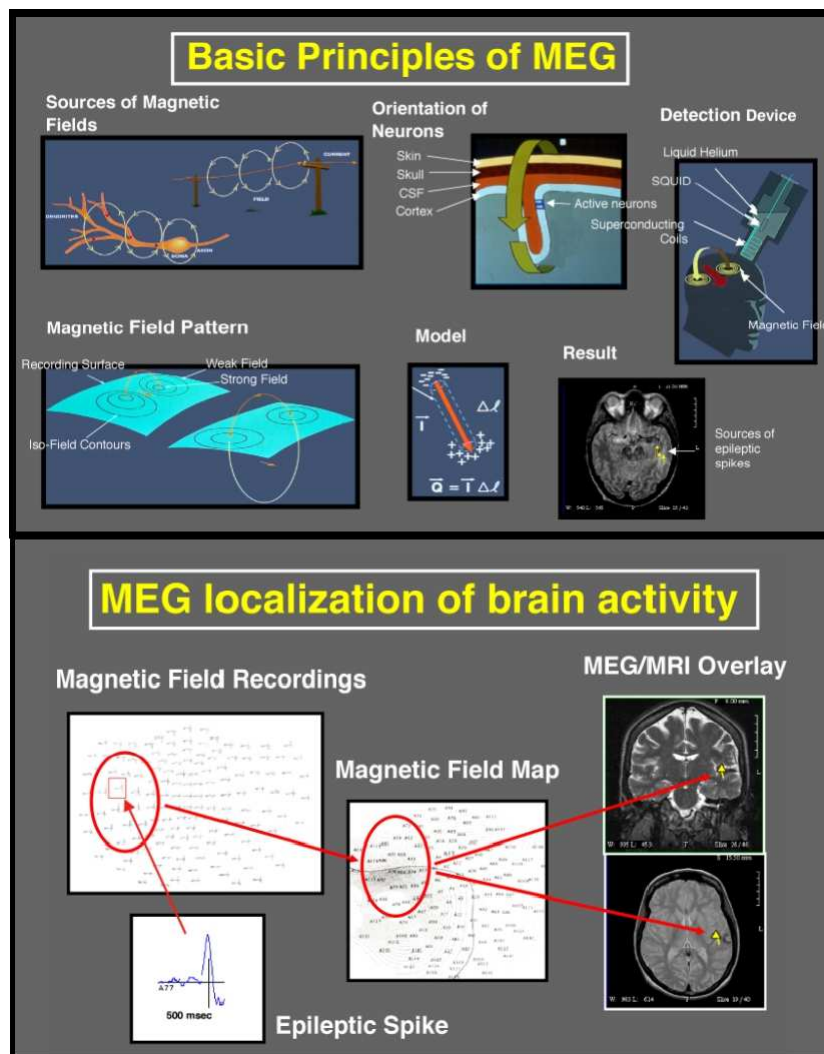
Figure 2.7: Sequence of steps to localise sources of brain activity using MEG [58].

MEG has better temporal resolution than indirect measurement techniques
and has plenty of advantages in a clinical setting. It provides high reliability

and greater accuracy of the estimated dipoles as it is not affected by volume conduction. Therefore, source estimation is easier than is found with EEG. For this reason, it is widely used when there is a need to identify and localise specific affected areas of the brain. For example, the mapping out of epileptic activity areas -one the most successful clinical applications of MEG - or when determining the dominant language area is needed [59]. The main disadvantage of MEG is the elevated cost associated with the instrumentation needed.

In addition, it possesses some advantages over other direct techniques such as EEG. For example:

- the magnetic fields are not affected by tissue conductivity as is the case with EEG

- subject preparation is reduced

- the measures are absolute and not dependent on the choice of a reference

- better spatial resolution of the source location (2-3 mm) than EEG (4-7mm)

EEG

The EEG is the recoding of the brain's electrical activity. As mentioned in section 2.1, neurons when activated, produce ionic currents at the level of cellular membranes (refer to Figure 2.1). It can be differentiated by two types of activation: one really fast depolarisation of the neuronal membranes inducing an action potential and slow changes, originated by synaptic activation. This process generates an impulse that can be propagated along axons and dendrites without being attenuated in its amplitude. This signal can be recorded when impacting with the electrodes placed on the head.

They are several procedures to record the electrical activity of the brain. They can be divided into two subgroups, invasive and non-invasive techniques. Invasive methodologies include the electrocorticography (ECoG). The term invasive is due to the necessity of a form of surgery, called a craniotomy, in order to place the electrodes directly into the brain cavity. Within the brain cavity the electrodes can be placed directly on the brain tissue (named parenchymal monitoring), below the layers covering the brain (subdural monitoring) or into one out of four ventricles (intra-ventricular monitoring) [60]. In this study the variation of EEG adopted to record brain activity is one that places the

electrodes on the surface of the subject's scalp. This will be explained in detail in the section 2.3.

ECoG also known as intracranial electroencephalography (iEEG) is based on the use of electrodes directly from the exposed surface of the brain to record its electrical activity. The standard procedure is the collocation of 16 electrodes placed in a grid, but this number can vary from 4 to 256 depending on the application [61]. During ECoG, electrical stimulation is frequently used to map critical areas such as the area of epileptic seizure onset. This stimulus is an electrical pulse applied during 1 to 5 seconds with an intensity of 0.5 to 2 mA and a voltage of 1 to 15V [62]. Among ECoG clinical applications the main one is in the treatment and detection of epilepsy, but also is used in research applications such as BCI systems [63].

2.3 Electroencephalography: background

The existence of electrical activity in the brain was discovered more than a century ago by Richard Carton [64]. However, it was not until the early 1920s when EEG was recorded from the human scalp for the first time [65]. Nowadays, EEG has become the most prevalent method for recording brain activity for BCI systems.

Scalp EEG recording displays the difference in electrical potentials between two different sites on the head, superimposing the cerebral cortex that is closer to the recording electrode. The problem of the extremely low amplitude values of the signals attenuated by the several brain layers which they have to cross until reaching the scalp is solved by the use of amplifiers. Modern technologies allow on-line filtering of the signals and other controls to regulate the signal output. Furthermore, data displays that follow acquisition, offer a wide range of options to represent the data for EEG interpretation. Figures 2.8 and 2.9 illustrate some types of EEG recorders available on the market. Figure 2.8 shows a simple Bluetooth based system with two channels plus another two channels for ground and reference. Figure 2.9 shows a more complex system of up to 256 channels composed by an EEG cap (A) and EEG amplifier (B).

2.3.1 EEG electrodes

Placement of the electrodes has been standardised by using a 10-20 system that uses anatomical landmarks on the skull. The name is based on the percentages used to determine the electrode installation. A total measure is divided into 10 or 20 percent segments. This system uses the distance between the subject's nasion and inion as a reference for longitudinal measurements, and the distance

between the subject's central coronal plane, the distance between both ears, for lateral measurements as is explained in figure 2.10. Nasion is the intersection of the frontal bone with the two nasal bones. It can be easily detected as its location is the depressed area between the eyes. Inion is the most prominent projection of the protuberance located on the lower rear of the human skull.

Figure 2.8: Example of EEG portable system to record EEG with 2 channels system plus ground and reference channels.

Each site has a name, a letter identifying the lobe and a number to identify the hemisphere. Even numbers correspond to the right hemisphere and odd numbers to the left hemisphere. The designations; F_p (frontopolar), F (frontal), T (temporal), O (occipital), C (central), and P (parietal) are utilised in the 10−20 system as shown in figure 2.10 [66].

Furthermore the letter 'z' makes reference to the central channels. For example, the name C_z corresponds to the position at 50% of the nasio-inion distance and at 50% of the distance between pre-auricular points. This means it is the exact centre point of the scalp. The letter C indicates "central" and 'z' makes reference to the 0% lateral offset from the Central coronal line.

Modern recording systems provide a cap where the electrode locations are already predetermined for ease of use as shown in figure 2.9. These modern systems require an increased number of electrodes and as a consequence they are placed on the 10-10 system, meaning that the distance between them is reduced to 10%.

(A)
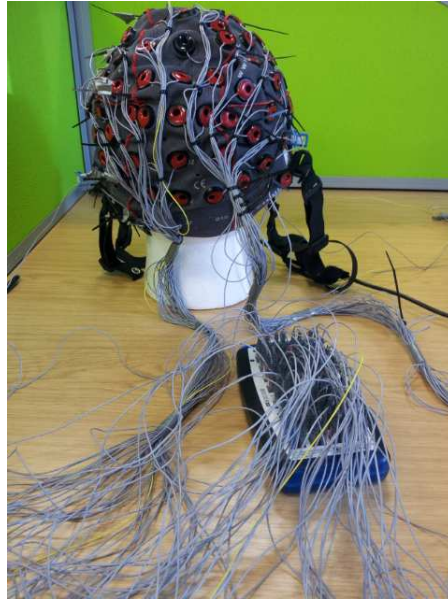
(B)

Figure 2.9: Figure (A) shows a g.tec g.GAMMAsys with 64 electrodes applied and (B) shows a g.tec g.HIamp amplifier [67]. The cap and amplifier are interconnected and linked via a USB to a computer.

Nowadays the variety of technologies used to develop EEG sensors covers a wide range, from wet and dry electrodes to wireless EEG sensors. However, all of them pursue the same objective: to be precise.

The term "wet electrodes" is related to the need to use conducting electrode gel to attach it to the scalp. The materials used for their construction are several: silver/silver-chloride (Ag/AgCl), tin (Sn), gold (Au) or platinum (Pt) [68]. The Ag/AgCl electrodes are considered the golden standard and they are used almost universally in clinical and research applications. The use of electrically conductive gel reduces the skin-electrode impedance leading to

higher quality signals [69]. However, the need to use a gel increases the time
needed to place the electrodes and an excess of it may create shorts between
sensors if it spreads out.
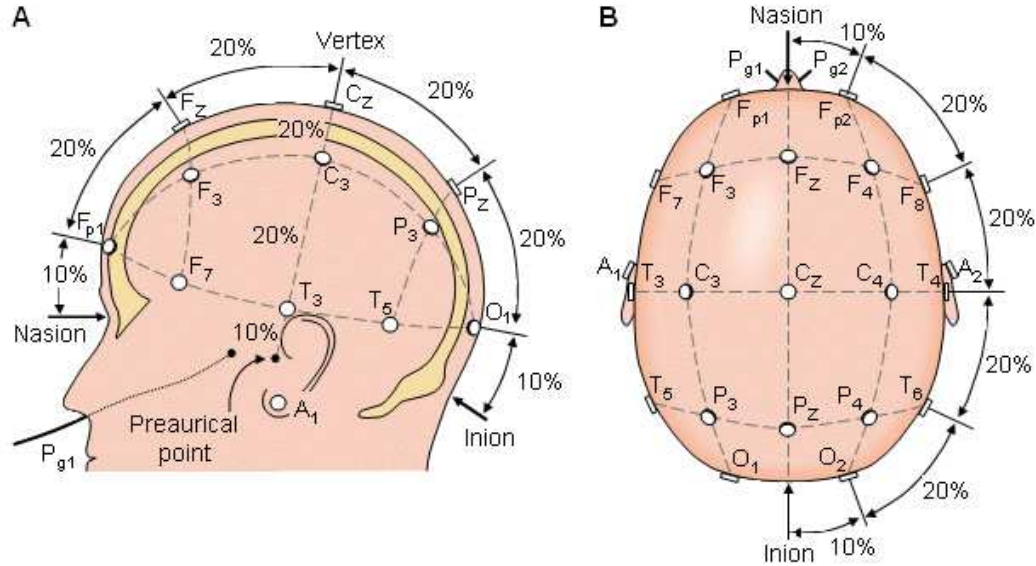
Figure 2.10: EEG 10-20 system electrode placement to show how the electrodes should be
placed using percentage values of the size of the person's scalp. Reference points are marked
in both views: nasion-inion and A_1-A_2. (A) Side-view of a person's head with the 10-20
coordinates overlaid. (B) Top-down view of a person's head with the electrode coordinates
overlaid [70].

By contrast, dry electrodes are designed to be efficient without the need
for conductive gel. The absence of gel is substituted by moisture on the skin,
mainly sweat [69]. Numerous variations of dry sensors exist on the market.
For instance: stainless steel discs or micro-fabricated silicon structures. This
type of electrode is used mainly in research as they present some problems of
usability for normal clinical applications due to their instability as they are
much more difficult to secure to the patient than wet electrodes [69].

There is another model of electrode that, contrary to a wet or dry electrode,
does not require direct physical contact with the skin. Some examples of
these non-contact sensors can be found in the literature [71, 69]. They are
comprised of a set of capacitive electrodes with a wireless transmitter to send
data to a computer. These systems have the advantage of being insensitive
to skin conditions and require zero preparation. However, their precision and
reliability still have not been proven.

For this research, the electrodes used are active Ag/AgCl ring electrodes
from g.Tec [67] implanted in the g.GAMMAcap previously mentioned and
shown in figure 2.9.

2.3.2 Clinical brain wave bands

Hans Berger was the first investigator to discover a rhythmic brain wave in the range of 8-12Hz that he named the α band [65]. Since his discovery, it has been demonstrated that, irrespective of the nature of the observed signal, most of the brain's activity has multiple frequencies that evolve over time. The most important are: the delta, theta, alpha, beta and gamma bands. They are identified according to their frequency and they possess different features that are described in table 2.2.

Table 2.2: Summary of the main EEG brain wave bands and their features [72].

Name	Frequency range (Hz)	Features
Delta (δ)	0.5-4	Occurs in sleep or a vegetative state of the brain, slow and high amplitude waves.
Theta (ϑ)	4-8	Occurs during light sleep, quiet focused meditation. They have been observed during memory retrieval.
Alpha (α)	8-12	Mediate level of consciousness, relaxed, awareness of the body, predominant with closed eyes, prominent above visual areas.
Beta (β)	12-30	Related to consciousness, busy or anxious thinking and active concentration. Low voltage waves.
Gamma (γ)	>30	With high level information processing, for learning and memory.

2.3.3 Artifacts

Recording electrical activity from the brain is subject to non-cerebral interference due to the high sensitivity of EEG systems. Those sources, named artifacts, can have a non-physiological origin. For example, electrical devices operating nearby or physiological interference signals originated from the subject's heart and muscle movements. Small movements such as blinking or frowning can introduce large spikes in the EEG signals and may deceive the interpreter to believe that the apparent sources are abnormal [73]. In [74], the authors performed a comparative study of the effect of blinking on the signal to noise ratio (SNR) for scalp EEG and iEEG simultaneously. This shows that peak amplitudes in EEG channels closest to the eyes related to blink artifacts that were also recorded from the same anatomical region in iEEG.

Another source of noise is the placement of the electrodes; if the references to place the cap or electrodes are not accurate, the recorded EEG signal will be affected by noise. In the same way, if one electrode is unsecure it can move during the experiment causing large artifacts. In addition to these types of noise, skin-electrode noise must be considered which strongly correlates with the skin impedance. The effect is reduced with the use of the gel in the case of wet electrodes but the issue still remains under-addressed in the case of dry electrodes, although efforts have been made to negate this effect [69]. Recognition and elimination of the artifacts in EEG recordings is an arduous task, but essential for the development of practical systems [73].

In the last decade several methodologies have been proposed to improve the SNR of EEG measurements, especially those comprising different signal processing techniques designed to reduce the noise using a range of temporal averaging schemes. For example median and weighted averaging, trimmed estimators, wavelet-based de-noising methods or spatial filters [75]. In order to effectively choose the most appropriate method to deal with noise, several aspects need to be considered in relation to the properties of the data and the research questions being asked [76].

2.4   Brain Computer Interfaces (BCI)

A BCI is a communication system that is non-dependent of the brain's normal output pathways. These systems provide its users with an alternative method of interaction with the world. Since the EEG was first described by Hans Berger in 1929 [65], BCI systems have made incredible progress. A variety of methods to monitor and record brain activity are available and include PET, fMRI, MRI, MEG or EEG. Refer to section 2.2 for further details regarding their basic function principles. Figure 2.11 depicts a classification of BCIs, also named a Brain Machine Interface [11].

The main concept of the BCI system is that the human brain reacts to a specific stimulus generating a specific evoked potential (EP) and the consistent response to this stimulus can be used to control a device or for any other purpose. An evoked potential, also known as event related potential or evoked response, is characterised by a series of fluctuations in the EEG that are time-locked to an event, including the onset of a stimulus or the execution of a physical response [77]. Beyond BCI, EP can be used to assess hearing or sight, especially in children, to diagnose disorders of the optic nerve or detect other

problems affecting the brain and spinal cord.
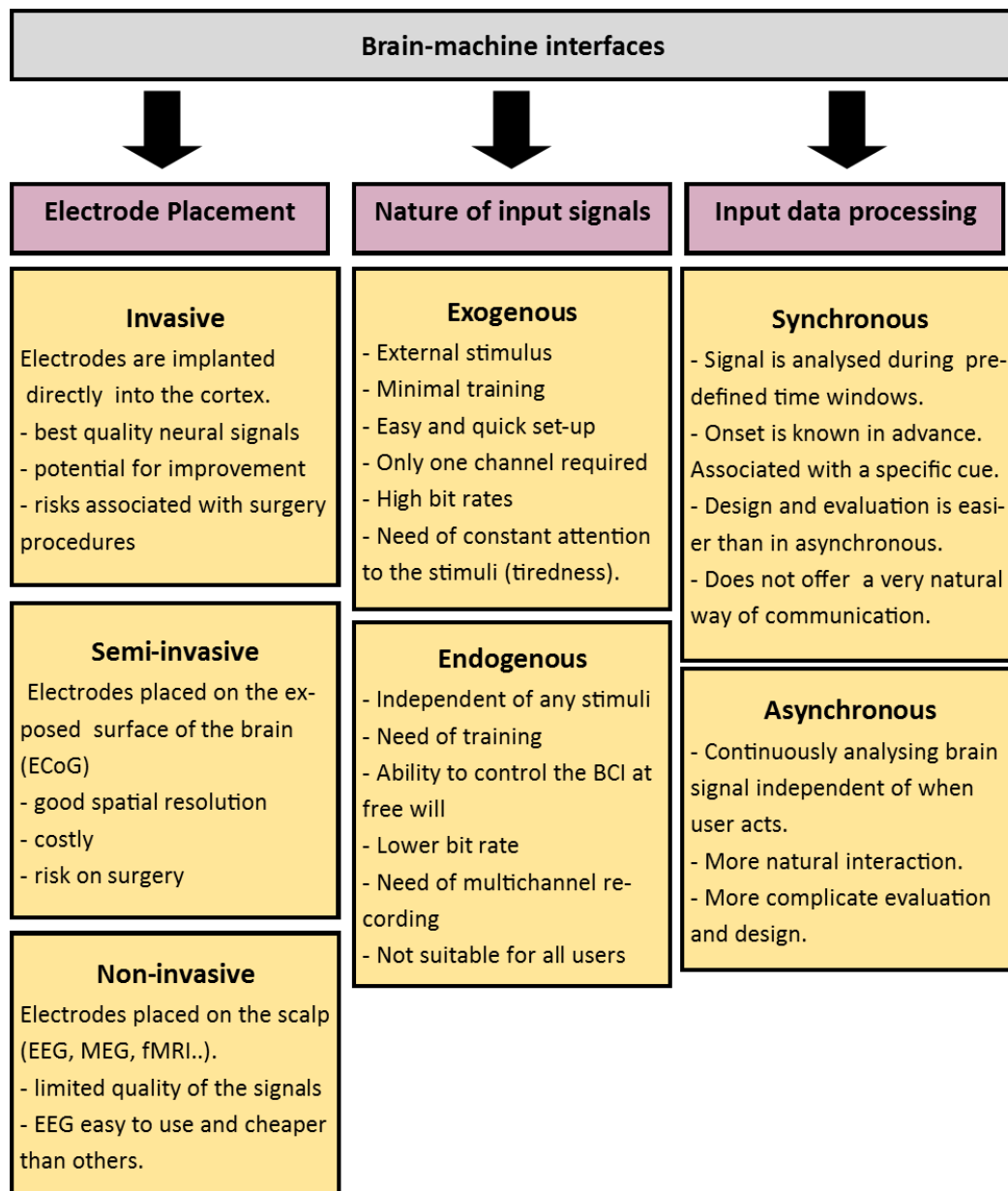
Figure 2.11: Classification of the different BCIs accordingly to different criteria: position of electrodes recording the brain activity according to the nature of the input signals of the BCI and according on how they are processed [11, 78].

EP represents the standard technique in cognitive neuroscience to investigate the temporal dynamics of cognitive processes. Through the electrode placements on the scalp, the brain activity is recorded before, during and after the stimulus onset. During the recording, several replications of the experimental conditions are conducted, named trials. Time-locked signal averaging is a necessary step to extract EP from the raw data. At each recording channel

and at identical times, from the beginning to the end of the trial, the signal is averaged across of all the trials recorded.

2.4.1   EP components

An EP consists of a series of positive and negative deflections, which will be denominated as components if they respond to a specific stimulus. These components are named by their polarity, positive (P) or negative (N), followed by a number referring to their latency (milliseconds). For example, the most popular component, P300, indicates a positive component for which peaks occur at around 300ms after the stimulus is presented to the user. Another widely used component is the N170, a negative peak around 170ms after the stimulus onset. Some more examples are: P100, P200, N270 or N400 [79]. Amplitude (of the order of μV) is usually defined as the difference between the mean pre-stimulus baseline voltage and the largest positive-going peak of the EP waveform within a time window determined by the stimulus modality, task conditions, subject age, and other factors. Latency is typically defined as the time from stimulus onset to the point of maximum positive amplitude within this same time window [79].

2.4.2   BCI paradigms

A paradigm is the mental or control task that the BCI users have to perform to induce a specific change in their brain activity. Recently, several mental tasks have been presented as appropriate as control strategies for this purpose. For instance mental rotation, auditory imagery, motor imagery, mental subtraction, silent singing or spatial navigation. BCI paradigms can be divided into two main classes:

1. *Exogenous, reflexive, synchronous or evoked paradigms* are those which have need of an external stimulus to produce a response in the brain that can be detected by the BCI system. The presence of a stimulus reduces the training time. In some cases eliminating it, completely as the response is induced automatically. However, it can induce false positives as the user has to focus attention on the stimulus or the absence of it.

   - *P300*: As mentioned before, P300 is one of the EP components. Its

name indicates a positive change in the amplitude of the signal achiev-
ing its maximum value around 300ms after stimulus onset. The first
studies regarding task relevance effects of the P300 component were
developed using the oddball paradigm, which consists of two stimuli
(commonly additive stimuli) in a random sequence where one occurs
less frequently than the other [77]. However the actual origin of the
P300 is unclear. It is suggested that it is related to the end of cog-
nitive processing, to memory updating after information evaluation
or to information transfer to consciousness. P300 has several applic-
ations. However, the most commonly used is the P300 Speller. The
first mention of a P300 based BCI speller was in 1998, when Farwell
and Donchin, proposed the FD-Speller [80]. It consisted of a mat-
rix of 6x6 cells displayed on the screen to represent 26 letters and a
few commands. The rows and columns where randomly highlighted,
when the column or row with the target cell was intensified, a P300
was elicited. An example of this matrix of letters is shown in the
figure 2.12.

Figure 2.12: Example of a typical matrix used in a P300 BCI Speller. When the row/column
where the target is highlighted, in this case letter D, the P300 is triggered [81].

- *Steady-State Evoked Potentials (SSEP):* these are a brain responses
  induced by a constant stimulation, usually repeatedly flickering a
  light at a constant frequency, approximately between 6 and 100Hz.
  Viewing this flashing light at a particular frequency stimulates the
  visual pathway. Consequently this frequency is radiated throughout
  the brain. The response manifests itself as an increase in amplitude
  of the stimulated frequency, for example, if the stimulus it is presen-
  ted on the screen flashing at a 5Hz frequency. Then the user's brain
  should produce frequencies at 5Hz, 10Hz, 15Hz, etc. Typical applic-

ations of SSEP in BCI systems are with the control of a computer
cursor [82] or the control of a simple computer program [83].

2. _Endogenous, self-generated, asynchronous or spontaneous paradigms_ are
   when the user voluntarily performs the mental task that activates a par-
   ticular area of the brain. The advantage of these paradigms is that they
   only appear when the user wants them to, reducing the possibility of false
   positives as there are no external stimuli to trigger a response. By con-
   trast, these systems are subject to the user's ability to control their brain
   activity. Consequently prior training is needed.

   - _Slow cortical potentials (SCPs)_: These are slow event-related, direct-
     current shifts in the EEG, originating from the large cell assemblies
     in the upper cortical layer. They last from 0.3s up to several seconds.
     Despite not being oscillatory in nature, they occur as a consequence
     of external or internal events. It has been demonstrated that SCPs'
     negative shifts reflect the depolarisation of the large cortical cell as-
     semblies, decreasing their excitation threshold. It is the case, for
     instance, for patients with epilepsy where large negative shifts have
     been detected seconds before a seizure [84]. SCPs are used in a wide
     range of treatments, such as hyperactivity disorder [85], including
     BCI system applications [86].

   - _Motor Imagery (MI)_: This paradigm relies on the fact that when a
     person physically moves a limb and imagines moving it without actu-
     ally performing any motor action it produces the same brain activity.
     When the stimulus is presented, a synchronisation of the active area
     occurs firstly as a large event denominated _as an event related de-_
     synchronisation (ERD) followed by an _event related synchronisation_
     (ERS) , please refer to figure 2.13. This event is locked to a specific
     frequency band: µ rhythm (8-12 Hz) and β rhythm (12-30Hz). This
     paradigm produces an asynchronous continuous output signal that is
     optimal for applications such as motor control. Wolpaw and his team
     were the first to use the concept of MI applied to BCI systems [87].
     They demonstrated for first the time that individuals can learn to use
     the µ rhythm to control movement of a cursor on a computer screen.
     They managed to translate the distributions of µ rhythm amplitudes
     to a series of parameters able to control the cursor movements with

control and high accuracy. Since then, the accuracy and diversific-
ation of MI based BCI has experienced an enormous improvement.
Nowadays, systems exist with the capacity to differentiate between a
relaxation state, imagination of left hand, right hand, feet or tongue
movements.
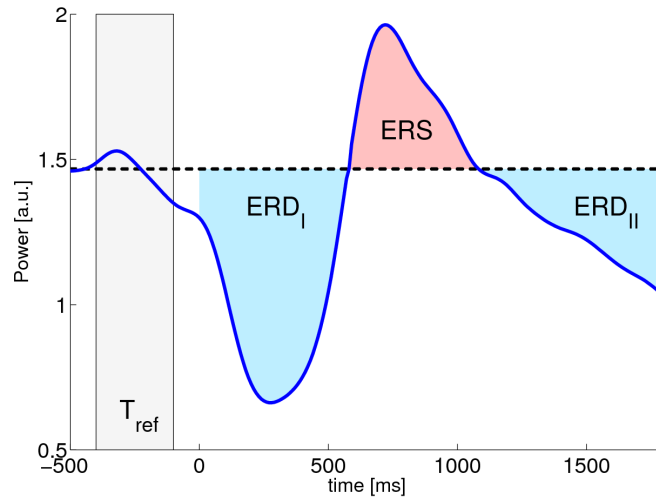
Figure 2.13: ERD/ERS process occurring after the stimulus onset during a MI paradigm
[88].

2.4.3   Emotional faces

There are different approaches to improve the performance of a BCI system.
Most research focuses on more accurate signal processing and classification
techniques. However, another interesting method to improve BCI performance
is achieved by means of optimising the user's control strategies [89]. Recently,
various mental tasks have been identified as suitable control strategies for
using BCIs. For example, Millán et al. [90] and Galán et al. [91] implemented
asynchronous BCI protocols in which participants successfully controlled a
wheelchair, a robot or a keyboard by the use of six different mental strategies to
choose between: relaxation, left and right hand motor imagery, cube rotation,
subtraction and word association. This study also highlighted the importance
of the choice of mental tasks in order to get good performance from the BCI
strategy. To perform the classification step, firstly a multiple discriminant
analysis is used to select the relevant EEG features and then, the authors used
the statistical Gaussian classifier algorithm. Friedrich et al. have completed
several studies comparing the effect and stability of seven different tasks on

classification performance for BCI [92, 89]. They selected tasks from different domains: mental rotation (imagine 3D L shape figure to rotate), auditory imagery of a melody (familiar tune without articulating the words), mental subtraction (successive elementary subtraction by a presented fixed number), spatial navigation through a familiar environment (your house or your room), imagery of familiar faces or motor imagery of the right and left hand. All of the mental tasks used in this work and their different combinations, obtained average accuracies comparable to the standard task left and right imagined movements using linear discriminant analysis as the classifier and common spatial patterns to compute the most discriminative features for classification [89]. In addition all of them were stable across sessions on different days. A novel P300-based BCI speller using faces as stimuli in conjunction with eyeless faces and eye only images was developed in the study presented in [93], resulting in a significant increase in performance when compared to the conventional P300-based BCI with stimuli of an intensification pattern. In this case, the feature extraction was made by extracting 8 spatio-temporal feature vectors from 8 flash sub-trials to later feed a linear discriminant analysis classifier. In the same manner, a novel stimulus for gazed-BCI is presented in [94] named the "coloured dummy face pattern". It is suggested that different colours and facial expressions could help users to locate the target and evoke larger ERPs. Bayesian linear discriminant analysis, an extension of linear discriminant analysis, was used in this research to classify between the different "coloured dummy faces".

Faces provide larger information with regard to diverse aspects such as intention, emotional state, age, gender or identity that make them play a vital role on a daily basis in social communications [95]. The process of facial emotion recognition is a complex task that comprises perceptual and memory skills, identification and analysis of the particular emotion in the face in view. In summary, facial emotion recognition requires integration of attention, perception, learning and memory [27]. Over the last decade face processing and recognition has been studied extensively. Bacsar et al. reported that different responses occur in different frequency bands when the participants were exposed to a picture of a loved person versus a picture of an unknown person [27]. Baumgartner et al. showed that EEG activity over the left hemisphere increases in happy conditions compared to negative emotional conditions [28]. Some investigations have been based on the study of different types of functional brain connectivity using emotions as stimuli to classify Parkinson's disease from healthy control patients [29] or Autism Spectrum Disorder from

typical developed children [30]. These important results open the door for
biomarker applications that can track emotional impairments.

Based on these findings regarding face emotion processing and perception, it
can be thought that the use of faces as stimuli may increase the classification
accuracies of MI-based BCI. In addition it is likely to help in tracing the
cognition patterns underlying MI tasks linked to a specific stimulus.

Another question to consider is the suitability of stimuli for patients with
different types of disability. ALS patients have impairments in working memory
but no prior studies have reported any impairments regarding facial perception
and structural encoding. In addition patients with different levels of cognitive
impairments may experience benefits from the use of emotional faces as stimuli
[93].

2.5   Brain connectivity

The human brain is a remarkably complex system of units interacting with each
other to incorporate and process both internal and external stimuli. This mar-
vel of neural wiring ranges from links between individual neurons to fibres that
meander through vast brain regions. Such a complex system cannot be stud-
ied by investigating individual units separately. These assemblies of specialised
neurons influence each other through a series of synaptic connections [96]. The
use of electroencephalography to attempt to measure the functional interactiv-
ity between different cortical regions has a long history [97, 98, 99, 100]. A
wide variety of methods has been used to uncover the underlying connectiv-
ity patterns across the brain in human [101] and non-human subjects [102].
Different methods are based on coherence, which is assumed to correspond
to synchronised activity between electrical activities across the different brain
regions in a specific frequency band [103]. Other techniques investigate the dy-
namics of the cross-correlation of the time series between a pair of electrodes.
This interaction across cortical regions has the name of *synchronisation* in
time scale and its dynamic is an essential instrument to understand how the
human brain performs a cognitive task given a particular stimulus [104]. Dir-
ect evidence supporting synchrony as a basic mechanism for brain integration
has been recently proven with studies based on visual binding [105], proving
that local integration -within neighbouring cortical areas- and large scale in-
tegration - concerning neural assemblies which are farther apart in the brain-
can be interpreted as a biological mechanism of integration.

There are several distinctions to considerer related to connectivity, the first
one is between *functional integration* and *segregation*. Functional integration
refers to the study of how the brain regions work together to process inform-
ation and effect responses whereas segregation suggests that a cortical area is
specialised for some aspect of perceptual or motor processing and this special-
isation is anatomically segregated within the cortex [106].

Another differentiation can be made between *functional* and effective *con-
nectivity*. Functional connectivity is defined as the statistical association or
dependency between two or more distinct time-series [107]. It reflects the stat-
istical dependency among remote neurophysiological events [106]. It is simply a
statement about the observed correlations against the null hypothesis; it does
not provide any direct information on how correlation is mediated between
these brain regions. To deal with this issue the concept of effective connectiv-
ity was developed. It is a more abstract notion than functional integration but
should be understood as a time-dependent experiment with the simplest pos-
sible neuron-like diagram that could produce the same temporal relationship
between the recorded neurons in a cell assembly [108, 103]. It is the direct or
indirect influence that one neural system exerts over another and consequently
it depends on some model to define this influence [107].

2.5.1   Analysing connectivity

The study of time varying functional connections will give us a pathway to
understand and quantitatively measure, brain connections happening across
the brain when a specific task is used as a stimulus.

Effective connectivity can be explained by means of *dynamic causal mod-
elling* (DCM). It models a network of discrete neuronal sources based on
Bayesian techniques. It relies on a biophysical model of neuronal dynamics
(neural-mass or conductance-based models) [109] and requires a priori defini-
tion of a large dataset of parameters [106]. Another drawback of this approach
is that crucial differences among a variety of analyses rest with the models on
which they are based [107]. DCM can be converted into a linear state-space
model by solving the series of Bayesian equations, which is known as Granger
causality [110]. *Granger causal modelling* (GCM) is explained in terms of lin-
ear vector autoregressive (VAR) models of stochastic time series data [111].
The multivariate VAR is defined as a set of k EEG channels as:

$$X(t) = \sum_{j=1}^{p} A(j) X(t-j) + E(t) \tag{2.1}$$

where:

- $X(t) = [X_1(t), X_2(t), \ldots, X_k(t)]^T$ is a vector of k signal values at each time t,

- $E(t) = [E_1(t), E_2(t), \ldots, E_k(t)]$ is a vector of noises at each time t,

- $A(j) = \begin{pmatrix} A_{11}(j) & \ldots & A_{1k}(j) \\ \vdots & \ddots & \vdots \\ A_{k1}(j) & \ldots & A_{kk}(j) \end{pmatrix}$ for j=1,...., p are the model parameters,

- p is called the model order.

The model order selection is important in order to obtain an appropriate fitting quality. If the order is too low, the resulting spectra can lack the necessary details for a correct analysis of the EEG connectivity. On the other hand, high model orders tend to create unwanted noise in the spectrum. Several criteria have been proposed to select the appropriate model order such as Akaike's information criterion or Bayesian-Schwart's criterion [112]. In the same manner, there are several algorithms of model parameter estimations differing in their ability to detect specific features of the spectrum of stability for shorter data segments [112, 113, 114].

The disadvantage of this technique is that it assumes that the EEG signals are stationary and it cannot provide information regarding phase coupling between electrode pairs. EEG signals are dynamic, and therefore non-stationary. This issue can be solved by applying windowing techniques. However, these windows should be short enough to treat the data within each of them as stationary. This approach performs well for shorter data epochs, however, the statistical significance of the estimates decreases with a shortening of the window size [112].

Functional synchronisation based on EEG signals can be measured in the time or frequency domains. EEG signals are non-stationary and the mutual influence of brain regions, hence EEG channels, do not always show a time invariant behaviour [115]. Several solutions have been proposed to address

this issue. Across them, one method to analyse the EEG signal in the time
domain, especially suitable for the study of spontaneous activity, is the de-
nominated microstate analysis [116]. Microstates are defined by topographies
of electric potentials recorded in a multichannel array over the scalp which
remain stable for a few milliseconds (80-120 ms) before transitioning sharply
into a new different microstate [117, 118]. The advantage of this technique is
that it simultaneously considers the signal from all the EEG electrodes to cre-
ate a more global representation of a functional state [118]. Researchers have
been using the concept of microstates for a vast diversity of studies since it was
first mentioned by Lehmann in his seminal paper [119]. For example, in study-
ing the impact on negative social information amongst depressed people [120],
the resting state from schizophrenic patients [121] or those with Alzheimer's
disease [122]. In addition, some investigations have been undertaken for aware-
ness, motor inhibition or grasping objects [123, 124, 125]. In [126], a test-retest
reliability study to assess the efficiency of the resting-state EEG microstates
analyses in healthy subjects over time was performed. The conclusions were
firm, this technique has a high test-retest reliability. In addition the consist-
ency of the most frequently used clustering algorithms (k-means clustering and
TAAHC) in extracting microstates maps was determined.

On the other hand, in the frequency domain, the classical methods to meas-
ure synchronisation are *correlation* and *coherence*. Coherence is a squared
correlation coefficient that estimates the consistency of relative amplitude and
phase between two pairs of signals in each frequency band [127]. EEG co-
herence depends mostly on the consistency of the calculated phase difference
between channels [128]. Transferring the equation 2.1 to the frequency domain,
the power spectrum can be calculated as [112]:

$$
\begin{aligned}
S\left(f\right) \; = \; & X\left(f\right)X^{*}\left(f\right) = \left(A^{-1}\left(f\right)E\left(f\right)\right)\left(A^{-1}\left(f\right)E\left(f\right)\right)^{*} = \\
& \left(H\left(f\right)E\left(f\right)\right)\left(H\left(f\right)E\left(f\right)\right)^{*} = H\left(f\right)E\left(f\right)E^{*}\left(f\right)H^{*}\left(f\right) = H\left(f\right)VH^{*}\left(f\right),
\end{aligned}
\tag{2.2}
$$

where V is the noise covariance matrix and $H\left(f\right) = \left(\sum_{m=0}^{p} A\left(m\right)e^{-2\pi imf\triangle t}\right)^{-1}$.
The coherence can be calculated by obtaining the cross-spectra of two signals
i and j from equation 2.2 [112]:

$$
K_{ij} = \frac{S_{ij}\left(f\right)}{\sqrt{S_{ii}\left(f\right)S_{jj}\left(f\right)}}.
\tag{2.3}
$$

EEG coherence is a sensitive measure that may yield important information about network dynamics and functional integration across brain regions. This metric is typically computed by first taking the Fourier Transform (FT) of the EEG signals recorded in two electrode sites and then taking the imaginary and real parts to calculate the phase of the transformed signals. EEG coherence has been widely used in research for a diverse range of topics such as the detection of Alzheimer's Disease, cognition, Parkinson's Disease or mental fatigue among others [129, 130, 131, 132]. However, the need to use FTs to calculate the coherence means that the temporal information of the phase dynamics is not preserved so therefore give an averaged synchronisation measure over a selected time window at each frequency band. Some investigations have solved this issue working with Short Time Fourier Transforms (STFT) [133]. However, stationarity is still required within each time interval for which coherence is calculated converting the selection of the time window length as a crucial step to achieve accurate results [109]. An alternative method for calculating coherence is the use of the Continuous Wavelet Transform (CWT). It is more flexible than STFTs but requires a-priori information about the coupling range in time and frequency in order to select optimal parameters [109]. This method has been applied to EEG and MEG signals for classification of autism disorder [134], learning study [135], brain computer interfaces [135, 136], performing a mathematical task versus a resting state or Alzheimer's Disease [137, 138].

Alternatively, with the aim of investigating the phase synchronisation at a finer scale than by means of coherence there is a series of non-linear methods based on deterministic chaos [109], namely phase *synchrony*. Phase synchrony index or the phase locking value (PLV) use wavelet coherence to measure the phase locking between two EEG signals. Synchrony measures depend only on the phase between signals, even when the amplitudes remain statistically independent [139].

Many other variations of these connectivity measures have been described: entropy and correlation entropy coefficients, partial and directed partial coherence, directed transfer functions, direct directed transfer functions, full frequency directed transfer functions, multivariate autoregressive (MVAR) models, omega complexity, mutual information or state space based synchrony [140, 141].

The aim of this work is not to study the underlying neuroanatomy of patients, but to investigate the transient dynamics of the information integration process across the different brain areas in a task-specific way. To this end, it is absolutely necessary to estimate the evolution of phase relationships along

with the task performance at different frequency bands and EEG electrodes. Deeper understanding of this integration process during a cognitive task may be useful in describing brain organisation [134].

2.5.2   Connectivity and volume conduction

Volume conduction is an undesired effect when an electric current passing through any biological matter from a source is recorded at the sensor. It has been argued that volume conduction in the head can lead to a high scalp correlation in the absence of significantly correlated sources [142].

Usually, the head is described by 3 or 4 concentric spherical layers, representing the brain, the skull, the scalp and the cerebro-spinal fluid (CSF). In each of these layers, the conductivity is assumed linear, isotropic, and homogeneous. Electrical currents spread nearly instantaneously throughout any volume, like membranes, skin, tissues, etc. Sudden synchronous synaptic potentials on the dendrites of a cortical pyramidal cell result in a change in the amplitude of the local electrical potential referred to as an "Equivalent Dipole" [143]. The signal recorded at the electrodes is smaller in amplitude since it has to travel through various media to be detected by the sensors. The potential recorded at the scalp is inversely proportional to the distance from the source. Under this premise, any small change in distance may cause a large change in the recorded signals. In addition, the polarity and shape of the electrical potential depends on the solid angle between the source and the electrode where it is recorded. The volume conduction can affect the synchronisation recorded from the EEG electrodes placed on the scalp, being corrupted or masked by linear mixing.

Volume conduction involves near zero phase delays between any two points within the electrical field as collections of dipoles oscillate in time [144]. Zero phase delay is one of the important properties of volume conduction and it is for this reason that measures such as the cross-spectrum, coherence, bi-coherence and coherence of phase delays are so critical in measuring brain connectivity independent of volume conduction [23]. Based on this idea, some works have recently been developed to mitigate the effect of volume conduction in the measure of phase synchronisation [145, 146, 147, 134]. In this particular case and based on these studies, it can be concluded that the synchrony derived from phase difference is not affected by the volume conduction effect as it does not report zero phase delay.

2.6   Graph Theory

The study of structural and functional connectivity within the living brain and the changes of the connectivity profiles over time can be done by means of a mathematical framework named network theory [148]. The use of network or graph theory relies on the definition of regions within the brain acting as nodes in a graph or network. The network is formed by nodes and edges. In this case the nodes are the EEG electrodes. The edges represent the connections between nodes and their representation is known as the adjacency matrix of the network. The structure of this matrix describes the communication pattern of the brain network.

Various theoretical aspects of brain connectivity have been investigated in recent years. For instance, rich club organisation [149] and segregation and integration [150]. Complex network measures have been used to explore and understand the brain network from EEG and fMRI signals recorded from patients with neurological conditions and to compare it against healthy patients. Conditions such as Alzheimer's disease [151], autism spectrum disorder [134, 30], attention deficit hyperactivity disorder (ADHD) [152] or developmental changes due to premature birth have been studied in this way [153]. The review revealed the potential of using graph theory for characterising group differences within the brain. The present work aims to use the principles and benefits of graph theory applied to EEG recordings to obtain quantitative metrics for further analysis to help in the characterisation of a more accurate MI-based BCI.

2.7   Summary

This chapter is a brief introduction to some technical concepts related to this thesis; making its reading more comprehensive. The chapter began by discussing the most complex organ in the human body, the brain. Furthermore, the formation of action potentials is explained and some techniques to measure brain activity are briefly mentioned. Among those techniques, we have focused on EEG as it was the methodology chosen to record the data used in this thesis. The idea of synchrony is key to understanding patterns of connectivity appearing across the neural assemblies and is highlighted in this chapter. In addition, a detailed review of the current techniques to measure synchrony in EEG is listed. During the review process the inability of state of the art techniques to preserve temporal information of the phase synchrony was noticeable, which

is fundamental information on how neural connections are constructed in a stimulus-specific way.

The concept of BCI, the base of this work, and state of the art EEG-based BCI is introduced in this chapter with special attention to MI-based BCI. The novelty of using faces as stimuli for the proposed MI-BCI is motivated by the fact that face stimuli may lead to a stronger response across participants [93]. Finally a brief introduction to graph theory has been provided in this chapter as a new and promising new approach for evaluating brain networks quantitatively. In the next chapter the proposed methods to measure phase synchrony without loss of temporal information is provided.

Chapter 3

Studying phase synchronisation: synchrostates

Complex systems such as the brain cannot be explained as individual units by themselves. For this reason, to understand brain cognition it is essential to study the connections across those individual units to give a sense of a global connectivity. In chapter 2, the need to investigate time varying phase synchronisation when a stimulus-based cognitive task is performed was highlighted. The method used to achieve this objective was firstly explained in [134] and applied in the study of autism disorder in children. The main idea of the developed methodology is to use the time and frequency information from the wavelet transform to understand the phase variations across time. Once these dynamics are obtained, this identifies the possible existence of specific patterns associated with a specific-stimulus-based MI task. There have been some attempts to study connectivity in the temporal domain by means of the concept of microstates leading to unique electric potential patterns. These patterns and their transitions are task specific. This concept of temporal switching of quasi-spatial stable states can be coupled to the idea that information processing between different areas of the brain within a similar dynamic functional state, is facilitated by the phase synchronised activity of the different neural groups. This allows us to obtain a more complete temporal-frequency representation of the EEG connectivity when a cognitive task is being performed. This focus on the transitions of phase synchrony provides a new perspective to gain a deeper understanding of the brain connectivity and its dynamics.

Based on this idea of microstates and their temporal switching patterns, instead of using electric potentials based topographies, the proposed method

divides the temporal line into a small number of states which are pre-defined by a clustering algorithm. This small number of states is phase synchronized and stable in time in the order of milliseconds before switching to the next state. The outcome of this methodology will help us understand the temporal sequence describing the temporal evolution of connectivity linked to specific tasks and stimuli. Consequently, it can yield a more accurate and reliable methodology for task classification on MI-based BCI uses.

In order to establish this finite number of states, namely synchrostates [154], two main steps are needed. Firstly, the extraction of temporal and frequency information is a fundamental stage of the algorithm to establish the temporal evolution of the connectivity from the EEG recordings. The next step forward is to try to understand the underlying patterns hidden in these phase difference sets of measures. For this purpose a clustering algorithm to determine the existence and number of synchrostates is applied. Throughout this chapter both methodologies are explained in detail together with the experimental protocol followed to obtain the EEG recording dataset.

3.1   Experimental protocol

So far, most of the research on MI based BCI is focused on finding new signal processing algorithms to enhance their performance. In this work, the aim is not only to find a novel signal processing methodology but also to discover a more user-friendly and feasible stimulus that will help to reduce the BCI illiteracy rate achieving comparable performance with classic paradigms. Face perception involves some specific proprieties not existing in other visual object perceptions. In addition, recent studies have demonstrated that face perception can enhance stronger responses across subjects [93] and achieve excellent performance when applied together to evoked EEG responses in BCI spellers' applications [155, 156]. After all these findings, the present study is devoted to investigating the effect of face perception for the first time on MI based BCI fusing stimuli of schematic emotional faces.

3.1.1   Subjects

Ten healthy volunteers, 8 males and 2 females, with an age range between 20 and 53 years (mean age of 31 ± 10.01) participated in the experiment. They were right handed with one exception and their sight was normal or corrected to normal. The recruitment was carried out by means of public announcements,

no selection criteria were used and no monetary compensations were offered to the participants. Written consent was signed by each participant after they were informed of the nature of the study. The University of Warwick Ethical committee approved the study (REGO-2014-821). In particular, the corresponding ethical committee for the studies undertaken within the Science and Medicine Faculties involving human participants at the University of Warwick is named Biomedical & Scientific Research Ethics Committee (BSREC).

3.1.2   Experimental montage and equipment

During the experiment participants were seated in a comfortable chair in a quiet room. A keyboard connected to a laptop was placed over their legs so they could use their left and right index fingers to press the appropriate keys during the task. In addition, a computer monitor was placed on a desk in front of them where the series of emotional faces was shown. This monitor was also connected to the laptop as the main part of the whole set up array controlling the timing and saving of the EEG recordings.

For recording the EEG signals, 62 active g.LADYbird electrodes already mounted in an electrode cap (g.GAMMACap) were used. The cap and the laptop were connected by an EEG amplifier (g.HIamp) all from g.Tec [67]. Electrodes on the cap were positioned accordingly to the 10-20 coordinates as shown in figure 2.10. To help the recording of the brain activity, a conductive gel was placed between the electrode and the user's scalp.

The amplifier and the cues program were managed by Simulink$^©$ with in-home specific libraries. Data were acquired at a sampling frequency of 256Hz and referenced online by subtracting the averaged references from two electrodes placed on the earlobes. Furthermore, an online notch filter (50Hz) and online Butterworth band pass filter (0.5-100Hz) were used to attenuate the effects of the power line and muscle artifacts respectively, both available from the recording equipment's Simulink libraries. EEG recordings were saved to a Matlab$^©$ file.

3.1.3   Experimental task

The stimuli used for this experiment are based on the schematic emotional facial expressions used by Babiloni and his team at [157], a study about conscious perception of emotional facial expressions. Simple schematic faces were used, widely known as emoticons, over a black background. The emoticons,

with a green background colour, could represent two types of emotions: happiness or sadness. These emotions are expressed by the function of the line describing the mouth: an up-right curve for happy and down-right line for sad expression.

Figure 3.1: Time-scheme of a typical trial. The sequence of the stimuli was as follows: (1) masked stimulus lasting between 1.25 to 1.75; (2) target stimulus (50% probability for each type of emotion) appearing on the centre of the screen for 0.5s; (3) black screen with a random duration of 1.25 or 3s where the participants were asked to perform the imaged movement of their right or left hand according to the emotion shown just before; (4) cross over a white circle indicating the end of the trial and where the participants have to press the 'm' or 'z' keys in consonance with the emotion, hence imagined movement, depicted by the schematic face.

The temporal scheme of the protocol for a typical trial is shown in figure 3.1. It was as follows: a masked face lasting from 1.25s to 1.75s on the screen, a target stimulus for half a second showing one of the possible emotions -happy or sad- with equal probability of appearance; a black screen with a random duration of 1.25 or 3s to avoid habitual prediction and finally a cross lasting one second to indicate to the participant the end of the trial and the beginning of the new one. Participants were asked to imagine the movement of squeezing a ball with their right hand or left hand during the duration of the black screen without actually performing any actual motor movement. The hand to perform the imaged movement was determined by the emotion shown by the schematic face, happiness means right hand and sadness left hand. In order to keep the attention of the participant to a maximum level they were asked to press the key 'z' or 'm' accordingly to the face they visualised previously,

happy and sad respectively. The tasks were divided into 4 blocks of 78 trials each, 312 trials in total, with around 5 minutes break between them as shown in figure 3.2.

The reaction time for each participant was recorded. It was measured as the time elapsed from the appearance of the cross, indicating the end of the trial and until the corresponding key is pressed. In order to verify the alertness of the participants during the experiment and any possible differences between happy and sad face responses a t-test was performed. The reaction times and the number of hits, indicated by the right key being pressed for each condition, were tested comparing across conditions without finding any relevant difference. Therefore, the response of the participants was similar for both conditions. The averaged number of hits, over 85% for both tasks, also indicates good concentration from the participants.
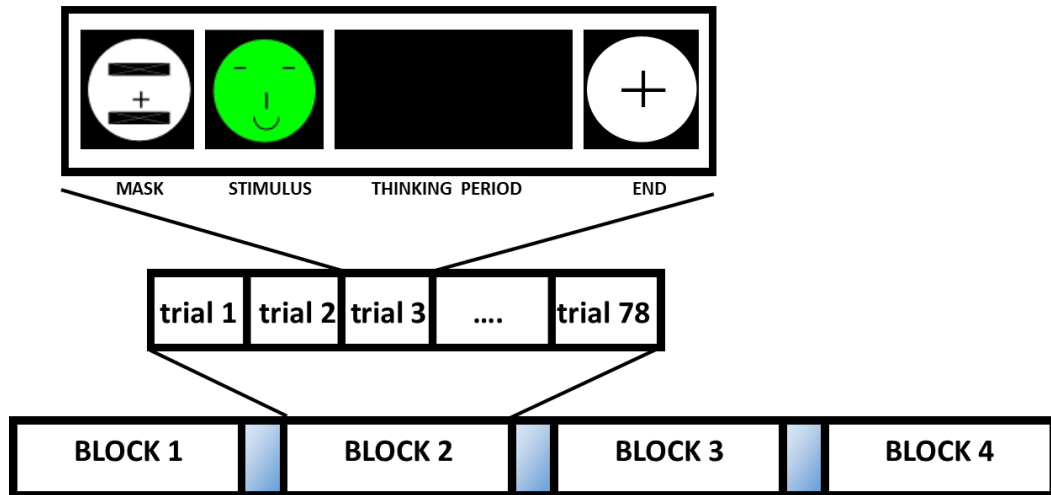
Figure 3.2: Experimental blocks sequence. The complete experiment was divided into four blocks of 78 trials each with same number of occurrences for both emotions, happiness and sadness. The breaks between blocks were around five minutes each and were dictated by the user. .

3.1.4   Pre-processing

EEG trials were divided into the two conditions, thinking right hand and left hand respectively. A first selection of trials was done by eliminating those ones with amplitudes over a 200μV threshold, a normal threshold used for EEG preprocessing. Afterwards a visual inspection of the whole dataset of trials and participants was performed and those trials affected by blinking or any other kind of muscle movement were eliminated because were considered artifacts. Finally artifact-free trials were divided into one second length epochs for each condition lasting from 100ms before stimuli onset to 900ms after.

3.2   Methodology

Synchronisation was originally introduced to describe two coupled harmonic oscillators into the physics field. Some examples can be extracted from daily life, such as two pendulum clocks moving in a common beam will finish synchronised. The idea of synchrony is also applied to the medical area. For example, to control abnormal respiration of a patient by forcing synchronisation with a ventilator moving at a specific frequency or for the study and comprehension of schizophrenia [158]. Despite the fact that the concept of synchronisation was earlier greatly highlighted by experimental results with micro electrodes placed on animals, it was not until 1999 that a practical method was proposed for the measurement of frequency-specific synchronisation between two neuroelectric signals [139]. Given two signals $s_1(t)$ and $s_2(t)$ and their corresponding instantaneous phase $\varphi_{s_1}(t)$ and $\varphi_{s_2}(t)$, it can be said that they are locked if:

$$\varphi_{s_1}(t) - \varphi_{s_2}(t) = constant \tag{3.1}$$

Two major approaches to extracting the instantaneous phase are the Hilbert transform and the complex Wavelet transform. Both methods transform real values from the time domain functions into complex values. Those methods based on Hilbert transforms obtain the analytical form of the signal. For a given signal, its analytic signal is a complex function of the time defined as [159]:

$$\hat{s}_1(t) = s_1(t) + j * \tilde{s}_1(t) = |\hat{s}_1(t)| e^{j\varphi(t)}, \tag{3.2}$$

where $\tilde{s}_1(t)$ is the Hilbert transform of the signal calculated as indicated in 3.3 and $|\hat{s}_1(t)|$ is the magnitude or envelope defined by 3.4.

$$\tilde{s}_1(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} s_1(\tau) \frac{1}{t-\tau} d\tau, \tag{3.3}$$

$$|\hat{s}_1(t)| = \sqrt{s_1^2(t) + \tilde{s}_1^2(t)}. \tag{3.4}$$

From equation 3.2, the estimation of the instantaneous phase is performed by $\varphi_1(t) = \arctan(\tilde{s}_1(t)/s_1(t))$ [160]. However, a fine bandpass filtering of the

signal around the frequency of interest is needed. On the other hand, complex-valued wavelet methods provide phase information in the time and frequency domains , therefore they are more suitable for non-stationary signals. This is particularly true for the scope of this work, where the temporal transitions of phase synchronisation are being studied in different frequency bands. There are several complex wavelet families such as the complex Gaussian, complex Morlet, complex Frequency B-Spline or Complex Shannon. Another approach similar to wavelet transform, also preserving time and frequency information, has been recently used in [21]. This new time-varying phase estimation method is based on the Reduced Interference Rihaczek (RID-Rihaczek) distribution belonging to Cohen's class. The main difference between this novel method when compared to the CWT is that the phase estimations area is uniformly distributed. For the proposed method here the well-known properties and extended use of the CWT based on complex Morlet mother wavelet is selected [154].

3.2.1   Instantaneous phase difference

The first step in quantifying the phase synchronisation between two signals consists of the estimation of the instantaneous phase of each individual signal for each time and frequency of interest [105]. Once those phase differences are calculated for each case, the phase difference between two signals can be calculated [21]. If the instantaneous phase difference between the two signals, $s_1(t)$ and $s_2(t)$ remains constant for a period of time it can be said that these two signals are in synchrony. The proposed method to calculate the instantaneous phase for each of the EEG electrodes used is the CWT, as mentioned above.

Across the different mother complex wavelets the Morlet wavelet is one of the most frequently used to work with biological signals. For example, in the study of cerebral flow auto-regulation [161] or EEG signals [162]. Another reason to use the complex Morlet wavelet is that it gives a good time location in the time domain. In addition, it can obtain the amplitude and phase of neural activity simultaneously. Therefore, it is able to investigate the synchronisation between neural activities simultaneously recorded at two different sites [163]. The complex Morlet wavelet function is defined as illustrated in the equation 3.5. A recent investigation of the use of harmonic wavelet function, which is similar to the ordinary discrete wavelet transform, was performed to study phase synchronisation which obtained similar results to when a Morlet wavelet

is used. However, this is not as widely tested as the complex Morlet Wavelet for the study of phase synchronisation [164]. The complex Morlet wavelet function is defined by equation 3.5. The values of the bandwidth parameter, F_b, and central frequency, F_c, were selected following the criteria adopted in [154], being 1Hz and 1.5Hz respectively.

$$\psi_M\left(t\right) = \frac{1}{\sqrt{\pi F_b}} e^{2j\pi F_c t} e^{-\frac{t^2}{F_b}}. \tag{3.5}$$

The methodology followed here is based on the proposed method in the study of Autism Disorders by Wasifa et al. [154]. In order to calculate the instantaneous phase of each signal in this case, each electrode used for recording the EEG signal, the complex Morlet wavelet is applied to each one of them. The result of this computation is a complex series of functions of time (t) and scales (s) for each one of the N available channels, $W_T\left(s, t\right)$. The relationship between the scales and their corresponding approximated frequencies can be calculated by3.6:

$$f = \frac{F_c}{s * \Delta}, \tag{3.6}$$

where Δ is the sampling period and f is the pseudo frequency associated to the scale s [165]. The phase of a complex signal can be determined by the inverse tangent between its imaginary and real part. If the imaginary part of $W_{Ti}\left(s, t\right)$ is expressed by $Im\left[W_{Ti}\left(s, t\right)\right]$ and real part as $Re\left[W_{Ti}\left(s, t\right)\right]$ the instantaneous phase $\varphi_i\left(s, t\right)$ of $W_{Ti}\left(s, t\right)$ can be computed by equation (3.7).

$$\varphi_i\left(f, t\right) = \arctan\left(\frac{Im\left(W_{Ti}\left(s, t\right)\right)}{Re\left(W_{Ti}\left(s, t\right)\right)}\right). \tag{3.7}$$

Once the instantaneous phase difference is obtained for each EEG channel, the next step is to calculate the instantaneous phase difference between each pair of electrodes, time sample and frequency bin. The phase difference between two channels, i and j, can be obtained following the equation in (3.8).

$$\triangle\varphi_{i,j}\left(s, t\right) = \left|\varphi_i\left(s, t\right) - \varphi_j\left(s, t\right)\right|. \tag{3.8}$$

When applying equation (3.8) for a time instant t_1 and a scale s_1 for each pair of EEG channels, the result is a symmetric square matrix. The main diagonal is zero as it represents the phase difference of a channel with itself. Figure 3.3 illustrates an example of these matrices for a specific time t_1 and scale s_k. The aim is to study the instantaneous phase difference over time for a specific band of interest. Consequently, an averaging step across all

the frequency matrices within the band of interest can be done, yielding an averaged difference of phases which is only a function of time. If the set of frequency bins or scales for a frequency band of interest F is defined by $F = \{s_1, s_2, ..., s_P\}$ then the averaged difference of phases can be calculated by (3.9).

$$\triangle \varphi_F(t) = \frac{1}{P} \sum_{i=1}^{P} \triangle \varphi(s_i, t) \tag{3.9}$$
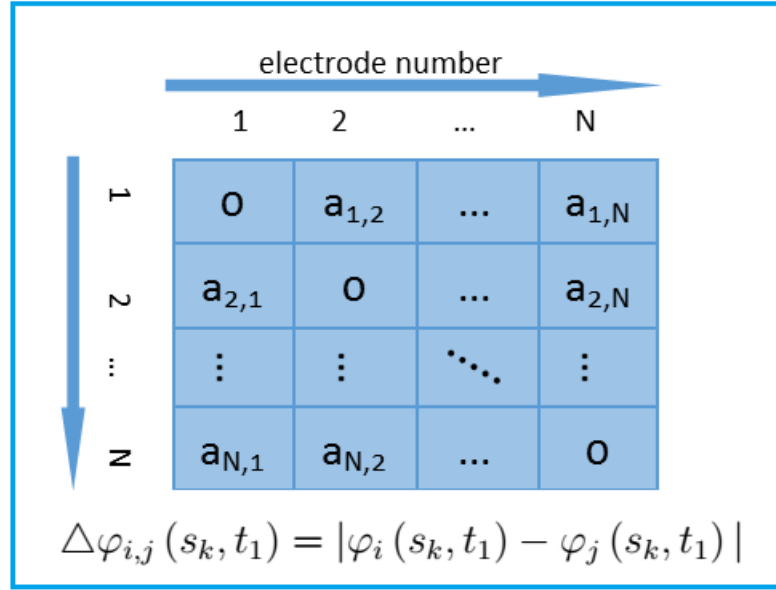

Figure 3.3: Structure of an instantaneous phase differences matrix at time t_1 and scale s_k. Adapted from [154].

A graphic schematic representation of this averaging step is illustrated on the top part of the figure 3.4. Subsequently, this averaging step should be repeated for each time instant $t = \{t_1, t_2, ..., t_n\}$. The result of repeating this averaging procedure is a set of matrices $\{\triangle \varphi_F(t_1), \triangle \varphi_F(t_2), ..., \triangle \varphi_F(t_n)\}$ describing the complete frame of the temporal evolution of the phase difference. Each of these symmetric matrices has dimensions of N by N, N being the number of channels as mentioned previously. Figure 3.4 illustrates a graphic summary of the averaging process for a time instant t and a frequency band F for each pair of electrodes and the result of replicating the process for each time t with the temporal range of interest. The result is a set of square symmetric matrices for each time instant t for a specific frequency band of interest F and for each pair of EEG channels than can be seen on the bottom of the figure.
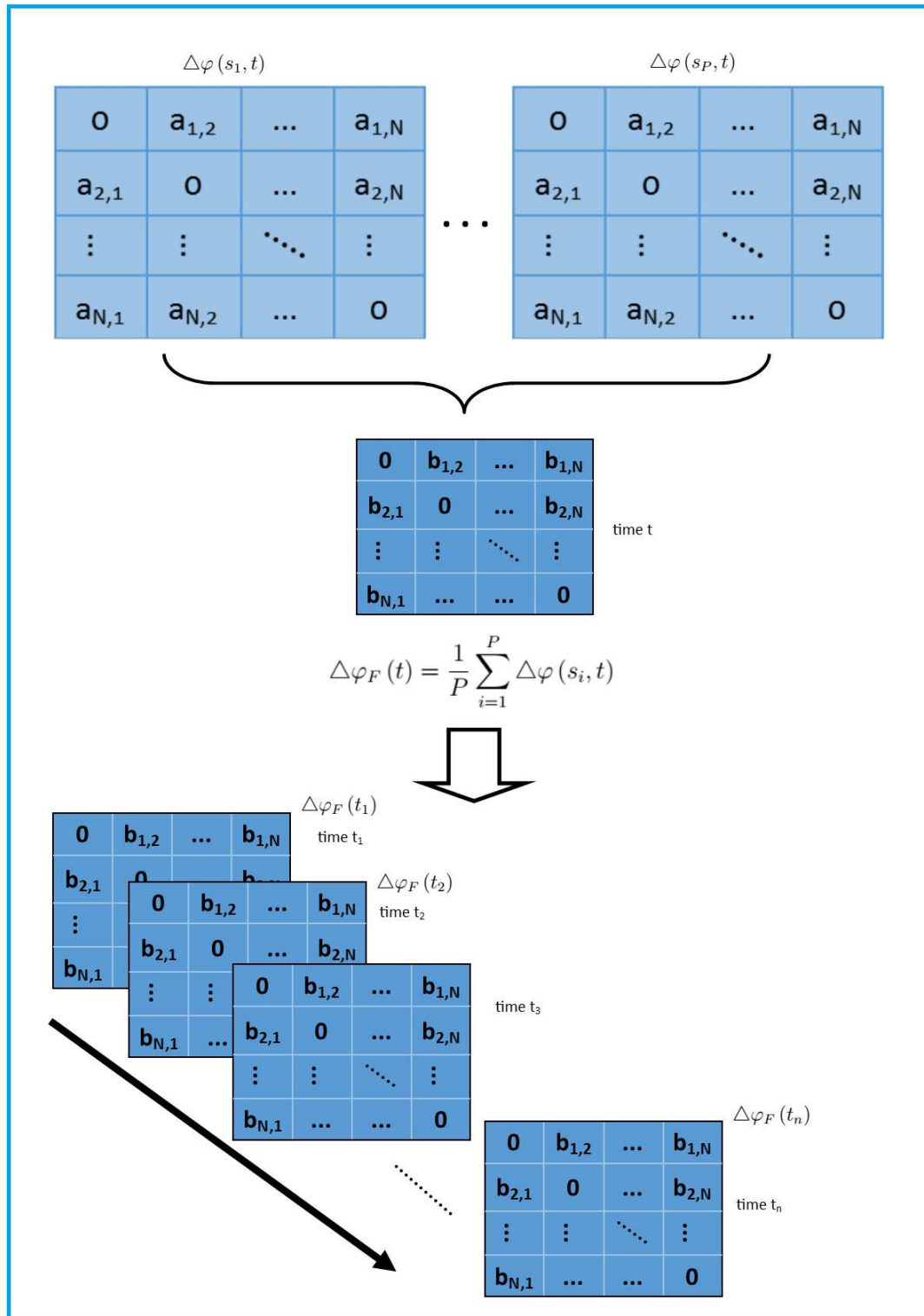
Figure 3.4: Scheme of the two final steps to calculate the instantaneous phase difference over time. The upper part represents the average step in order to calculate the phase differences in a specific band of interest F at a time t $\triangle\varphi_F(t)$. The repetition of this step for each time instant $t = \{t_1, t_2, ..., t_n\}$ is shown at the bottom of the figure. The final result is a set of symmetrical square matrices for each time instant t giving the complete information of the phase differences over time for each EEG channel pair. N is the number of EEG channels. Adapted from [154].

3.2.2   Clustering

Once the phase difference matrices along a specific frequency band and time interval are obtained, the next logical step is to investigate if there is any underlying pattern in the phase differences. In order to discover significant patterns of features in the phase synchronisation obtained from the algorithm described in the previous section, a pattern recognition technique is needed. The *k-means* [166] clustering algorithm is the most widely used partitional clustering algorithm. It has applications across a broad range of data mining problems [167] as it is one of the simplest and most efficient clustering algorithms that exists in the field of data clustering.

k-means clustering assumes that the number of underlying clusters is known. It starts by randomly choosing k points as the initial centroids. Posteriorly each point of the initial dataset is assigned to the closest centroid based on a specific proximity measure, widely known as a cost function. Once the clusters are formed, the centroids for each one of the clusters are updated. These two steps will be iteratively repeated until the centroids do not change any more or a chosen convergence criterion is achieved.

The cost function selected for this particular study is based on the Euclidean distance as dissimilarity measure. Some other proximity measures which can also be used are Manhattan distance or Cosine similarity [168]. The choice can significantly affect the centroid assignment and the quality of the final selection. In this case Euclidean distance was selected as it is the most popular choice and consequently the more tested option [169].

$$J\left(\theta, U\right) = \sum_{i=1}^{N} \sum_{j=1}^{k} u_{ij} \|x_i - \theta_j\|^2 \qquad (3.10)$$

The cost function used within this chapter is defined as equation 3.10, where $\theta = \left[\theta_1^T, \ldots, \theta_k^T\right]^T$ are the cluster representatives or simply representatives corresponding to points of the given dimensional space, $\|.\|$ stands for the Euclidean distance, x_i is the ith element of the dataset $\chi = \{x_1, x_2, \ldots, x_N\}$ and $u_{ij} = 1$ if x_i lies closer to θ_j; otherwise $u_{ij} = 0$ [170]. In this case the dataset χ is the complete range of instantaneous phase differences for each pair of EEG electrodes as a function of time and averaged over a particular frequency band of interest calculated as explained in section 3.2.1.

The two major factors that can affect the *k-means* algorithm and consequently may have an impact on its performance are: the choice of the initial centroids and the estimation of the number of clusters. *k-means* cannot guarantee the convergence to a global minimum of the cost function, returning instead

the clusters corresponding to the local minima [170]. To avoid this initialisation issue, several initialisation methodologies have been proposed. Hartigan and Wong proposed a method based on the nearest neighbour density, Milligan used the results obtained by means of agglomerative hierarchical clustering or the popular *k-means++* which carefully selects the initial centroids following a simple-probability approach [169]. In this work, the criterion adopted to circumvent this handicap is to establish a number of random initialisations for each one of the cluster numbers selected to run the clustering algorithm. The best results of the *k-means* algorithm for each choice of k are selected from the n different random initialisations. The number of random initialisations were selected as 10, 50 and 100. The choice of 3 different numbers of randomisations were considered to study the influence of the initial centroids estimation on the final result of the clustering algorithm.

Table 3.1: k-means clustering algorithm pseudocode.

k-means clustering algorithm
1. Select number of clusters range, m=[2 10]
2. Repeat for each m_i
• Repeat for each n_j *(n=1 to 10, 1 to 50 or 1 to 100)*
- Random initialisation of initial centroids
- Form clusters by assigning each point to its closest centroid (cost function $J(\theta, U)$)
- Re-compute the centroids
• Until convergence criterion is met
• Select and storage the minima of $J(\theta, U)$
3. Plot $J(\theta, U)$ versus m
4. Select the m_i value showing the most significant "knee".

In order to deal with the second problem, the selection of the number of clusters, an initial range of possible clusters $m = [m_{min}, m_{max}]$ that can define perfectly the data set χ is defined [170]. This initial range is set between 2 and 10 clusters. For each one of this possible range of clusters, the algorithm is randomly initialised n times, calculating and saving the minimum value of the cost function $J(\theta, U)$. The simplest way to estimate the right number of clusters is by plotting the stored values of the cost function against the corresponding number of clusters m. If the plotted graphics shows a significant local change, popularly known as significant "knee", at a clustering number m_i, it can be said that the optimal number of clusters for the studied dataset will be m_i. The absence of a significant "knee" on the graph is a clear indicator of the non-clustering structure of the particular dataset [170]. Another issue than can emerge when using this methodology to determine the right number of clusters is the possibility that more than one local change or "knee" can

appear in the graphic of the cost function versus the number of clusters m. In this case the convention followed within the machine learning literature is to select the earliest and most prominent "*knee*" as the likely one to determine the right number of clusters [134, 154]. The steps of the *k-means* clustering algorithm described in this section are listed in table 3.1.

Prior to the application of the incremental clustering algorithm to the instantaneous phase differences dataset, a process of unwrapping needs to be done. It is demonstrated that the phase is circular in nature, consequently phase differences are circular too. The wavelet based instantaneous phase differences should always be between $\pm \pi$ to avoid this problem [134]. In addition, a normalisation process is performed across all of the electrode pairs by means of the maximum and minimum values of the instantaneous phase difference. As a result of the normalisation process, all the transformed values will be within the range [0,1]. After these unwrapping and normalisation steps, the instantaneous phase differences are ready to feed into the clustering algorithm.

The dataset χ is formed from all the instantaneous phase differences as calculated in 3.9, $\chi = \{\triangle\varphi_F\left(t_1\right), \triangle\varphi_F\left(t_2\right), \ldots \triangle\varphi_F\left(t_n\right)\}$. Once this dataset is unwrapped and normalized it is clustered along each time instant t to investigate the possible underlying patterns within a specific frequency band. The clustering algorithm results yield a right number of clusters k, those minimising the cost function, and for each one of these clusters, information regarding the centroids and cluster labels is saved. The cluster labels with a length of n, one label for each time instant t, hold information about the state transitions. Whereas the centroids give the averaged information for each one of the k states defined by the clustering algorithm [134].

Using this information, two types of graphics can be drawn. On one side, the clustering labels for the k different states explaining the dataset can be plotted versus the time instants $t = \{t_1, t_2, \ldots, t_n\}$ to explain in which temporal instant each state occurred and the transitions of such states along time. On the other hand, the clustering centroids can be used to translate the unique states into topographic maps. To outline these head topographies, firstly an average of the phase difference matrices is calculated. As it is a symmetrical matrix the average can be done equally, taking a row or column average. Each value of this averaged step will be assigned a colour after a normalisation process by means of maximum and minimum values. The assignation of the colours is magenta tones for higher values, meaning larger averaged phase difference, and orange for values showing a lower phase difference with the rest of the electrodes.

3.3   Results

The EEG data were collected from 10 participants, 8 males and 2 females with an age range between 20 and 53 years (mean age of 31 ± 10.01) during a motor imagery task with emotional schematic face as stimuli. Data were recorded from 62 EEG channels at a 256Hz sampling frequency. Recordings were online averaged and filtered with a notch filter (50Hz) and a Butterworth filter (0.5-100Hz). The motor imagined tasks was squeezing the ball with the right or left hand when the schematic face showed happiness or sadness respectively as shown in figure 3.1. Posteriorly, the complete dataset was visually inspected to eliminate those trials with artifacts, baseline corrected and divided into epochs for the two task conditions, Thinking Right and Thinking Left. The epoch length was 1s, starting from 100ms pre-stimulus to 900ms post-stimulus.

The following step, once all the epochs with artifacts are rejected, is to obtain the instantaneous phase difference between each one of the pair of electrodes used for the register of the EEG as indicated in section 3.2.1. The CWT is applied to the dataset and the instantaneous phase for each electrode is obtained as the argument between the imaginary and real parts of the transformed signal. The phase difference between a particular electrode for each time instant and frequency bin is obtained by subtracting the instantaneous phase of this electrode from the rest of the electrodes. This procedure yields a square symmetric matrix as shown in figure 3.3. These matrices are then averaged across the number of trials under consideration.

The study of MI tasks, in accordance with the literature, is always performed in a specific frequency band way. For this reason the set of instantaneous phase matrices is averaged across all the frequency bins within a chosen frequency band of interest as indicated in 3.9. Repeating this averaging step for each time instant of the epoch length results in a set of time-dependant matrices as shown in figure 3.4. The frequency bands selected to perform the described algorithm are the alpha (α), beta (β), theta (ϑ) and gamma (γ) bands. Refer to table 2.2 for more information regarding these frequencies. The δ band is not considered for this section as it does not present useful information for motor imagery tasks as it gradually diminishes with the age of the subject. However, it is an important biomarker of a functional deficit of cerebral coordination and therefore as a state of functional disconnection in patients suffering from schizophrenia [171].

The results in this section are presented firstly as an average across participants and secondly as individuals belonging to a population. For the first

case, the averaged instantaneous phase matrices for each participant were calculated by averaging across trials. Posteriorly, a second average was performed across all subjects and time instants within a frequency band of interest. For the second case only those averaged across trials are needed. After this step the procedure for both cases is the same, running the *k-means* clustering algorithm explained in section 3.2.2. The method of using a population average and also considering this as an individual participant gives a wider point of view regarding the temporal evolution of the phase differences and their interrelation.

The clustering algorithm applied to the dataset yielded a small number of clusters explaining the underlying phase information along the length of the epoch of the EEG data. The particularity of this finite number of clusters is that they remain stable in the order of milliseconds, they are phase synchronised, and suddenly change to a new, completely different configuration where they remain stable again. This is clearly identified when the cluster labels are plot versus the time instants t. This transition plot reflects the time evolution of the states resulting from the clustering, when they remain stable and when they switch to the next state. These semi-stable states were named synchrostates, described for the first time by Wasifa et all at [134]. During this work the nomenclature developed in that study will be followed.

3.3.1   Averaged population

In order to select the optimal number of states underlying the dataset the incremental *k-means* algorithm was run for an incremental number of clusters. Figure 3.5 illustrates the results of performing the clustering algorithm in order to calculate the optimal number of clusters k for the four frequency bands under consideration (ϑ, α, β and γ bands) and the two conditions or MI tasks performed by the complete population (Thinking R and Thinking L for the right and left hand imagined movements respectively). The value of the cost function, J, is dependent on the number of clusters and the dataset under consideration. A higher value indicates a less compact cluster. At this point it is worth mentioning that the absolute value of the cost function $J(\theta, U)$ is not relevant. The value however, of the number of clusters m at which the cost function presents a minimum value is the important parameter indicating the optimal number of underlying clusters [23].
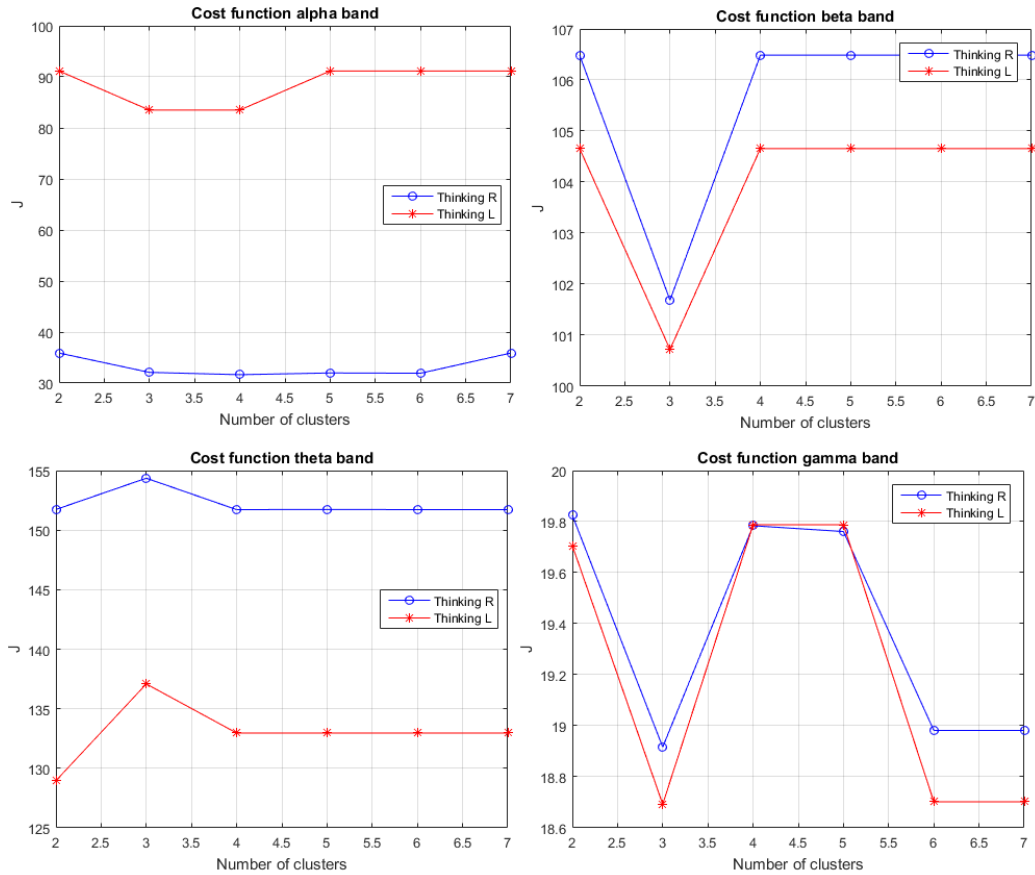
Figure 3.5: Determination of the optimal number of underlying clusters k for an averaged population. The different plots show the clustering algorithms for the four frequency bands considered for this study. For each frequency band the algorithm was run for each one of the conditions or MI tasks, Thinking R and Thinking L.

The election of the optimal number of clusters will be determined by the first and most significant local minima or "knee" shown in the picture and as discussed in section 3.2.2. According to this selection criterion, the number of unique states for the α, β and γ bands are determined by the first local minima of the cost function $J(\theta, U)$ at $k=3$. In the case of the ϑ band, the knee is not as clear as in the other three frequency bands studied. The first minimum occurs for $k=2$, however it can be argued that the most clear knee occurs at $k=4$. A higher number of clusters will increase the complexity of a dataset that can be explained with only two clusters. The larger the number of clusters used, the more likely that at least one of the "physical" clusters will be split into two or more "sub-clusters" [170]. Under this consideration, the optimal number of clusters for ϑ band is set to $k=2$. This almost negligible variation in the number of states across the frequency bands, between 2 and 3, may be explained by the fact that they represent different background processes

executing in parallel but that may not be related to these particular cognitive task [134].

The existence of a finite number of synchrostates with a small variability across frequency bands, between 2 and 3 for both conditions, may lead to the confirmation of the existence of synchrostates for the explored bands.

Once the optimal number of clusters has been decided, the cluster centroid points for the selected k are used to generate head plot topographies for the different states determined by the clustering algorithm. This will graphically illustrate how the electrodes having a similar difference in phase are connected across different brain regions. The values are normalised between 0 and 1 for visualisation purposes; where values near to 1 (magenta tones) mean a higher phase difference for this electrode with respect to the rest of the electrodes. By contrast, lower values (closer to 0, orange colours), indicate that the electrode has smaller phase difference to the rest of the electrodes. In order to obtain these phase difference contour style topographies, firstly it is necessary to reduce the cluster centroid matrix dimensions for each state into an averaged phase difference; for example row-wise. This highlights the fact that these topographies are not the standard qualitative EEG plots showing an averaged power over the scalp. Both graphics are completely different concepts. A standard qualitative EEG plot reflects the electrical amplitude in voltage. By contrast, topographies illustrated in this chapter represent the averaged difference of phase within a specific frequency band, represented by the centroids resulting from the clustering algorithm.

In order to determine the influence of the number of randomisations from the results of the clustering algorithm, three different initialisation times were selected n=10, 50, 100. The optimal number of clusters remained the same for the three different values under test for the averaged population. However, the clustering centroids change with the number of initialisations as can be seen from figures 3.6 and 3.7 showing the clustered synchrostates topographies for the α band with 10 and 100 random initialisations of the clustering algorithm. A similar result is observed for the β band in figures 3.8 and 3.9. It can be seen that when a lower number of initialisations of the clustering algorithm is used, n=10, the differences among the topographies of the three different states are larger, with one of the states representing those averaged phase differences closer to zero (orange colours). The other two states hold the higher gross phase difference values (magenta tones).
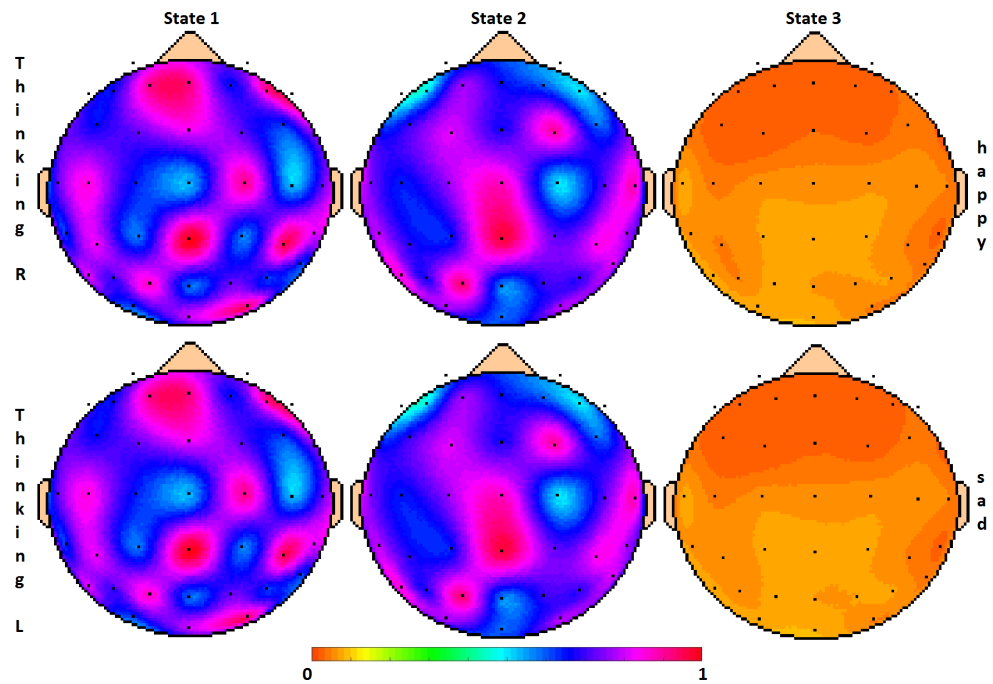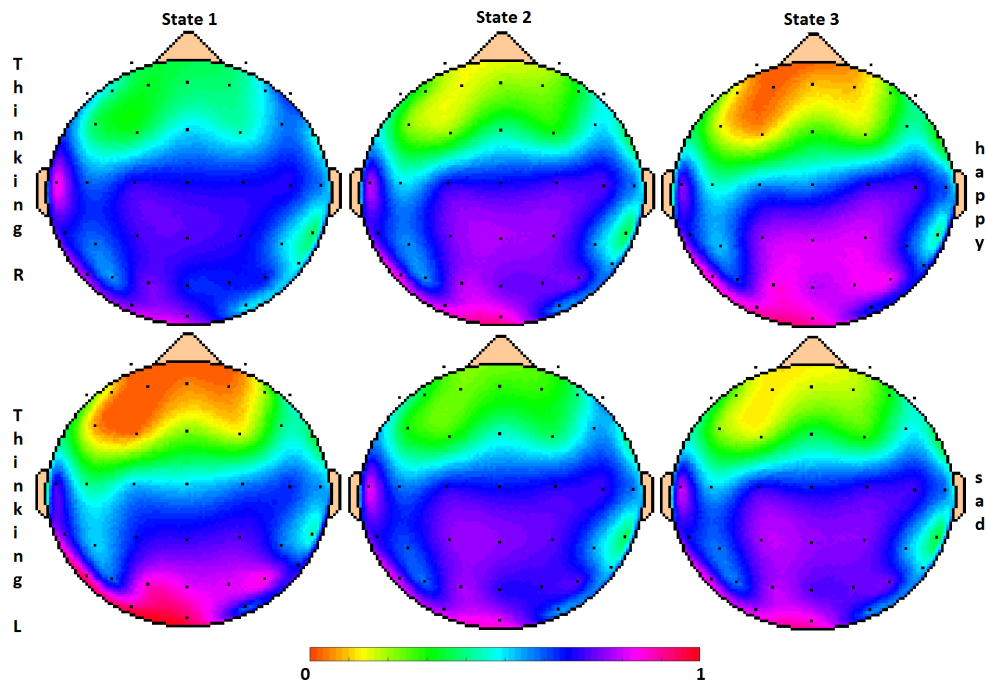
Figure 3.6: Clustered synchrostates topographies showing the averaged phase differences across the different areas of the brain related to a stimulus -specific task for the α band and n=10 random initialisations of the clustering algorithm. The headplots are illustrated for two different MI-tasks, Thinking R and Thinking L.
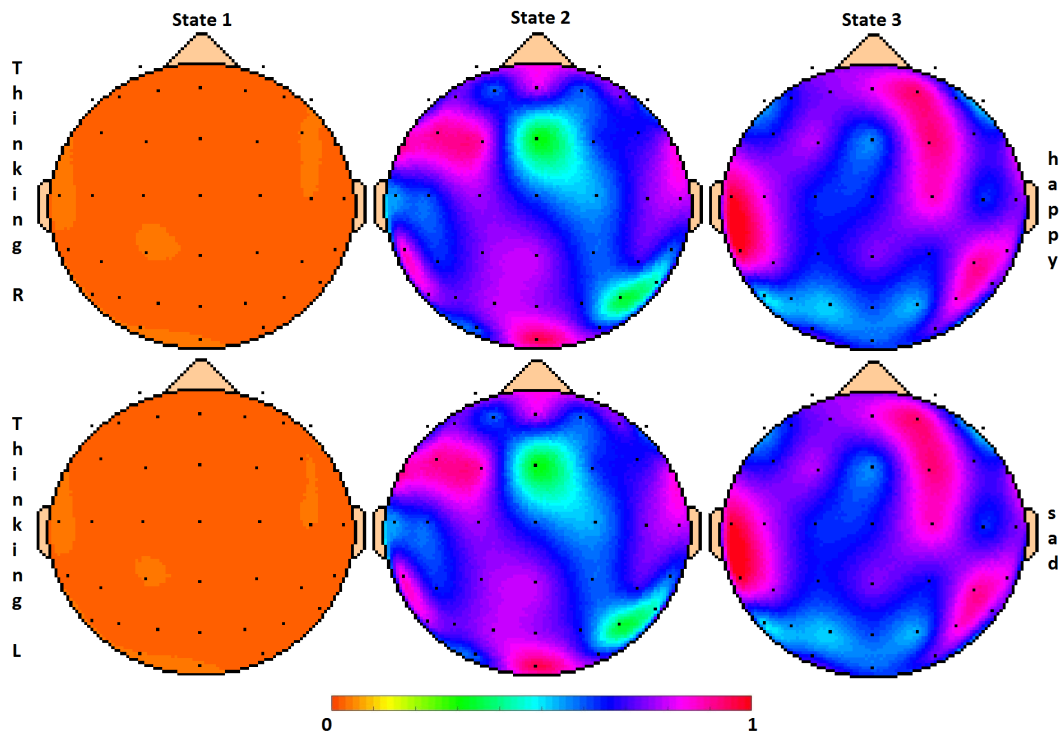

Figure 3.7: Clustered synchrostates topographies showing the averaged phase differences across the different areas of the brain related to a stimulus -specific task for the α band and n=100 random initialisations of the clustering algorithm. The headplots are illustrated for two different MI-tasks, Thinking R and Thinking L.

On the other hand, when 100 randomisations of the algorithm are performed those differences among the states within the same stimulus tend to become more diluted. Finally, the case with 50 random initialisations of the *k-means* clustering algorithm generated similar topographic plots to when the number of initialisations is set to 100. As the optimal number of synchrostates resulted in the same number across the four frequency bands, for all the randomisations, adding to the fact that the computational cost of using 50 or 100 different initial values for the representatives was considerably more expensive to lead the same number of optimal clusters and that the aim of this work is to study the temporal evolution of the states rather than the topography head maps, the selection of $n=10$ for the rest of the calculations was clear.
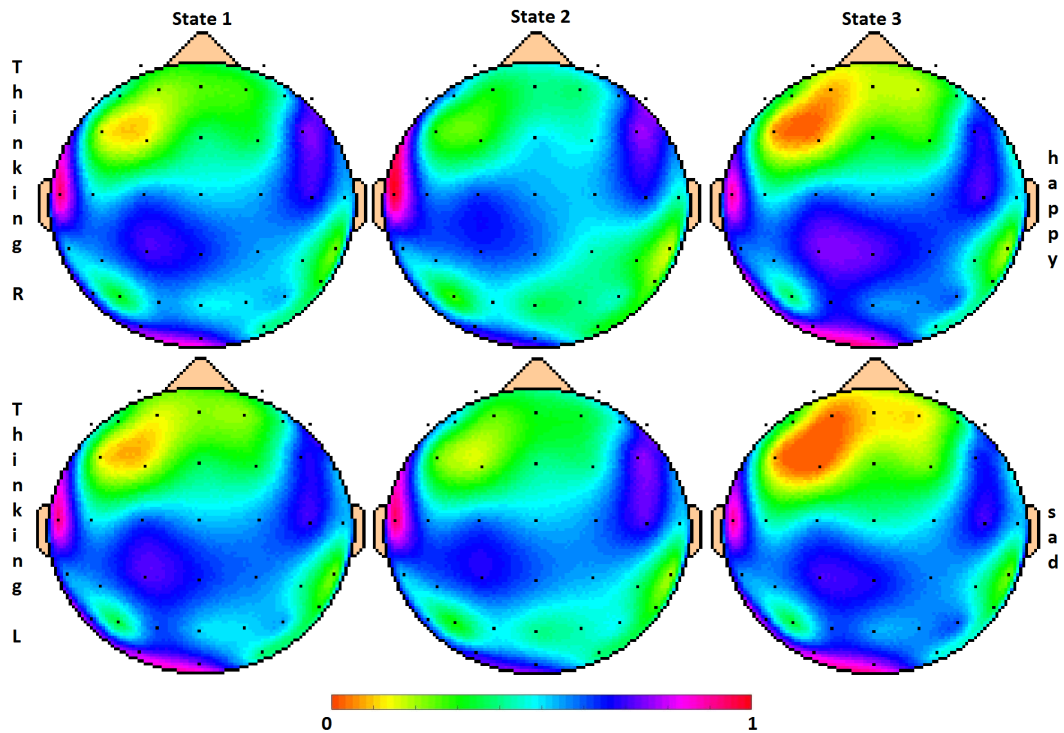

Figure 3.8: Clustered synchrostates topographies showing the averaged phase differences across the different areas of the brain related to a stimulus -specific task for the β band and n=10 random initialisations of the clustering algorithm. The headplots are illustrated for two different MI-tasks, Thinking R and Thinking L.

Regarding the synchrostates results, from figure 3.6 it is evident that the topographies of all the synchrostates are very similar for the two different stimulus based tasks in the α band. A similar result is observed for the β band in figure 3.8 and the γ band in figure 3.10 where the synchrostates plots for both conditions are analogous. A result that can be expected as both tasks, Thinking R and Thinking L, are basically within the same category. There

is a similarity for all frequency bands regarding the distribution of the phase difference values between the three synchrostates with one of the states having the lower values as mentioned before. Between the other two states, it may be appreciated that one of them contains slightly more magenta tones colours than the other (state 1 for the α band and state 3 for the β and γ bands). It is noticeable that the synchrostates topographic plots for the γ and β bands are practically a replication, only differing slightly in some of the magenta tones numerical values. This is consistent with findings in [154, 172]. The small difference across frequency bands may be due to the variability across trials, mental state of the subjects or any other similar background process independent of the tasks.
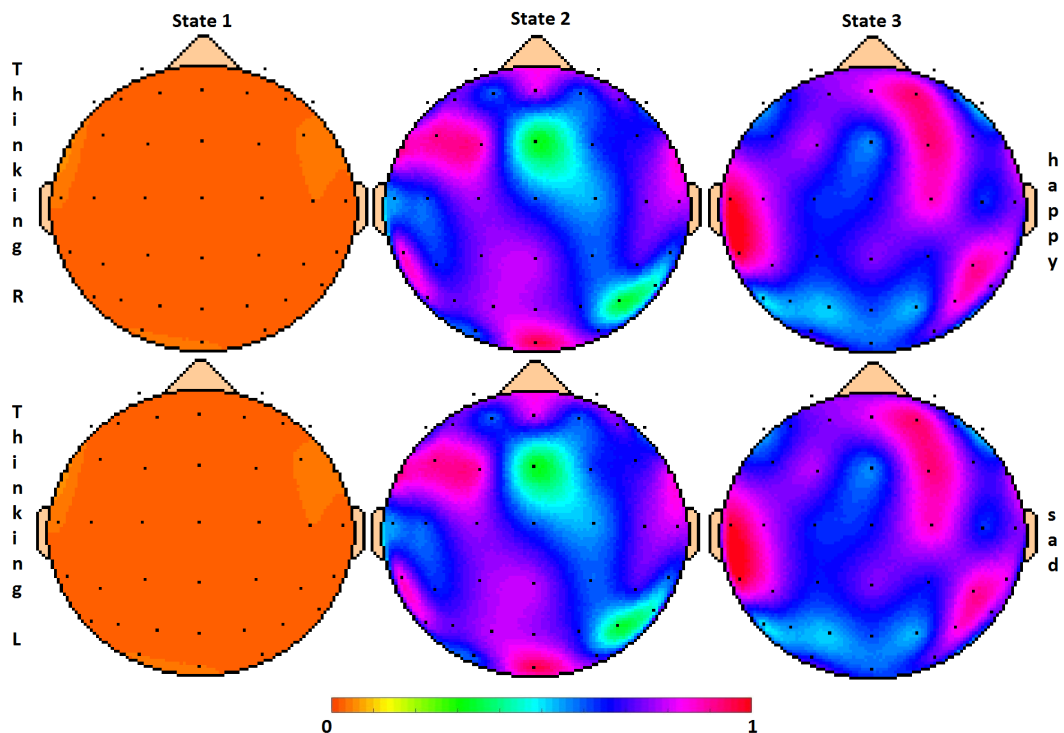

Figure 3.9: Clustered synchrostates topographies showing the averaged phase differences across the different areas of the brain related to a stimulus -specific task for the β band and n=100 random initialisations of the clustering algorithm. The headplots are illustrated for two different MI-tasks, Thinking R and Thinking L.

3.3.1.1   Exploring transition features of synchrostates

The exploration performed on the synchrostates led to a finite number of synchrostates for all frequency bands. The topographic maps of the averaged phase difference clustered synchrostates were similar across the two task conditions and between the various ranges of frequency bands studied. In this section an examination of the time transitions for each one of the three syn-

chrostates linked to a specific stimulus is performed. In addition, the number
of times that each one of these states occurs for each band and task condition
is calculated. In order to perform this investigation, the cluster labels resulting
from each condition and frequency band are plotted versus time for the whole
length of the epoch, ranging from 100ms before the stimuli onset to 900ms
afterwards. In the same way, based on the cluster labels, the number of tem-
poral instants for each state are listed for different groups of time windows.
Firstly the number of occurrences for the 1s length are computed. Then the
time interval is divided into three smaller windows: 100ms pre-stimulus, 500ms
representing the time when the corresponding stimulus is shown on the screen
and the last, 400ms where the participant is asked to perform the imagined
movement. The third group of time intervals is focused on the first 500ms
after the stimulus onset, this elapsed time is grouped into 100ms sub-intervals.
As an illustrative example of the last two cases, we refer to figures 3.11 and
3.12 for the three time windows whose divisions are indicated by the red lines
and figure 3.13 for the last case where the extra dotted red lines indicate the
100ms time sub-intervals.

Figure 3.10: Clustered synchrostates topographies showing the averaged phase differences
across the different areas of the brain related to a stimulus -specific task for the γ band and
n=10 random initialisations of the clustering algorithm. The headplots are illustrated for
two different MI-tasks, Thinking R and Thinking L.

The previously mentioned figures, 3.11, 3.12 and 3.13 represent the switching pattern over time for the α, β and γ bands respectively. For all three figures, the top graphs refer to the Thinking R condition and the bottom one, to the Thinking L condition. Contrary to the results of the head-plot maps, the transition patterns over time are clearly different between the two stimuli and also across the different frequency bands. These findings ratify the theory that the synchrostates are stimuli-specific in nature [172].
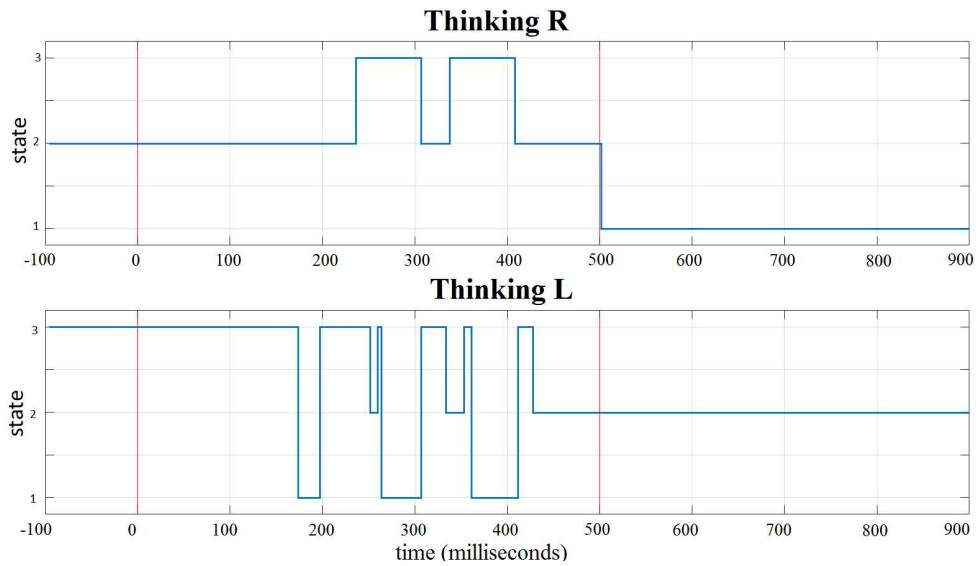

Figure 3.11: Temporal evolution of the clustered synchrostates for the α band for both conditions, Thinking R (top) and Thinking L (bottom). The red lines indicate the time instants of the stimulus onset and offset respectively.
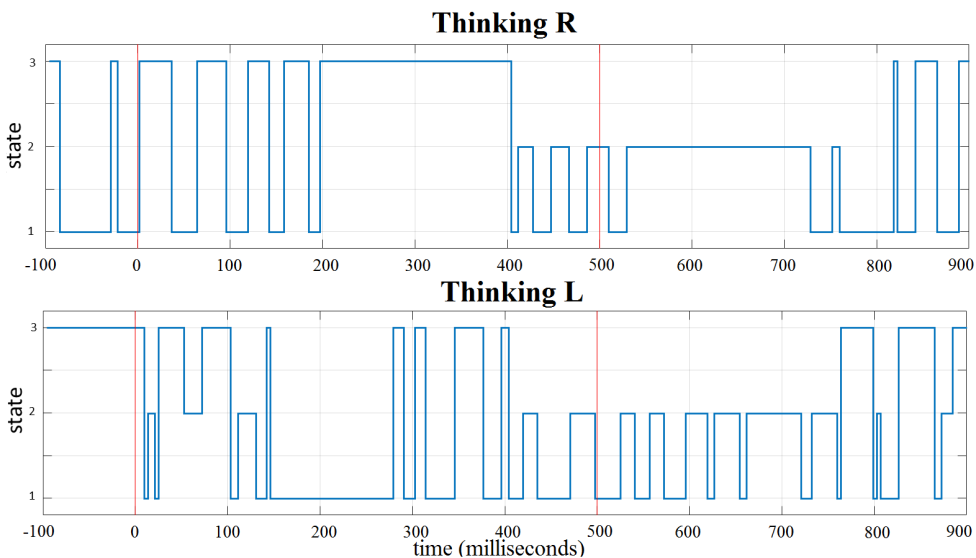
Figure 3.12: Temporal evolution of the clustered synchrostates for β the band for both conditions, Thinking R (top) and Thinking L (bottom). The red lines indicate the time instants of the stimulus onset and offset respectively.
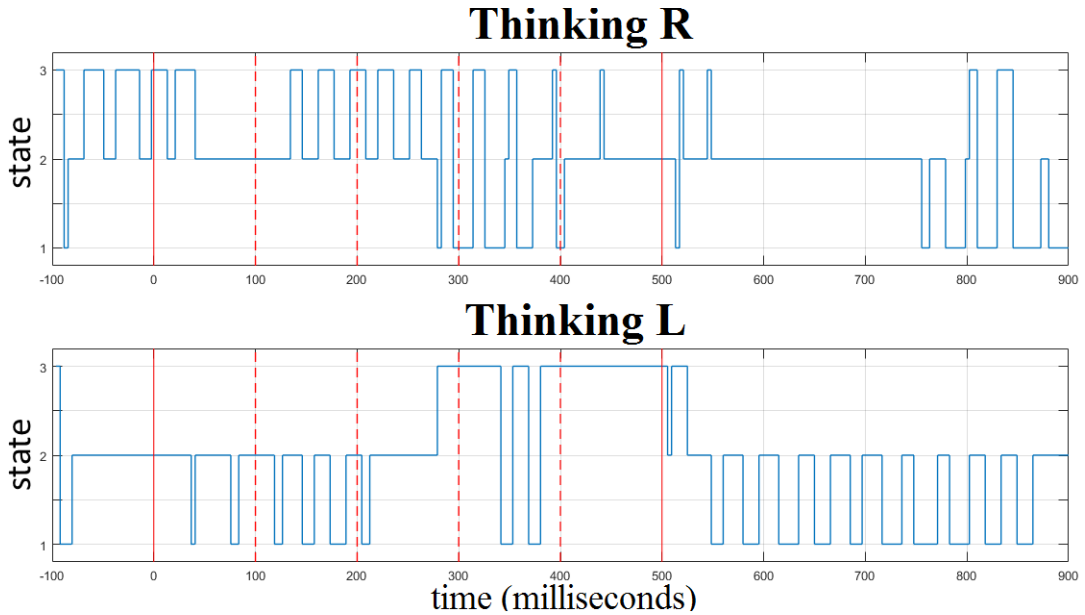
Figure 3.13: Temporal evolution of the clustered synchrostates for γ band for both conditions, Thinking R (top) and Thinking L (bottom). The continuous red lines indicate the time instants of the stimulus onset and offset respectively and the dotted red lines reflects 100ms intervals within the 500ms that the stimulus is shown on the screen to the participants.

It can be seen from these figures that the synchrostates remain stable for a few milliseconds and suddenly change to the next state. This abrupt transition between the different states is in agrement with the results concerning the behaviour of microstates in [126]. In [23] it is suggested that time duration patterns for each synchrostate may be indicative of the time required for the brain system to perform a subtask, assuming beforehand that a task can be split into a set of subtasks.

Comparing the transition plots (figures 3.11, 3.12 and 3.13) across the ranges of frequency bands analysed in this work, the reduced number of changes between states is noticeable in the α band versus the other two range of frequencies. It is confirmed when the number of transitions across synchrostates is computed, as can be observed from table 3.2. In the α band, (figure 3.11) both conditions start in different states, 2 and 3 for Thinking R and L conditions respectively. The different initial states, for the different emotional faces, is in line with previous findings [134], where face perception tasks were studied from patients suffering from Autism Spectrum Disorder and a control group. It may be explained by the different processing pathways across participants of the image prior to the stimulus onset or it may be because of any other underlying process occurring in parallel and not related to the task. However, the β and γ bands transition patterns start at state 3 for both conditions. However,

the transition scheme between frequencies soon diverges, remaining considerably different over time. In the same manner, there is a marked difference between conditions within a specific frequency band. In the α and γ bands, it can be seen from sub-tables 3.2b and 3.2c than the state were both conditions spend more time is the second one. In contrast for the β band the maximum occurrence state varies between both stimulus-linked conditions, being state 1 for Thinking L and state 3 for Thinking R.

Table 3.2: Number of occurrences for each one the three synchrostates for the α, β and γ bands with two different conditions, Thinking R and Thinking L when the complete length of the epoch is considered.

(a) α band

condition	state 1	state 2	state 3
Thinking R	103	117	36
Thinking L	30	129	97

(b) β band

condition	state 1	state 2	state 3
Thinking R	89	68	99
Thinking L	103	70	83

(c) γ band

condition	state 1	state 2	state 3
Thinking R	44	154	58
Thinking L	66	133	57

Table 3.3 illustrates the same idea as Table 3.2, but this time the number of occurrences for each synchrostate, condition and frequency band are counted by dividing the length of the one second epoch into three different intervals to study the synchrostate transition scheme evolution before the stimulus onset (-100ms to 0ms), during the time interval the stimulus lasted on the screen (0ms to 500ms) and during the period of time given to the user to perform the corresponding imagined hand movement (500ms to 900ms). It is remarkable that for the pre-stimulus interval, the transitions are almost zero with the exception of the γ band Thinking R condition where the transitions alternatively change between states 2 and 3. This behaviour may be explained by the demand for complex processing being lower during the pre-stimulus period than afterwards when a cognitive task is performed by the user.

Table 3.4 shows the number of times that each one of the three synchrostates occur in intervals of 100ms from the stimulus onset (0ms) to the stimulus offset (500ms). It is noticeable that usually only two out of three synchrostates are present in each one of the 100ms intervals for both conditions and frequency bands.

Table 3.3: Number of occurrences for each one the three synchrostates for the α, β and γ bands with two different conditions, Thinking R and Thinking L when the epoch length is divided into three different time intervals: -100ms to 0ms, from 0ms to 500ms, 500ms to 900ms. The second time interval is the total duration of the presence of the stimulus -schematic emotional face- is in the screen.

	-100 to 0ms	0ms to 500ms	500ms to 900ms

(a) α band

condition	state 1	state 2	state 3
Thinking R	0	25	0
Thinking L	0	0	25

condition	state 1	state 2	state 3
Thinking R	0	91	36
Thinking L	30	26	71

condition	state 1	state 2	state 3
Thinking R	103	0	0
Thinking L	0	103	0

(b) β band

condition	state 1	state 2	state 3
Thinking R	20	0	5
Thinking L	0	0	25

condition	state 1	state 2	state 3
Thinking R	32	13	82
Thinking L	71	23	33

condition	state 1	state 2	state 3
Thinking R	37	55	11
Thinking L	32	47	24

(c) γ band

condition	state 1	state 2	state 3
Thinking R	1	10	14
Thinking L	3	21	1

condition	state 1	state 2	state 3
Thinking R	17	75	35
Thinking L	20	56	51

condition	state 1	state 2	state 3
Thinking R	26	69	8
Thinking L	43	55	5

Table 3.4: Number of occurrences for each one the three synchrostates for the α, β and γ bands with two different conditions, Thinking R and Thinking L when the the period of the stimulus on the screen is divided into 100ms intervals.

	α band	β band	γ band

(a) 0ms to 100ms

condition\state	1	2	3
Thinking R	0	26	0
Thinking L	0	0	26

condition\state	1	2	3
Thinking R	9	0	17
Thinking L	2	7	17

condition\state	1	2	3
Thinking R	0	18	8
Thinking L	3	23	0

(b) 100ms to 200ms

condition\state	1	2	3
Thinking R	0	25	0
Thinking L	6	0	19

condition\state	1	2	3
Thinking R	11	0	14
Thinking L	19	5	1

condition\state	1	2	3
Thinking R	0	16	9
Thinking L	9	16	0

(c) 200ms to 300ms

condition\state	1	2	3
Thinking R	0	9	17
Thinking L	10	2	14

condition\state	1	2	3
Thinking R	0	0	26
Thinking L	23	0	3

condition\state	1	2	3
Thinking R	3	11	12
Thinking L	2	18	6

(d) 300ms to 400ms

condition\state	1	2	3
Thinking R	0	8	18
Thinking L	12	5	9

condition\state	1	2	3
Thinking R	0	0	26
Thinking L	13	0	13

condition\state	1	2	3
Thinking R	14	6	6
Thinking L	6	0	20

(e) 400ms to 500ms

condition\state	1	2	3
Thinking R	0	24	1
Thinking L	2	19	4

condition\state	1	2	3
Thinking R	12	13	0
Thinking L	14	11	0

condition\state	1	2	3
Thinking R	0	24	1
Thinking L	0	0	025

3.3.2   Individual subjects

The results presented in the previous sections, topographic maps and transitions between synchrostates plots, are based on the averages across participants' phase differences. The aim of this section is to explore the inter-subject variability regarding the optimal number of synchrostates across the diversity of frequency bands under scrutiny. The procedure is exactly the same to the one followed for the averaged population case. For each one of the participants, the phase difference matrices were obtained as explained in 3.2.1. Then, the iterative incremental *k-means* clustering algorithm was applied to the set of matrices for each individual instead of on the averages as before. The clustering algorithm explained in section 3.2.2 is run for the ten participants, both conditions and the four frequency bands. The results for each one of these combinations is analysed based on statistical methods such as median and quartile ranges to illustrate the variability in a box plot.

Figure 3.14 shows the number of optimal synchrostates for each one of the 10 participants in the study for both conditions and all frequency bands. The box plot indicates with a red cross the excluded outliers. The blue box reflects the inter-quartile ranges. The solid red line in the middle of the blue box is the median and the black dashed lines indicate the maximum and minimum values for each case. It can be seen from the figure that the number of synchrostates remains between 3 and 6 for this assembly of subjects, for both conditions and three out of four of the frequency bands. In the case of the ϑ band the finite range of synchrostates lies between 3 and 6 for the Thinking R condition but a reduced range for the Thinking L, between 2 and 4. According to these numbers, it can be said that the variability in the optimal number of synchrostates for the individual subjects when compared with the averaged case results is not significant. These slight variations can lead to the idea that the number of synchrostates is subject-specific. Although this may be explained by the fact that the processing pathways for each individual are different.

The reduced number of synchrostates for each individual participant is in accordance with the results achieved for the averaged population case. Therefore it reinforces the theory of synchrostates.

3.3.3   Volume conduction effect and artefacts

Volume conduction is an undesired effect that can affect the EEG recording analysis in an undesirable way as mentioned previously in section 2.5.2. Even

if the EEG recordings have been carefully inspected to reject all possible trials affected with artifacts and the participants were asked to blink, if possible, in a specific time range within each trial to minimise its effect, it is still worth verifying their non-existence before going further in this research.
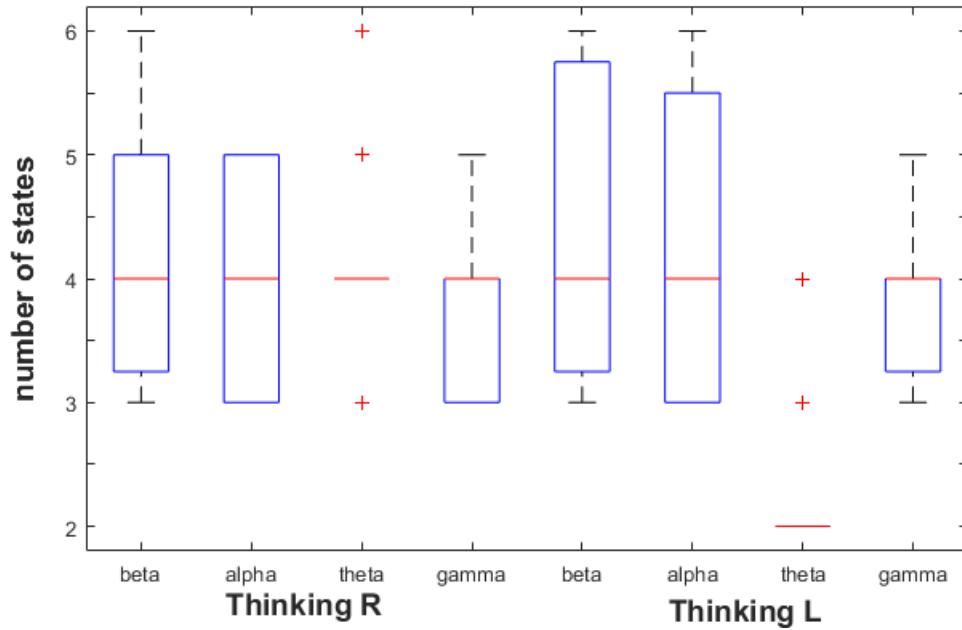
Figure 3.14: Variability in the optimal number of synchrostates for each subject during two MI-tasks, Thinking R and Thinking L. The groups were formed of 10 subjects and were replicated in the four frequency bands (α, β, γ, ϑ).

The strongest physiological artifacts stem from eye blinks, eye movements and muscle movements. Blinking and eye movements are associated with low frequencies (below 7Hz) and muscle movement with high frequencies (over 30Hz). The methodology applied in this chapter is based on the complex Morlet wavelet transform which performs a filtering process in the frequency range selected beforehand. It means that this artifact effect is probably minimised, particularly in the α and β bands and after a complex Morlet wavelet is applied.

The topographic head maps presented throughout this chapter, as well as the transition plots between synchrostates, are assembled over the order of milliseconds. According to the literature, artifacts occur generally over a time range of seconds. Putting together these two ideas, if artifacts were present in the data, all of the clustered synchrostates would be affected by this phenomenon and consequently all the states should present equal phase relation across the scalp electrodes [23] which is not the case as the figures in this chapter demonstrated. If artifacts cannot be present in such a small time window it can be considered that synchrostates are not affected by this unwanted

effect.

In addition, as mentioned in section 2.5.2, volume conduction is linked to zero phase delays between any two points across the scalp. Following previous studies that investigate how to minimise the volume conduction effect when measuring phase synchronisation [147, 145, 146, 154, 134], this study is only based on non-zero phase differences. These studies are based on the idea that if an intermediate point between two sources is not at phase zero it cannot be explained by volume conduction.

3.4   Conclusion

The existence of a finite number of unique phase difference patterns, denominated synchrostates, was consistent for both an averaged population and for individuals belonging to the population. These synchrostates remain stable of the order of milliseconds, then suddenly and abruptly change to the next one. The switching pattern across the different synchrostates over time gives information regarding the evolution of the phase synchrony across the different regions of the brain over time. The optimal number of synchrostates varies slightly across individual subjects, but is within a small range, from 3 to 6. This may be explained by the inter-variability across individuals as the processing pathways change among them. Therefore, the similarity in the results obtained from the averaged group and the individual subjects gives consistency to the theory of synchrostates.

The topographical maps of the clustered synchrostates showed differences between frequency bands and conditions. In addition to the head-plot map, graphics illustrating the temporal transition between the synchrostates were displayed. These resultant transition patterns are unique for each of the conditions considered in this study. These findings showing a noticeable differentiation between both MI tasks (imagined movements of the right and left hands) leading to the idea that these dissimilarities can be translated to a more quantitative metric and subsequently applied to a classification problem.

The concept of synchrostates, firstly introduced by Wasifa et al. at [134] is similar of the well-extended concept of microstates [116, 125, 120, 123, 119]. The physical explanation to the quasi-stable states, or temporary stable states, given by [23] is by assuming that a cognitive task such as the one presented in this work can be subdivided into smaller tasks. The time duration of each one of these states may mean the necessary time for the brain system to perform

each one of the mentioned subtasks, as long as the task can be divided into subtasks.

Once the temporal transitions of the synchrostates has been determined for one of the motor tasks, a further step needs to be done to quantify their temporal stability. The following chapter explores the possibilities that neural networks offer and enable us to understand the communication processes between different areas of the brain and establish a quantitative system to evaluate synchrostates.

Chapter 4

Network analysis from synchrostates

The logical step that follows in the analysis of synchrony based on synchrostates is investigating their temporal stability. As studied in Chapter 3 of this thesis, the existence of synchrostates and their temporal transitions have been demonstrated for two MI tasks. However, their temporal stability needs to be more deeply studied and in a more quantitative manner. In order to achieve a finer understanding of the synchrostates topographical maps and transitions patterns between the different synchrostates, a series of complex networks metrics, will be employed throughout this chapter. These complex network measures can yield a quantitative pathway to allow us to work with cognitive functionalities.

The transformation from synchrostates form towards complex network measures needs an intermediate step, a weighted measure of connectivity representing the temporal stability intervals (transition patterns) in conjunction with the phase difference values (head-plot maps). This matrix is commonly known as the adjacency matrix in graph-theory nomenclature and in this case will be represented by the synchronisation index. This index will provide information regarding the stability of the synchrostates as opposed to the incremental *k-means* algorithm that only provides information about the optimal number of synchrostates and their temporal switching transition.

An introduction to the concept and formulation of the synchronisation index will be provided in the coming section 4.1. Section 4.2 will introduce some of the terminology used for complex networks, commonly investigated network measures and a brief overview of some underlying principles. Finally the analysis procedure, results and conclusions will be explained.

4.1   Synchronisation Index

Once the phase difference based clusters have been identified for each stimuli, their temporal stability needs to be quantified. This is because the clustering algorithm cannot provide information relative to how long the stability of each one of the unique clusters lasts. This can be done by means of a synchronisation index which has been used before in the study of cerebral blood flow auto-regulation [161, 173] and Autism Disorders [134].

Using the definition of synchronisation index given by [161], this parameter can be understood as an inverse circular statistical analogue of variance and is described by equation 4.1 [134],

$$\Upsilon_P(F) = \frac{1}{N}\sqrt{\left[\sum_t \cos\left(\triangle\varphi_F(t)\right)\right]^2 + \left[\sum_t \sin\left(\triangle\varphi_F(t)\right)\right]^2}, \qquad (4.1)$$

where N is the number of time points associated with a specific synchrostate, $\triangle\varphi_F(t)$ is the phase difference matrices averaged across a specific frequency band obtained as explained in section 3.2.1. The synchronisation index $\Upsilon_P(F)$ lies in the interval [0,1]. Higher values of the index means that the phase differences in the specific frequency band F are in synchrony as they present low variation over time, hence, they can be considered synchronised. This index quantifies the averaged temporal stability of the synchrostates in that frequency band, which is information that the clustering algorithm cannot provide. In addition, this index is capable of capturing the band-specific temporal behaviour in contrast to coherence based measures [23].

4.2   Network analysis

After obtaining the synchronisation index for each pair of EEG channels, synchrostate, condition and frequency band these can be translated into a complex network. This process contributes by providing more information and new perspectives regarding the evolution of phase synchrony over time and across the different areas of the brain. Network theory is a hot topic area in modern science and it has been successfully applied to such diverse topics as flavours in recipes, social interactions or biomedical applications [22].

A network based on a graph theory approach is a collection of nodes and the links connecting those nodes, named edges. Figure 4.1 is a simple example

of different types of existing graph models, illustrating the difference between undirected versus directed graphs or a weighted network based on a graph theory concept. A weighted network, in the case of a sociological study can represent the strength of a friendship where stronger affection will be reflected as a thicker link. When this concept is translated to the EEG area, the nodes are represented by the electrodes used to record the brain activity. The links or edges encoding a weight and a direction of connection are characterised by the synchronisation index matrix.
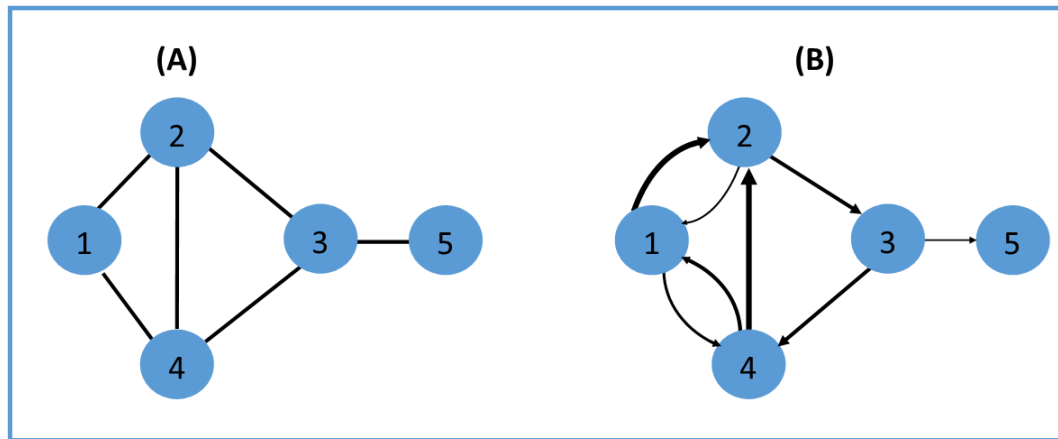
Figure 4.1: Representation of two types of basic general graphs. Nodes are represented by circles and edges by lines. (A) An example of an undirected graph with 5 nodes and 6 edges. (B) A representation of a directed and weighted graph. The thickness of the connections or edges indicates the weight, while the direction is represented by arrows.

Networks can be characterised at different levels ranging from properties explaining the whole network at the global scale to properties of the network components at a local scale. The networks measured and used throughout this study can be divided into individual network measures, measures of functional segregation and measures of functional integration [174].

4.2.1   Individual network measures

These measures are focussed on assessing the importance of individual nodes over the whole network. There are many measures of centrality. In this section only two of them are explained: degree and density.

The *degree* of an individual node is one of the most common measures used in graphs and many other graph metrics are based on the concept of degree. Therefore, degree is likely, the most fundamental network measure. It can be defined as the number of connections that link a node to the rest of the network [148]. Depending on the types of network graph, it can be obtained in

different ways. For instance, in a directed graph, the concept of degree can be divided into in- and out-degree, this being the number of links with directions towards or from the node respectively. In the case of binarised graphs, the degree is simply the sum of the links. For weighted graphs the degree can be considered as the sum of the weights of the weighted links [22].

Density is the ratio of the actual number of edges presented in the graph as a proportion of the total number of edges possible. It can be seen as the probability of existence of an edge between a randomly chosen pair of nodes [175]. It is the simplest estimator of the physical cost of a network[174], and help us to understand the physical cost of a network as its wiring cost which is directly proportional to the number of connections, length and their cross-sectional area. As these metrics are difficult to measure in a large-scale brain network, most studies approximate the physical cost by using the density value [176].

4.2.2   Measures of functional segregation

Segregation metrics in a network refer to local connectivity and are usually associated to pairs of nodes. A complete summary of the available range of measures of functional segregation can be found in [174]. Here only the concepts of transitivity and modularity will be explained as both will be used within this work.
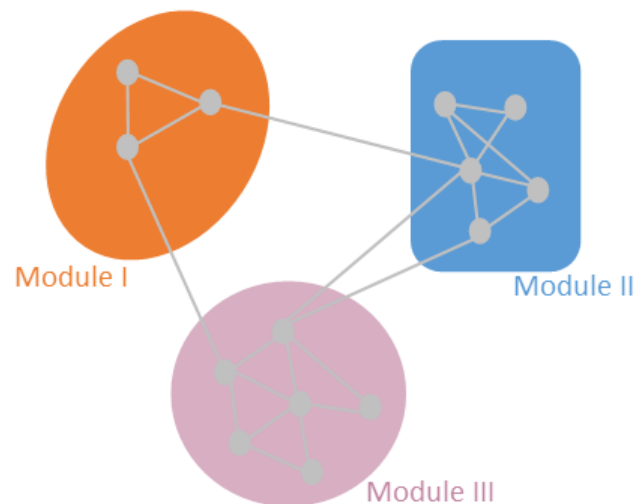
Figure 4.2: Sample graph illustrating the concept of modules in a structure.

Transitivity (T) reflects the ratio of closed triangles or triplets in the network to the total number of closed triangles possible [175]. The concept is similar to the average clustering coefficient, the difference being that, the normalisation process to obtain the transitivity value is done collectively instead of individually. Transitivity of a network can be calculated as

$$T = \frac{\sum_{i \epsilon N} 2t_i}{\sum_{i \epsilon N} k_i (k_i - 1)}, \tag{4.2}$$

where N is the set of all nodes in the network and k_i is the degree of a node i and t_i the number of triangles around a node i [174].

Modularity (Q) of a network is a reflection of the natural segregation within a network. The underlying aim is to try to determine how well a network can be separated into individual modules or communities as can be seen from figure 4.2 [22]. Modularity shows how well a given separation into modules performs. The definition of these modules is not always unique. In addition, this metric does not include information regarding how many modules exist [175]. The modularity can be defined explicitly by 4.3 [175],

$$Q = \frac{1}{l} \sum_{ij} \left[a_{ij} - \frac{k_i k_j}{l} \right] \delta_{m_i m_j}, \tag{4.3}$$

where m_i is the module containing the node i, $\delta_{m_i m_j} = 1$ when $m_i = m_j$ and 0 otherwise, l is number of links and $a_{ij} = 1$ when a link between i and j exists and $a_{ij} = 0$ otherwise [174].

4.2.3   Measures of functional integration

Measures of functional integration, also known as global measures, look at the network as a whole and in macroscale. These measures try to estimate the ease with which brain regions communicate and the capacity of a whole integration of the network [174]. Characteristic path length and global efficiency are two common measures of integration and both will be used within this investigation.

The *path* of a network is a set of distinct nodes and links representing potential routes of information flow between pairs of brain regions [174]. Path length consequently estimates the minimum number of edges that must be crossed from one to another so giving an idea of the potential for functional integration between brain regions. For instance, random and complex networks have short mean path lengths, by contrast, regular networks have long mean

path lengths [175]. Based on these two concepts the *Characteristic Path Length (CPL)* of a node can be defined as the average shortest path-length between the node and all other nodes in the network [177]. The equation to calculate the CPL for an unweighted network is given in 4.4 [22]. The mean measure of the CPL taken over all nodes of the network is referred to as a global characteristic path length.

$$CPL = \frac{1}{n} \sum_{i \epsilon N} \frac{\sum_{j \epsilon N, j \neq i} d_{ij}}{n - 1}, \tag{4.4}$$

$$GE = \frac{1}{n} \sum_{i \epsilon N} \frac{\sum_{j \epsilon N, j \neq i} d_{ij}^{-1}}{n - 1}, \tag{4.5}$$

where d_{ij} is the shortest distance between nodes i and j belonging to N, the set of all nodes in the network. *Global Efficiency (GE)* is calculated similarly to the CPL and also based on the shortest paths. In this case, however, the average is computed by selecting the reciprocals of the shortest paths as given in equation 4.5 [174].

Based on the concept of paths and path lengths another network metric can be found, *eccentricity*. This yields two important measures of complex networks, *radius* and *diameter*. The eccentricity of a given node is defined as the longest of all the shortest paths connecting it to the remaining nodes in the network [22]. Consequently, a global measure of eccentricity can be obtained by the average of all nodal eccentricity values. Once, this value is obtained the diameter and radius of the network can be determined as the maximum and minimum values of the global eccentricity respectively.

4.2.4   Small-word brain connectivity

The first time the term *small-world* appeared was in a social-network research study conducted by Milgram [178]. It was related to the idea that a person could reach any other person through a relatively short chain of acquaintances, known as the "six degrees of separation". Subsequently, Watt and Strogatz defined and described the underlying principles of small-networks problems [177]. Small-world networks are defined as networks that are more clustered than random networks but not as much as completely ordered ones [22]. It can be seen that networks representing small-world properties may be simultaneously segregated and integrated networks [174], sharing properties of both types of networks. Small-world networks generally present large local clus-

tering as completely ordered networks and short path lengths, typically char-
acteristically represented in random networks [177]. Network small-worldness
has been quantified by a small coefficient, σ, calculated by comparing cluster-
ing and the path length of a given network with a random network having the
same degree on average as described by 4.6. With this definition, a small-world
network will have $\sigma > 1$, numerator also >1 and denominator ~1 [179].

$$\sigma = \frac{clustering_{brain}/clustering_{random}}{length_{brain}/length_{random}}, \qquad (4.6)$$

There is not a clear criterion to classify a network as a small-world network.
Several proposals have been proposed to measure *"small-worldness"* [180, 181].
It has been demonstrated that small-world phenomena are found in several real
life networks as they are found in the brain [22]. The human brain is prone to
exhibit some of the features typical of small-world phenomena mentioned be-
forehand. In a recent study where small-world was applied to a neural network
machine learning algorithm, it has been proven that small-world networks have
influence in achieving greater rates of information processing [182]. For their
peculiar characteristics, small-world phenomena have become a hot topic in
the research of the human brain.

4.3   Results

The phase difference matrices result in a finite number of clustered synchro-
states after running the algorithm described in chapter 3 for each condition,
Thinking R and Thinking L and for the frequency band studied. This pro-
cedure provides information regarding the temporal evolution and transition
patterns of the synchrostates. However, it does not provide details about their
temporal stability. In order to circumscribe this issue and gain more knowledge
of the meaning of synchrostates, complex networks based on graph theory are
used. The transformation into complex brain networks was done using the
clustering results of the averaged population for each condition and frequency
band. The nodes of the network were the EEG electrodes and the synchron-
isation index $\Upsilon_P(F)$, calculated as indicated by equation 4.1, were used as the
weighted edges connecting each pair of nodes.

(a) No threshold   (b) Threshold of 60%
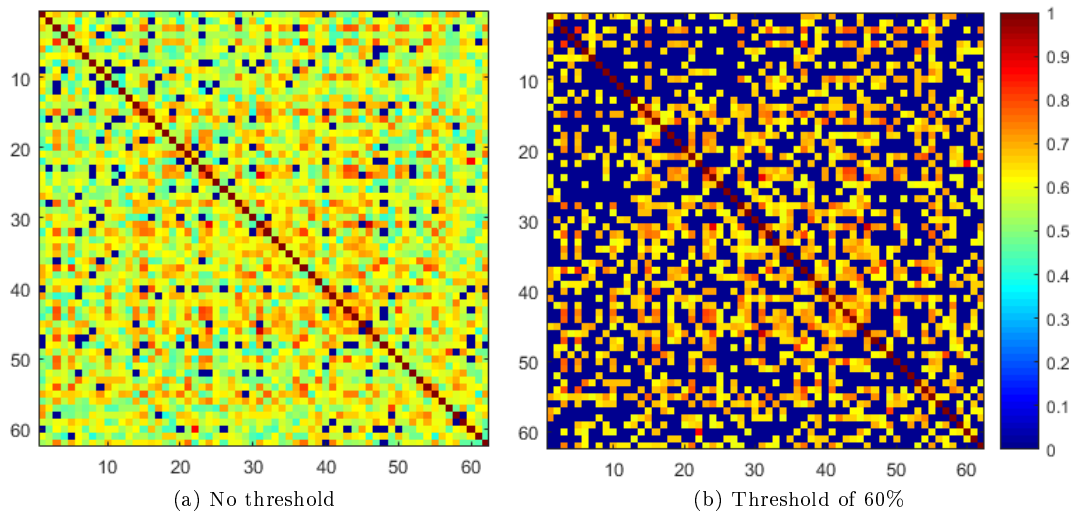
Figure 4.3: Example of the synchronisation Index $\Upsilon_P (F)$ matrices for clustered synchrostate 1, the Thinking R condition and the γ band. The left matrix is without threshold (a) and the right one is the result of using a threshold of 60% as an example of the effect of using threshold (b).

Figure 4.3 depicts an example of the results of transformation from the synchrostates into synchronisation index matrices. This particular image is of the synchrostate 1, Thinking R condition and γ band under two different circumstances; the normal synchrostate index matrix and when a threshold is applied to it. Therefore only a percentage of the data is used for further calculations. The complex networks resulting from the synchronisation matrix will vary depending on the percentage of the threshold used.

Afterwards, once the synchronisation index matrices $\Upsilon_P (F)$ have been obtained for each condition and frequency band, they can be used as the weighted values for the edges between pairs of nodes. This will contribute to a further understanding of how well connected the different pairs of electrodes or nodes are across the areas of the brain and also about the temporal stability of each state. All the brain connectivity plots and values of the networks metrics have been obtained using the EEGNET [183] free software. Complex network figures have been plotted using a 5% threshold for ease of visualisation, meaning only 5% of the most highly connected edges are considered to plot the graphics. The colours and sizes of the nodes are based on their degree, meaning that a bigger diameter of the node contains a higher number of links connected to this node. In the same way, reddish colours of the nodes signify more links connected to the node; Bluish, a lower number of links connected to it. Similarly, the edges thicknesses are based on the weighted values of the synchrostate index matrices. Consequently, thicker lines connecting electrode pairs mean

higher values than thinner ones of the adjacency matrix. The results of these complex network plots are shown on figures 4.4, 4.5 and 4.6 for the α, β and γ bands respectively. Each one of the figures contain the head plots of the three clustered synchrostates resulting from the incremental *k-means* algorithm and both conditions, Thinking R and Thinking L. From the figures, it can be seen that two channels, Tp9 and Tp10, are located outside of the physical area of the brain. This is due to the plotting function of the software used, EEGNET, that performs a 2D projection directly over the head outline.

It can be observed that the synchrostates present clear differences among them, across conditions and frequency bands, according to the findings described in chapter 3. From comparison across frequency bands, it can be said that the γ band connectivity plots are less dense than for the other two bands. This agrees with the literature regarding MI based BCI, where the α and β bands are those most commonly used for representing higher activity during a MI task.
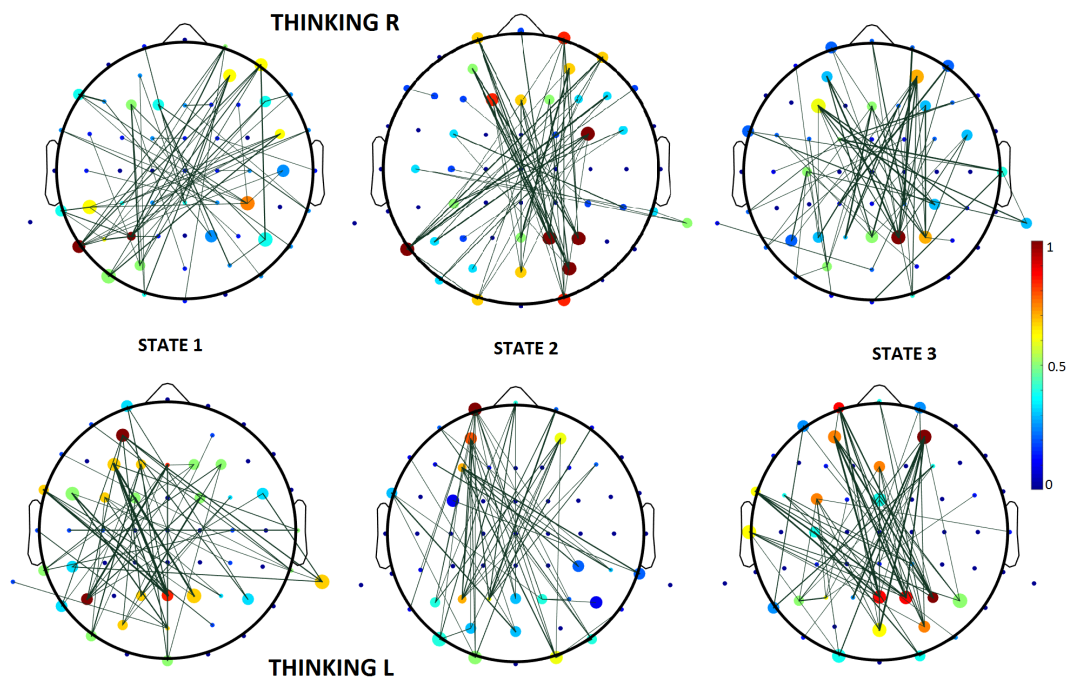
Figure 4.4: Brain connectivity plots from the three clustered synchrostates for both conditions, Thinking R and Thinking L, the α band and a threshold of the 5% highest connected edges.
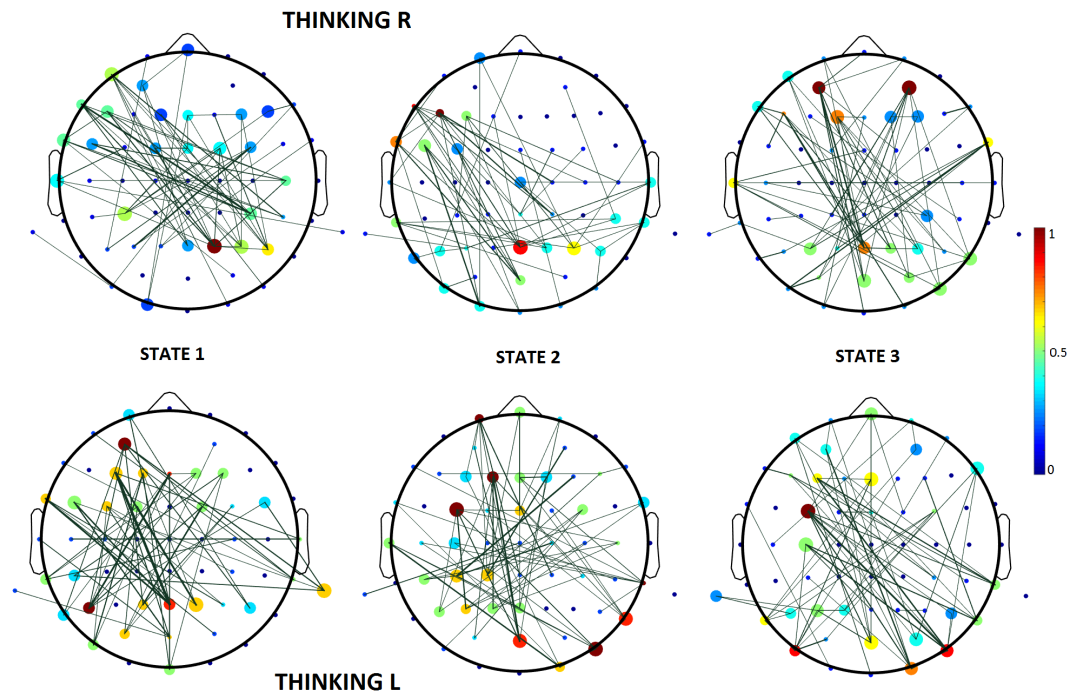
Figure 4.5: Brain connectivity plots from the three clustered synchrostates for both conditions, Thinking R and Thinking L, the β band and a threshold of the 5% highest connected edges.
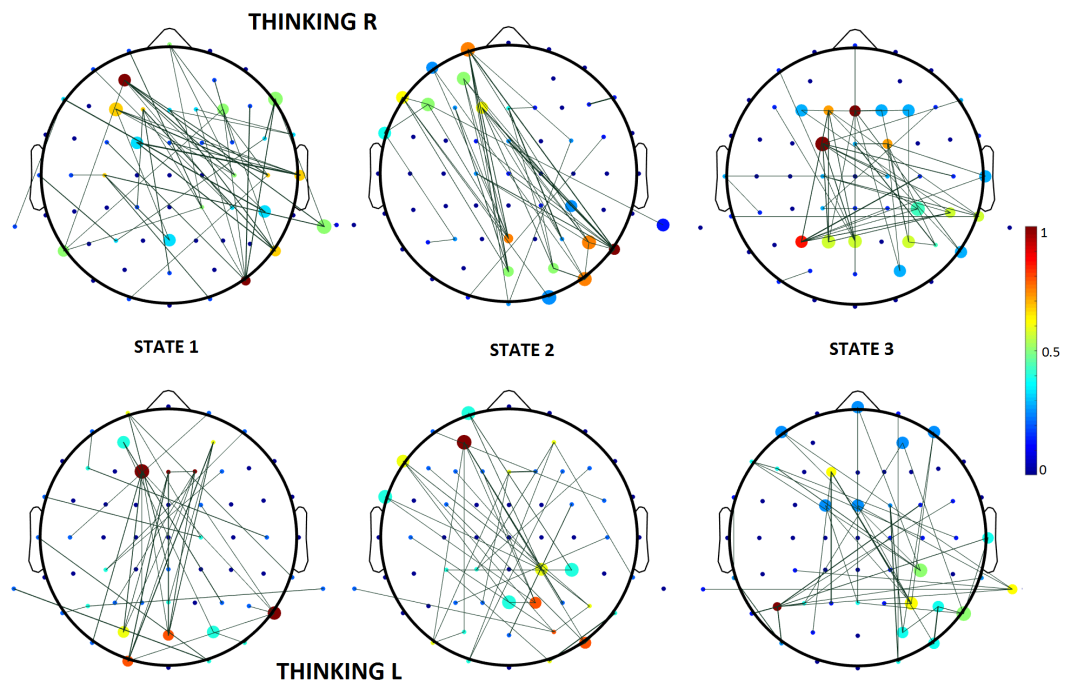
Figure 4.6: Brain connectivity plots from the three clustered synchrostates of both conditions, Thinking R and Thinking L, the γ band and a threshold of the 5% highest connected edges.

For the α band, state 3 displays some similarities for both conditions, where the stronger connections are between parietal and frontal electrodes. States 2 are clearly differentiated in both cases. For the Thinking R task there is a clear dense region in the right hemisphere, with parietal electrodes showing a higher degree. This can be seen from the presence of reddish, larger circles. By contrast, the Thinking L condition is not as densely connected with most electrodes presenting bluish colours, indicating that the number of links connecting these electrodes is smaller.

Table 4.1: Network metrics for the each of the clustered synchrostates for the two conditions, Thinking R and Thinking L, and the α band.

Network measures	Thinking R			Thinking L		
	state 1	*state 2*	*state 3*	*state 1*	*state 2*	*state 3*
CPL	4.4539	3.3924	4.3540	4.0154	3.2635	3.3616
G.E.	0.1404	0.0823	0.1422	0.126	0.1047	0.1145
Diameter	11	10	10	9	7	7
Highest degree	9	7	11	7	11	9
Density	0.0682	0.0624	0.0672	0.0624	0.0624	0.06272
Num. of edges	129	118	127	118	118	127
Modularity (Q)	0.6917	0.6496	0.6812	0.6539	0.5763	0.57
Transitivity (T)	0.8213	0.7401	0.7009	0.7984	0.6246	0.6443
Num. of components	14	28	16	23	29	27

For the β band state 1 and Thinking R condition, the nodes with a higher number of links, are those placed in the frontal area of the brain. This contrasts with the Thinking L condition where the higher nodes are distributed between the frontal, parietal and occipital areas of the left hemisphere. State 2 exhibits a lower number of connections for both conditions when compared with the other two states. State 3, the Thinking R condition, shows two dominant electrodes in the frontal area sharing connections with the parietal area of the right hemisphere.

However, the connectivity graphs (figures 4.4-4.6) cannot offer any measurable and evident conclusions regarding the differences across synchrostates, so a more quantitative methodology is needed. To this end, a series of connectivity metrics is calculated. Together with the complex network head plots, the measures mentioned in section 4.2 are obtained once more for each synchrostate, MI task and frequency band by means of the EEGNET software [183]. The measures were obtained using a binarised network with a 5% threshold. As mentioned before this type of network only report a connection if the weight value is above the threshold, otherwise it is reported as an absence of connection [175]. The estimation of the right threshold is a crucial parameter as this

factor directly affects the resulting network. There has been a long debate on how to fix the threshold with no consensus on the solution for the optimal way to set this threshold [184]. A range of thresholds from 3 to 10% was examined as they are the most widely used in the literature. However, as the main concern is to standardise the comparison across the different synchrostates and conditions, a threshold of 5% was selected.

Tables 4.1 to 4.3 represent a list of the measures obtained with in the Matlab$^{(C)}$-based tool EEGNET for the α, β and γ bands respectively. Noteworthy is the fact that EEGNET calculates the percentage of the highest connected edges from those elements of the adjacency matrix that are different from zero, not from the total number of nodes. The metrics are in agreement with the connectivity head plots. For instance, the G.E. value from table 4.1 for the Thinking R condition is a maximum for State 3 whereas for the Thinking L condition it is State1. In addition, modularity values for the Thinking R condition remain almost stead, contrasting with the values of the Thinking L condition yielding values of a order lower. In both cases the dominant state with maximum values is State 1, this may mean that for both cases State 3 is the one reflecting the minimal specialised segregate processing [134]. The diameter is remarkably lower for the Thinking L condition than for the R condition and it is also noticeable that for both MI tasks, States 2 and 3 present the same value which is smaller than the one for state 1. A smaller diameter indicates a more robustly connected network.

From the β band, table 4.2, a similar situation is shown when the CPL metric is observed. States 2 and 3 give the maximum values for the Thinking R condition, which reflect a maximum information integration ability for these two states. However, for the Thinking L condition, the state showing considerably lower values of CPL and T is State 3, when compared with the other two which may indicate that the minimum integration information takes place in this state for the Thinking L MI task. The behaviour of the diameter value for the Thinking L condition is similar to that found in the α band, with States 2 and 3 yielding the same value which is lower when compared with State 1. However, for the other condition the states presenting lower and identical values are States 1 and 2.

It can be seen from table 4.3 that State 1 presents the maximum CPL for Thinking R condition and the minimum value for the other one with a noticeable difference in values. Similar behaviour can be seen from the global efficiency metric, but with inverse values. Also noticeable is that the differences between diameter values across states and between conditions are clearly

reduced in comparison with the other two frequency bands.

It is evident from figure 4.4, corresponding to the α band, that State 2 of the Thinking L condition, associated with sad faces, shows a clear difference to the rest of the network parameters from the whole set of synchrostates. Its lowest transitivity, minimum CPL, a high G.E. and minimum diameter lead to the theory that participants need higher attention and processing information to understand and distinguish a sad face or the left hand MI task. This confirms that the process of information integration is task specific in nature [23]. Similar behaviour is shown for State 3 of the Thinking L condition in the β band (table 4.2) and State 1 again for the γ band (table 4.3).

Table 4.2: Network metrics for each of the clustered synchrostates for the two conditions, Thinking R and Thinking L, and the β band.

Network measures	Thinking R			Thinking L		
	state 1	state 2	state 3	state 1	state 2	state 3
CPL	4.0947	4.3663	4.3553	4.4131	4.2138	3.8054
G.E.	0.1487	0.1435	0.1461	0.1208	0.1673	0.1337
Diameter	10	10	11	12	9	9
Highest degree	12	9	9	8	7	9
Density	0.0672	0.0629	0.0672	0.0619	0.0672	0.0672
Num. of edges	127	119	127	117	127	127
Modularity (Q)	0.6384	0.6728	0.7043	0.6996	0.6969	0.64
Transitivity (T)	0.7127	0.8088	0.8	0.8643	0.8013	0.7508
Num. of components	19	18	16	21	14	20

Table 4.3: Network metrics for each of the clustered synchrostates for the two conditions, Thinking R and Thinking L, and the γ band.

Network measures	Thinking R			Thinking L		
	state 1	state 2	state 3	state 1	state 2	state 3
CPL	4.4850	3.4989	3.6708	2.9444	4.2125	4.4462
G.E.	0.0997	0.072755	0.0953	0.0371	0.0534	0.1056
Diameter	10	8	9	8	9	9
Highest degree	7	9	8	6	6	9
Density	0.0571	0.0576	0.0576	0.0523	0.0523	0.0571
Num. of edges	108	109	109	99	99	108
Modularity (Q)	0.7141	0.643081	0.6762	0.7852	0.7956	0.7103
Transitivity (T)	0.8942	0.8099	0.8127	1.0486	1.143	0.8793
Num. of components	24	32	27	32	29	24

4.4   Conclusions

In the previous chapter, the presence of synchrostates was detected for the first time during the execution of different MI tasks. In this chapter the combination

of synchrostates and their temporal stability, by means of a synchronization index, gave us more insights about the brain functions of segregation and integration within the human brain for a specific task. This combination, used for first time for the study of MI tasks can pontentially lead to a more person-specific characterisation of the brain and its functioning and consequently, to a more personalised BCI systems.

Accordingly, the findings from the translation of synchrostates information into complex brain networks for both imagined tasks has been confirmed that the information integration in the brain is task-specific. These results are in line with previous findings [134], hence, validating the use of the synchrostate information to perform a comparison between two MI tasks. The differences of the values across the synchrostates and between conditions of the main network metrics, such as modularity, transitivity or CPL, may be useful as biomarkers to differentiate between both motor imagery tasks to control a MI-based BCI. However, extended analysis needs to be done regarding the possibility of using graph theory metrics as classification features for individual participants. The next chapter investigates the ability of connectivity network metrics to distinguish between the two proposed motor tasks.

Some recent studies have been undertaken using graph theory metrics, especially small-world and motif properties, to analyse the dynamic brain networks of the brain based on EEG recordings [21, 185, 186]. In addition, some more investigations based on spectral power analysis in conjunction with a network approach, trying to identify time-frequency dynamics [187, 188]. The main difference with respect to these studies and the present one is that this work explores the temporal evolution of clustered synchrostates linked to a specific cognitive task.

Chapter 5

Classification of MI tasks from synchrostates

In chapter 3 the concept of synchrostates associated to a specific stimulus was introduced together with their transition patterns over time and head map topographies. In order to gain more insight into their temporal stability, in chapter 4, the synchronisation index was used as a mechanism to translate the information from the clustered synchrostates into complex networks based on graph theory. Both connectivity head maps and network metrics, applied to synchrostates, give more quantitative information regarding the information flow across the different areas of the brain. The graph theory metrics for each synchrostate demonstrated that it is a good indicator of how the underlying connectivity works, how the local and global segregated and integrated information is distributed over the synchrostates for a specific task and frequency band.

Network metrics and connectivity plots have recently been used to explore and understand how the brain works within pathological conditions and to try to use this information to find distinctive features to differentiate them from control populations [151, 152, 153, 30]. In addition, some initial explorations have been undertaken regarding the use of graph theory metrics to control BCI systems. For example, Stefano Filho et al. presented a MI-BCI system based on motif as a graph theory metric to feed a classification algorithm [189, 190]. A MI-based BCI online game has been produced, based on a graph lifting transform where the participant has to use left and right MI movements in catching a coin [191]. Based on these findings, and always working from the synchrostates approach, the possibility of using graph theory metrics to discern between the two MI tasks is explored within this chapter. Aiming

to determine the best combination of graph metrics, classifiers and frequency bands, a diverse range of cases is considered. Firstly, the feature selection criterion used is explained, afterwards a set of classifier algorithms is selected. Finally, the performance of the different options are listed. Figure 5.1 shows the steps involved in the process followed from the original EEG recordings dataset to the connectivity metrics to feed the classifiers based on the clustered synchrostates from the phase differences.
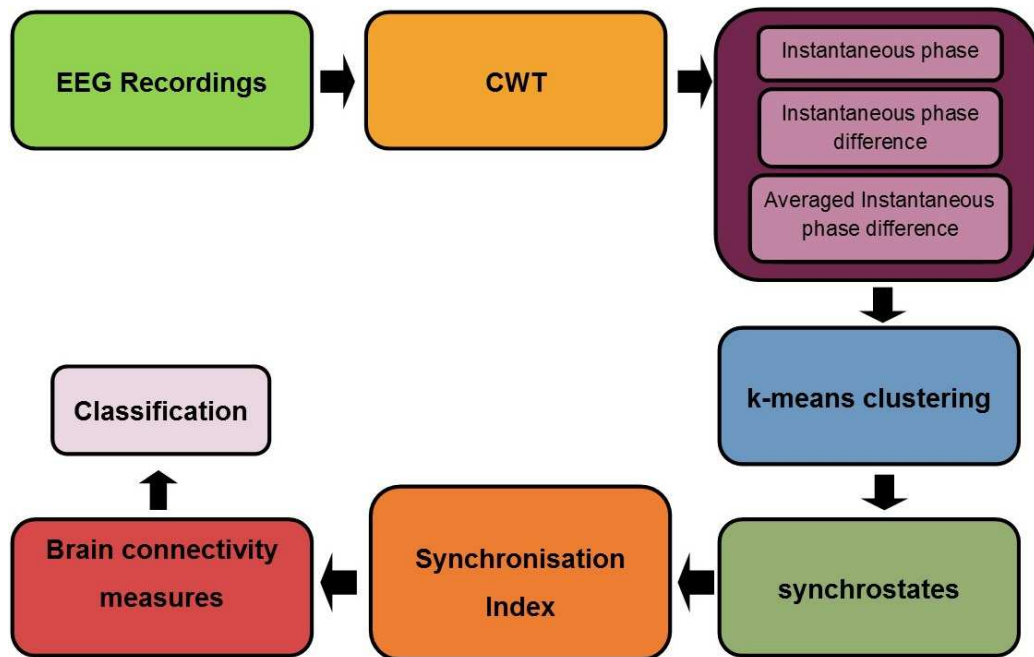
Figure 5.1: Graphical flowchart covering all of the steps followed from EEG recordings to the classification step.

5.1   Dataset

The EEG recordings considered were obtained from 10 healthy volunteers when performing two MI tasks using schematic emotional faces as stimuli as explained in section 3.1. The aim of this chapter is to investigate the theory that graph theory metrics derived from the translation of synchrostates into brain complex networks can be efficiently used as features to classify between the two tasks.

Network measures were obtained from the wavelet phase response and averaged across trials for each one of the participants for each of the stimuli within a specific frequency band. The resulting clustered synchrostates from the incremental k-means algorithm for each subject were translated into connectivity measures using the synchronisation index as weighted edges connecting each

pair of electrodes. The number of synchrostates varies slightly across participants as can be seen from figure 3.14 within a finite number of synchrostates. It can be seen from the transition plots in figures 3.11-3.13 or from table 3.2 that in the averaged population case, some of the states occur more than others. Extrapolating this concept to individual subjects of the population, the maximum and minimum number of occurrence states will be used from here on in for this chapter as an equitable selection criterion across participants. Selected maximum and minimum occurring synchrostates for each participant are converted to networks as shown in figure 5.2 the for β band. The literature has demonstrated that MI leads to attenuation/accentuation short-lasting peaks in the α and β bands, widely known as μ and β rhythms in the MI-based BCI area [192, 193]. Therefore, here the α and β bands brain connectivity metrics are used to accomplish the classification algorithm described in this chapter.
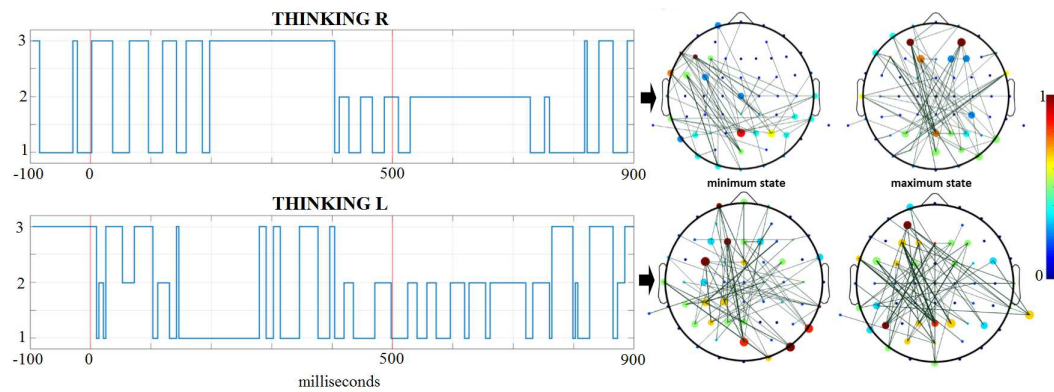
Figure 5.2: Synchrostates transition plots for both conditions in the β band and the translation to brain connectivity plots for the maximum and minimum number of occurrence synchrostates for each condition with a threshold of the 5% highest connected edges.

5.2   Separability criterion

One of the most important properties of a classification system is its ability to find the most informative features describing the objects that are classified because this guarantees as compact a decision rule as possible. In the machine learning literature there is a wide range of feature selection techniques, each with their own pros and cons in solving this issue. Two main, different approaches, can be highlighted, *scalar feature selection* and *feature vector selection*. The former is independent of the classifier used where the features are ranked using a specific score metric. Some examples of this methodology are the t-test, receiver operating curve (ROC) or the Fisher's discriminant ratio

(FDR) [170]. By contrast, feature vector selection is classifier dependent and the aim is to find the optimal combination of features to select the optimal sub-set of features. This can be done exhaustively, considering all possible combinations of available features and for each one of this combination, the class separability is computed. This search is computationally expensive so normally a suboptimal search technique is used. Some examples of such techniques are sequential backward selection, sequential forward selection, floating forward feature selection or plus-l-takeaway-r feature selection [194]. The separability criterion preferred to represent the difference between the analysed classes in this work is FDR. This separability criterion has been selected for two main reasons. Firstly, it is independent of the classification algorithm selected. This is an essential feature in this case where a set of different classifiers will be compared. Secondly, the result of applying this algorithm will give us a detailed list of the most discriminative features, those with a higher FDR value and not only the optimal number of them. In this case, the FDR is applied to a two class problem but it can be extended to problems with any number of classes [195].

FDR is a measure of the distance between two normal distributions inspired by the *z-score*. The *z-score* is a statistical value representing the number of standard deviations by which the value of a data point is above or below the mean value (μ) of the total population being observed. The *z-score* of a raw score x is defined by $z = (x - \mu)/\sigma$, where σ is the standard deviation of the population. It is widely used in normalisation processes, as the resulting data retain the same properties as the original data but their mean is zero and the standard deviation is 1. Therefore, two or more datasets with different units can be compared [196, 197].

$$FDR = \frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)} \qquad (5.1)$$

FDR is defined by using the mean $\{\mu_1, \mu_2\}$ and variance $\{\sigma_1, \sigma_2\}$ of each class as described in 5.1. FDR has large values when the mean difference between the two populations is large with a small within-class variance. Features presenting a higher FDR value will be more compact and distantly located. This meaning that there is a better discriminant power. On the other hand, if there is little difference between two populations, FDR presents a score close to 0 [140]. Prior to computing the FDR for each of the obtained features, a data normalisation step is performed to avoid bias from larger values within the set of features. Normalisation scales the feature vectors so they lie between

the minimum and maximum value for each feature.

5.3   Classifier algorithms

The brain complex network measures can be used as features to feed a classification algorithm to differentiate between the two MI tasks. The aim of this study is to explore the diverse features and classification algorithms to identify the best combination in order to achieve higher performance between the Thinking R and Thinking L conditions. The election of the right features pool and classifier type is crucial in order to obtain significant and consistent classification accuracy rates.

Supervised learning can be divided into parametric and non-parametric learning. The basic assumption of parametric learning is that the only unknown factors are the parameters of the probability densities involved. On the other hand, non-parametric methods are learning techniques for which prior knowledge of the conditional distribution is not available or not used explicitly [194]. Both types have advantages and disadvantages and the choice will be dataset dependent. The limited number of participants in this study makes it more suitable for the use of non-parametric learning methods for offering a higher flexibility in comparison with the parametric classifiers [198]. For this particular work three different approaches are used: nearest neighbour classification, discriminant analysis and support vector machines. The three of these are explained in the following sub-sections.

5.3.1 k-nearest neighbours classifier

This classification technique is very popular due to its simplicity, excellent empirical performance and the ability to handle binary and multi-class data [199, 200, 201]. Nearest neighbour estimation has no training phase. In order to classify a vector z_{new} the k samples from the training dataset closest to the new value are selected. Following this, a majority voting is performed and the class with the maximum number of votes coming from these k samples is assigned to the new value z_{new}. In order to select the k nearest neighbours, a distance measure is needed. In this case, the Euclidean distance measure is chosen. The euclidean distance is the most commonly used distance metric, particularly for continuous datasets. This is as opposed to other metrics such as the Hamming distance which are more commonly utilised for discrete variables [168, 202]. Therefore, it is one of the most reliably tested distance metrics.

One drawback of this algorithm is the selection of the optimal value of k. If it is too small, the classification results will be affected by noise. By contrast, if it is too large, the computational cost will increase [194]. In this case the value selected for the algorithm is $k=3$ as a good compromise between computational cost and accuracy rates.

5.3.2   Discriminant analysis based classification techniques

Linear discriminant analysis (LDA) classifiers are based on the assumption that the samples in the training dataset can be classified correctly by means of linear decision boundaries [194]. LDA assumes a normal distribution of the data with an equal covariance matrix for both classes $\{C_1, C_2\}$. If the samples of the two class sets C_1 and C_2 are defined as $\{x_{11}, x_{12}, \ldots, x_{1m}\} \, \epsilon C_1$ and $\{x_{21}, x_{22}, \ldots, x_{2n}\} \, \epsilon C_2$ the the simplest representation of a linear discrimination function is given by 5.2,

$$y\left(\mathbf{X}\right) = \mathbf{w^T}\mathbf{X} + w_0,\tag{5.2}$$

where $\mathbf{w}$ is a weight vector, y is the predicted class label and w_0 is a bias. Using the vector notation and a least squares error learning estimation to train the vector $\mathbf{w}$, the optimal solution of equation 5.2 is given in the form of the pseudo-inverse of the input features as can be appreciated in 5.3 [203].

$$w_{opt} = \left(X^T X\right)^{-1} X^T y.\tag{5.3}$$

For conventional LDA the classification for an input x is based on the comparison of $y\left(x\right)$ and a threshold or decision boundary. A simple decision boundary can be set as: if y is greater than zero, the object is assigned to class C_1 and consequently, if y is less than zero the input is assigned to class C_2. If the data are not linearly separable across classes, the least squares error based LDA will not perform properly. Maybe a more complex decision boundary is necessary. Quadratic discriminant analysis (QDA) classifiers are used for heterogeneous variance and covariance matrices. QDA produces a higher dimensional space where the decision function is quadratic, including the original features, their cross product and the quadric features [194]. For instance, in the simplest case of two classes the QDA transforms the space into a five-dimensional space $\{x_1, x_2, x_1 x_2, x_1^2, x_2^2\}$. Both methods will be compared in this chapter.

5.3.3   Support Vector Machine

Support Vector Machine (SVM) classifiers have been successfully used in MI based BCI applications [204, 205, 206, 207] due to their excellent empirical performance and because the number of parameters that must be set for the algorithm is related to the number of training objects instead of the number of attributes [198]. In opposition to discriminant analysis, the SVM selects one particular solution; the one that separates the classes with maximal margin. For example, the distance from the nearest training points. Maximising the marginal solution approach is proven to achieve the highest generalisation ability [194, 182]. In order to find the decision boundary that maximises this margin an optimisation routine is needed. In the case of linear decision boundaries the equation is defined by 5.2. If the values of y lie between $[-1, 1]$, then this means that the optimisation problem should find the solution to $| \mathbf{w^T X} + w_0 |= 1$. Using the geometrical definition of distance between a point $\mathbf{X}$ and a hyperplane $(\mathbf{w}, w_0)$ given by 5.4,

$$distance = \frac{| \mathbf{w^T X} + w_0 |}{\| \mathbf{w} \|}, \tag{5.4}$$

it can be seen that the maximisation of the margin is reduced to the minimisation of $\| \mathbf{w} \|$. Such a minimisation process can be done by means of Lagrange multipliers as explained in [194].

In the case that the data are not linearly separable it can be mapped into a higher dimensional space. The idea is that if the data cannot be separated by a linear function they may be separated by a quadratic or cubic function by means of more general kernel functions. The fact of replacing the inner product in 5.2 by a kernel function is known as a kernel trick [194, 198]. Increasing the order of the kernel also increases the computational cost of the algorithm and can lead to overlapping classes. In this work linear and second order polynomial (quadratic and cubical) SVM algorithms will be used.

5.3.4   Cross-validation and performance measure

In order to avoid the problem of over-fitting the classifier and reducing the sensitivity regarding the selection of training and testing sets, a cross-validation technique is needed. In this particular case, having a reduced size dataset, a *leave one out* cross-validation method is the most suitable to overcome the issue of over-fitting. It has been empirically demonstrated that this method has better bias-variance trade-off than those based on *k-fold* cross validation

for small datasets [208]. There is always a slight bias when using a bootstrap validation such as *leave one out*, but in practice, it is not meaningful [209]. It is the most extreme case of a *k-fold* validation scheme where each observation of the data is left out for validating the model and the remaining *n-1* observations are used to train the algorithm. Posteriorly, the accuracy obtained for each one of the data-points is averaged to obtain the classifier accuracy. All the classification algorithms and performance measures for all of the classification algorithms have been calculated by means of the free machine learning tool for Matlab© called PRTools [194] .

The performance of each of the classifier methodologies used throughout this thesis are calculated using the standardised measures of accuracy (*acc*), the true positive rate (TP_r) or sensitivity and the true negative rate (TN_r) or specificity. TP and TN are correct classification, by contrast false positive (FP) and false negative FN are miss-classifications. FP is when the outcome is incorrectly predicted as positive when it is actually negative and FN is the opposite, when the outcome is labelled as negative when it is actually positive. According to this nomenclature, TP_r is defined as TP divided by the total number of positives ($TP + FN$). Consequently, TN_r is FP divided by the total number of negatives ($FP + TN$). Finally the overall classification success rate or accuracy is defined as the number of correct classifications ($TP + TN$) divided by the total number of classifications ($TP + TN + FN + FP$)[210]. All these measures are listed in equation 5.5,

$$
\begin{aligned}
Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
Specifity = TN_r &= \frac{TN}{TN + FP} \\
Sensitivity = TP_r &= \frac{TP}{TP + FN}.
\end{aligned}
\tag{5.5}
$$

5.4   Results

The dataset of features for classification between the two MI tasks, Thinking R and Thinking L, is composed of two states (maximum and minimum number of occurrence states) and nine complex network measures (refer to section 4.2 for more details) giving a total of 18 possible features. The individual and collective discriminative power of all of them will be determined by the FDR. However, it maybe be worth exploring which synchrostate (minimum

or maximum number of occurrences) and which network parameter among the nine selected graph metrics is the combination with the highest power of discrimination, hence a larger accuracy rate [30]. In order to achieve the greater combination of features, the complete range of 18 features was split into three different cases as listed in table 5.1. Dividing the whole dataset into different groups, not only gives information about the optimal combination of features that can be obtained but also provides an insight regarding the behaviour of the max and min synchrostates. Therefore, a deeper understanding about brain functions and synchrostates will be gained. For the first case (named case I) all vectors of features (18 in total) were considered. In the next two groups (case II and case III), only the maximum and minimum state network measures were considered respectively (9 features in each case). In addition to these situations, synchrostates from two frequency bands were considered, the α and β bands.

Table 5.1: List of the different cases for classification based on a range of combination between the maximum and minimum number of occurrence states and network measures for each condition (Thinking R and Thinking L).

Cases	Combination
case I	both synchrostates metrics
case II	maximum state metrics
case III	minimum state metrics

5.4.1   Results with 5% threshold

In chapter 4 the connectivity plots from the synchrostates and the network metrics were calculated using a 5% threshold, meaning that only the highest 5% of the edges matrix, based on the synchronisation index for each synchrostate were used. In this section, the performance of the different classification algorithms is shown when this threshold is applied. Afterwards, in the next section a comparison of accuracy rates of the classifiers when no threshold is used is performed.

5.4.1.1 α band results

As mentioned in section 5.2, FDR is the separability criterion selected to elaborate a ranking of the different features according to a descending order of discriminative power. The results of this decreasing order step for cases I to III are illustrated in figures 5.3 to 5.5 respectively. The names of the features

are listed on the horizontal axis of the figures, ending in min or max referring to the minimum and maximum occurrence state respectively. The corresponding FDR value for each one of the features is shown on the vertical axis of the whole set of figures. The set of figures is transformed into line charts to aid easy visualisation, this is to facilitate the aim of grouping these features into smaller sets according to their FDR values.

Figure 5.3: FDR ordered values for the network metrics selected for case I, all possible combinations are considered, and the α band.

From figure 5.6 it can be seen that the set of features for cases I and II can be divided into five groups and for case III only four groups were formed. The fewer number of groups formed means that features will have closer discriminative ability as their FDR values lay closer to each other. It is also noticeable that the FDR values for case III are considerably higher than those for case II when the maximum occurrence synchrostate features are used. It can also be seen that the top four features for all the cases are clearly separated from the rest of the features within the same range. It is noticeable that the FDR values for the bottom half of the case III features present a remarkably low FDR value in comparison with the top half, meaning that their ability to discriminate between the two MI tasks should be lower. Maximum values for case I are similar to those for case III as it is a combination of both cases. For case I, the first group is formed for the top feature, the following groups are

integrated by 2, 4, 9 and 18 features respectively. Following the same criterion, the first group of case II is formed by the top 2 features, then 3, 5, 8 and 9 features. Finally, for case III, only the minimum state features are used, the four groups have the top 1, 3, 4 and 9, respectively.
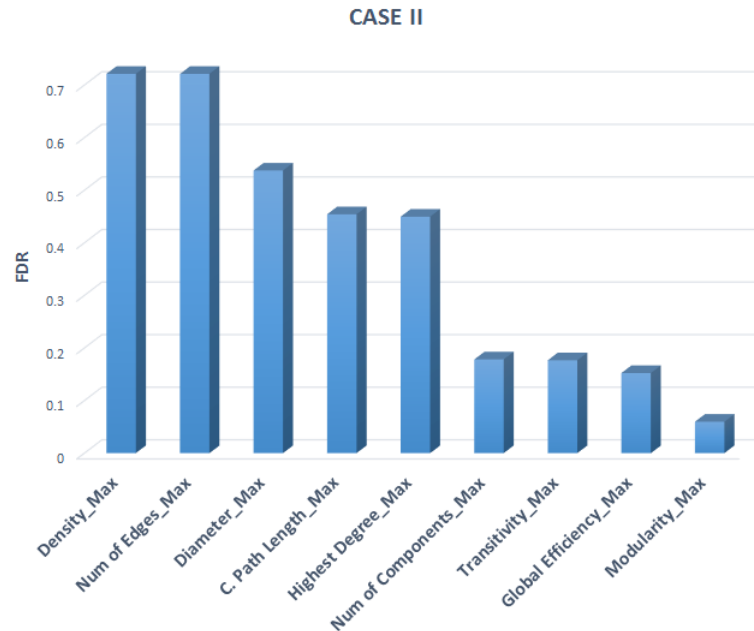
Figure 5.4: FDR ordered values for the network metrics selected for case II, only maximum state network metrics are considered, and the α band.
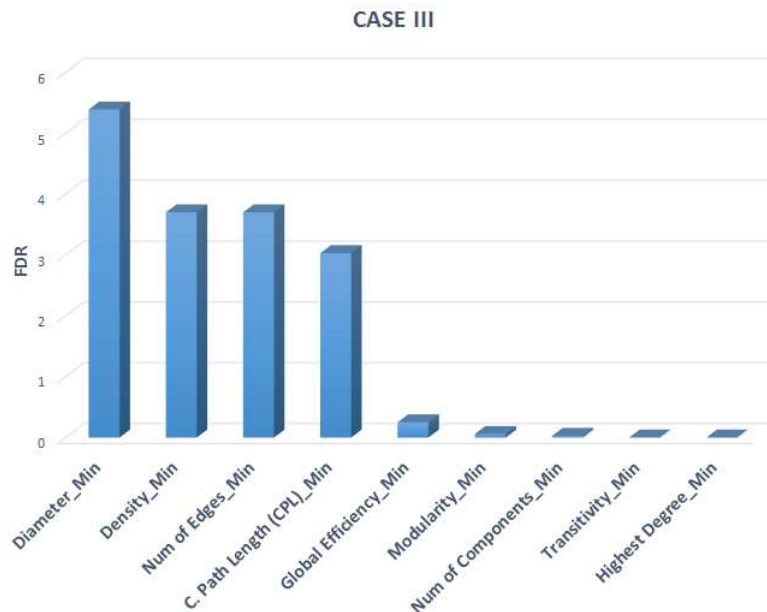
Figure 5.5: FDR ordered values for the network metrics selected for case III, only minimum occurrence state network metrics are considered, and the α band.

Performance for the six different classifiers using the *leave-one-out* validation method for case I is illustrated in figure 5.7. It can be seen that for linear discriminant analysis, the fact of adding more features to the classifier means lower performance. This agrees with the theory, that a higher number of features can lead to over-fitting [194]. It can be seen that the three different kernels of the SVM algorithms performed similarly, obtaining the best accuracy when only one feature, having the highest FDR value, is used. The accuracy achieved in this situation is 74% (71% TNr and 76% TPr). In general, the nearest neighbour (*3-nn*) algorithm showed the worst accuracy rates, which can be expected as it is the simplest classifier from the ones selected. The highest performance among all the classifiers is for the quadratic discriminant classifier (qdc) when the top four features are used. This highest accuracy is 83% (83% TNr and TPr). The top four features used to train the qdc classifier are diameter, density, number of edges within the 5% threshold and CPL for all metrics from the minimum number of occurrence synchrostates as can be seen in figure 5.3.
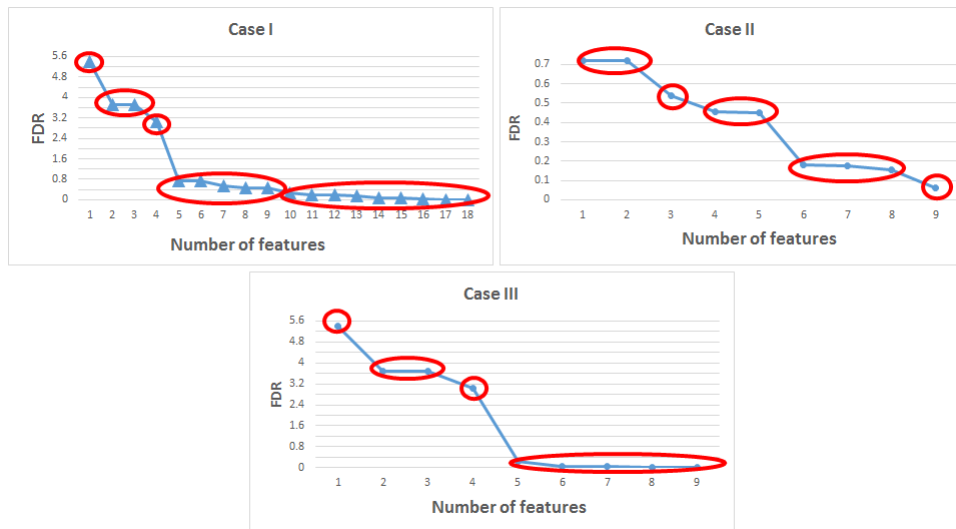
Figure 5.6: Grouped features by their FDR values ranking for cases I to III.

For case II, only the graph network metrics corresponding to the maximum number of occurrence synchrostates are considered. The performance of the different classifiers for this case are evident in figure 5.8. It can be seen that the overall performance of the whole set of classifiers is lower that the rates achieved by case I when all of the features were under consideration. The accuracy percentages are between 50% and 60% for all types of classifiers and groups. This result was expected as the values of the FDR ranking were

conspicuously lower that those presented for case I or case III as can be clearly
seen from figure 5.6. It is noticeable that the performance rates for the three
types of SVM classifiers are quite unbalanced, obtaining a specificity or true
negative rate considerably higher than the sensitivity. The same behaviour is
shown for the discriminant analysis classifiers for the first and second group
of features, containing 2 and 3 metrics respectively. All the classifiers, with
the exception once more of the nearest neighbour algorithm, show the highest
accuracy when the top two features were used. The performance rates for this
case were constant across all of the classifiers, reaching an accuracy rate of
73% with 89% for TNr and merely 58% for TPr. The top two features used
were density and the number of edges remaining after the threshold limit used,
in concordance with the top four features of case I.

Figure 5.7: Comparison of the performance of six different classifiers for case I in the α
band. Features grouped accordingly to their FDR values. For each group of features the
accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.
The classifiers from top to bottom and left to right: 3-nearest neighbour, linear discriminant
classifier, quadratic discriminant classifier, SVM linear kernel, SVM kernel order 2 and SVM
kernel order 3.

Figure 5.8: Comparison of the performance of six different classifiers for case II in the α band. Features grouped accordingly to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown. The classifiers from top to bottom and left to right: 3-nearest neighbour, linear discriminant classifier, quadratic discriminant classifier, SVM linear kernel, SVM kernel order 2 and SVM kernel order 3.

The last situation under consideration, case III, only uses the minimum state graph metrics to train the classifiers. As mentioned before, the first three groups of features corresponding to the top 1, top 3 and top 4 features are the same for case I than for case III: diameter, density, number of edges and CPL of the minimum number of occurrence synchrostates. Consequently, the performances of these three groups are the same in both cases. Therefore, the highest accuracy is shown by the qdc option, being 83% with 83% for both, TNr and TPr. The last group of features, when the whole set of them - 9 measures in total- are used to feed the classification algorithm, performed worse than the other cases for all of the algorithms. This can be attributed once more to the over-fitting effect which is especially marked for the qdc classifier where the accuracy dropped to a poor 48%, highlighting the fact that the performance obtained for all classifiers, except the 3-nn classifier, when the

top feature, is used is basically the same for all the whole variety of algorithms used (74% accuracy, 71% TNr and 76% TPr). This accuracy is not as great as the 83% obtained when the top four features are utilised to feed the algorithm but it can also be argued that the computational cost of using only one feature instead of four is considerably lower.
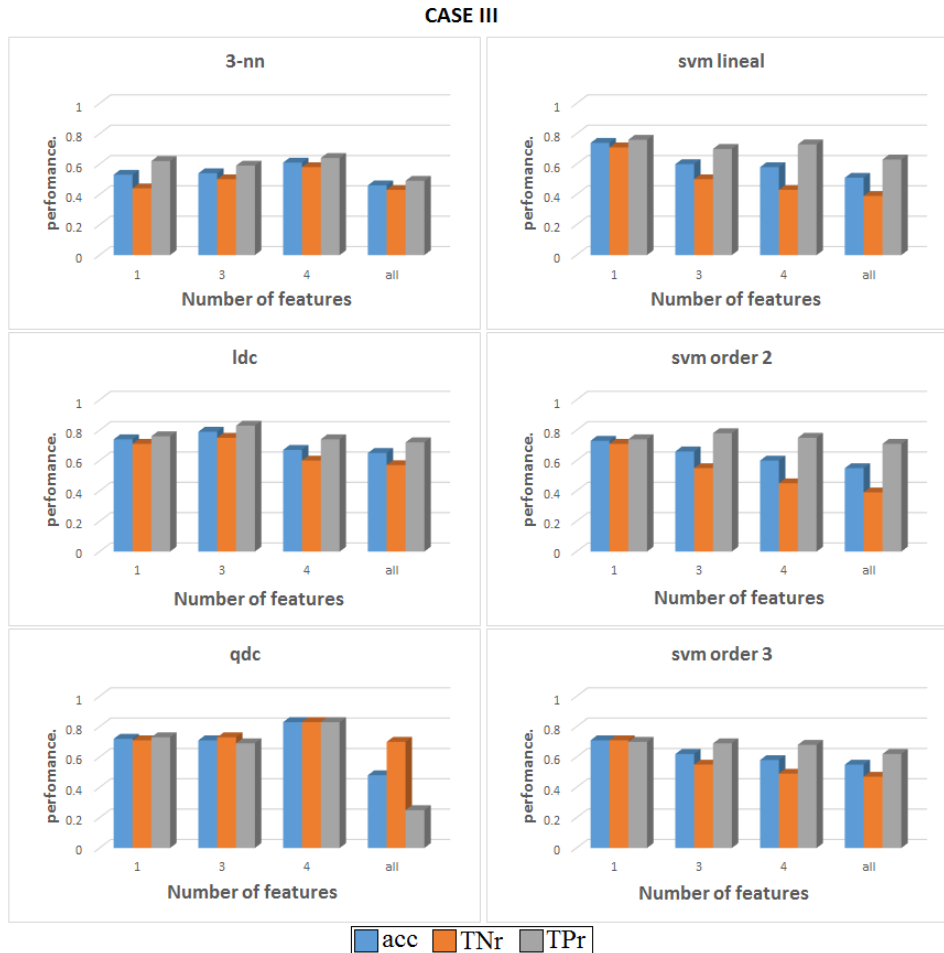
Figure 5.9: Comparison of the performance of six different classifiers for case III in the α band. Features grouped accordingly to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown. The classifiers from top to bottom and left to right: 3-nearest neighbour, linear discriminant classifier, quadratic discriminant classifier, SVM linear kernel, SVM kernel order 2 and SVM kernel order 3.

This summarises the comparative study for classifying two MI tasks using network metrics from the maximum and minimum synchrostates in the α band. The exploration of these three different scenarios yielded a profound understanding of which combinations from the assembly of features and synchrostates considered have the higher discriminative ability. The maximum accuracy rate of 83% accompanied by a TNr and TPr of 83% indicates that

this methodology is a valid resource to distinguish between the two MI tasks to control a BCI.

According to the results shown in figures 5.7, 5.8 and 5.9, it can be confirmed that for most of the cases, increasing the number of features to feed the algorithm causes a decrease in the general performance of the classifiers. In addition, it can be said that the more complex classifiers, for instance the SVM of kernels 2 and 3, do not lead to a higher performance rate in general, being the discriminant analysis based algorithms that achieved higher accuracy rates.

Regarding the comparison between the maximum and minimum number of occurrence states, the latter had better discriminant abilities than the maximum state for this frequency band. The average performance for the maximum state remains at under 60% for most of the groups of features and classifiers. Another characteristic present in the maximum number of features case, is the remarkable difference between the sensitivity and specificity rates, making the classifiers unreliable for classifying one of the two MI-tasks.
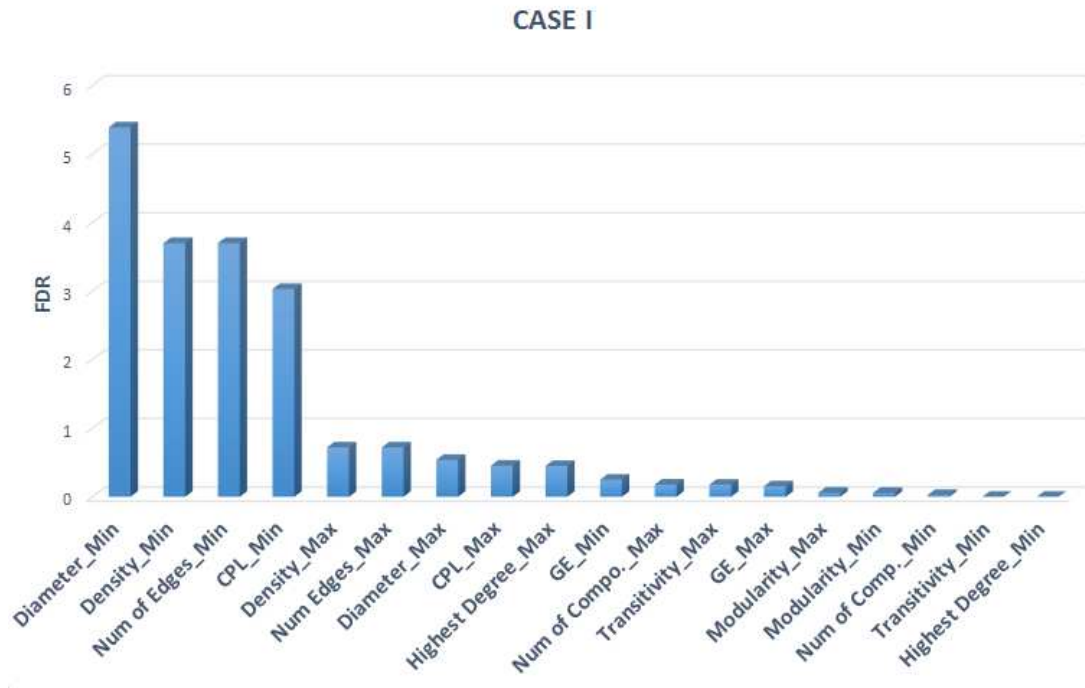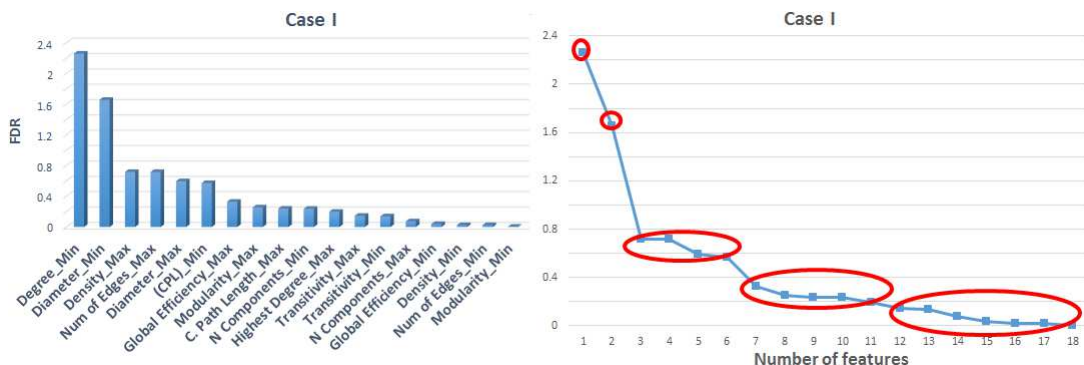

Figure 5.10: FDR ordered values for the network metrics selected for case I (all possible combinations are considered) for the β band (left side). The right graph has same descendent ranked values, but the features are grouped in agreement with their FDR values to feed the classification algorithms.

5.4.1.2   β band results

The same methodology as the one explained in section 5.4.1.1 is applied here, but this time for the β band. Firstly the FDR values are ranked and ordered for the three cases and the groups of features are used to feed the ensemble of classification algorithms used. The results are represented in figures 5.10 and 5.11. The graphs on the left side of figure 5.11 represent the feature-FDR value pairs ordered in a descending order to ease visualisation. Similarly, by the process described above, these values are transformed into a line graph to group the features having similar FDR values, hence, similar discriminative

capacities, to train the set of classifiers selected. The classification algorithm families were the same as those selected for the α band.

Network metrics features in cases I and II were divided into five groups, the same number of divisions as in the α band. However, for case III the number of groups is bigger, rising to six instead of the four groups for the same scenario with the α band. Case I groups were formed by the highest FDR value feature, followed by the top 2, 6, 11 and finally all of the 18 features. For case II, the initial group is composed of two features and the next groups by 3, 4, 7 and 9 respectively. Case III has the first and second group in common with case I, with the top one and top two features belonging to the minimum number of occurrence states. It is noticeable that case II shows similar FDR values than in the α band. In contrast, the highest values of the ranking are clearly lower in the β band where the maximum FDR value is 2.2; half that of the top value of the previously studied frequency band. One more similarity between both frequency bands for case I is that the features obtained from the minimum synchrostate have the highest and lowest values of FDR.
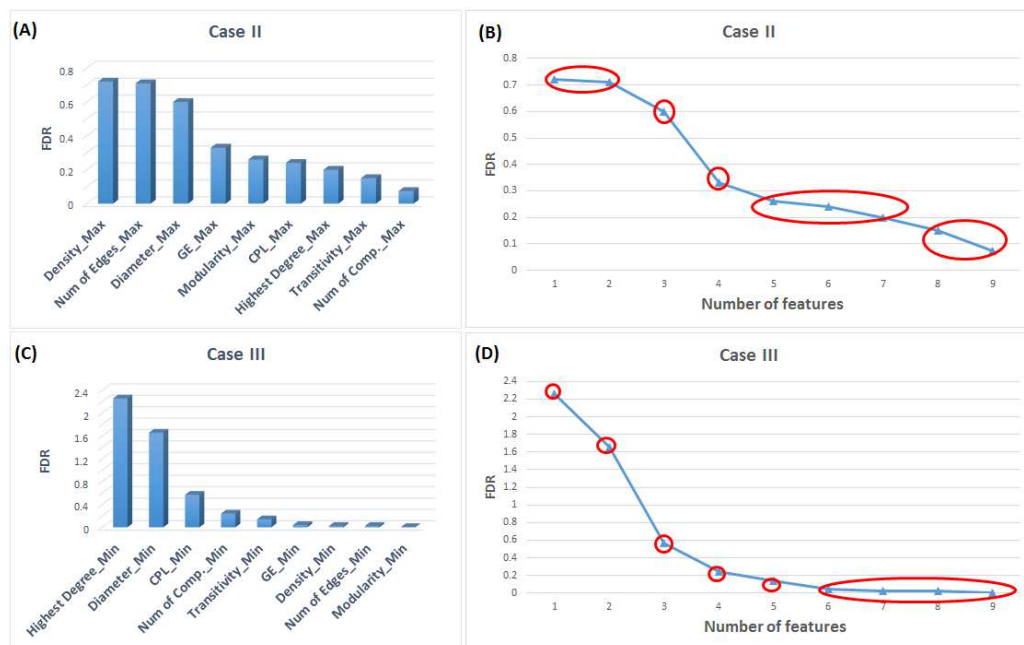
Figure 5.11: Figures (A) and (B) are the ranking of decreasingly-ordered FDR values for case II and (B) and (C) for case III. The left hand side graphs, A and C, are the features for the maximum (max) and minimum (min) synchrostates respectively. B and D graphs are the same pair of the metric-FDR values, but grouped accordingly to feed the classification algorithms.

Figures 5.12, 5.13 and 5.14 present the classification performance rates for the cases I, II and III respectively. For case I, with all the features included, the

average performance for all classification algorithms is over 60% except when all the features are used due to the over-fitting effect. The highest accuracy rate achieved is 80% for three different classifiers when the top FDR value feature is used. However, the performance varies across the three algorithms. For the qdc option, the TNr is 100% and the TPr 60.1%. Similar behaviour is found for the linear SVM. In contrast, the SVM kernel order 2, presents a more balanced performance with a percentage of 80% for TNr and TPr; contrary to the α band scenario where the performance of both discriminant analysis classifiers were higher than the SMV algorithms. For this situation, the accuracy rates have similar values across classifiers.
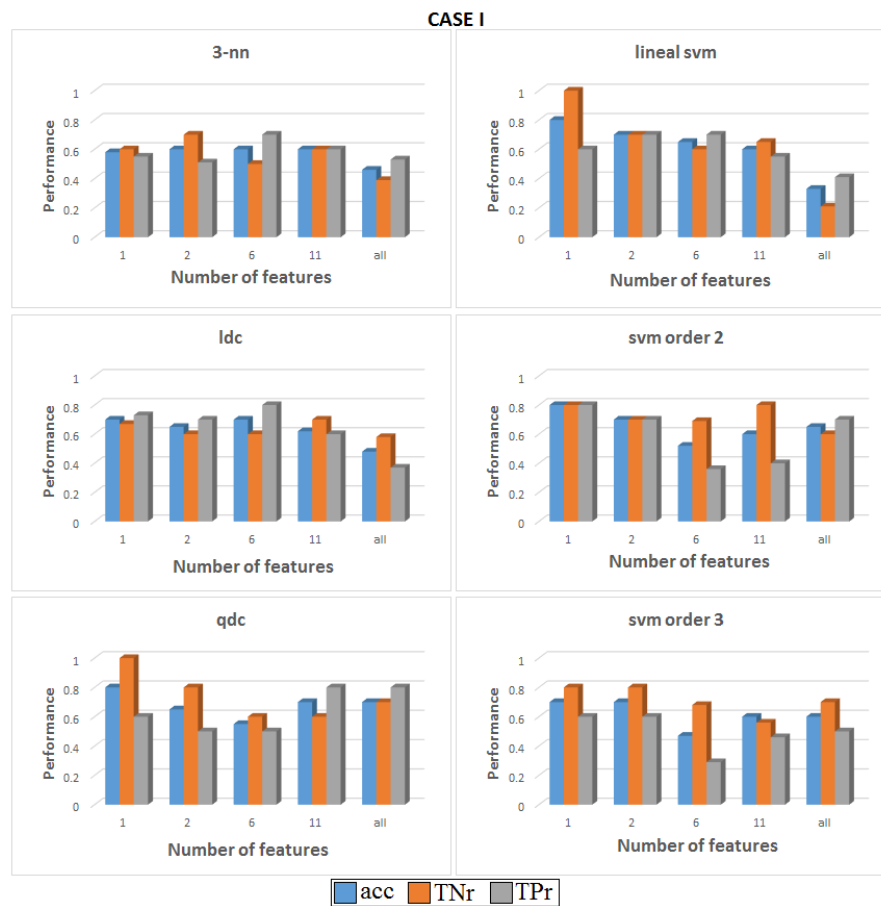
Figure 5.12: Comparison of the performance of six different classifiers for case I in the β band. Features grouped according to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown The classifiers from top to bottom and left to right: 3-nearest neighbour, linear discriminant classifier, quadratic discriminant classifier, SVM linear kernel, SVM kernel order 2 and SVM kernel order 3.

For case II, figure 5.13, when only features from the synchrostate with maximum number of occurrences are used to feed the algorithms, the best

performance for all of the classifiers is achieved when the top two features are used. These top two features are density and the number of edges remaining after the 5% threshold is applied; highlighting that these two features are precisely the top two for case II in α band. The highest accuracy percentage is for the 3-nn classifier with an 85% (85% TNr and 85% TPr). Followed by the three types of SVM algorithms with a 75% accuracy (90% TNr and 61% TPr). This discrepancy between true positive and negative rates shown for the different varieties of SVM was also present in the the α band for the same case. The poor performance of the qdc algorithm is noticeable when compared to the others with a maximum accuracy of 58% when top three features are used.
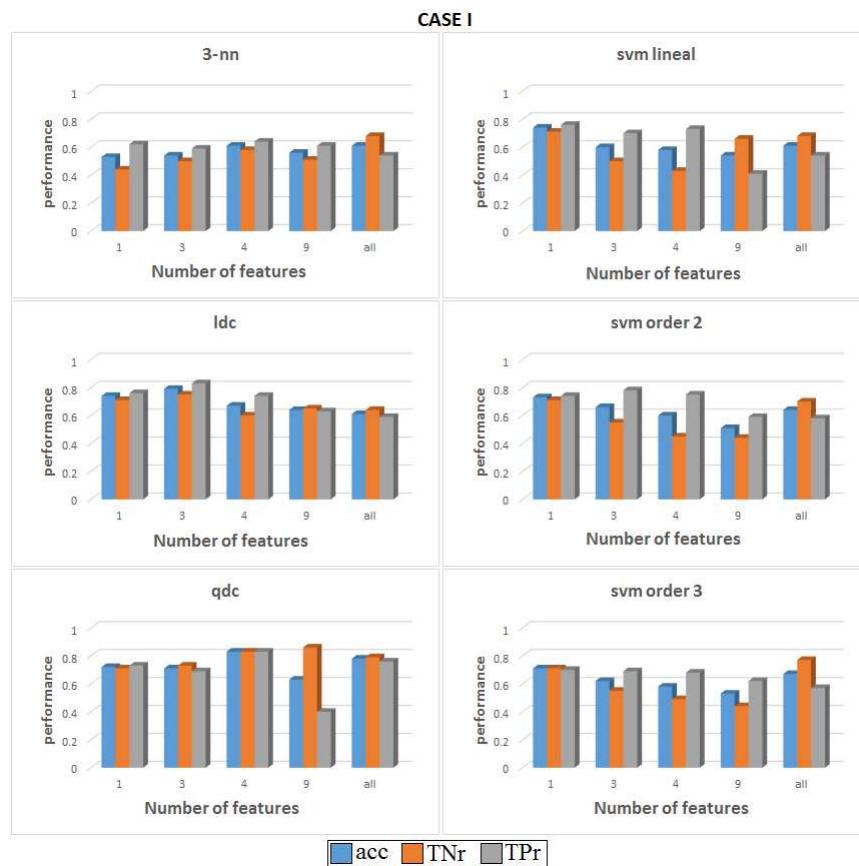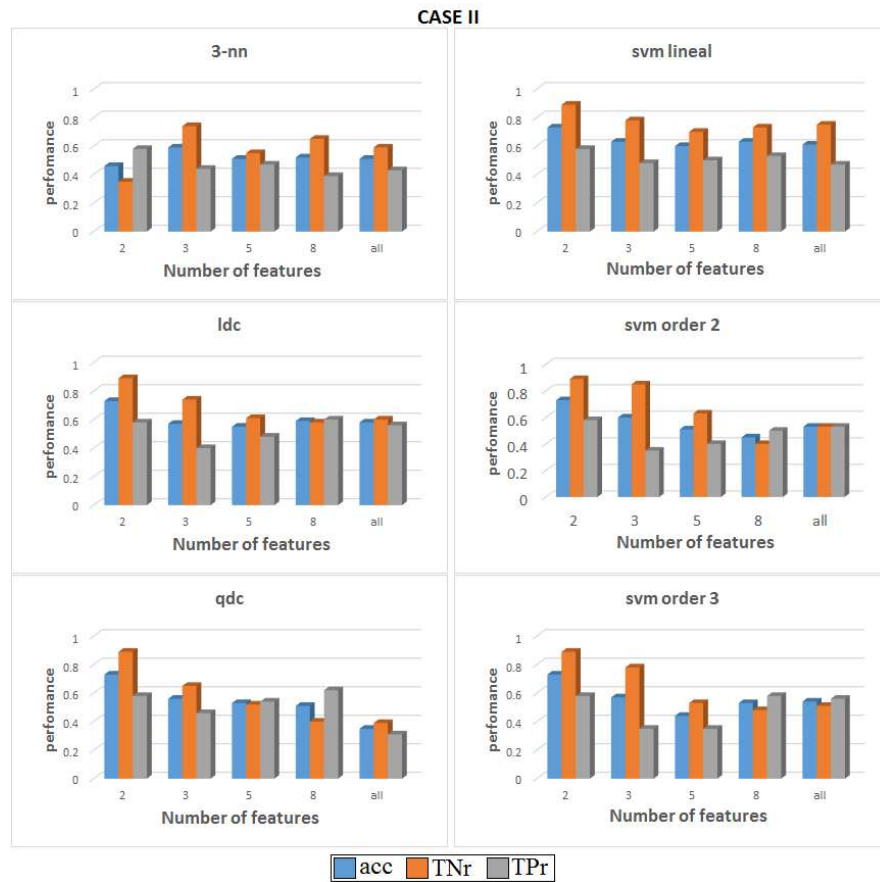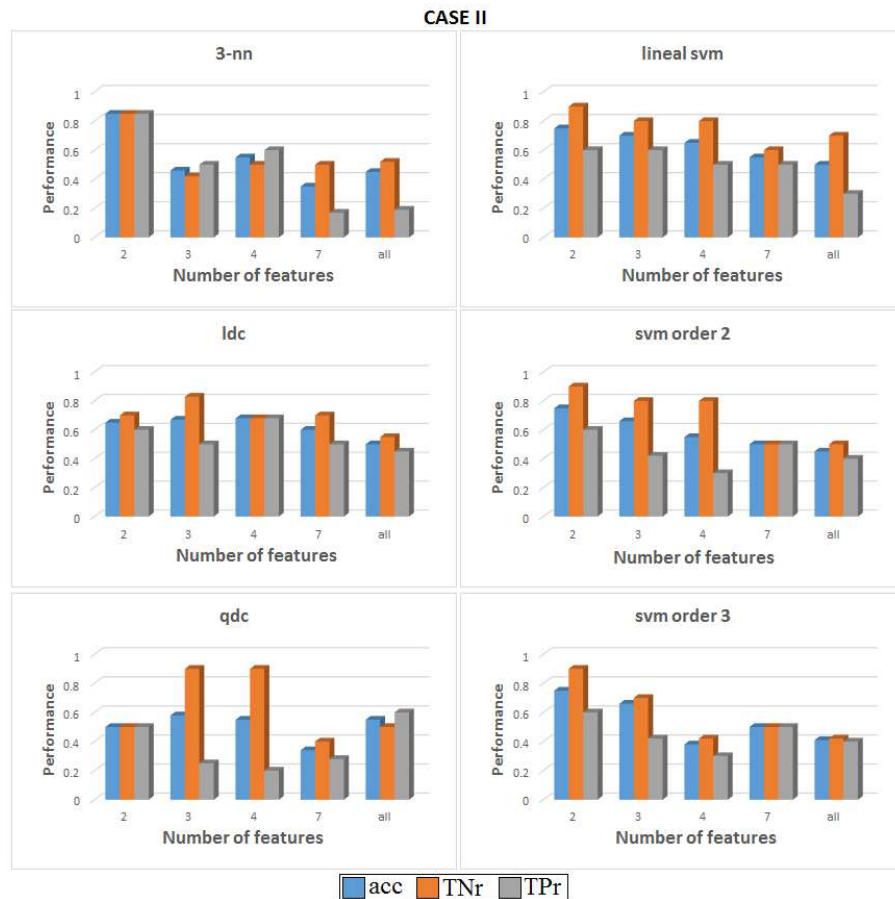

Figure 5.13: Comparison of the performance of six different classifiers for case II in the β band. Features grouped according to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

For case III, shown in figure 5.14, only the minimum state features are considered. The performance for the two first groups is the same as for case I as the top two features used are the same in both cases. The SVM with

order 2 kernel the algorithm achieved the greatest performance with 80% of accuracy and same values of TNr and TPr. Similar values were shown for the linear SVM and qdc algorithms. However, the unbalanced rate between true positive and true negative makes them less suitable for this application.
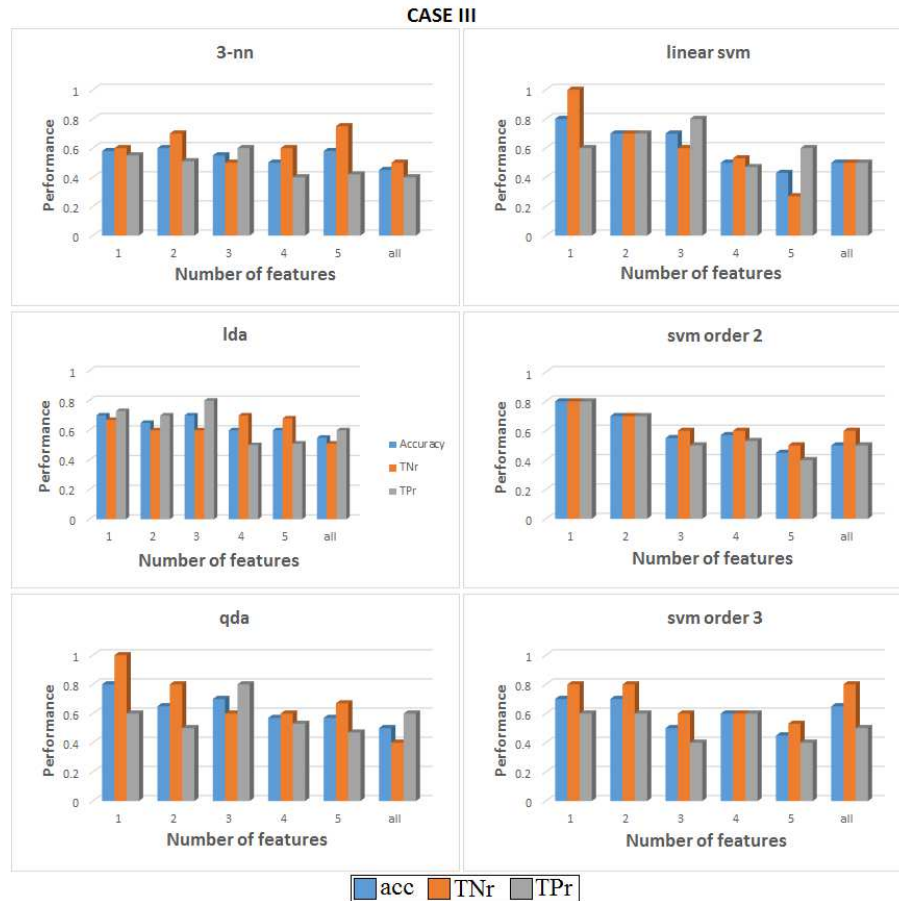
Figure 5.14: Comparison of the performance of six different classifiers for case III in the β band. Features grouped according to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

5.4.1.3 Summary

Table 5.2 shows a list with the best performance results for each case and frequency band considered during this section. The classification algorithm reaching the maximum accuracy rate for each scenario is also listed with the number of features used to feed the algorithm. It can be seen that the performance for both frequency bands is in the same range and can be said to be the same. However, the algorithm and number of features needed to achieve higher accuracy rates are clearly different. For cases I and III in the α band, four features are used by only 1 in the β band, both being a small number of features, the computational cost will be lower for the last scenario. By con-

trast the winning classifier in the α band is the qdc algorithm, it is a simpler algorithm than the SVM with kernel order 2 that is the one selected for the β band. In general, it can be said from the results obtained that this methodology is a promising tool for distinguishing between the two MI-tasks under consideration.

Table 5.2: Summary of the best classifier performance for each case and frequency band when a 5% threshold is used. Note the bad performance of case II with a TP_r of only 58%.

	α band					β band				
	acc	TNr	TPr	classifier	No. Features	acc	TNr	TPr	classifier	No. Features
case I	83%	83%	83%	qdc	4	80%	80%	80%	SVM order 2	1
case II	73%	89%	58%	SVM (all)	2	85%	85%	85%	3-nn	2
case III	83%	83%	83%	qdc	4	80%	80%	80%	SVM order 2	1

5.4.2   Results without threshold

This uses the same dataset and the same procedure of feature selection and classification algorithms than the previous section but no threshold is applied before the calculation of the network metrics. In this case, only two of the nine network measures used in sections 5.4.1.1 and 5.4.1.2 present a FDR value different from zero and so can be used as a feature to feed the algorithms. These two features for both frequency bands under consideration were modularity and transitivity. As explained in table 5.1, three cases will be studied. Case I includes the features for both states, the maximum and minimum number of occurrences of synchrostates. Case II considers only the features for the maximum state and case III for minimum state. In addition, two more scenarios are included using only each of the network metrics from both states. This means that case IV comprises only modularity values from both states and case V only transitivity values to try to understand which features have the largest capability of discerning between the two MI-tasks.

5.4.2.1   α band results

Figure 5.15 showscase I for the α band when no threshold is used. The general performance is substantially lower when compared with the threshold case. Especially noticeable for the range of SVM used is that the majority of the accuracy rates are under 50%. The two highest performances are for the linear discriminant algorithm with 60.8% (70% TNr and 51.3% TPr) with only one

feature used and quadratic discriminant algorithm with 66.3% accuracy (67.2% TNr and 65.4% TPr) and three features used. The feature used to feed the ldc algorithm was the modularity of the maximum state. This feature plus the transitivity of the maximum and minimum states were the network metrics used to feed the qdc algorithm.
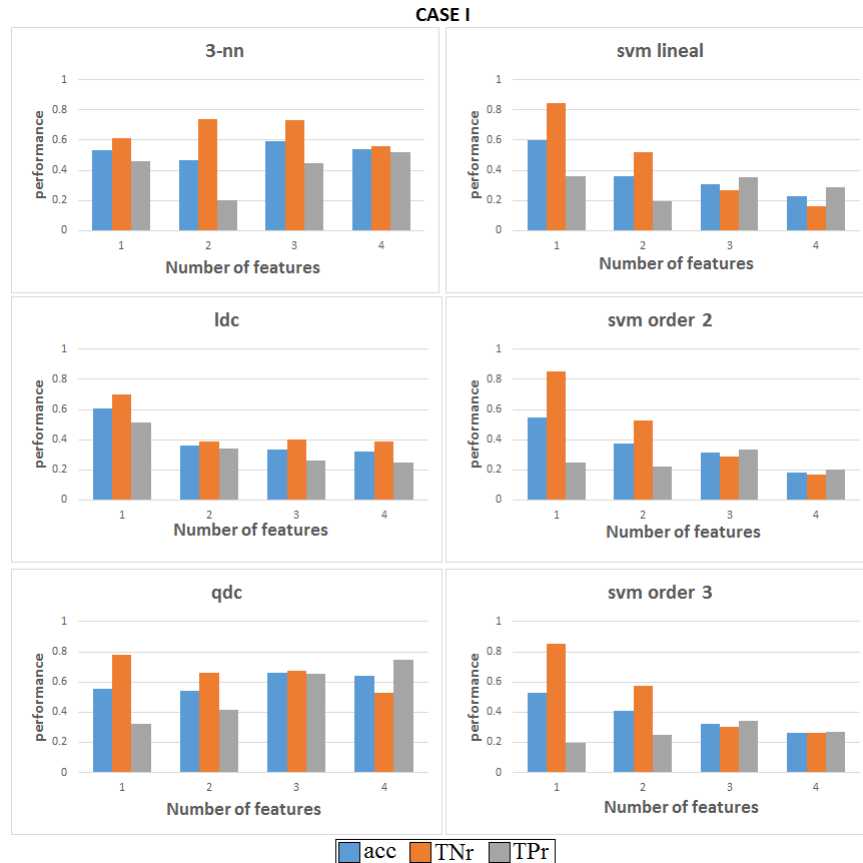
Figure 5.15: Comparison of the performance of six different classifiers for case I in the α band when no threshold is applied. Features grouped according to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

Figure 5.16 illustrates the rest of the cases, from II to V, when no threshold is applied. The results are no more encouraging than for case I, with the highest accuracies around 60% for all classifiers and the average under 40%. In particular, this highlights the poor performance of case V, when only transitivity is used as a feature. Consequently, it can be said that the modularity (case IV) has a larger ability to distinguish between both conditions than transitivity does. In addition, from figure 5.16, it can be said that the max state metrics have more powerful discriminating capacities than the min state.
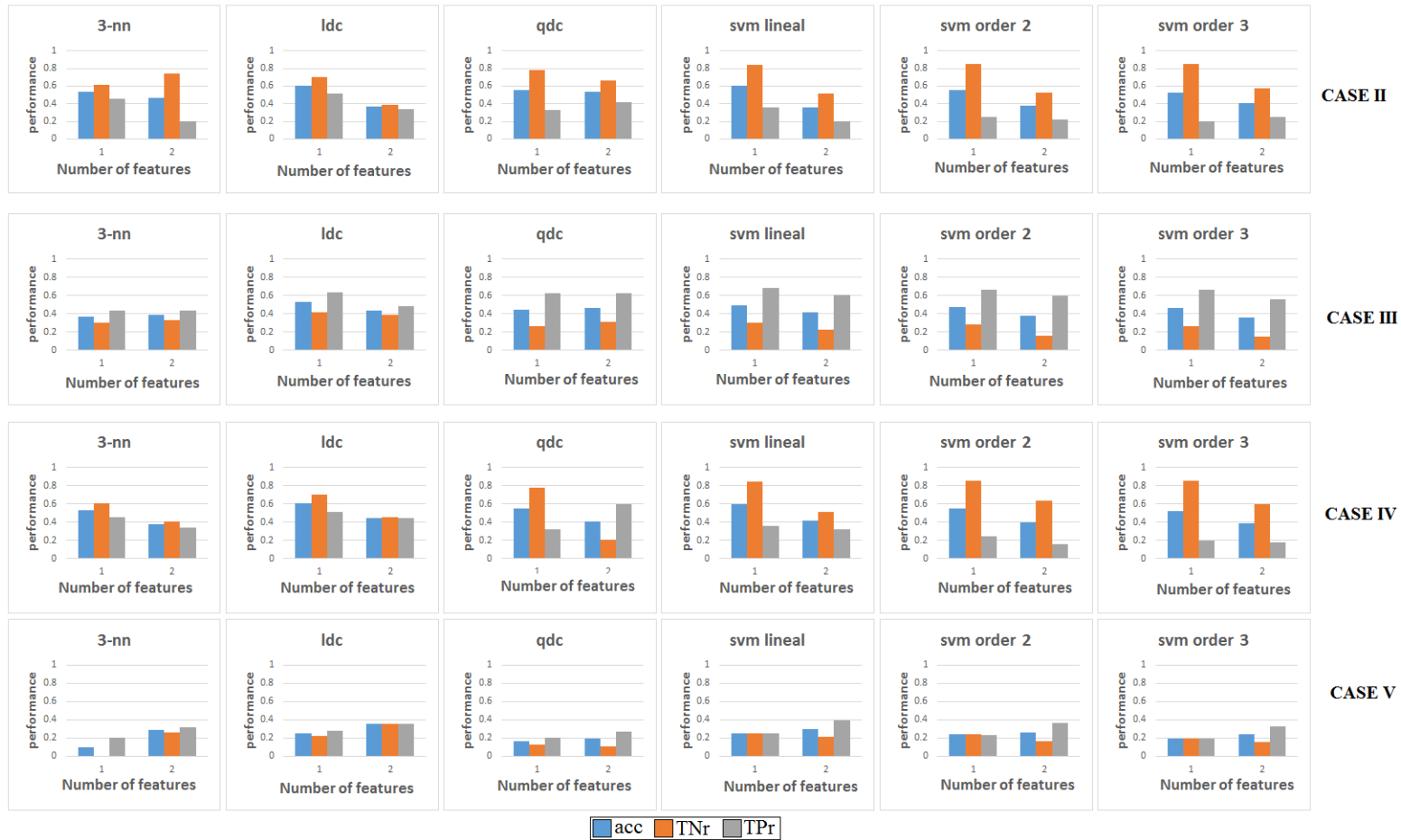
Figure 5.16: Comparison of the performance of six different classifiers for cases II to V in the α band when no threshold is applied. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

5.4.2.2   β band results

Figure 5.17 shows the case I, when all the features and synchrostates are considered. It can be seen that the performance is similar across classifiers and clearly superior to those results obtained from case I for the α band (figure 5.15). The highest performance is achieved for the ldc algorithm with 75.1% accuracy (67.3% TNr and 81.9% TPr) when three features are used. The three features used to train the algorithm are transitivity and modularity of the minimum state and transitivity of the maximum state. The same classifier, ldc algorithm, with only one feature -transitivity minimum synchrostate- obtained an accuracy rate of 70.1% (71.2% TNr and 69% TPr). The lowest accuracy, once more, is for the 3-nn classifier with a maximum accuracy of 58% (57.3% TNr and 58.8% TPr) when the top FDR value feature is used.
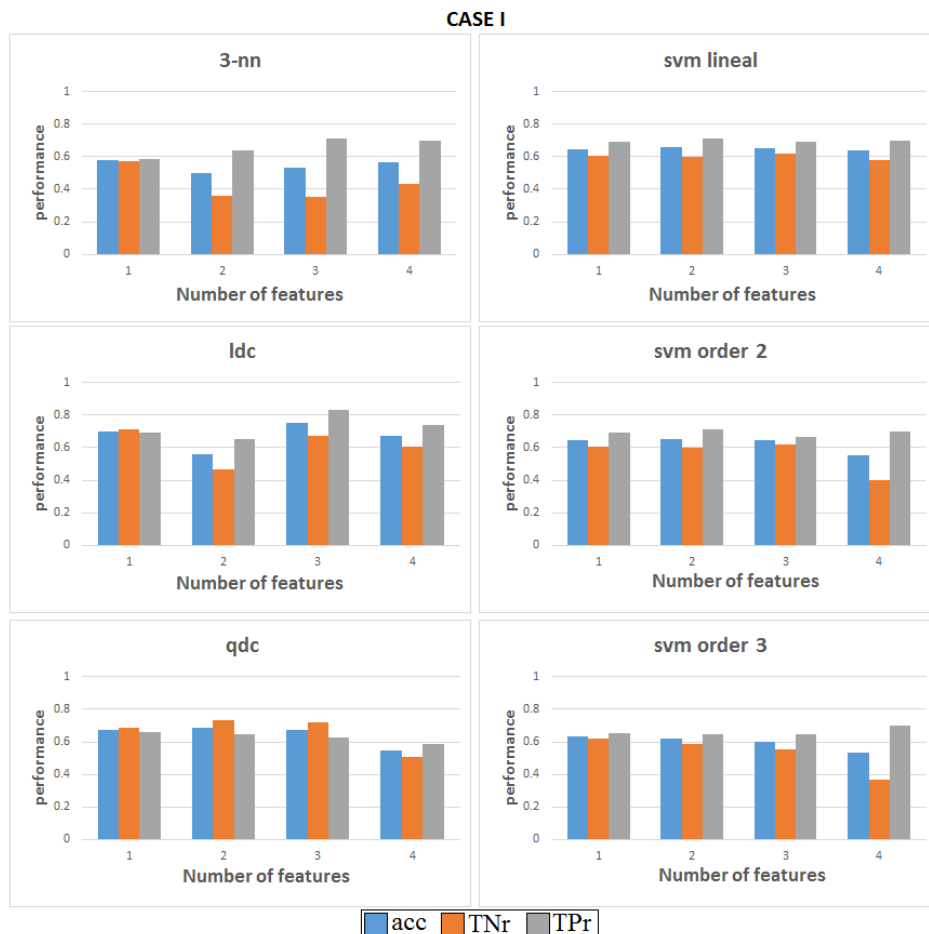
Figure 5.17: Comparison of the performance of six different classifiers for case I in the β band when no threshold is applied. Features grouped accordingly to their FDR values. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

The best result is obtained with 3 features, similar to the α band for case I. However, the network metrics are not the same, when in the α band the highest FDR values were for the modularity and transitivity of the maximum synchrostates in the β band. Conversely, the top features belong to the minimum state. This behaviour is similar to the situation when a threshold is utilised previously to calculate the network metrics, where the minimum state always offers a higher discriminative ability.

As mentioned before, the accuracy of the results for the β band are clearly improved in comparison to those obtained from the α band. Nevertheless, the accuracy rates are inferior when compared to the scenario when a threshold is applied with percentages of accuracy which are over 80%.

The rest of the cases, from II to V, are shown in figure 5.18. In opposition to the results for the α band for cases II to V (figure 5.16), transitivity (case V) has higher accuracy rates than modularity (case IV). Case IV, only modularity of the maximum and minimum synchrostates present a top accuracy of 65% (54.8% TNr and 75.2% TPr) using only the feature from the minimum state. By contrast, case V achieved the highest performance overall with 77.2% accuracy (70% TNr and 84.2% TPr). This performance was obtained when two features were used to feed the ldc algorithm. In general, it can be said that the performance between all of the classifiers for cases IV and V remained constant across the different groups of features.

Similar behaviour is demonstrated in cases II and III (top two cases in figure 5.18) where the minimum occurring state metrics (case III) performed remarkably better than the maximum state (case II). From the figure it can be seen that case II is the worst case among the five studied in the β band, presenting accuracies between 20 and 30% for the vast majority of the classifiers and groups of features. The highest accuracy for case III is achieved by the ldc classifier with a percentage of 70.1% (71.2% TNr and 69% TPr) as mentioned before. Once more, the performance across classifiers for this scenario when only one feature was used were similar, ranging from 65% to 70.1%, with the exception of the 3-nn classifier with only 58% accuracy (57.3% TNr and 58.8% TPr).
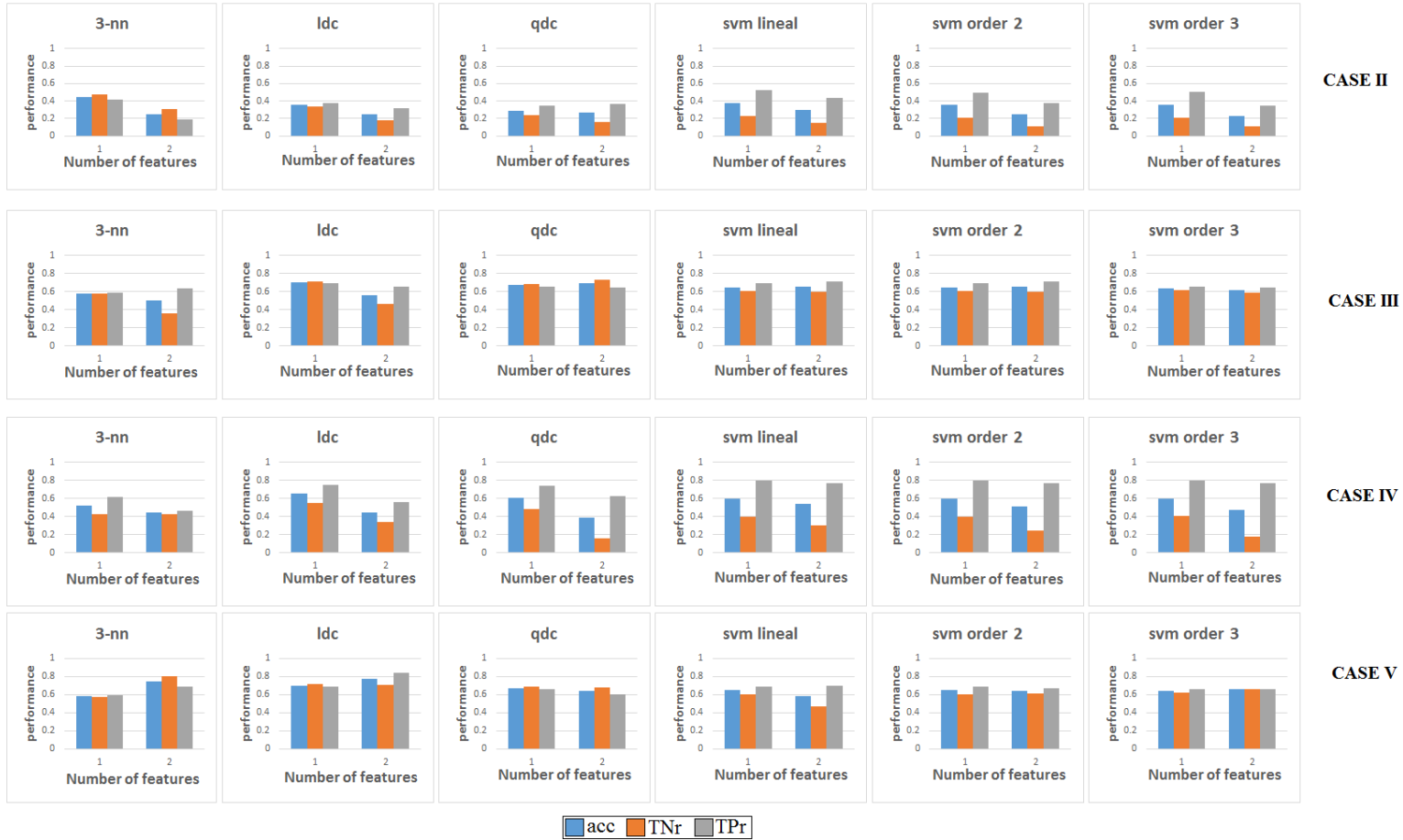
Figure 5.18: Comparison of the performance of six different classifiers for cases II to V in the β band when no threshold is applied. For each group of features the accuracy (acc), the true negative rate (TNr) and the true positive rate (TPr) are shown.

5.4.2.3   Summary

Table 5.3 shows a list with best performance results for each case and frequency band considered during this section, similar to the table presented in section 5.4.1.3. The classification algorithm reaching the maximum accuracy rates for each scenario is also listed with the number of features used to feed the algorithm. The performance when no threshold is used is remarkably lower than when it is applied as can be seen when compared to tables 5.2 and 5.3. Similar to the previous scenario, the performance of the β band is higher than in the α band, especially for case V where the performance in one band is half that of the performance of the other. This highlights that the number of features and classifiers is similar for both frequency bands, needing only one feature to achieve the highest accuracy rate in most of the cases and being the discriminant analysis the classification algorithm with higher success.

Table 5.3: Summary of the best classifier performance for each case and frequency band when no threshold is used.

	α band					β band				
	acc	TNr	TPr	classifier	No. Features	acc	TNr	TPr	classifier	No. Features
case I	66.3%	67.2%	65.4%	qdc	3	75.1%	67.3%	82.9%	ldc	3
case II	60.8%	70.3%	51.3%	ldc	1	44.5%	47.2%	41.8%	3-nn	1
case III	52.8%	41.7%	63.9%	ldc	1	70.1%	71.2%	69%	ldc	1
case IV	60.8%	70.3%	51.3%	ldc	1	65%	54.8%	75.2%	ldc	1
case V	35.5%	35.8%	35.3%	ldc	1	77.2%	70.3%	84.2%	ldc	2

5.5   Discussion

The aim of this chapter was to investigate the possibility of using complex brain networks from the synchrostates to distinguish between two MI-tasks, the imagined movement of the right and left hands. In the process of probing this theory, the idea of defining which synchrostate and network metric possess the most discriminant properties to differentiate between both conditions emerged. Finally, comparison between the two typical frequency bands was performed within this chapter in conjunction with a comparison of classifier performance using a threshold or not.

The maximum and minimum number of occurrence states was used to extract a set of network measures based on graph theory. Afterwards, these networks metrics were used to feed six different classification algorithms to differentiate between the two conditions. Maximum accuracy rates of over 80% for both bands when a 5% threshold is applied and over 70% when no

threshold is used before calculating the network metrics, indicating that this is a promising technique to classify MI-tasks.

Regarding which synchrostates have better discriminative abilities, it can be said that the minimum number of synchrostates in general obtained better classification performance for all scenarios. There was only one exception, where the maximum number of synchrostates presented better results than the minimum state, this is for the α band when no threshold is applied.

Using a threshold of 5% increased the general performance of the classification algorithms for all possible combinations under consideration, with the highest accuracy for the α band of 83% (83% TNr and 83 TPr) and 85% accuracy for the β band with the same values for TNr and TPr. By contrast, the highest accuracy when no threshold is applied is 66.3% (67.2% TNr and 65.4% TPr) for the α band and 77.2% accuracy (70% TNr and 84.2% TPr) for the β band. It can also be said that results of the β band were higher than the performance in the α band for both situations (threshold and no threshold) as can be seen from tables 5.2 and 5.3. The use of the threshold also increased the number of features with a FDR value different from zero. For this reason, when it is used, the number of network metrics used to feed the classification algorithms was nine in total. However, when the complete adjacency matrix of the synchronisation index was used to calculate the metrics for each synchrostate, only two measures were found with FDR values different from zero, modularity and transitivity. Maybe this is one of the reasons that led to a lower performance rate for the classifiers when no threshold is applied.

The features with a higher discrimination capability when a 5% threshold is used, common to both frequency bands, are density and the number of edges within the threshold. In the other scenario, when the whole weighted matrix is used, for the α band the top feature is modularity of the maximum state and the top accuracy rate for the β band was achieved when the transitivity of both states was used.

Regarding the classifiers, not one of them performed clearly better than the others, being discriminant analysis algorithms in some cases and the SVM classifiers in others. Even if in table 5.2 it seems that SVM algorithms achieved better results and in table 5.3 discriminant analysis classifiers, the accuracy rates remained within a small range across classifiers in most of the cases, making the selection of only one of them difficult.

Several examples can be found in the literature of investigations trying to classify MI-tasks to control BCI based on EEG and machine learning as mentioned within this work [211, 212, 213]. The main differences in the research

generated until now in MI-based BCI are that in this work, firstly the extraction of synchrostates based on phase differences is done for each subject and frequency band. Secondly, the maximum and minimum occurring synchrostates are selected and used to calculate complex network metrics by means of a synchronisation index. These network metrics based on graph theory such as density, diameter, CPL, GE, modularity or transitivity are becoming a hot topic in neural engineering as they have strong features to characterise several combinations of brain connectivity such as segregation and integration.

5.6   Conclusion

In this chapter the study of connectivity network metrics has been proposed, derived from the maximum and minimun occuring synchrostates, for the classification between two motor imagery tasks. The classification algorithms gave state of the art results for small datasets [214]. Connectivity metrics, always from the synchrostates perspective, have demostrated the ability to quantify the underlaying functional connections between the different areas of the brain during motor imagery tasks.

This novel application of synchrostates led to a new classification methodology for motor imagery based BCI systems, and it can be the base for new applications. However, even if the results are promising further investigations should be done. Mainly, a larger number of participants should be used to reduce to a minimum any possible misclassification effects.

Chapter 6

Synchrostates: emotions from human faces

The results obtained in chapter 3 demonstrated the existence of quasi-stable states in the order of milliseconds when participants are performing an MI task based on a schematic faces paradigm. This phenomenon was named synchrostates. The number and the switching patter among synchrostates are specific for each one of the motor imagery tasks and frequency band. Furthermore in chapter 4, research regarding their temporal stability was performed by means of a synchronisation index and complex network analysis.

The aim of this chapter is validating the existence of synchrostate phenomena using images of human faces as stimuli instead of schematic emotional faces to indicate to the participants which motor imagery task they should perform. The use of human face images is aligned with the idea stated in chapter 3 that emotions can enhance a stronger response across subjects and consequently can be used as a novel paradigm in motor imagery based BCIs. Furthermore, the number of cognitive tasks has been increased from two to four with the intention of testing and validating the presence of synchrostates over a larger amount of motor imagery tasks.

Following the same steps described in chapters 3 and 4, the phase information will be analysed for the different tasks and frequencies. The temporal evolution and stability of the task-specific synchrostates will be explored in the different areas of the brain by using complex network metrics.

6.1   Experimental protocol

Emotions play a vital role in the communication and interaction between humans. The ability to recognise the emotional state of the surrounding people is a natural step in human communication. The recognition of a facial emotion is a highly complex task utilising a wide range of skills. Emotional recognition requires attention, perception, learning and memory skills. Numerous studies have been performed using EEG as a basis for automatic emotion recognition in the past few decades. It has been proven that EEG signals can highlight important information regarding emotional states [215]. This information associated with the emotions can be used as stimuli in brain computer interfaces. For instance, this is the case for novel BCI speller systems that using faces as stimulus achieved a considerable incremental improvement in their performance when compared to conventional BCI spellers [93, 156]

6.1.1   Subjects

Fifteen healthy volunteers, 10 males and 5 females, with an age range between 24 and 39 years participated in the experiment. The participants in this experiment were different from those who joined the previous experiment explained in chapter 3 to avoid any familiarity with the experimental set-up and so, avoid a possible influence in the classification results. Their sight was normal or corrected to normal. The recruitment was carried out by means of public announcements, no selection criteria were used and no monetary compensations were offered to the participants. Written consent was signed by each participant after they were informed of the nature of the study. They fully understood all of the processes and were comfortable with them. The University of Warwick Ethical committee, named Biomedical & Scientific Research Ethics Committee (BSREC), approved this study (REGO-2014-821).

6.1.2   Experimental montage and equipment

The experimental montage and the equipment described in section 3.1.2 are also used for this experiment. Please refer to the relevant section for more information about the EEG cap and electrodes used for recording the data.

6.1.3   Paradigm

The stimuli for the experiments were taken from a widely used database of

human facial expression images [216, 217]. The Karolinska Directed Emotional Faces (KDEF) database is a set of 4900 pictures in total of human facial expressions of emotion. Forty faces were selected from 5 female and 5 male subjects, representing 4 emotions each: happiness, sadness, neutrality and surprise.

The subjects and emotions were selected as being the most voted for from a ranking performed by thirty people. The participants in the ranking were shown the total number of frontal pictures from the database after being normalised using Matlab[C] software. They had to select the emotion expressed by the person in the picture and voted (from 1 to 5 -really good-) on how clear the emotion was. More details about the selection procedure can be found in Appendix A.

The experiment was conducted in 4 blocks of 80 trials each. In each block 20 happy, 20 neutral, 20 sad and 20 surprised faces were presented at random. The temporal scheme of a typical trial is as follows: a scrambled face is presented on the screen for half a second, followed by one of the selected emotion images lasting on the screen for 0.5 seconds. Afterwards, a black screen is presented to the participant for 3 seconds. During this time, the participant is asked to perform the imagined tasks associated to the emotions shown just before on the screen. Finally a cross is presented lasting for 1 second indicating the end of the trial. The motor imagery tasks that participants were asked to perform were: imagined movement of the right hand when a happy face is shown, left hand movement when a sad face is presented. This selection was specifically chosen to follow the same criteria as in the experiment performed in chapter 3. In addition, for emotions of surprise the participants were asked to perform the imagined movement of both feet. Finally when a neutral face is shown the participants were requested to relax, and not do perform any type of movement. Further details regarding the sequence and temporal scheme of the experimental set up can be found in appendix B. The movement of both feet, together with the movement of left and right hands are the typical movements associated with MI based BCI systems [218]. Regarding the discrimination between right and left foot imagery, there are some studies covering the topic, which uniformly show poor results with accuracies of around 60% [219]. For this reason, in this thesis, the movement of both feet was selected.

6.1.4   Pre-processing

EEG data are acquired at 512 Hz from 62 channels + 2 reference channels

placed on the earlobes. The continuous EEG data are divided into the four conditions: right and left hand movements, feet and relax. A first selection of trials was done by eliminating those with amplitudes over a 200μV threshold, following the same criterion as in chapter 3. Afterwards a visual inspection of the whole dataset of trials and participants was performed and those trials affected by blinking or any other kind of muscle movement were eliminated and considered as artifacts. Finally, artifact-free trials were divided into one second length epochs for each condition lasting from 100ms before stimulus onset to 900ms after and baseline corrected. Due to the low amount of artifact-free trials, one of the male participants was discarded from future analysis.

6.2 Synchrostates from the averaged population

Similar to the process described in chapter 3, the study of temporal evolution of the phase is developed first in the averaged population and posteriorly, the variability within the subject is studied. In order to obtain the synchrostates from the averaged population the phase difference matrix for each individual subject has to be calculated as described in section 3.2.1. Afterwards, the average of the phase matrices of each subject at each time instant and frequency band is formulated. The resultant averaged matrices are clustered by means of the *k-means* algorithm as described in section 3.2.2.

Figure 6.1 shows the results of the incremental *k-means* algorithm for a number of clusters between 2 and 10 for all the frequency bands of interest (ϑ, α, β and γ) and for the four different motor imaginary tasks from the averaged population (R hand, L hand, Feet and Relax).

The top row (blue colour) represents the cost function for the four imaginary tasks in the ϑ band. It can be seen that the first clear dominant knee occurs when the number of clusters is two for all of the cases. Although in some of the cases, the cost functions have fluctuations, increasing or decreasing, the earliest knee will be the only one to consider as previously discussed in section 3.2.2. It means that for the averaged group of 14 people the number of unique synchrostates is $k=2$ for the ϑ band.

In the next row of the figure the results of the α band-cost functions are illustrated (red colour). It can be seen that the optimal number underlying the phase difference matrices is $k=3$ for the four different situations under consideration. This is the same number of synchrostates obtained for the α band for the right and left hand imagined movements experiment designed

and explained in chapter 3. The same situation is not sustained for the β band (black colour), where only the R hand case presents a number of clusters similar to the previous study, $k=3$. Being only $k=2$ for the other three conditions (L hand, Feet and Relax). Finally the last row (green colour) shows the results of the algorithm for the γ band. It can be clearly seen that the optimal number of cluster is $k=4$ for all situations.
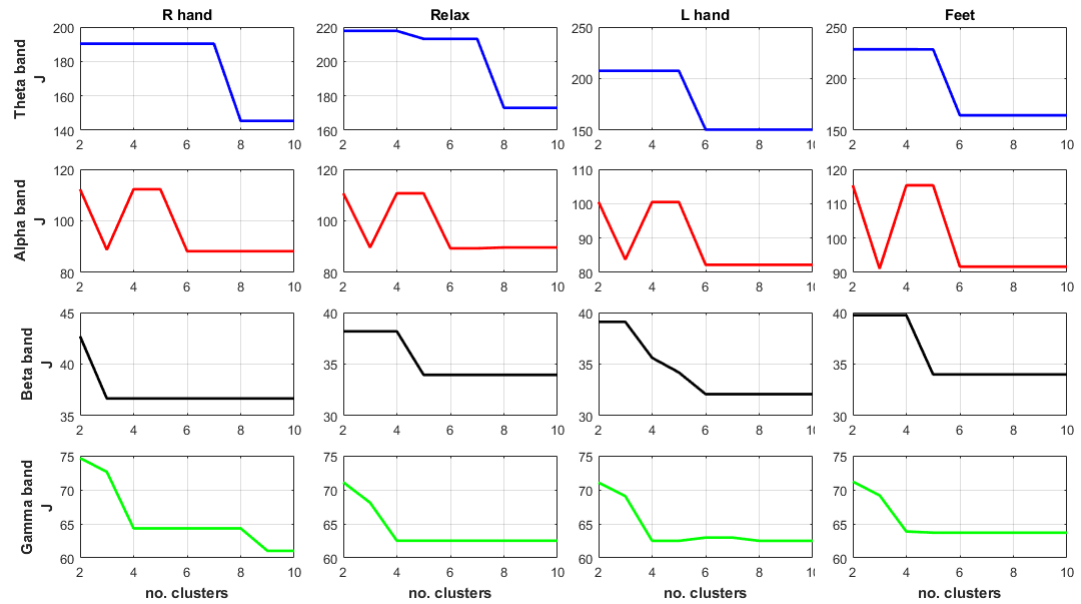
Figure 6.1: *k-means* clustering cost function results for the four conditions under analysis (R hand, L hand and feet movements and the no-movement or relax task). The upper row (blue colour) corresponds to the ϑ band. It can be seen that the optimal number of clusters is 2 for each one of the tasks. Second row from the top, red colour, represents the cost function for the α band where a minimum can be found corresponding to 3 clusters for all the cases. Third row, in black, represent the results for the β band. In this case the number of clusters selected for the R hand movement task is 3, being only 2 for the other three tasks. Finally the last row, green colour, represent the γ band results where a clear minimum can be found for a number of clusters equal to 4 in each one of the performed tasks.

As discussed in previous chapters, the α and β bands will be further studied, as they are the typical frequency bands used for BCI system investigations according to the literature. In addition, it has been demonstrated that the β band is related to face processing and cognition [134]. In this chapter, further research on the γ band is also performed as recent discussions argued that the γ band is induced by different sensory stimuli or cognitive tasks. Therefore, the γ band synchronisation is an elementary and fundamental process in the functioning of the brain [220].

Figure 6.2 shows the head plots of the increasing *k-means* algorithm synchrostates from the averaged group in the α band. At first glance, it can be appreciated that the topographies of the different states are very similar along
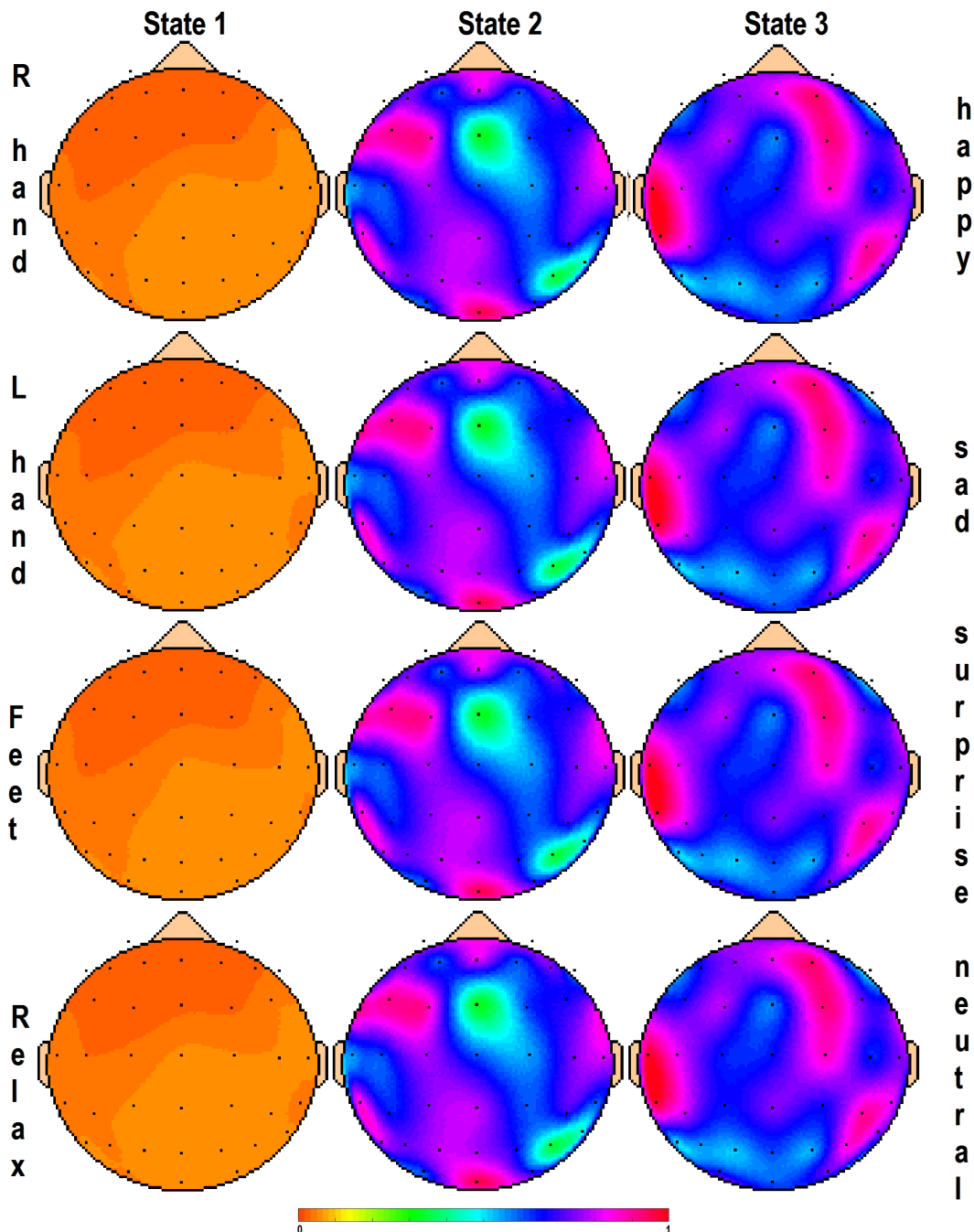
Figure 6.2: Topographies from the synchrostates of the averaged population for the α band. Columns represent the synchrostate (1 to 3 in this particular case) and rows indicate the performed task. From the top to the bottom: R hand imagined movement, L hand, Feet and Relax. On the right side, the emotion linked to each one of the tasks is indicated: happiness, sadness, surprise and neutral respectively.

the different stimuli presented to the participants. A similar result is shown by the other frequency bands, β and γ, as can be seen from figures 6.3 and 6.4 respectively. The synchrostates of the four different tasks are similar within a frequency band, which is in line with the results found in chapter 3 for the
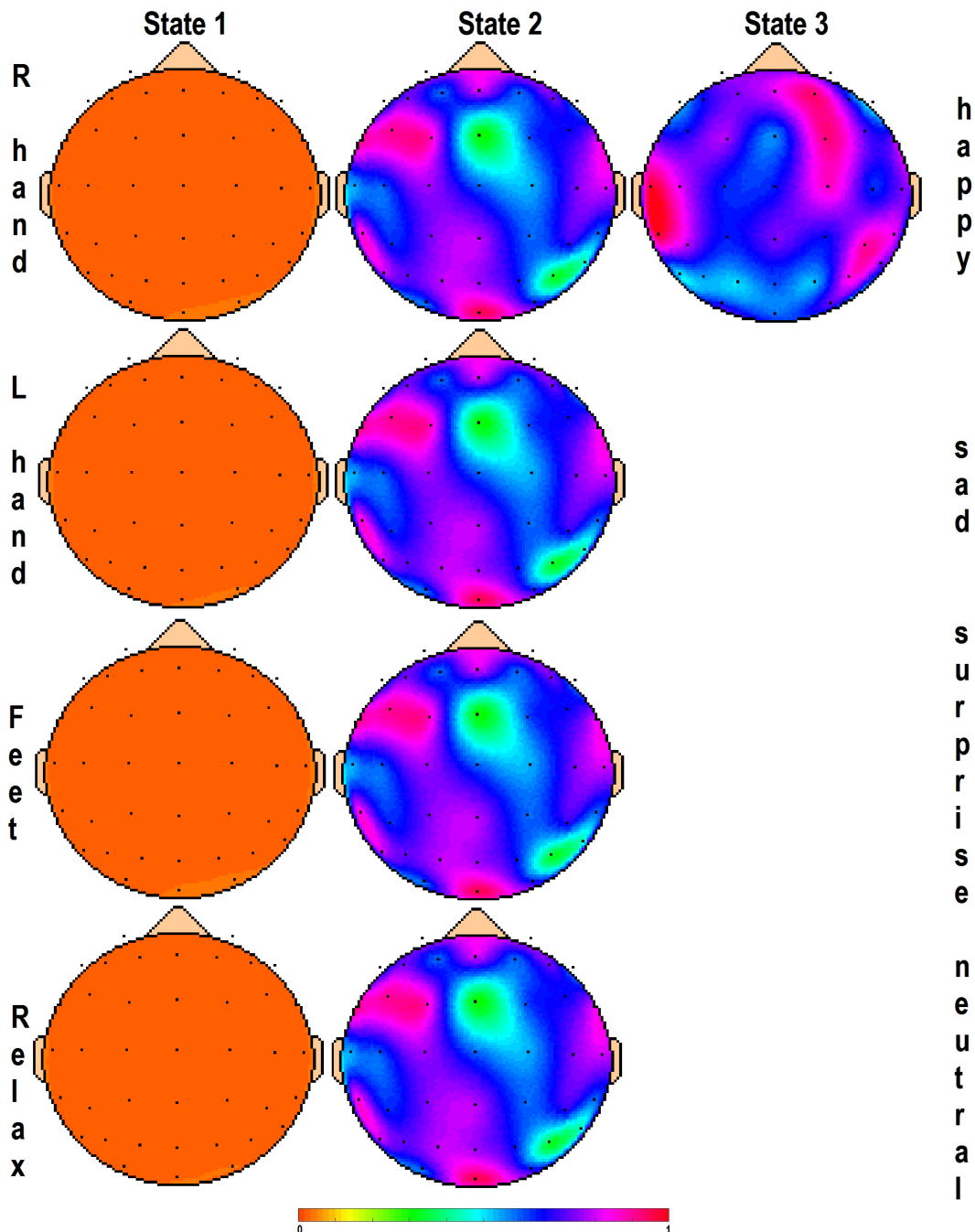
Figure 6.3: Topographies from the synchrostates of the averaged population for the β band. Columns represent the synchrostate number, up to 3 for the R hand movement and 2 for the rest of conditions. Rows indicate the performed task. From the top to the bottom: R hand imagined movement, L hand, Feet and Relax. On the right side, the emotion linked to each one of the tasks is indicated: happiness, sadness, surprise and neutral respectively.

schematic emotional faces stimuli. Another important observation is that the synchrostates head plots of the α band closely resemble those obtained in the β band. This behaviour is compatible with a previous study [134]. In the case of the γ band the synchrostates topographies still present some similarities, but

the fact that the number of clusters is larger than in the other two frequency bands makes a possible comparison difficult from just visual observation of the graphs.

Following the same steps explained in chapter 3, the temporal evolution of the synchrostates switching patterns are studied to understand the differences across stimuli and between frequency bands. Figures 6.6 to 6.7 illustrate the time-course plots for the α, β and γ bands respectively for the four stimuli. From the figures, it can be said that the overall transition patterns between frequency bands and also between the four stimuli are markedly different. This demonstrates the specific nature of the phase synchrostates.

From the topographies of state 1 in figures 6.2 and 6.3, it can be argued that the averaged phase difference for these synchrostates remains almost uniformly distributed across all the electrodes. The same behaviour can be found in states 1 and 4 in the γ band (figure 6.4). This may be explained by looking at the temporal switching pattern of the synchrostates (figures 6.6 to 6.7). From these time plots it can be seen that the averaged population remain for most of the time in the state 1, therefore, it can be said that the averaged population resides at a state of phase homogeneity for a longer period of time.

Observing the transition plots for the α band (figure 6.6) it can be observed that for the movement related tasks, R hand, L hand and feet, the transition starts in state 1, but for the relax task, it starts in state 2. The different initial states, for the different emotional faces, is in line with previous findings [134], where face perception tasks were studied from patients suffering from Autism Spectrum Disorder and a control group. Once more, as explained in chapter 3, it may be explained by the different processing pathways across participants of the image prior to the stimulus onset or it may be because of any other underlying process occurring in parallel and not related to the task. The solid red line indicates the stimuli onset and the dotted red lines are set to 300 and 600 milliseconds to help the visualisation process. This highlights the similar behaviour of the time course plots for all the stimuli between 100 and 300 milliseconds approximately, changing quickly between states at almost the same instant in time. It may be indicative of a similar processing pathway of the brain when analysing the image presented on the screen and the working memory process to remember the tasks linked to each face, even for the neutral face/ relax task. Afterwards, the transitions are slower, remaining longer in one state. This highlights the differences between the tasks. It is in line with the argument that the synchrostates are task-specific. The time period

between 300 and 600 milliseconds (between the dotted lines) shows a stable transition for the imagination of the feet movement and relax tasks, staying in state 1 and 2 respectively. Contrarily, the right and left hand imagined movement's transition plots remain switching between state 1 and 2 during this time period. It can be seen that the number of switches across states is considerably lower for the relax task than for the other three conditions. This can be expected as the participants were asked not to perform any movements. Consequently the brain processing flow should be lower.
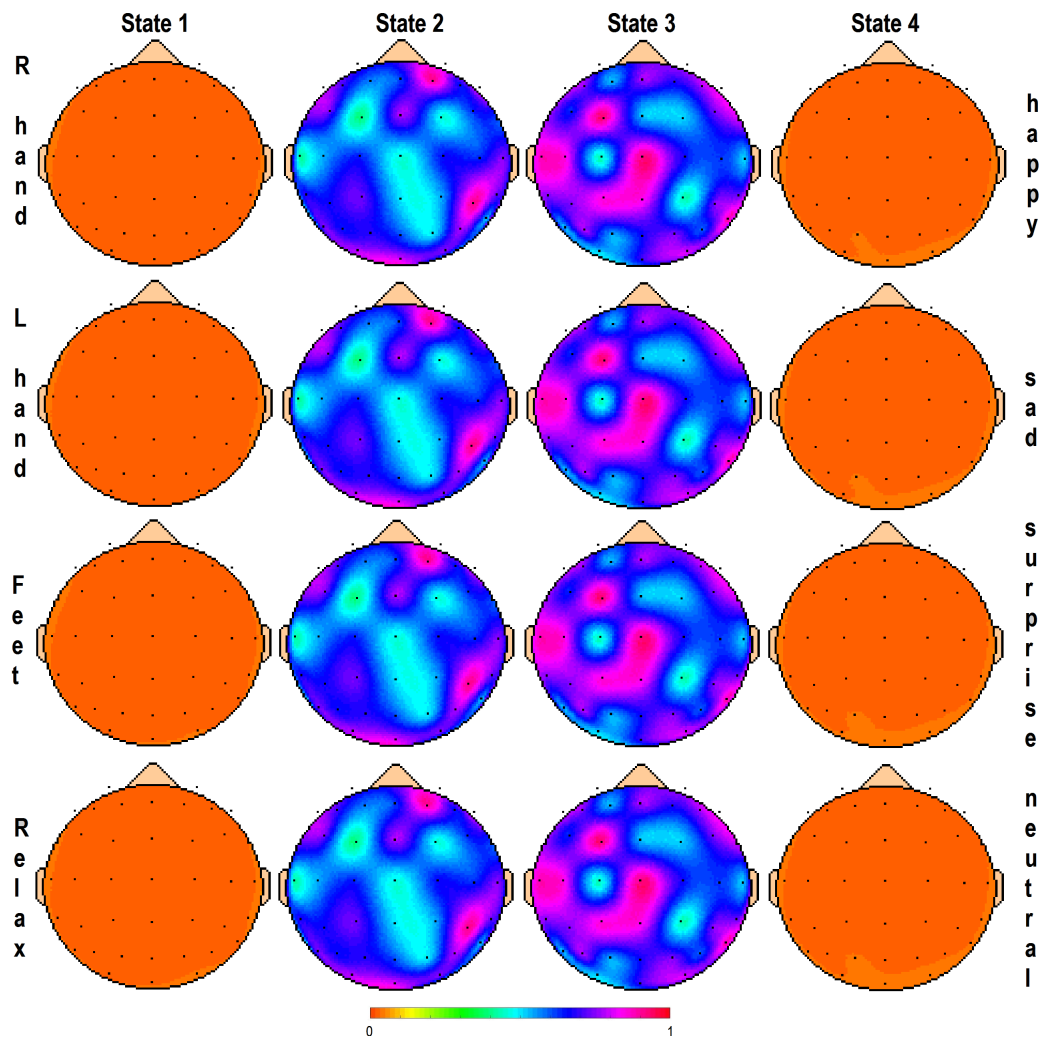
Figure 6.4: Topographies from the synchrostates of the averaged population for the γ band. Columns represent the synchrostate number, from 1 to 4 for all conditions. Rows indicate the performed task. From the top to the bottom: R hand imagined movement, L hand, Feet and Relax. On the right side, the emotion linked to each one of the tasks is indicated: happiness, sadness, surprise and neutral respectively.
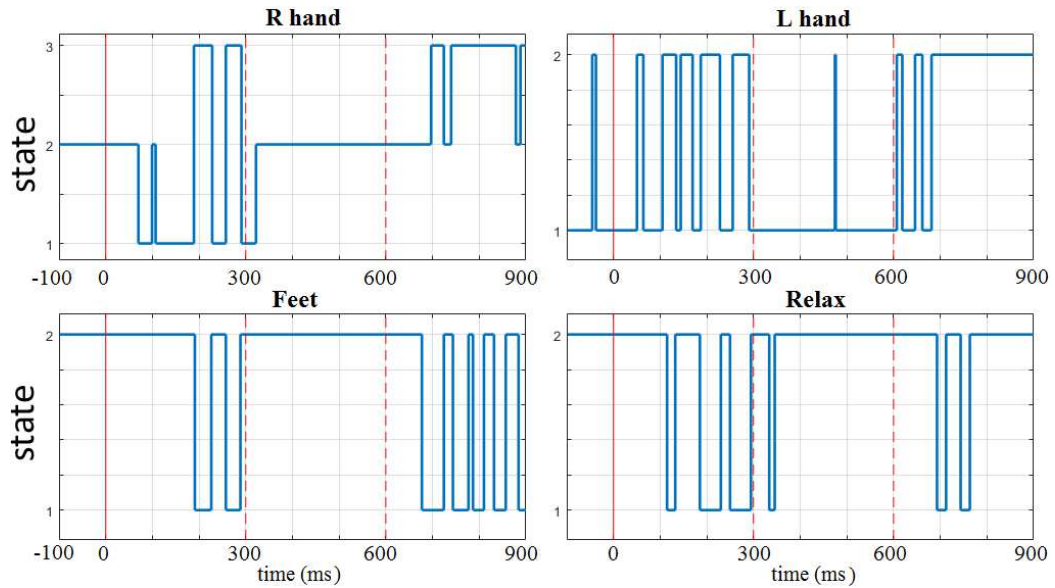
Figure 6.5: Temporal evolution of the clustered synchrostates for the β band for the four conditions: R hand (top left), L hand (top right), Feet (bottom left) and Relax (bottom right). The solid red line indicates the time instants of the stimulus onset and the dotted red lines are placed at 300 and 600 milliseconds after the stimuli onset.
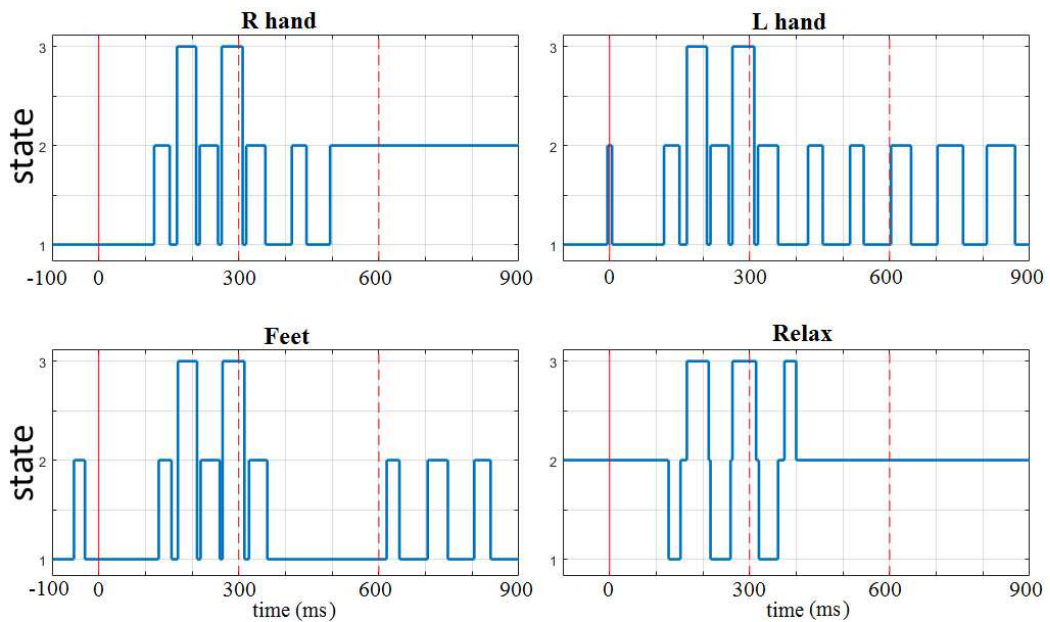
Figure 6.6: Temporal evolution of the clustered synchrostates for the α band for the four conditions: R hand (top left), L hand (top right), Feet (bottom left) and Relax (bottom right). The solid red line indicates the time instants of the stimulus onset and the dotted red lines are placed at 300 and 600 milliseconds after the stimuli onset. It seems that the period that elapses between the stimulus onset and 300 milliseconds, the temporal switching for all the cases presents similar pattern.

From figure 6.5 it can be observed that the L hand task switching pattern transitions are quicker than in the others, remaining for most of the time

in state 1, in contrast to the other three tasks where the most visited state
is state 2. It is noticeable that during the period of time between the two
red dotted lines, from 300 to 600 milliseconds, the four stimuli remain almost
completely stable in one state, number 2, for all the cases except for L hand
tasks which is state 1. In the γ band (figure 6.7) the transitions occur almost
constantly and occur in larger numbers than for the previous frequency bands.
This quick transition, the larger number of them and the presence of one more
synchrostate can lead us to think that maybe the γ band synchronisation can
reflect more complex tasks happening in parallel to the tasks presented to the
users. This is in line with the argument that γ band synchronisation is a fun-
damental process that follows any elemental operation of cortical computation
previously mentioned.
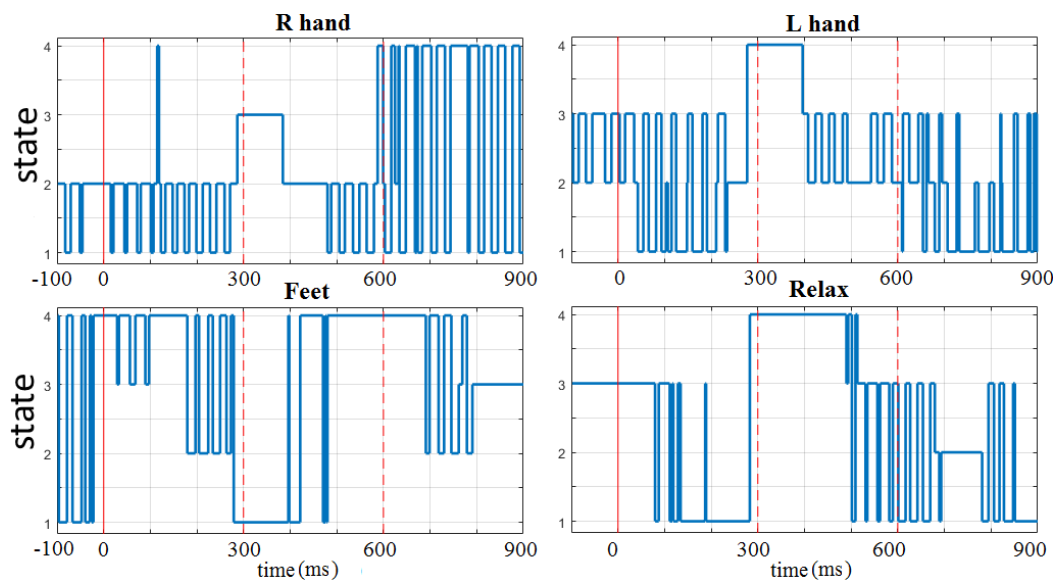
Figure 6.7: Temporal evolution of the clustered synchrostates for the γ band for the four
conditions: R hand (top left), L hand (top right), Feet (bottom left) and Relax (bottom
right). The solid red line indicates the time instants of the stimulus onset and the dotted
red lines are placed at 300 and 600 milliseconds after the stimuli onset.

6.3   Variability across subjects

The results presented so far, topographic maps and transitions between
synchrostates plots, are based on the averages across participants. The aim
of this section is to explore the inter-subject variability regarding the optimal
number of synchrostates across the diversity of frequency bands under scrutiny.
The differences in the phase matrices for each participant were clustered by
the *k-means* algorithm as explained in section 3.2.2 for each condition and

frequency band. Finally, statistical measures such as median and quartile ranges are used to illustrate the variability of the number of synchrostates as explained in section 3.3.2.
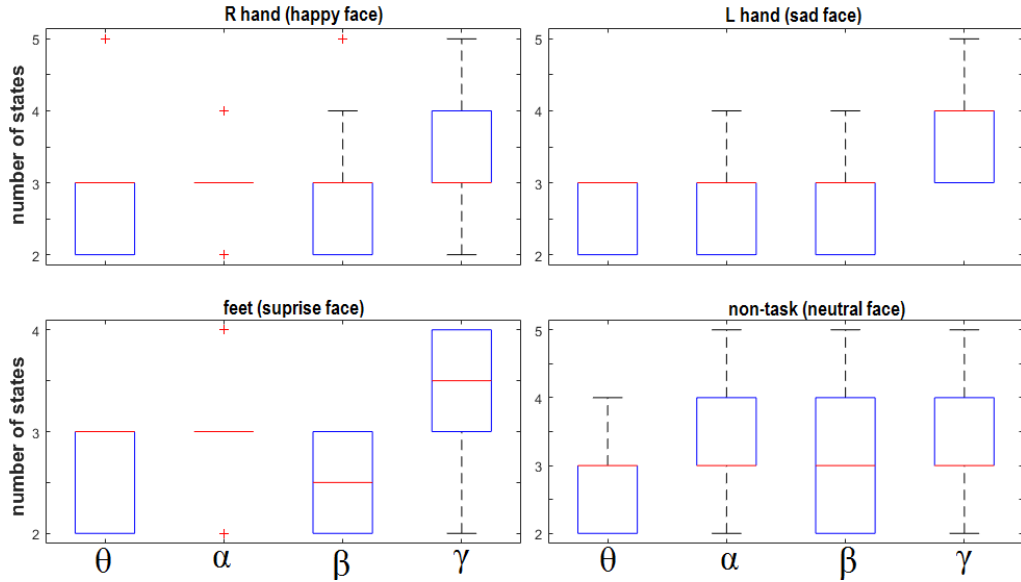
Figure 6.8: Variability in the optimal number of synchrostates for each subject during four different tasks: R hand, L hand, Feet and Relax. In brackets, the emotion stimulus linked to each task. The group was formed by 15 subjects and was replicated in four frequency bands (α, β, γ, ϑ).

The number of synchrostates across subjects varies between 2 and 5, 2 to 4 or 2 and 3 depending on the task and frequency band as can be seen from figure 6.8. This is in line with the results described in section 3.3.2, sustaining the existence of task-specific synchrostates. Once more, it can be mentioned that the small variation between the number of synchrostates can be due to the fact that the processing pathways are unique for each person and they can be influenced by several factors.

6.4 Variability due to other factors

Changing some parameters during the EEG recording such as the number of electrodes or sampling frequency may affect the final result on the head plots topographies or in the temporal resolution of the synchrostates transitions. It is likely that a higher sampling frequency can trace with more detailed granularity the phase evolution of the signals. In the same manner, a larger number of electrodes will probably lead to more consistent results than a reduced amount of EEG channels.

However, there is also a small fluctuation due to the variability from trial

to trial and from person to person, leading to a small variation in the number of synchrostates (see figure 6.8). This variation is reduced enough to consider the presence of synchrostates validated across participants. In this work, a number of different random initialisations of the *k-means* clustering algorithm have been used in the aim of obtaining more consistent results. Also the sampling frequency was increased from 256Hz in the first study (chapter 3) to 512Hz to gain in time-scale granularity.

The dependence on the type of experiment has been tested with two different sets of participants and two different types of stimuli, schematic and human faces, with regulated recording equipment and in a controlled environment. In addition, similar results were previously detected in autistic children and a control group [134]. However, the presence of synchrostates using other types of stimuli, different from the processing of emotional faces, has not being tested yet as it is not within the scope of this thesis.

6.5 Network analysis

Synchrostates have been defined as states where the phase difference across electrodes remained constant over time, specifically of the order of milliseconds. To study their stability in a time context and keeping the information related to their phase difference values at the same time, the synchronisation index is selected as the suitable measure. Afterwards, once the synchronisation matrices are formulated for each one of the synchrostates and frequency bands of interest, those matrices can be translated into complex network measures. The aim of this process, fully described in chapter 4, is to gain further information about the behaviour of the phase synchronisation amongst different areas of the brain. Therefore, a deeper insight into the temporal dynamics of the connectivity patterns for a specific task can be obtained.

6.5.1 Synchronisation index

The synchronisation index for each synchrostate, task condition and frequency band has been calculated according to 4.1. The cross electrode plots resulted from the synchronisation index in the α band are shown in figure 6.9. Similar to the process described in section 4.1, the values closest to 1 (reddish colour on the graph) indicate a larger degree of synchronisation.

After the analysis of the results from chapter 4, it was determined that those states having the largest number of occurrences were the most stable and probably will carry vital information. In the same manner, the states with the lowest number of occurrences were also included in the connectivity analysis as

they may hold some specific information related to the brain connectivity. In line with the nomenclature followed in chapter 5 the maximum and minimum number of occurrences will from now be named the 'max' and 'min' states respectively.
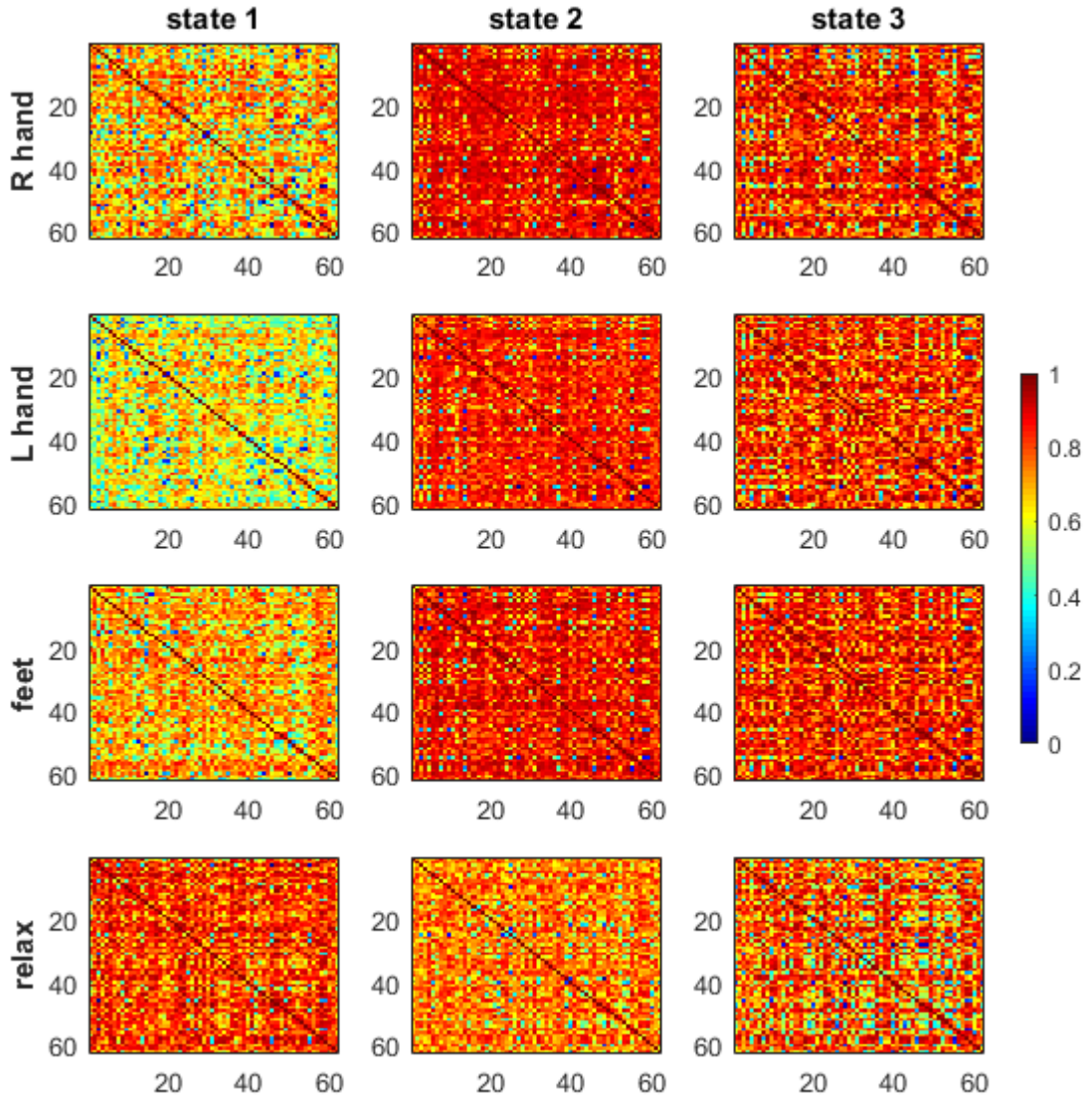
Figure 6.9: Synchronisation Index $\Upsilon_P(F)$ matrices for the three synchrostates obtained for the α band and task condition.

6.5.2 Network measures

The synchronisation index obtained for the max and min states for each condition and frequency band will be used as a weighted adjacency matrix indicating the degree of connectivity between nodes or EEG electrodes. This will contribute to a further understanding about how well connected the different pairs of electrodes are or nodes across the areas of the brain and also about

the temporal stability of each state. The process is described in chapter 4.

Figures 6.10 - 6.12 show the resulting connectivity maps for the max (left column) and min (right column) states for each condition and frequency band: the α, β and γ bands respectively. The rows represent the different tasks: (from top to bottom) R hand, Left hand, Feet and Relax linked to happy, sad, neutral and surprised face stimuli respectively. The resulting connectivity graphs for the max and the min states are shown with only 5% of the strongest connections retained for further analysis. This was performed to maintain a similar average degree distribution between subjects, since the betweenness centrality can be affected by the degree of a network [221]. A similar threshold was selected to perform the analysis as detailed in chapter 4. A range of thresholds from 3 to 10% was examined as they are the most widely used in the literature. However, those thresholds were not found to significantly change the properties of the examined connectivity networks.

Following the same criteria than as in previous chapters, the colours and sizes of the nodes are based on their degree, meaning that a bigger diameter of the node contains a higher number of links connected to this node. Similarly, a reddish colours of the nodes, signifies more links connected to the node; Bluish, a lower number of links connected to it. Furthermore, the edge thicknesses are based on the weighted values of the synchrostate index matrices. Consequently, thicker lines connecting electrode pairs mean higher values than thinner ones in the adjacency matrix. As mentioned in chapter 4, the nodes situated outside of the brain contour correspond to the nodes labelled as Tp9 and Tp10, situated by the ears. This is due to the way EEGNET software plots the connectivity graphs.

An interesting observation from Figures 6.10 to 6.12 is that the relax task for all frequency bands shows, in general, the less localised connectivity compared to the other tasks linked to an imagined movement. In addition, the strength of the edges represented by the thickness of the connections between nodes is lower in relation to the other three stimuli. It may mean that the number of processes ongoing within the brain when this stimulus is presented to the user is lower, consequently the need of specialised information integration operations are less that those needed for a motor imaginary task.

From the connectivity figures it can also be noticed that there is a clear difference between the max and min states for the motor related tasks, especially in the α and γ bands. The min state is the one presenting more segmented connectivity in relation to the max state. This may mean that most of the specialised information integration operations occur during the min state.
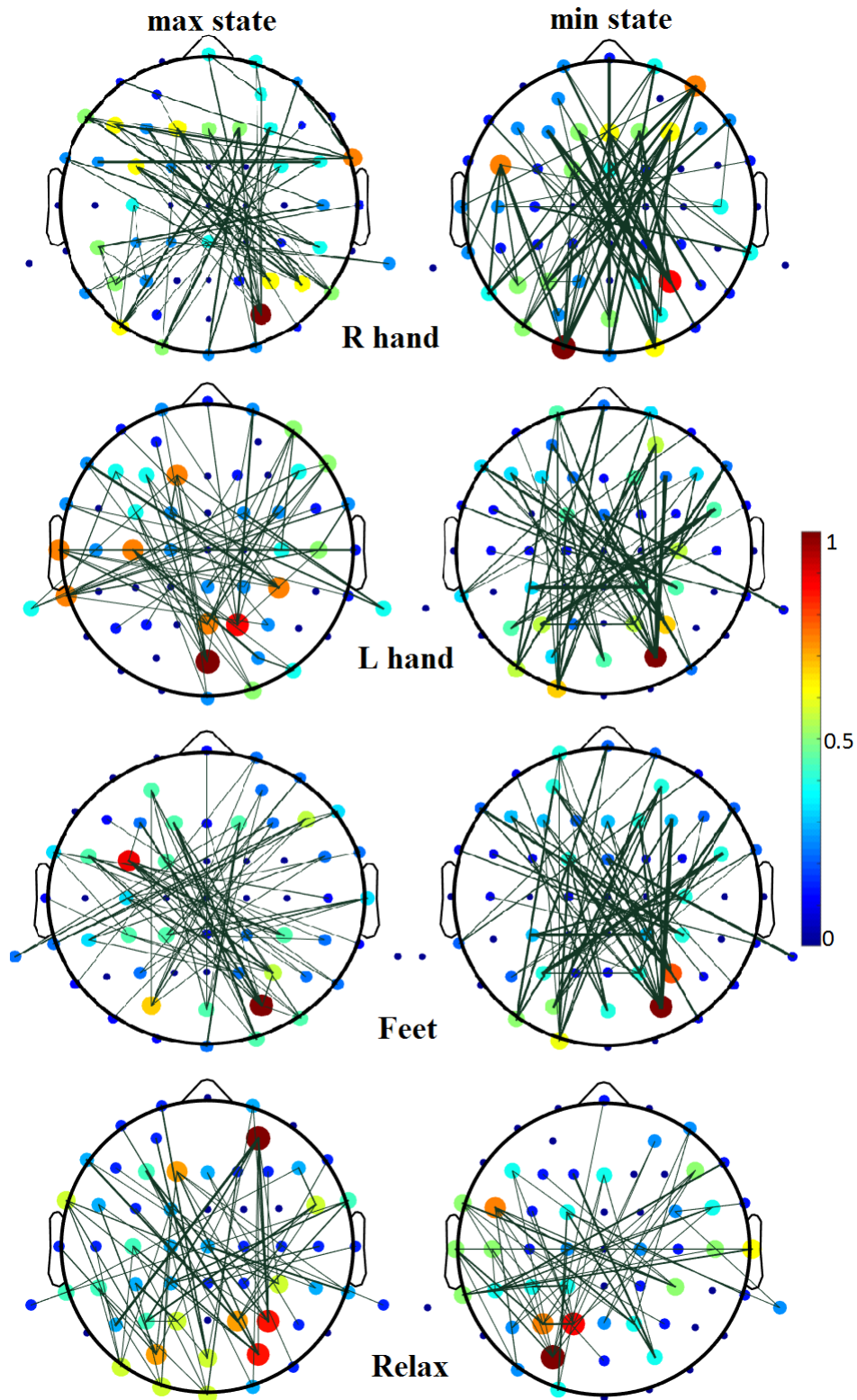
Figure 6.10: Brain connectivity plots from the maximum and minimum number of occurrences synchrostates in the α band for all four conditions: R hand movement, L hand movement, Feet movement and relax. The colours and sizes of the nodes are based on their degree: bigger diameters and reddish colours of the node meaning a larger number of links. The edge thickness is based on the weighted values of the synchrostate index matrices.
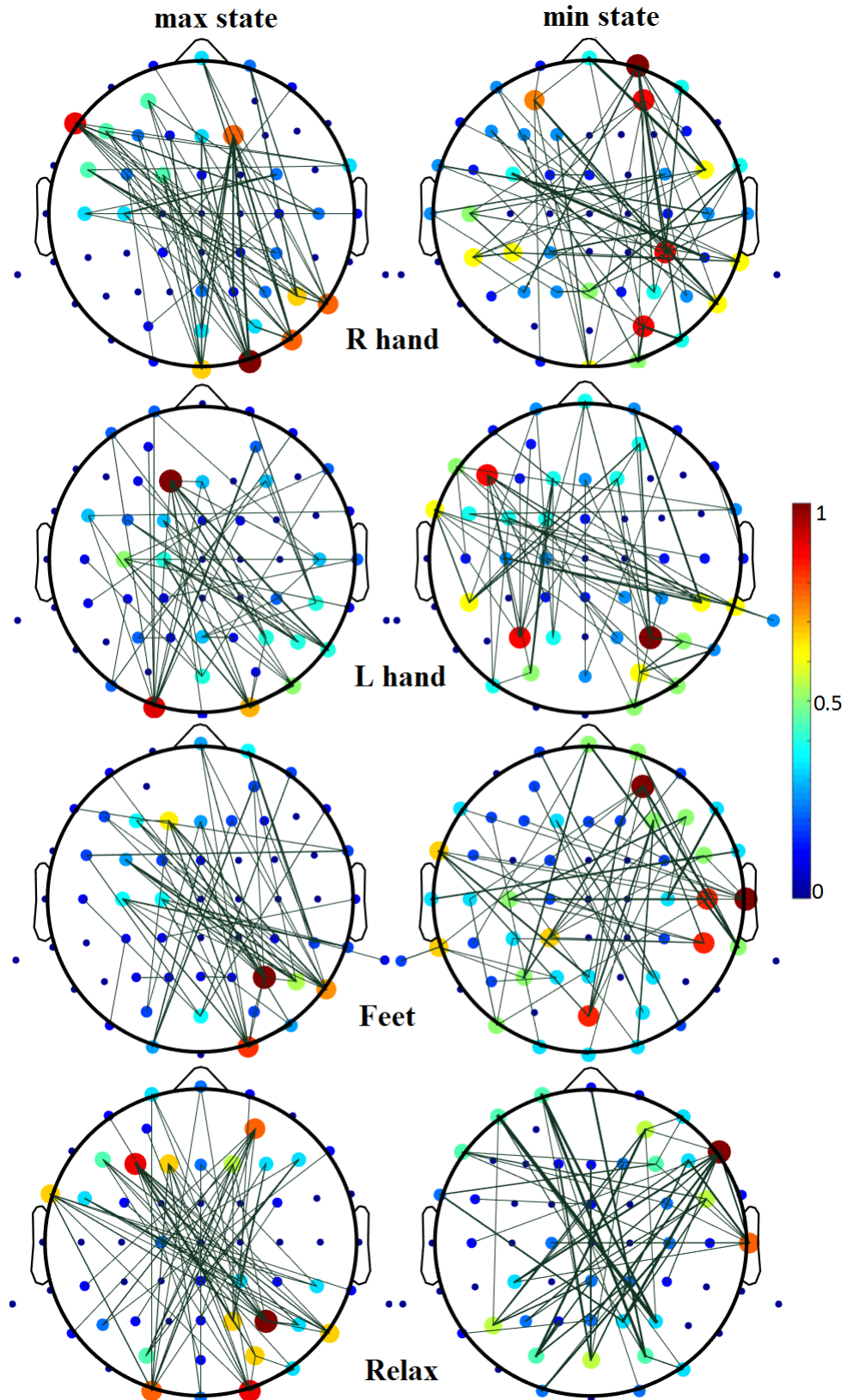
Figure 6.11: Brain connectivity plots from the maximum and minimum number of occurrences synchrostates in the β band for all four conditions: R hand movement, L hand movement, Feet movement and relax. The colours and sizes of the nodes are based on their degree: bigger diameters and reddish colours of the node meaning a larger number of links. The edge thickness is based on the weighted values of the synchrostate index matrices.
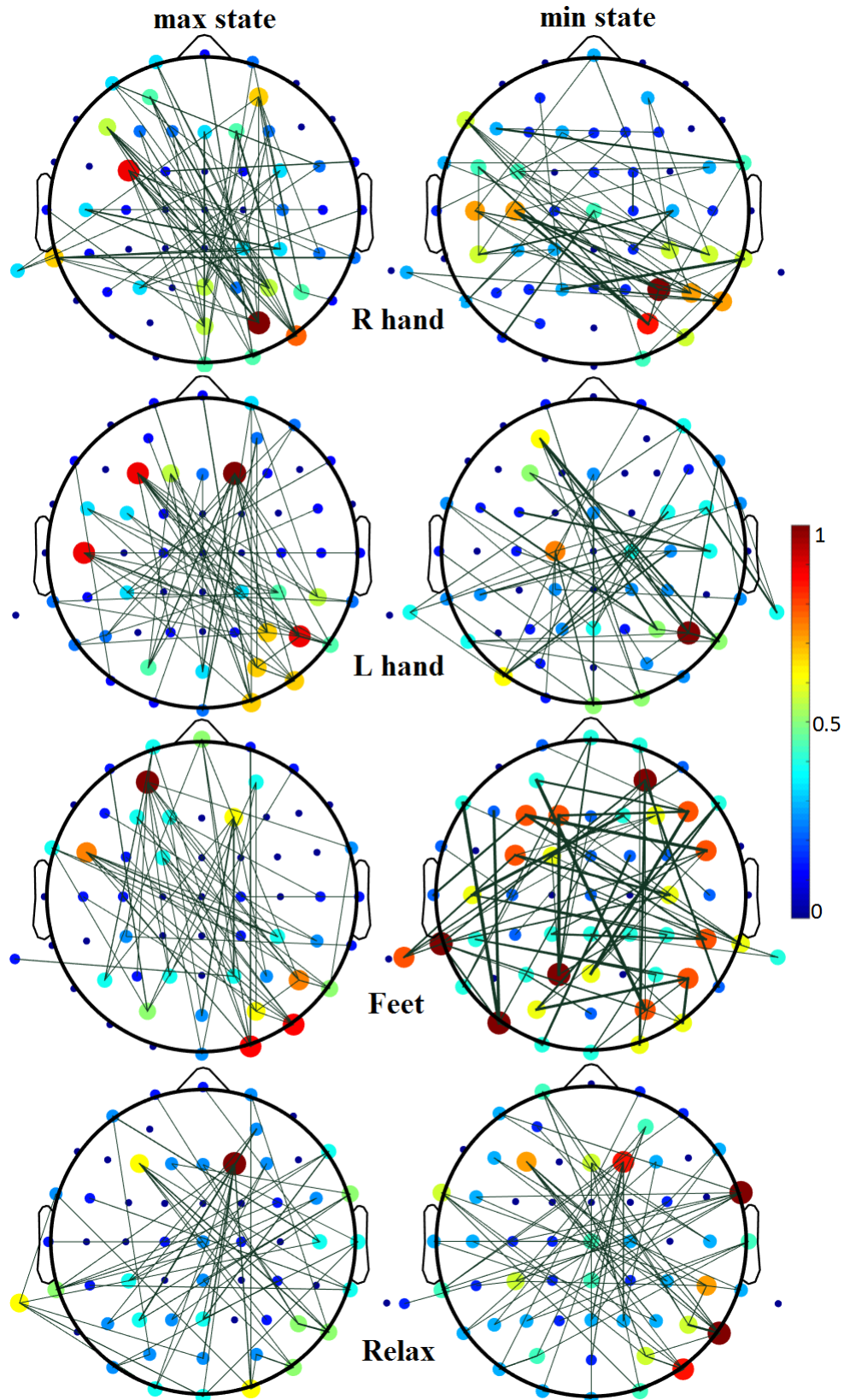
Figure 6.12: Brain connectivity plots from the maximum and minimum number of occurrences synchrostates in the γ band for all four conditions: R hand movement, L hand movement, Feet movement and relax. The colours and sizes of the nodes are based on their degree: bigger diameters and reddish colours of the node meaning a larger number of links. The edge thickness is based on the weighted values of the synchrostate index matrices.

Therefore, min state quantitative characterisation may lead to a higher power of discrimination between tasks. This is according to the results found in the previous chapters. This phenomenon is not as clear in the β band where the differences between max and min states are less intuitive from a visual point of view.

It can be said that there are some similarities in the α band (figure 6.10) between the min states for the three motor imaginary tasks (the top three topographies of the right column), specially between L hand and both feet movements. This behaviour cannot be observed in the other frequency bands. This may be explained by the influence of other processes occurring simultaneously in the brain that can be frequency specific [134]. Therefore, affecting differently to the range of frequencies observed.

From the above connectivity plots, as mentioned in chapter 4, objective information cannot be obtained, only some visual conjectures and a more quantitative evaluation mechanism is needed to yield any solid conclusions. For this reason, in addition to the connectivity network topographies showed in figures 6.10 to 6.12 a set of connectivity metrics was also obtained for each frequency band, condition and max/min synchrostates. For a more exhaustive explanation about the meaning and formulation of each one of the graph theory based measures please refer to chapter 4. Tables 6.1 and 6.2 show two examples of the network measures calculated for the α and γ bands respectively.

Table 6.1: Example of two network metrics, characteristic path length (CPL) and global efficiency (G.E.), for the max and min synchrostates for the four conditions (R hand, L hand, Feet, Relax) for the α band.

measures	R hand		L hand		Feet		Relax	
	max state	*min state*	*max state*	*min state*	*max state*	*min state*	*max state*	*min state*
CPL	4.402	4.29	3.515	3.868	4.451	4.50	4.017	4.52
G.E.	$151*10^{-3}$	$165*10^{-3}$	$130*10^{-3}$	$129*10^{-3}$	$156*10^{-3}$	$136*10^{-3}$	$114*10^{-3}$	$143*10^{-3}$

Table 6.2: Example of two network metrics, modularity and transitivity, for the max and min synchrostates for the four conditions (R hand, L hand, Feet, Relax) for the γ band.

measures	R hand		L hand		Feet		Relax	
	max state	*min state*	*max state*	*min state*	*max state*	*min state*	*max state*	*min state*
Q	$61.4*10^{-2}$	$71.8*10^{-2}$	$63.7*10^{-2}$	$70.0*10^{-2}$	$66.4*10^{-2}$	$71.6*10^{-2}$	$70.9*10^{-2}$	$70.3*10^{-2}$
T	$66.5*10^{-2}$	$86.9*10^{-2}$	$61.4*10^{-2}$	$80.6*10^{-2}$	$59.9*10^{-2}$	$84.3*10^{-2}$	$70.0*10^{-2}$	$69.3*10^{-2}$

From table 6.1 it can be seen that the max state for the L hand and Feet movement tasks present higher GE values and lower CPL than their corresponding min states, meaning that the min states for L hand and feet conditions have a larger information integration ability than their max states. However,

for the R hand tasks the max state boasts this feature. Furthermore, despite some states leading the integration capacity of the brain for the specific task, all of the motor tasks have a similar value of those features, meaning that all of them perform to a comparable level of integration process. This can explained by the notion that at the end, although different stimuli have been shown to the participant, all belong to the same cognitive task. It means that similar information integration specific to this task and stimuli is performed. This behaviour also explains the similarities between the head plot topographies for the different states within a frequency band.

Also noticeable are the higher values of the transitivity in table 6.2 of the min state compared with the max state for the motor imaginary tasks (feet and R and L hands), This is not the case for the Relax condition, where min and max states show similar values for both connectivity metrics. These results corroborate the previously mentioned regarding the proposition that the min state can lead the segregated specialised processing.

6.6   Four class classification from synchrostates

In chapter 5 the possibility of using connectivity network metrics from the synchrostates to differentiate between two motor imaginary tasks, right and left hand, has been clearly demonstrated. In this section the pursued aim is similar, using the metrics from the max and min synchrostates as features to classify between different tasks to control a BCI system. However, two main differences can be highlighted. Firstly, from a two class classification problem, right and left hand, it has increased to a four class problem: right hand, left hand, feet and no movement or a relax task. The other important difference is the type of stimuli used, previous chapter results are based on schematic emotional faces as stimuli, now the emotions are transmitted by images of human faces.

The steps involved in the classification process are detailed in chapter 5. Briefly, the individual and collective discrimination power of the features coming from the max and min synchrostates are determined by the FDR. Three cases will be considered: only min state features, only max state features and all together. Following the same nomenclature as in chapter 5, the different cases will be named case I (all features), case II (max state features) and case III (min state features). In addition to the α and β bands studied in chapter 5, γ band has also been included in this study. Nevertheless, only the connectivity metrics when a threshold is used will be considered as they demonstrate having a stronger discriminative capacity. The number of and type of classifi-

ers remain identical to the previous study. Please refer to section 5.3 for more details.

6.6.1   Results for the α band

The FDR separability criterion results for cases I to III are illustrated in figures 6.13-6.15 respectively. The whole set of features is divided into smaller subsets according to their FDR values. These reduced groups of features will be used later on to feed the different classification algorithms. For case I, figure 6.13, the number of groups after the grouping process is 9.

The first group is formed by the top one feature, degree network metric from max synchrostate. The following groups are formed by 2 features, the GE value and also from the max synchrostate and modularity (Q) of the min synchrostate. The next four groups are formed by only one feature, followed by 3 and 2 features groups and the last ensemble includes all the available features.

Figure 6.13: Grouped features by their FDR values ordered in decreasing order for case I, all network measures are considered, and the α band.

Following the same criteria, the process of dividing into smaller sets of features to feed the classifiers is applied to cases II and III where only the max and min synchrostates network measures are considered respectively. The number of groups for case II is 6 as can be seen from figure 6.14. The first one is formed by the top one feature which is the same as in case I. The next two sets are also formed by only one feature, followed by groups of 3 and 2 elements respectively. Finally, the 6th group that includes all features.

Case III's situation, that only considers the network metrics obtained from the min synchrostate, is illustrated in figure 6.15. The number of groups that resulted in this particular case is only 5 and except for the last one that includes all of the range of features, the rest are formed from a single metric. A big gap is noticeable between the FDR values of the top four features and the rest.

This means that their ability to discriminate between the different MI tasks will be remarkably lower than for the top four features.

Figure 6.14: Grouped features by their FDR values ordered in decreasing order for case II, only max synchrostate network measures are considered, and the α band.

It can been observed from figures 6.13, 6.14 and 6.15 that the connectivity measures with higher FDR values, hence, a larger discriminative power, are modularity, transitivity, global efficiency or the degree of the network. This is in line with the expected behaviour as commented in previous sections of this chapter.

Figure 6.15: Grouped features by their FDR values ordered in decreasing order for case III, only min synchrostate network measures are considered, and the α band.

The different subsets of features for cases I to III were used to feed six different classifiers using *leaving-one-out* validation following the criteria established in chapter 5. For illustrative purposes, only the top two classifiers showing the highest performance among the different cases and frequency bands are presented in this section. Figure 6.16 illustrates the performance of the algorithms. The first one is a linear discriminant analysis classifier (ldc, top graph) and

the second one is a quadratic discriminant classifier, (qdcbottom graph). This behaviour is similar to the previous chapter, being also those two algorithms which achieved the best performances classifying between two MI tasks. It can be seen that the higher average accuracy for the linear discriminant algorithm is obtained when the top three features are used, being 90% with true positive values of 100% for the R and L hand tasks, and 84% for the non-task situation and 76% for the feet task.
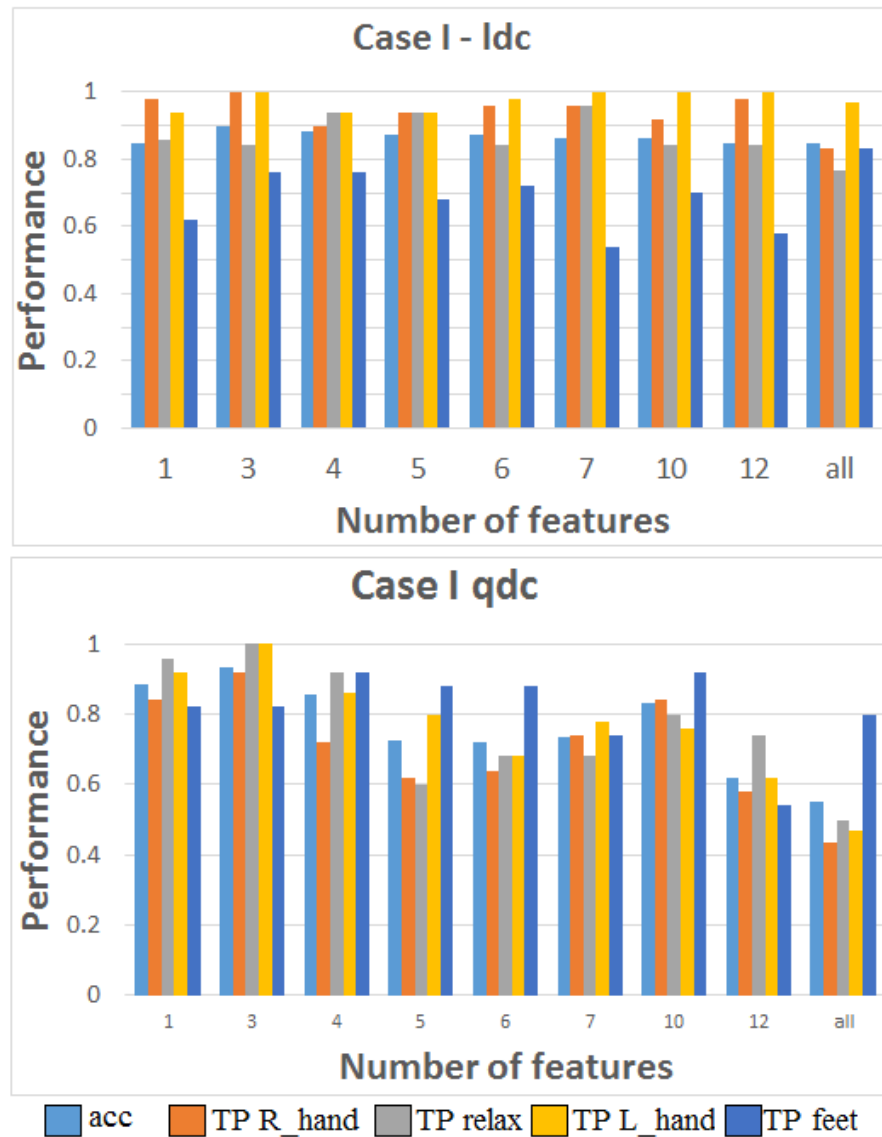
Figure 6.16: Comparison of the performance of the top two classifiers for case I in the α band. Features grouped according to their FDR value are fed to the different classification algorithms. For each one of the formed subsets, the average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand, relaxing or no-task, L hand and feet) are illustrated. The upper graph shows the performance results for the linear discriminant analysis classifier (ldc) and the bottom graph shows the results obtained for the quadratic discriminant classifier (qdc).

Similar results are obtained for the first two groups of features when the quadratic discriminant algorithm is used , the bottom graph of figure 6.16, achieving a maximum averaged accuracy of 93% when the top three features are used with a true positive rate of 92%, 100%, 100% and 82% for R hand, no-task, L-hand and feet respectively. The addition of more features to both classifiers leads to a lower averaged performance that once more, can be attributed to the over-fitting phenomenon. It can be seen that for the linear discriminant algorithm, that the worst performance is obtained for the last MI tasks, consisting of imagining moving both feet. This behaviour is not present when the quadratic discriminant algorithm is used and where the true positive rate between the four tasks is more balanced. Maybe this is due to the lower general averaged performance of this classifier when compared to the linear discriminant classifier.
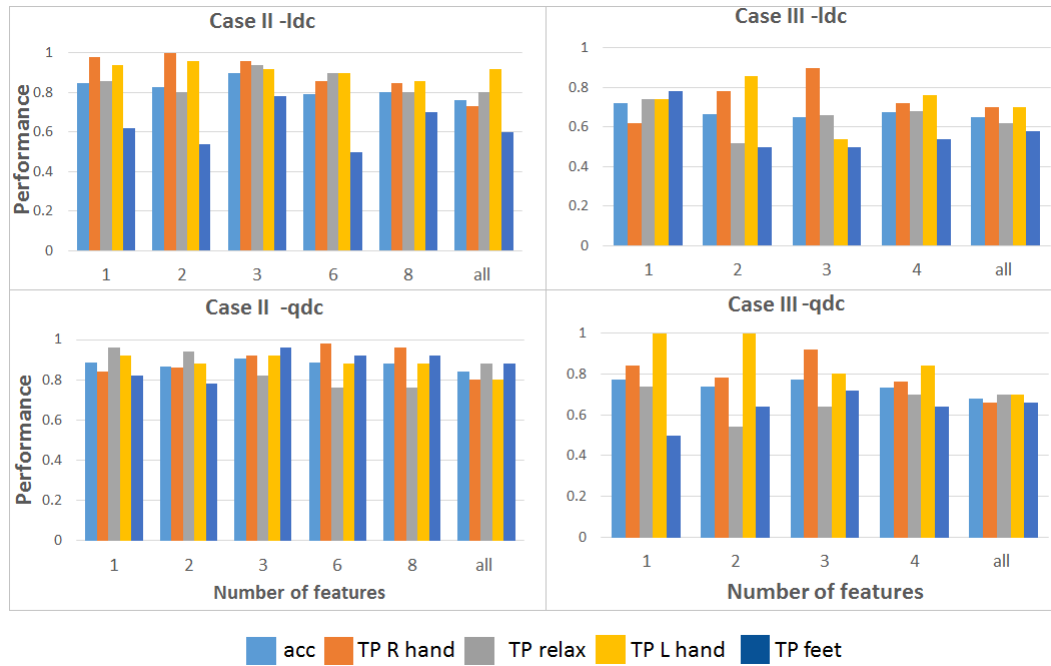
Figure 6.17: Comparison of the performance of the top two classifiers for case II (left column) and case III (right column) in the α band. Features grouped according to their FDR value are feed to the different classification algorithms. For each one of the formed subsets the average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand, relaxing or no-task, L hand and feet) are illustrated. The upper row shows the performance results for the linear discriminant analysis classifier (ldc) and the bottom row shows the results obtained for the quadratic discriminant classifier (qdc).

The performance for case II and case III is displayed in figure 6.17. The left column of the figure illustrates case II, when only network metrics corresponding to the max synchrostate are considered. The right column, consequently reflects the performance obtained for case III, when only the min synchrostate

connectivity measures are used to feed the classification algorithm. The higher performance obtained in this case for the linear discriminant algorithm is similar to that in case I, 90% with TP rates of 96%, 94%, 92% and 78% for R hand, non-task, L hand and feet respectively. This averaged accuracy is scored when three features are used to feed the algorithm. The top two features are the same as for case I so it is logical to think that the level of accuracy should be similar. Once more the MI task of moving both feet is the one obtaining the lower performances. Similar numbers resulted from the second classifier under study, quadratic discriminant, with an averaged accuracy of 90.5% (92%, 82%, 92%, 96%) when the top three features are used. In this case the poorest performance is linked to the non-task (resting) situation.

Finally, performance results for case III are visibly lower than the two previous cases. This behaviour can be explained by the lower values of the FDR when compared to the values of case II, see figures 6.14 and 6.15. In this case, for both classifiers, the greatest performance achieved is when the top feature, modularity is used to feed the algorithm. The averaged accuracy obtained is 72% and 77% for linear and quadratic classifiers respectively.

The results obtained for the α band after classifying the four different tasks are considerably higher than those obtained in section 5.4.1.1 when only two tasks were classified (highest accuracy of 93% versus 83% respectively). This can be explained through the FDR values which are almost double for the 4-tasks classification problem than for the 2-tasks problem as can be seen from figures 6.13 and 5.7 respectively. The higher results shown in this section may be explained by other factors such as the larger number of participants or that a higher number of tasks to perform during the same experiment induced an increment in the concentration of the participants. Another difference with respect to the 2-tasks classification problem is that the connectivity measures obtained from the min synchrostates present lower FDR values, hence poorer discriminative abilities than the measures from the max synchrostates. This is in line with the previous existing literature where the maximum number of occurring state network metrics have better discriminative capabilities than the minimum number of occurring state metrics [30, 134].

6.6.2   Results the β band

Following the same steps used for the α band the FDR values of the different network metrics obtained from the max and min synchrostates are arranged in descending order to subdivide the whole set into smaller groups to feed

the different classifiers as depicted in figures 6.18 to 6.20. The same number of cases are studied for the β band, starting from case I, where all available features are included (figure 6.18). Only metrics of the max synchrostates are considered for case II (figure 6.19) and only measures of min synchrostates for case III (figure 6.20).
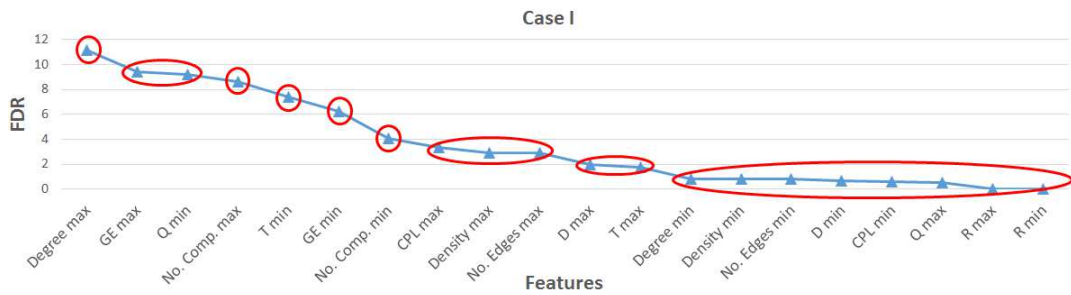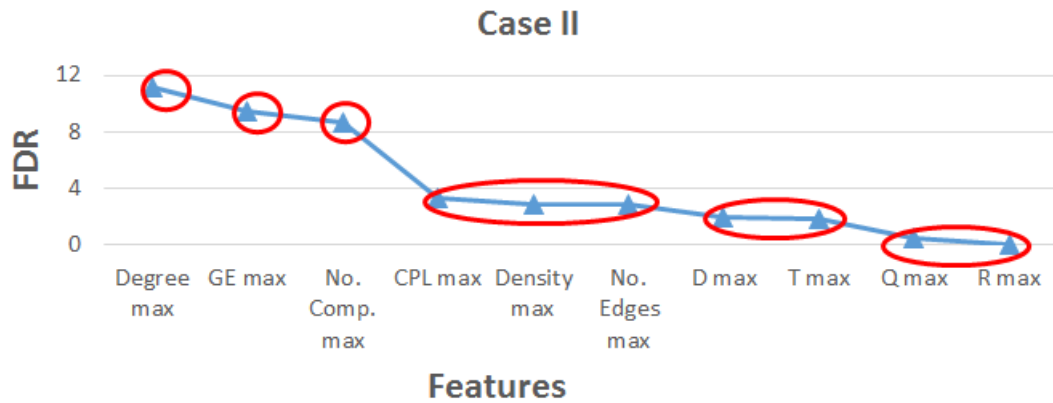

Figure 6.18: Grouped features by their FDR values ordered in decreasing order for case I, all network measures are considered, and β band.

The number of groups formed for case I is eight, one less than for the α band, as can be seen in figure 6.18. The first four groups are formed by a single feature: CPL max, diameter (D) max, transitivity (T) min and T max respectively. The successive groups are formed by 6, 8, 13 and all possible features respectively. Similar to the previous section, the top two features are from the max synchrostate set of metrics. In addition, it is noticeable that the FDR values are considerably lower than in the α band which means that the power of discrimination of the β band is in general, worse than that of the α band.
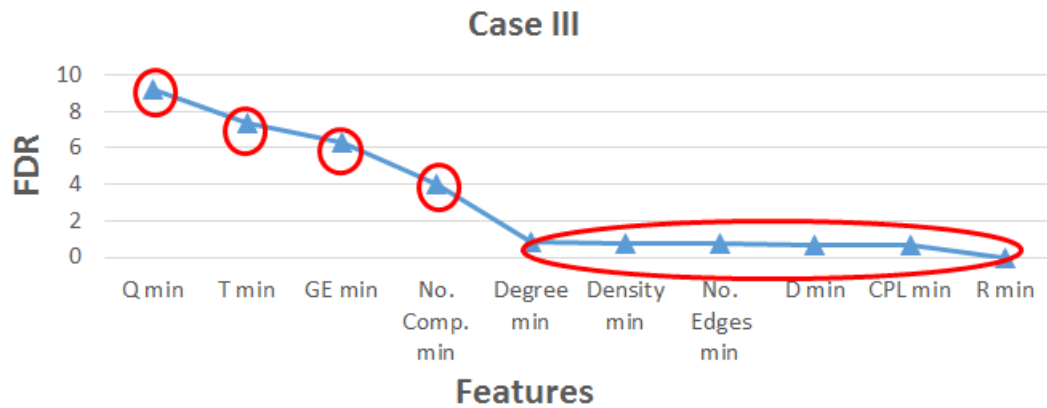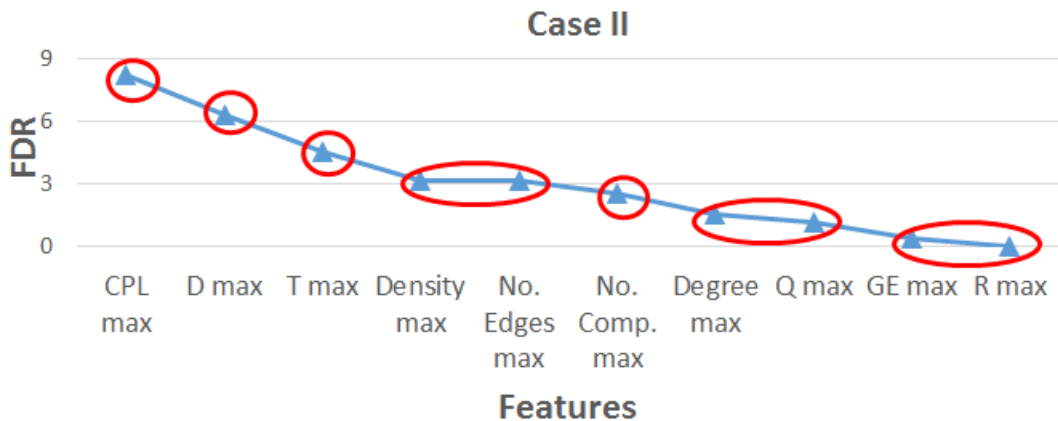

Figure 6.19: Grouped features by their FDR values ordered in decreasing order for case II, only max synchrostate network measures are considered, and the β band.

For case II the number of groups has increased from six in the α band to seven in the β band as illustrated in figure 6.19. The first three groups are formed of only one feature, with the same features as in case I: CPL max, D max and T max. The subsequent subgroups are formed by 5, 6, 8 and the whole set of features respectively. Figure 6.20 shows the last situation, case III. The number of groups resulting in case III is only four: 1, 3, 5 and all features respectively. Similar to case III of α the band, the FDR values in this scenario are significantly lower than for case II and case I.
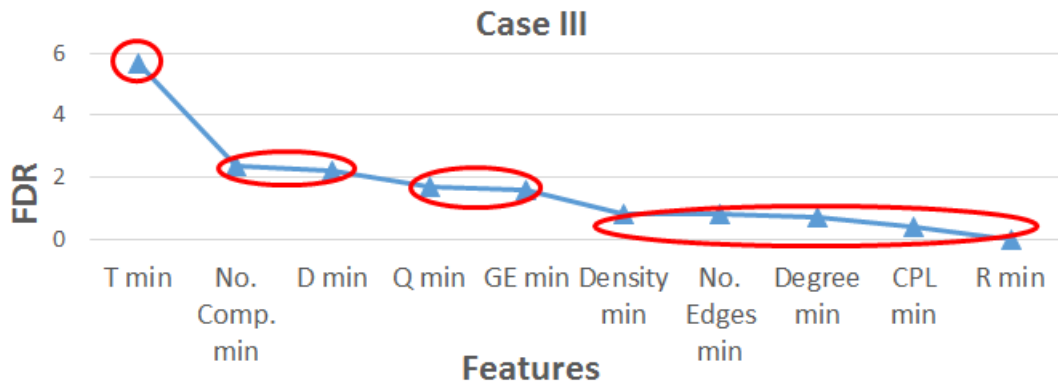
Figure 6.20: Grouped features by their FDR values ordered in decreasing order for case III, only min synchrostate network measures are considered, and the β band.

Figures 6.21 and 6.22 show the classification performance rates for case I and case II and III respectively for the 4-tasks classification problem. For case I, the highest averaged accuracy for both classifiers is obtained with only one feature, the one presenting the largest FDR value, CPL max. The accuracy rate is 73% with true positive rates of 100%, 70%, 70% and 52% for R hand, non-task, L hand and feet respectively for the linear discriminant algorithm and 82.5% (100%, 100%, 50%, 80%) for the quadratic discriminant classifier. This performance is substantially lower than in the same case for the α band where the averaged accuracy was over 90%. This behaviour is in line with the lower FDR values for this frequency band as commented beforehand.

In addition, the true positive rates for both classifiers are clearly unbalanced. Both algorithms show strength in detecting some tasks especially R hand movement with 100% of hits, but are really bad for others such as feet or L hand movements with rates near the chance level. It can be seen that the over-fitting phenomenon is more pronounced in this case than in the same scenario of the α band. For both classifiers the performance drops to 50% or less when more than 9 features are used.

Figure 6.21: Comparison of the performance of the top two classifiers for case I in the β band. Features grouped according to their FDR value are fed to the different classification algorithms. For each one of the formed subsets, the average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand, relaxing or no-task, L hand and feet) are illustrated. The upper graph shows the performance results for the linear discriminant analysis classifier (ldc) and the bottom graph shows the results obtained for the quadratic discriminant classifier (qdc).

For case II, illustrated in the left row of figure 6.22, the first two groups are formed by the same features than for case I, consequently the performance of the classifiers is the same as in case I, 73% and 82.5% for linear and quadratic algorithms respectively. The second classifier, quadratic discriminant, shows slightly higher averaged accuracies in general than linear discriminant algorithm, with the exception of the last group. When all of the features are used to feed the quadratic discriminant algorithm, it cannot classify correctly

any feature but feet movement which reveals a true positive rate of 100%. Similar to the other cases studied so far the two tasks obtaining lower classification performance are, in general, the neutral face linked to the non-task or relax situation and the movement of the feet.

Figure 6.22: Comparison of the performance of the top two classifiers for case II (left column) and case III (right column) in the β band. Features grouped according to their FDR value are fed to the different classification algorithms. For each one of the formed subsets, the average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand, relaxing or no-task, L hand and feet) are illustrated. The upper row shows the performance results for the linear discriminant analysis classifier (ldc) and the bottom row shows the results obtained for the quadratic discriminant classifier (qdc).

Finally, case III presents the slightly lower performance than in case II and case I, with 71% of the averaged accuracy for both classification algorithms. Similar to case I and II the true positive rates for the different tasks are unbalanced. But in this case the performance is extremely poor for the non-task scenario. These results can be explained by the fact that the synchrostates are task and frequency specific as has been demonstrated earlier in this thesis. In addition, the connectivity plots and consequently, connectivity measures show clear differences across states and frequencies. This means that they process different information, for example the minimum number of occurrences state performs a more specialised processing, leading to a different classification performance according to the tasks and frequency band being executed.

6.6.3   Results for the γ band

The FDR results for the last frequency under study are shown in figures 6.23 to
6.25 for cases I to III respectively. For case I, only seven groups were formed,
being the first four individual feature groups formed with the top highest
values of the FDR. The rest of the groups for case I were composed of 9, 13
and all sets of features respectively. Three out of the top four are features from
the min synchrostates: T, degree and CPL. This behaviour is the opposite to
the other two frequency bands where the max synchrostate presented higher
FDR values than the min synchrostate network metrics, therefore a larger
discriminative ability. This in line with the findings for the previous chapter
for the α and β band, when the stimuli presented to the user were schematic
emotional faces. The mechanisms to process schematic emotional faces and
real human emotional faces are obviously not the same. However, the aim is
to recognise the emotion shown on the screen and translate it to an imagined
movement. The fact that connectivity metrics from the min synchrostate of
the γ band for human faces presents higher discrimination power similar to the
α and β band in the schematic emotional faces problem may indicate that some
of the processing pathways are shared for both paradigms, but at a different
frequency level. Maybe this is because the processing of human emotional
faces is a more complex process.
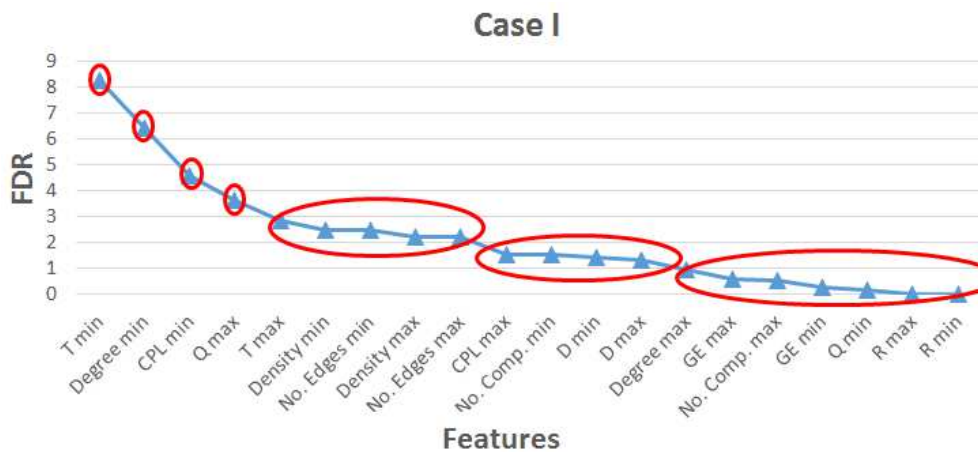
Figure 6.23: Grouped features by their FDR values ordered in decreasing order for the γ
band. and case I, when all available network metrics are considered.

Case II presents the lowest peak among all of the top FDR values from all
of the studied scenarios with a value of 3.6 versus 8.28 for the case III. As
mentioned previously, this means that the ability to discriminate between the
different tasks of the max synchrostate is lower than the min synchrostate for

this frequency band. The number of groups formed for this particular case is six: top one feature, top two features, followed by a group of 4 features, 5 , 6 and the whole set of network metrics as can be seen from the figure 6.24.
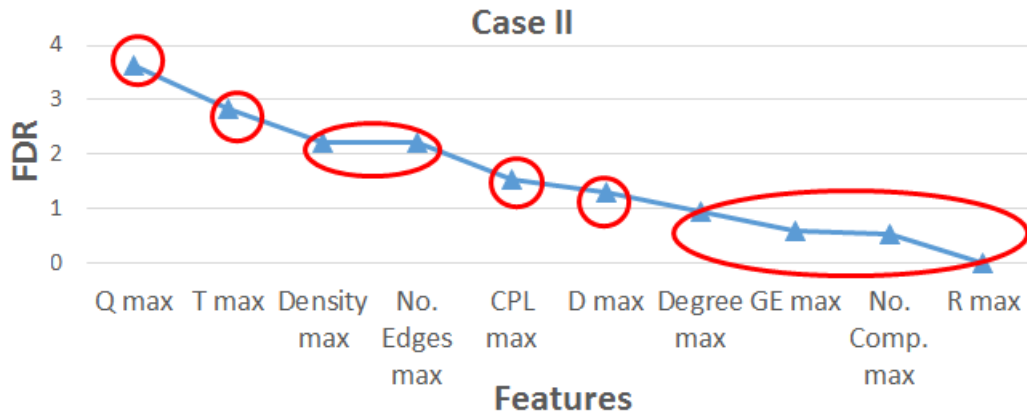
Figure 6.24: Grouped features by their FDR values ordered in decreasing order for the γ band. and case II, when only max state network metrics are considered.

Finally, case III (figure 6.25) presents FDR values of the same order as case II of the β band (see figure 6.19). The groups resulting from this selection criterion are also six, as in case II. The three first groups are one feature subsets: T min, degree min and CPL min. These three groups are the same first three groups as for case I. The next groups are integrated by 5, 7 and all features.
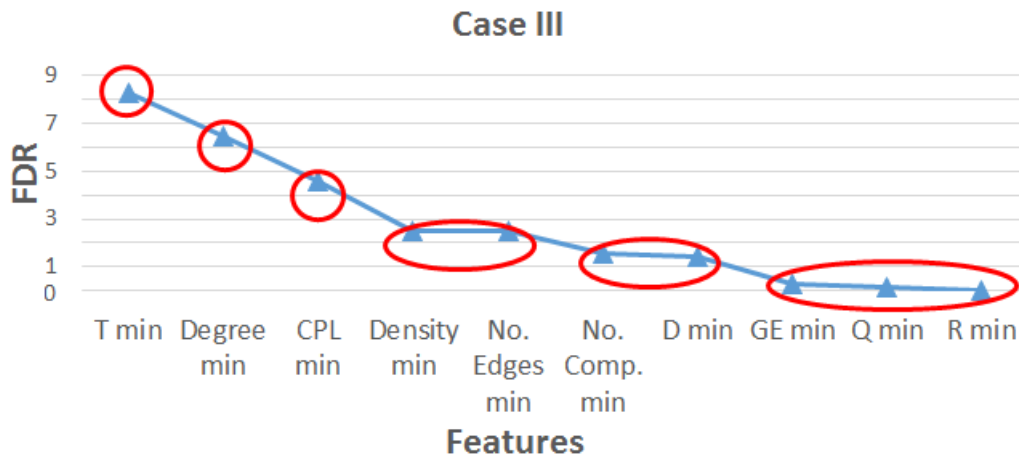
Figure 6.25: Grouped features by their FDR values ordered in decreasing order for the γ band. and case III, when only min state network metrics are considered.
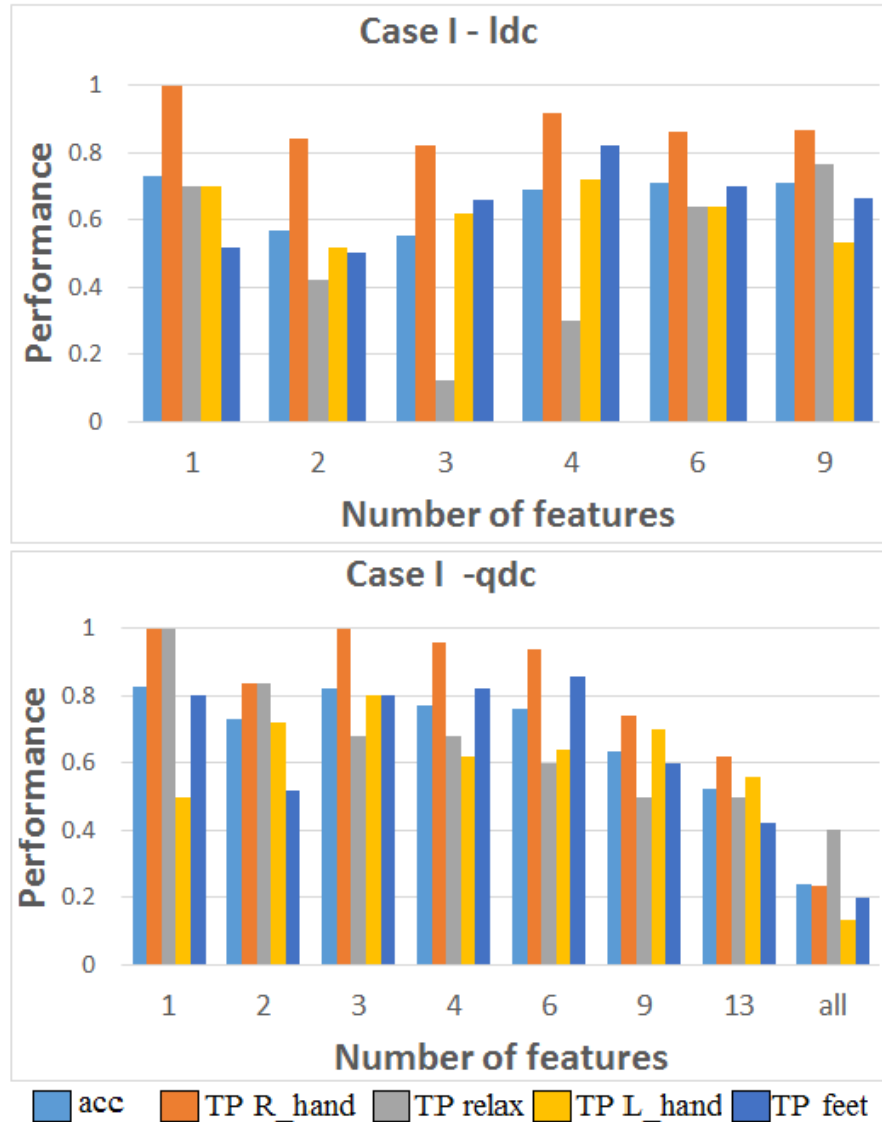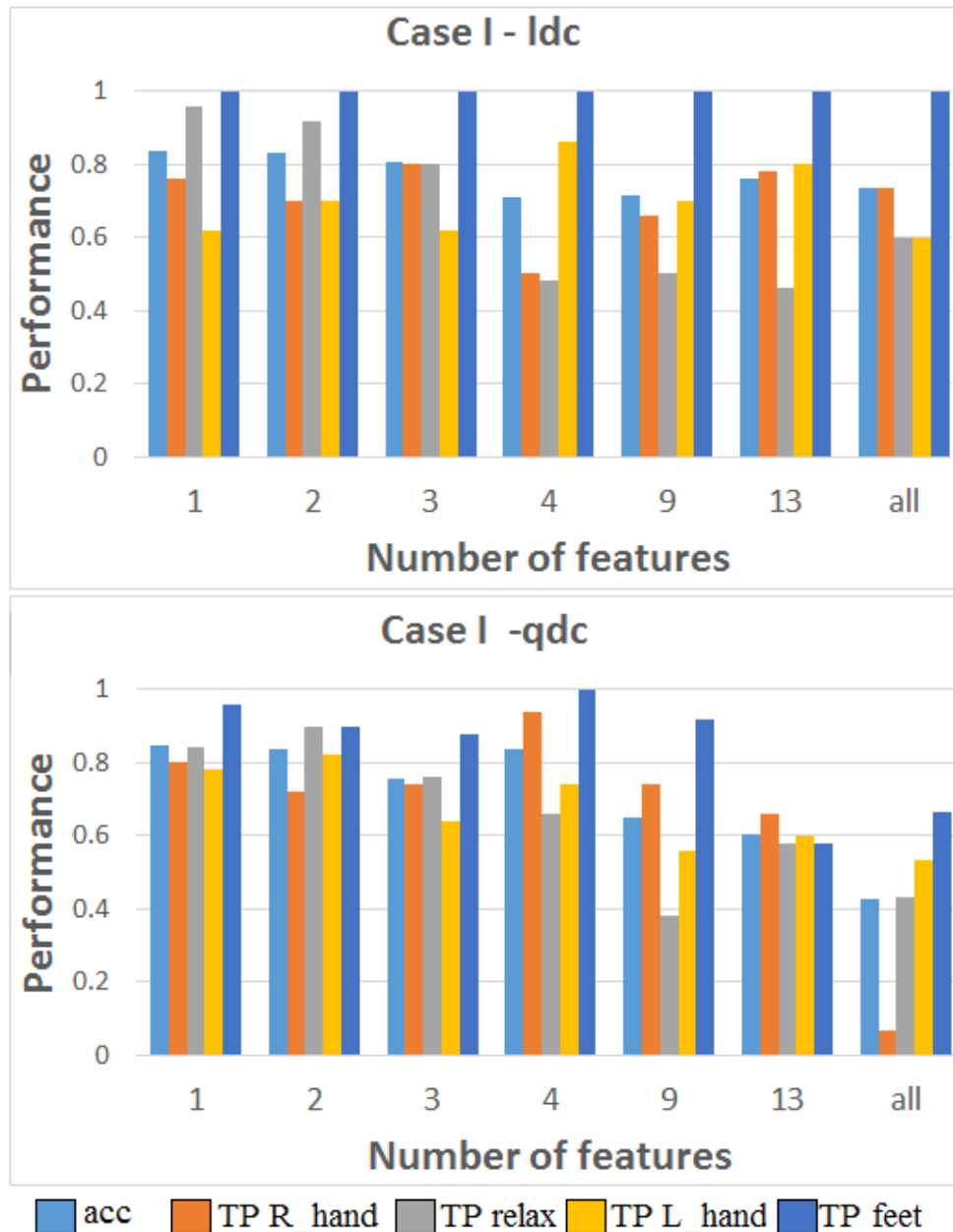
Figure 6.26: Comparison of the performance of the top two classifiers for case I in the γ band. Features grouped according to their FDR value are fed to the different classification algorithms. For each one of the formed subsets, the average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand, relaxing or no-task, L hand and feet) are illustrated. The upper graph shows the performance results for the linear discriminant analysis classifier (ldc) and the bottom graph shows the results obtained for the quadratic discriminant classifier (qdc).

Classification results for the three cases are shown in figures 6.26 and 6.27. The average performance for case I (figure 6.26) is similar to case I for the β band as the values of FDR for both cases are similar in range. The similarities are present also in the behaviour of the classification algorithms. The highest

accuracy is obtained with only one feature and the additional metrics used
to feed the classifiers only provokes a progressive decreasing of the averaged
accuracy. However, the top features are different, CPL max for case I, for the
β band and T min for case I for the γ band. The best performance rate for
the linear discriminant classifier is 83.4% with true positive rates of 76%, 96%,
62%, 100% respectively. In the case of the quadratic discriminant algorithm,
the averaged performance is slightly higher at 84.5% with true positive rates of
80%, 84%, 78% and 96%. For case II, represented in the left column of figure
6.27, the accuracy results are notably low for both classifiers with averaged
accuracies under 50%. Finally, in case III, illustrated in the right column of
figure 6.27, the greatest results are the same as for case I as the first three
groups are formed by the same features. The highest accuracies are 83.4% and
84.5% for the linear and quadratic discriminant algorithms respectively.
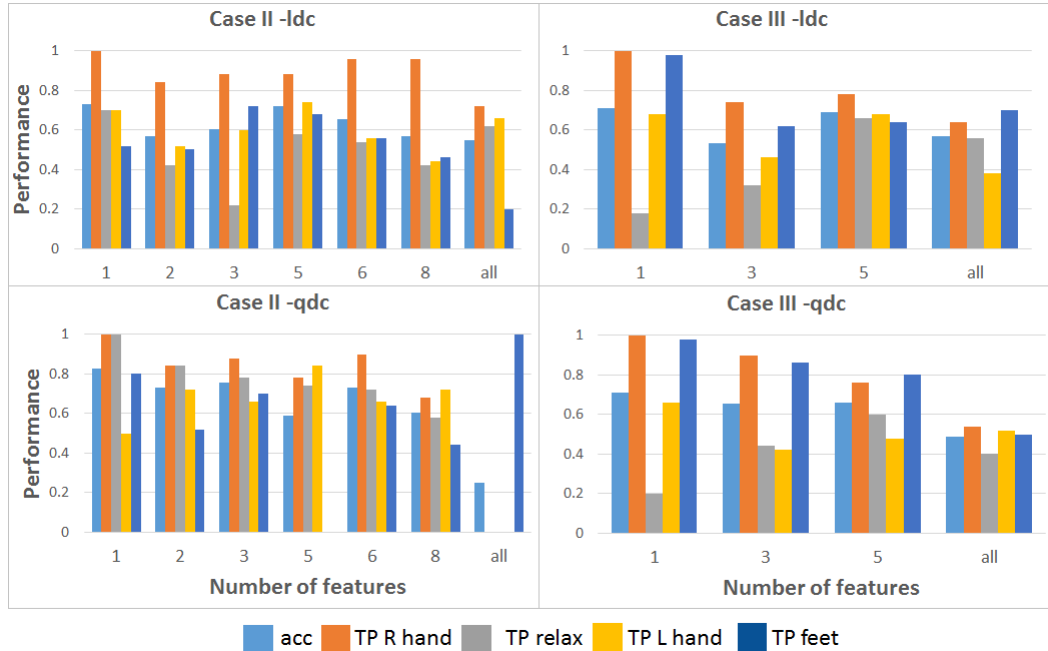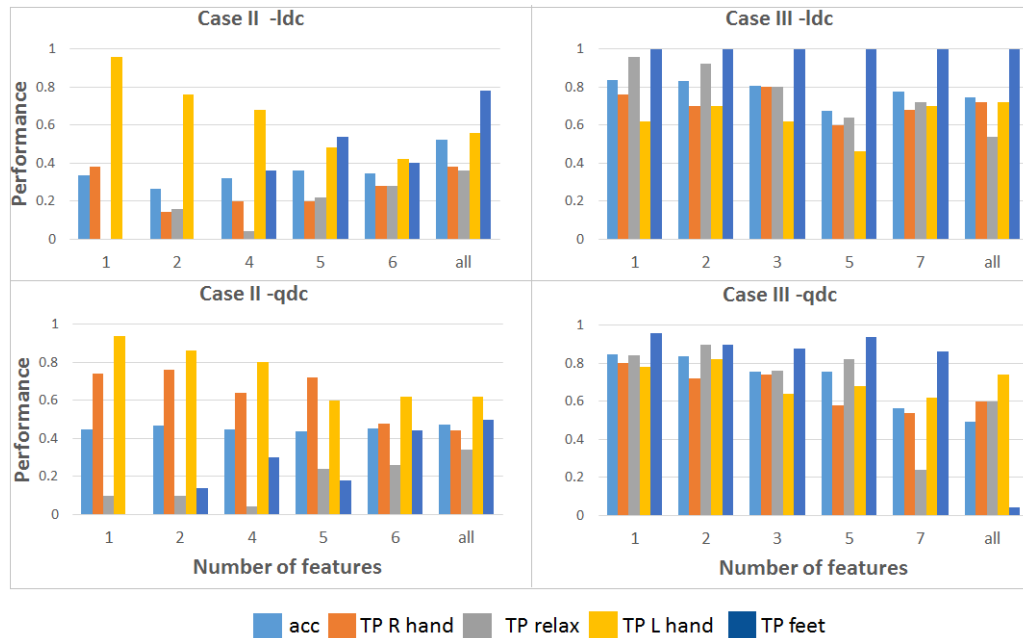

Figure 6.27: Comparison of the performance of the top two classifiers for case II (left column)
and case III (right column) in the γ band. Features grouped accordingly to their FDR value
are fed to the different classification algorithms. For each one of the formed subsets, the
average accuracy (acc) and the true positive (TP) for each one of the four tasks (R hand,
relaxing or no-task, L hand and feet) are illustrated. The upper row shows the performance
results for the linear discriminant analysis classifier (ldc) and the bottom row shows the
results obtained for the quadratic discriminant classifier (qdc).

6.7   Conclusions

The existence of synchrostates defined as quasi-stable periods of the phase
difference has been identified for MI tasks based in emotional human face

images. The study has been performed as an averaged population and as individuals belonging to a group. From the group results it can be seen that the phase topographies of the different synchrostates within the same frequency band present similarities. However, the temporal switching pattern across the different synchrostates showed the differences among stimuli within a frequency band. These differences have been quantified by means of connectivity network measures based on graph theory. For each MI task performed by the user and frequency band of interest, different network metrics have been calculated to re-affirm the findings of the previous chapters, that the information interpreted from the brain, is task-specific.

The study of each subject of the cohort of participants resulted, as in chapter 3, in a small variation of the optimal number of synchrostates due to the variability inter-trials and inter-subjects. However, the existence of a reduced number of synchrostates across the participants is consistent and compatible with previous results.

The differences observed in the connectivity network measures have been further used to classify between four different MI tasks with accuracies of 93% using only three features or less to feed the classifiers. Simpler classifiers such as linear and quadratic discriminant analysis performed similarly or even better than more complex kernels. The excellent classification accuracy achieved for the multi-class problem presented in this chapter is superior to the performance achieved in the previous chapter, when schematic emotional faces were used as stimuli. This increment in the performance can be explained by the use of human faces instead of emoticons. However, a deeper study with a larger population and using both types of stimuli should be performed in order to quantitatively measure any improvement regarding the use of human face images.

Another conclusion that can be obtained from the classification process is that for the α and β bands, the set of features coming from the maximum number of occurrences poses a considerably higher discriminative ability than the minimum number of synchrostates metrics.

In general, it can be said according to the findings presented in this chapter that:

- A reduced number of synchronised states exist during the performance of different motor imaginary tasks.

- These states have different switching patterns demonstrating that they

are task-specific.

- The combination of these states with a synchronisation index gave us a deeper understanding of the brain segregation and integration functions.

- The information from these synchrostates and their temporal stability can be measured and quantified by a set of connectivity network metrics.

- There is a set of network metrics perfectly capable of identifying and characterising the underlying mechanisms associated to a cognitive task.

The validation of this of this procedure over a larger number of participants and with a different variety of stimuli, would result in a significant move-forward in the world of neuroscience.

Chapter 7

Conclusions

BCI systems are still a challenging topic of study in neural-engineering. Despite being used in assistive devices for people with motor impairments, for instance, BCI-based wheelchairs, or in the rehabilitation process after a stroke the amount of people unable to use this assistive technology or that need a long period of training is excessively high [222]. In order to increase the reliability of BCI systems, it is necessary to deepen understanding of how the brain reacts to a specific stimulus over time. To fulfil this aim, this work has focused on the temporal resolution characteristic of EEG based BCI to determine the phase synchronisation variation over a period of time between the different areas of the brain when a specific stimulus is presented to the user. Phase synchronisation is an important key to understanding the underlying mechanisms of the brain in exchanging information across the different regions.

This work has identified the existence of a finite number of unique phase different patterns, named synchrostates that share similarities with the concept of microstates [119]. These synchrostates remain stable in the order of milliseconds, suddenly and abruptly changing to the next one during the processing of a specific stimulus. The switching pattern across the different synchrostates provides information regarding the evolution of the phase synchrony across the different regions of the brain over time. This information leads to a greater understanding of the information exchange occurring in the brain. A higher understanding of the information exchange dynamics of the brain will help the specialist to create more accurate and personalised BCIs. Despite the fact that the topographical maps of the clustered synchrostates demonstrated that those synchrostates remain topographically similar within a frequency band, their occurrence and switching pattern is clearly dependent on the stimulus

presented to the user. This means that the information processing pathways within the different regions of the brain are task-specific and each task can be divided into smaller sub-tasks represented by the synchrostates. The presence of synchrostates was identified using emotional schematic faces linked to the imaginary movement of right and left hand and also with human emotional faces images as a stimulus representing four different tasks: right hand, left hand, feet and non-task or relaxing task.

In order to quantify the differences among the task-specific synchrostates and consequently establish a possible method to accurately identify the variety of motor imaginary tasks performed by the user, they were translated into a connectivity network based on graph theory. This step added a further and deeper understanding of the function integration and segregation features of the synchrostate phenomenon, therefore a better understanding of the information processing of the human brain linked to a specific task. Connectivity network measures not only added a deeper knowledge regarding the underlying mechanism of the brain, but also provided a mechanism of classification between the different motor imaginary tasks. The network metrics obtained from the maximum and minimum number of occurrences of the states were used to feed a wide range of classification algorithms among different supervised learning techniques: discriminant analysis and support vector machines with different levels of kernel complexity and nearest neighbours with $k=3$. The performance results obtained for the two task problem, right and left hand, were as high as 83% accurate with 83% sensitivity and specificity for the α band.

Results for the four tasks problem were of 93% of averaged accuracy with sensitivity of 92%, 100%, 100% and 82% respectively for the right hand, non-task, left hand and feet respectively for the α band. The proposed method of classification of a multi-class MI based BCI system gives excellent accuracy rates and exceeds some of the contemporary proposed systems [223].

In parallel to the development of this methodology based on connectivity metrics obtained from synchrostates and with the aim of increasing the reliability of the MI-based BCI systems, a novel stimulus was introduced, faces showing different emotions. Nowadays, there is a wide variety of stimuli used in conjunction with EEG recordings and BCI systems: sounds, images, videos, navigation, letters, arrows in static position or moving across the screen, etc. However, the use of emotional faces as a stimulus for a MI based BCI has been introduced for the first time in this work. Firstly, schematic emotional faces, popularly known as emoticons used in social networks, were used as stimuli.

Participants were asked to perform the imaged movement of the right hand when a happy face appeared on the screen and a sad face were associated to the imaged movement of the left hand. After the high accuracy results obtained in the classification process, the next logical step was to demonstrate that the human face images displaying emotions can provoke the same or higher reaction to the BCI users. In summary, the use of emotional faces has been validated; Both schematic and human can be used as stimuli for MI-based BCI systems.

The number of electrodes, the sampling frequency, nature of the participants or the performance of the participants on a particular day are, among others, factors that can influence in the effectiveness of any BCI system. However, to minimise those effects, the standard protocols to record and pre-process the EEG signals for MI tasks have been followed. In addition, two different sets of experiments were undertaken with different participants and different sampling frequencies to avoid any possible bias.

Furthermore, the presence of synchrostates was identify for first time during a face processing task with autistic children (6-13 years old) using 128 channels. In this thesis, the age of the participants ranged from 20 to 53 years and the number of electrodes was reduced to 62. Further research has to be done regarding the minimum number of electrodes that can be used without losing essential information, nevertheless, it can be said that the presence of synchrostates is consistent for a wide range of subjects and number of electrodes.

Two main aims were addressed in this thesis, the validation of faces showing different emotions as stimuli for a MI based BCI and the study of phase synchronisation information to increase the performance of the current BCI systems. The classification results for both experimental set-ups indicates that both aims were successfully accomplished and always considering the limitation of the small datasets used. In addition, in order to prove and quantify the improvement of using emotional faces, a further study comparing the classification performance using emotional faces versus more traditional stimuli should be performed.

7.1   MI-based BCI systems: state of the art

One of the main aims of this thesis was to design a novel BCI system with improved performance in comparison to the state of the art MI-based BCI systems. In this section, a comparison with some of the latest classification algorithms for two and multi-class problems is performed.

A great number of algorithms have been proposed for feature extraction and classification of EEG data. In [224], the authors performed a comparison between six different feature selection algorithms (correlation-based feature selection, minimum redundancy-maximum relevance, *ReliefF*, consistency, decision tree C4.5 and a genetic algorithm approach) and five commonly used classifiers (probabilistic neural network, SVM, radial basis function network, linear discriminant analysis and *k*-nearest neighbours) for a two-class problem. The MI-tasks to classify were the right and left hand from the publicly available dataset IIIB from the BCI Competition III [225]. The highest performances were over 90%, clearly superior to the performance described in chapter 5. However, the number of features needed to obtain this performance ranged between 11 and 99. These numbers are extremely high compared with the range of 1 to 4 features used in the proposed algorithm. Another study, using the same dataset from the BCI Competition III [225], proposed a novel algorithm combining a cross-correlation technique for feature extraction and a modified SVM as the classifier [226]. The averaged performance rate was 95.72%, using a total of 235 feature vectors for each participant. Once more, an elevated number of features was needed to achieve higher accuracies.

The common spatial pattern (CSP) is a technique based on spatial filtering that has been proven as being effective in classifying multichannel EEG signals [227]. However, this technique offers limited performance when the number of training samples is reduced [228]. To address this issue in [214], an adaptation named "subband regularized CSP" was proposed. The mean performance was 82.69% for the two tasks problem when at least more than 100 training samples per class were used.

On the other hand, some other adaptations to the traditional CSP methods have been proposed to extend its use from a two-class problem to a multi-class situation. For example, in [218] a "one versus one" and "one versus the rest" CSPs are proposed in combination with a hierarchical SVM to classify between left hand, right hand, both feet and tongue movements of nine subjects. In this case the averaged classification accuracy was 64.4%. Higher accuracy was achieved in [229], with a mean classification performance of 82.39% for a three-class problem (right hand, left hand and both feet). The authors proposed a method called "a composite kernel support vector machine based stationary-CSP". In the first step, the electrodes are divided into five groups accordingly with their neurological anatomy (frontal, central, temporal, parietal and occipital). Secondly, a stationary CSP is used to extract the features for each of the five regions. Finally these features are fed to a composite kernel SVM, a

non-linear supervised method based on composite kernels. This approach only highlights the relevant brain regions for a particular task, discarding the least informative regions in each iteration of the classification algorithm. A similar averaged accuracy rate was shown in [230], 83.06%, for a four-class problem (right and left hands, tongue and both feet). Once more, the CSP approach was selected for the feature extraction step. For classification, the authors compared three strategies: a Gaussian mixture model, linear discriminant analysis and the ridge regression algorithm. The ridge regression algorithm being the technique achieving the best results.

In summary, recent classification methods for a two- MI tasks classification achieve performances ranging from just over 80% to more than 95%. The accuracy of the proposed method in this thesis for a two class problem is within this performance range. By contrast, for the multi-class case, the performance of the proposed method, based on human faces showing different emotions clearly exceeds the state of the art classification methodologies. However, a larger number of participants is needed in order to consolidate the results shown for the proposed method and determine the level of improvement of using emotional faces as stimuli for MI-based BCI systems.

7.2   Limitations and future work

This work has given more evidence of the existence of synchrostates linked to a specific MI-task and the ability to translate their information into connectivity metrics to establish a classification procedure. To extrapolate this classification mechanism to a clinical situation, a more extensive EEG data collection should be obtained with a wider variety of participants to obtain a more generalisable and definitive result. One of the limitations of this work is the collection of participants for the experiments. This methodology can be used in the future with a larger EEG dataset to classify between more MI tasks such as right and left foot or tongue among others.

Another logical step derived from the results obtained in this thesis is pursuing an online classification. In order to achieve this aim, the idea can be studied of performing the clustering of the phase synchronisation, trial by trial instead of on the grand average. Also, in line with this objective, it can be studied if there is an optimal assortment of electrodes than can help the clustering process to be more efficient. For example, re-organising them by areas, all electrodes from the frontal area together, the occipital area electrodes grouped

and so on, might improve the computational cost of the clustering algorithm. The clustering algorithm chosen in this work has been the *k-means* algorithm, which is a hard clustering method as one data point only belongs to one of the clusters. The idea of using another type of algorithm, not as rigid as those based in fuzzy logic or neural networks may also facilitate the process. Some recent investigations have demonstrated that the use of fuzzy logic improves the performance of BCI systems in terms of accuracy and system stationarity [231]. Therefore, this technique could also be used to improve the classification performance of the synchrostates-based method proposed in this thesis.
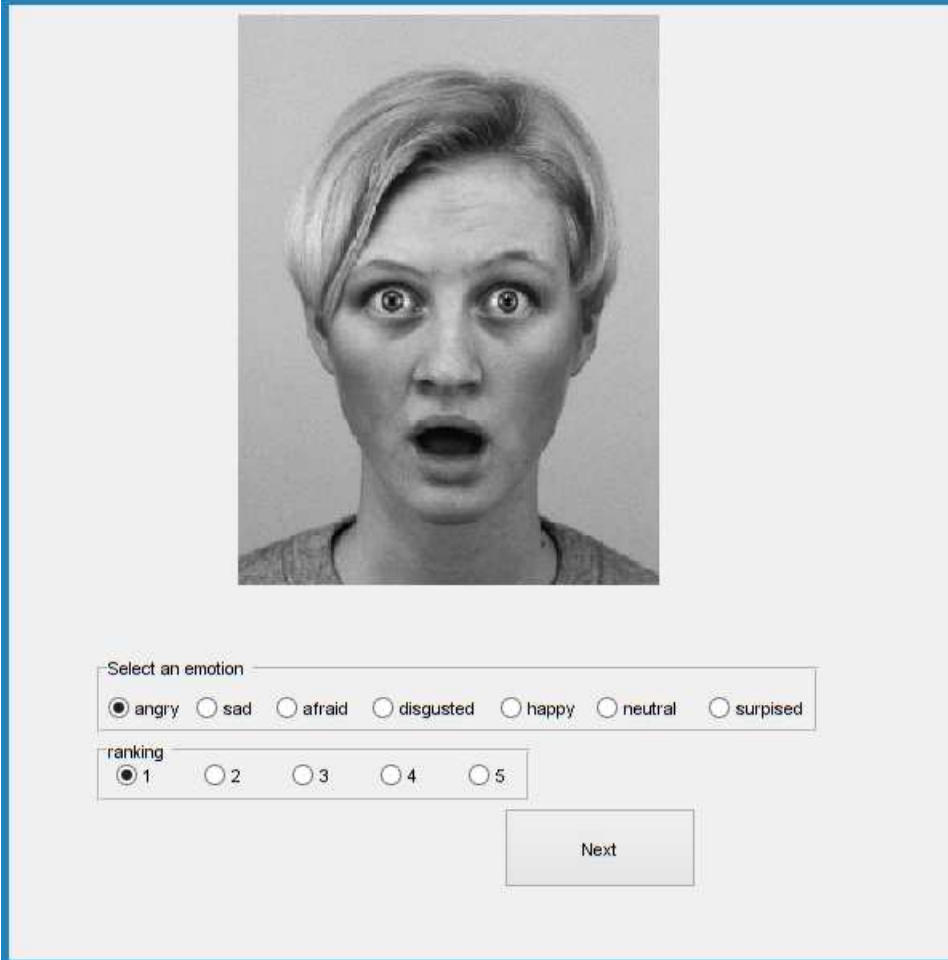
As an outcome of this work, a Matlab-based toolbox is being developed with the aim of facilitating the use of this mechanism to other researchers interested in this topic. The toolbox will allow users to estimate phase synchronisation for each one of the frequency bands, the optimal number of synchrostates and the topographic maps of the resulting synchrostates.

Appendix A

Face Database Selection

HAPPINESS   NEUTRAL   SADNESS   SURPRISE

Figure A.1: Example of a female (top row) and a male (bottom row) emotional faces of the Karolinska Directed Emotional Faces (KDEF) database [217] used for the experimental design detailed in chapter 6. The selected emotions for the experiment were happiness (left), neutral, sadness and surprise (right).

Figure A.2: Example of the interface created for the selection process of the emotions and the actors performing such emotions. Participants have to select which emotion in performing the actor/actress between seven possible: angry, sad, afraid, disgusted, happy, neutral and surprised. Once they have selected the appropriated emotion they should rank how well the emotion is transmitted from 1 (not too good, not clear) to 5 (really good, clear). The process goes through all the frontal images of the dataset, a total of 490 images. After discarding those images which conducted a high number of errors, the top 4 emotions were selected and following the same criteria 5 female and 5 male images for each emotion were selected from the same actor/actress.

Appendix B

Temporal scheme of a typical trial
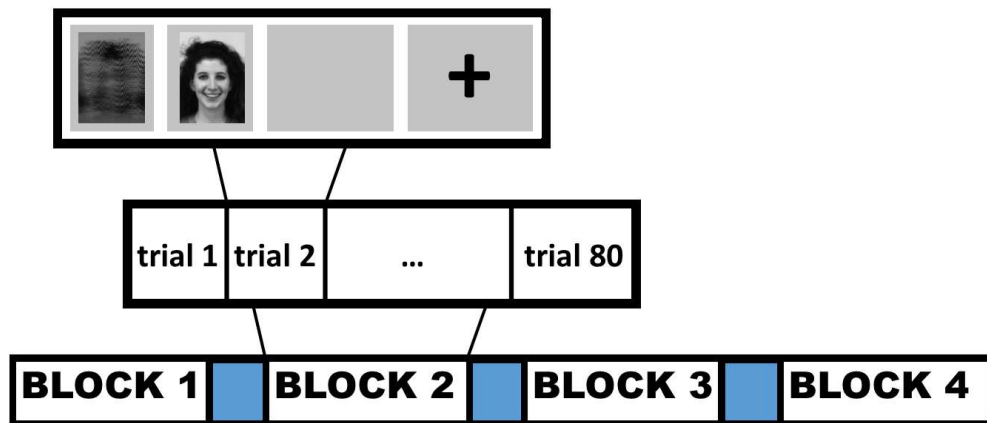
Figure B.1: Blocks scheme of the experimental design developed in chapter 6. Each participant performed a total of 4 blocks of tasks. There is a variable break period between trials according to the participant needs. For each block there are a total of 20 trials for each one of the 4 possible tasks, meaning 80 trials per block, with a total of 320 trials per participant. The details regarding time and sequence of a typical trial are shown in the figure B.2
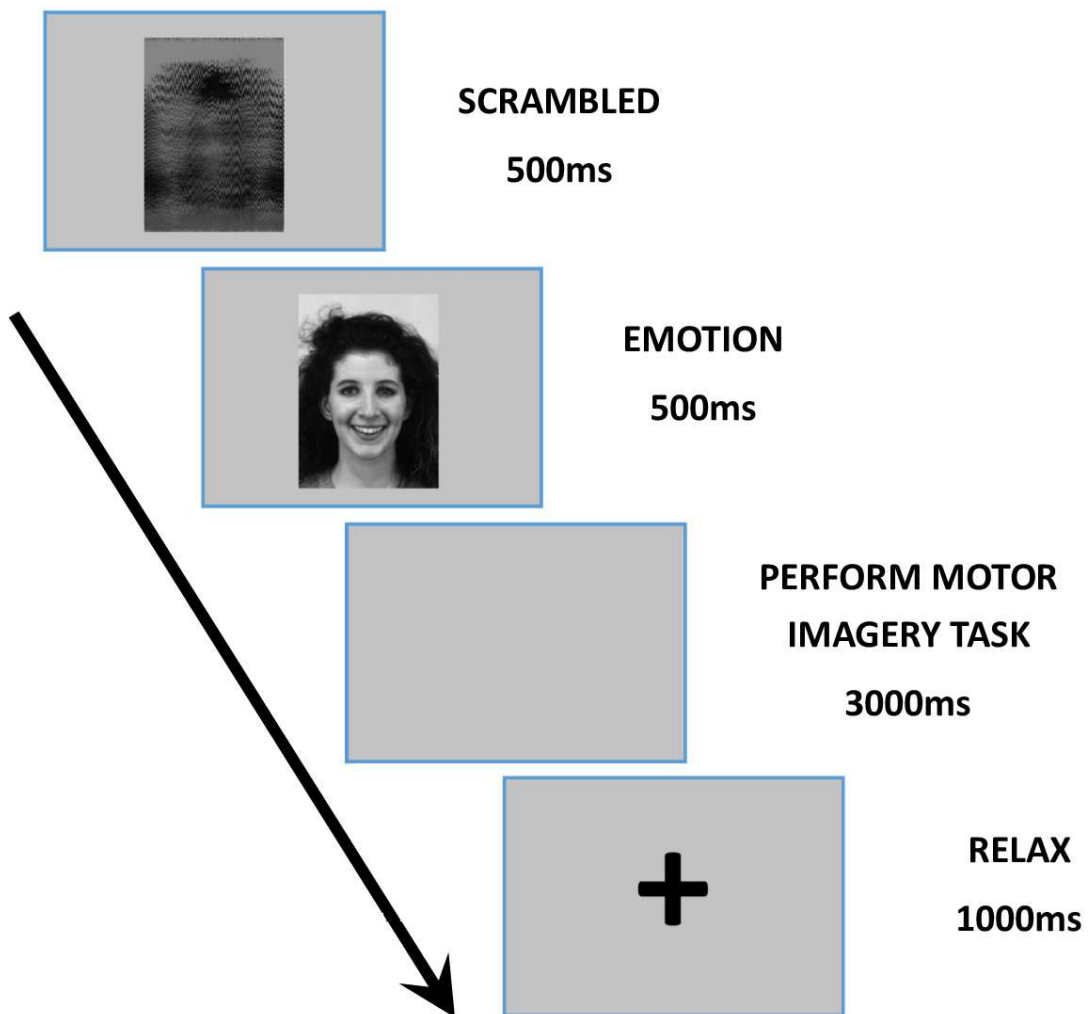
Figure B.2: Temporal scheme of a typical trial for the experiment developed in chapter 6. The starting of the trial is marked by a masked face lasting on the screen for 500ms. Afterwards, the image of an actor/actress showing one of the four emotions: happy, sad, neutral or surprised face is shown. The images of the emotions are shown in a random order with equal probability each. After 500ms a grey screen (Matlab default figure colour) indicates the period to perform the tasks associated with the emotion previously shown. This is imagined movement of the right hand for a happy faces, imagined movement of left hand for sad faces, surprised faces are linked to the imagined movement of the feet and neutral faces means no movement at all, just doing nothing. The end of the trial is illustrated by a black cross in the middle of the screen lasting for 1000ms.

Appendix C

Newton fund

This collaboration project between University of Warwick (UK) and Tsinghua University (China) was funded by the British Council via the "Newton Fund". It has an objective to study novel classification algorithms from high density iEEG. In addition, it emerged the idea of exploring the connectivity patterns between the different areas of the brain in the β band instead of the widely used high γ band, aimed at finding alternative features for motor tasks classification.

C.1 Dataset

Twenty participants with intractable epilepsy form the dataset for this research study. The electrode placement were placed accordingly to the patient's clinical needs. The performed task blocks of 20 seconds each interleaved with resting period of 8 seconds. Each task block consisted of the movement of either tongue or hand indicated by a picture on the screen. An auditory cue indicates the start of the movement that was repeated four times per block. At the end of the experiment participants performed 10 repetitions of each task block. More details of the dataset and the paradigm can be found in [232].

C.2 Time frequency analysis

Firstly, a preliminary investigation is performed to determine which channels and frequency bands can be the most sensitive to changes between the resting state and the motor task period. The power spectrum of each participant, channel and task for both periods, resting (one second before stimuli onset) and motor action (0.3 to 1.3 seconds after stimuli onset) were contrasted. The significance of the modulation of the movement period versus resting was to quantify by means of a *t-test*. Participants showed a significant difference

between both period during high γ band and also during α and β bands as can
be seen from figure C.1. This is in line with previous studies [232].
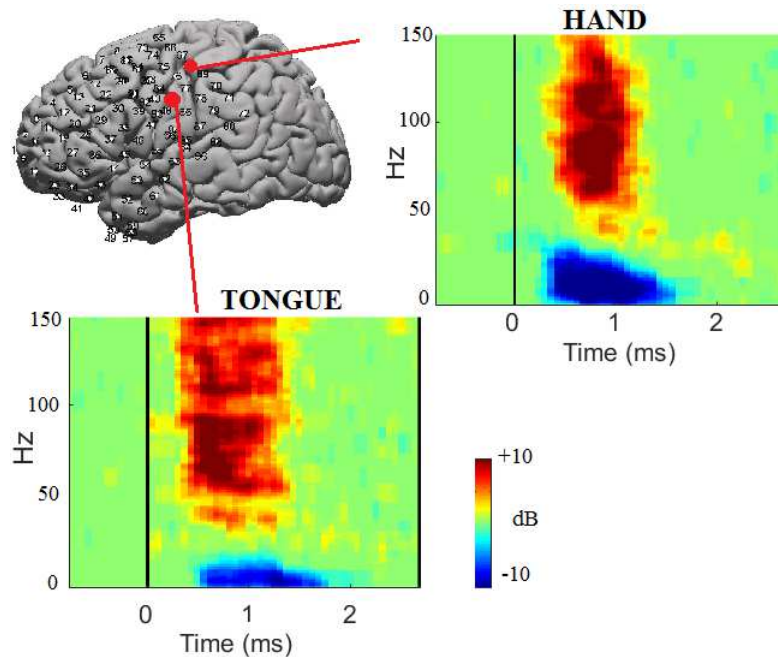
Figure C.1: Power spectrum showing significant differences between resting and motor task
states after a t-test. The upper right graph shows the power spectrum of the hand movement
from channel 68. Bottom spectrum graph is the resulted t-test of tongue movement from
channel 85 for the same participant. Electrode location for this participant is illustrated in
the top left corner. It can be seen that high γ and α/β band showed the highest differences
for both tasks.

C.3   Classification

Once the power spectrum is obtained for each participant and task, the next
step is trying to find a classification method to distinguish between two motor
tasks -hand and tongue. The selected method is based on Pearson correlation
[233], which is a measure of the linear dependence or correlation between two
variables. The classification process of a trial is as follows: the average of the
power spectrum values of each one of the tasks is averaged without including
the trial to be classified. Afterwards the trial is correlated with the averaged
model of hand and tongue of power spectrum and labelled as the task it had
the highest correlation score with. This is performed in a specific frequency
band, counting in total six different canonical bands: 4-8Hz, 8-13Hz, 13-24Hz,
24-35Hz, 60-90Hz, 110-130Hz [234]. Performance of each one of the studied fre-
quency bands is illustrated in figure C.2. It can be seen that the performance
for the 60-90Hz frequency band is, on average, higher than the others present-

ing accuracies over 90% in half of the participants. However, the performance of participant number 13 and participant number 14 remained specially low for all the frequency bands under study.
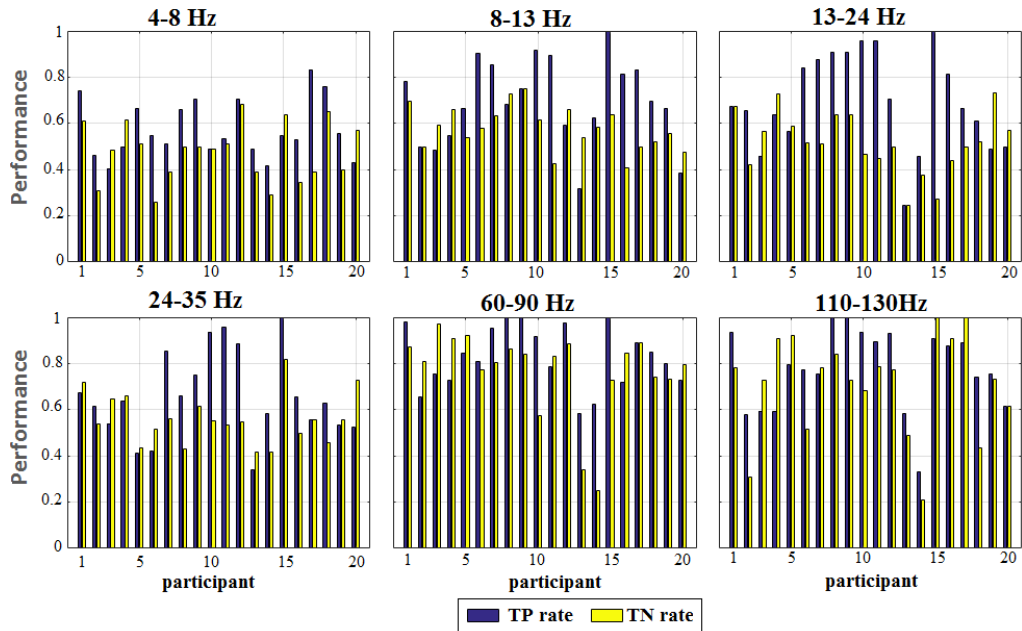
Figure C.2: Classification performance using Pearson correlation of two motor task, hand and tongue. Classification is performed in six different frequency bands, listed from top left corner to right bottom corner: 4-8Hz, 8-13Hz, 13-24Hz, 24-35Hz, 60-90Hz and 110-130Hz. Each graph shows the true positive (TP) rate in blue colour and true negative (TN) rate in yellow colour for each one of the twenty participants. It can be seen that for the two sub-bands within γ band, the performance is the highest among the different studied bands.

C.4   Future work

A logical further step after the findings of the previous section is to improve the classification algorithm to increase the accuracy rate for each one of the participants. In addition, identify the number of channels required for an optimal classification and in which area of the brain these electrodes are located, that will lead to improved algorithms, therefore their efficiency will be increased.

Additionally, in line with the scope of the main work of this thesis, the possibility of using connectivity measures as a biological markers to classify the two motor tasks will also be investigated.

Appendix D

Matlab code

```matlab
1   function DiffphaseN=Synchrostates(dataset,FileName,freq_band,fs)
2   %constants
3   Nchannels=size(dataset,1);Ntrials=size(dataset,3);Nsamples=size(dataset,2);
4   % instantenous phase CMW
5   [phase,scales]=InstantaneousPhase(dataset,freq_band,fs);
6   disp('instantaneous phase calculated');
7   % diff phase
8   DiffphaseN=InstantaneousPhaseDiff(phase,scales);
9   % save diff phase file
10  save([FileName '_DiffPhase_' freq_band '.mat'],'DiffphaseN');
11
12      function [ph2,scales]=InstantaneousPhase(data,freqband,fs)
13          fc=1.5;
14          % frequency bands
15          fbands.theta=8:-1:4;fbands.alpha=12:-1:9;
16          fbands.beta=30:-1:13;fbands.gamma=40:-1:31;
17          % chosen freq band
18          switch freqband
19              case 'theta'
20                  fband=fbands.theta;
21              case 'alpha'
22                  fband=fbands.alpha;
23              case 'beta'
24                  fband=fbands.beta;
25              case 'gamma'
26                  fband=fbands.gamma;
27              otherwise
28                  disp('not valid');return;
29          end
30          scales=zeros(1,length(fband));
31          for i=1:length(fband)
32              scales(i)=round(fc/(fband(i)*(1/fs)));
33          end
34          % phase
35          ph2=zeros(Ntrials,Nchannels,length(scales),Nsamples);
36          for trial=1:Ntrials
37              parfor ch=1:Nchannels
38                  d=data(ch,:,trial);coeff=cwt(d,scales,'cmor1-1.5');
39                  ph2(trial,ch,:,:)=angle(coeff);%angles between +-pi
40              end
41          end
42      end
43      function DiffphaseN=InstantaneousPhaseDiff(ph2,scales)
44          %% Instantaneous phase Difference
45          difff=zeros(length(scales),Nchannels,Nchannels);
46          temp=zeros(Ntrials,Nchannels,Nchannels);
47          Diffphase=zeros(Nsamples,Nchannels,Nchannels);
48          for t=1:Nsamples
49              for trial=1:Ntrials
```

```matlab
50                    parfor f=1:length(scales)
51                        ch=squeeze(ph2(trial,:,f,t)); %vector of channels
52                        for i=1:Nchannels
53                            difff(f,i,:)=abs(bsxfun(@minus, ch, ch(i)));
54                        end
55                    end
56                    temp(trial,:,:)=mean(difff,1);%average across freq
57                end
58                Diffphase(t,:,:)=mean(temp,1); % average across trials
59                per=(t*100/Nsamples);
60                disp(['percentage: ' num2str(per)]);
61            end
62            % Normalize
63            DiffphaseN=Diffphase./(2*pi);
64        end
65    end
```

```matlab
1    function [theta,bel,J,iter]=k_means_book(X,theta)
2    %function taken from
3    %Introduction to Pattern Recognition: A MATLAB based approach,
4    %S. Theodoridis, A. Pikrakis, K. Koutroumbas, D. Cavouras,
5    %Academic Press, 2010. ISBN: 9780123744869
6    [¬,N]=size(X);[l,m]=size(theta);e=1;iter=0;
7    while(e≠0)
8        iter=iter+1;
9        theta_old=theta;
10        dist_all=[];
11        for j=1:m
12            dist=sum((((ones(N,1)*theta(:,j)'-X').^2)'));
13            dist_all=[dist_all; dist];
14        end
15        [¬,bel]=min(dist_all);
16        J=sum(min(dist_all));
17
18        for j=1:m
19            if(sum(bel==j)≠0)
20                theta(:,j)=sum(X'.*((bel==j)'*ones(1,l))) / sum(bel==j);
21            end
22        end
23        e=sum(sum(abs(theta-theta_old)));
24    end
```

```matlab
1    function [results]=cost_function_kmeans(alpha2,nruns,opt)
2    [l,¬]=size(alpha2); %number of initializations
3    results=struct();
4    switch opt
5        case 1
6            for m=2:10
7                J_temp=inf;theta2=[];label2=[];
8                for t=1:nruns
9                    rand('seed',100*t);
10                    theta_ini=rand(l,m);
11                    [theta1,label1,J1,¬]=k_means_book(alpha2,theta_ini);
12                    if (J_temp>J1)
13                        J_temp=J1;theta2=theta1;label2=label1;
14                    end
15                end
16                results(m-1).J=J_temp;results(m-1).clusters=theta2;
17                results(m-1).labels=label2;
18                disp(['cluster number: ' num2str(m)]);
19            end
20        case 2
21            for m=2:10
22                J_temp=inf;theta2=[];label2=[];
23                for t=1:nruns
```

```
24                theta_ini=randn(l,m);
25                [theta1,label1,J1,¬]=k_means_book(alpha2,theta_ini);
26                if (J_temp>J1)
27                    J_temp=J1;theta2=theta1;label2=label1;
28                end
29            end
30            results(m-1).J=J_temp;results(m-1).clusters=theta2;
31            results(m-1).labels=label2;
32            disp(['cluster number: ' num2str(m)]);
33        end
34    otherwise
35        disp('error');
36        return
37 end
```

Bibliography

[1] L. Santamaria and C. James, "Use of graph metrics to classify motor imagery based bci," in *International Conference for Students on Applied Engineering (ICSAE)*, pp. 469–474, IEEE, 2016.

[2] L. Santamaria and C. James, "Classification in emotional bci using phase information from the eeg," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 371–374, IEEE, 2016.

[3] L. Santamaria and C. James, "Electroencephalogram analysis in emotional bci," in *8th IEEE EMBS UK and RI Postgraduate Conference in Biomedical Engineering and Medical Physics (PGBiomed)*, pp. 45–46, 2014.

[4] L. Santamaria and C. James, "Single phase locking value classification algorithm for motor imagery based bci," in *2nd WIN Conference*, p. 57, 2016.

[5] L. Santamaria and C. James, "Classification in emotional eeg-based bci using connectivity measures," in *WMG Doctoral Research and Innovation Conference*, p. 14, 2016.

[6] P. N. Leigh and K. R. Chaudhuri, "Motor neuron disease," *J. Neurol. Neurosurg. Psychiatry*, vol. 57, no. 8, pp. 886–96, 1994.

[7] D. Bertens, N. Frankenmolen, D. Boelen, R. P. Kessels, and L. Fasotti, "Validity of an adapted scoring method for a modified six elements test in individuals with brain injury," *Archives of Clinical Neuropsychology*, vol. 30, pp. 122–29, March 2014.

[8] C. Shears and M. Gauvaina, "Acquired brain injury results in specific impairment of planning knowledge," *Brain Impairment*, vol. 16, pp. 28–45, May 2015.

[9] S. T. Vistisen, J. Jensen, J. Fleischer, and J. F. Nielsen, "Association between the sensory-motor nervous system and the autonomic nervous system in neurorehabilitation patients with severe acquired brain injury," *Brain Inj.*, vol. 29, pp. 374–79, May 2015.

[10] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, pp. 767–91, June 2002.

[11] M. A. N. Mikhail A. Lebedev, "Brain machine interfaces: past, present and future," *Trends Neurosci.*, vol. 29, pp. 536–46, September 2006.

[12] J. J. Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.*, vol. 2, pp. 157–80, 1973.

[13] T. W. Berger, J. Chapin, G. A.Gerhardt, D. McFarland, J. Principe, W. Soussou, D. M. Taylor, and P. A. Tresco, *Brain Computer Interfaces: An International Assessment of Research and Development Trends.* Springer Netherlands, 2008. ISBN 978-1-4020-8704-2.

[14] C. Brunner, N. Birbaumer, B. Blankertz, C. Guger, A. Kubler, D. Mattia, jose del R. Millan, F. Miralles, A. Nijholt, E. Opisso, N. Ramsey, P. Salomon, and G. R. Muller-Putz, "Bnci horizon 2020: towards a roadmap for the bci community, brain-computer interfaces," *Brain-Computer Interfaces*, vol. 2, no. 1, pp. 1–10, 2015.

[15] J. R. Wolpaw, "Brain computer interfaces as new brain output pathways," *The Journal of Physiology*, vol. 579, pp. 613–19, March 2007.

[16] D. S. Tan and A. Nijholt, eds., *Brain Computer Interfaces: Applying our Minds to Human-Computer Interaction.* Human-Computer Interaction Series, Springer-Verlag London, 1 ed., 2010. ISBN: 9781849962728.

[17] C. Vidaurre and B. Blankertz, "Towards a cure for bci illiteracy," *Brain Topogr.*, vol. 23, pp. 194–98, June 2010.

[18] C. Vidaurre, C. Sannelli, and K.-R. M. B. Blankertz, "Machine learning based co-adaptive calibration: A perspective to fight bci illiteracy," in *Hybrid Artificial Intelligency Systems (HAIS). Lecture Notes in Artificial Intelligence.* (M. Romay, E. Corchado, and M. Garcia-Sebastian, eds.), vol. 6076, (Berlin), SPRINGER-VERLAG BERLIN, June 2010. 5th International Conference on Hybrid Artificial Intelligence Systems.

[19] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery," *PLoS One*, vol. 8, p. no. e80886, Nov 2013.

[20] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "Bci demographics: how many (and what kinds of) people can use an ssvep bci?," vol. 18, pp. 107–16, Apr 2010.

[21] A.-Y. Mutlu, E. Bernat, and S. Aviyente, "A signal-processing-based approach to time-varying graph analysis for dynamic brain network identification," *Comput. Math. Methods Med.*, vol. 2012, p. no. 451516, 2012.

[22] M. D. Schirmer, *Developing Brain Connectivity: Effects of Parcellation Scale on Network Analysis in Neonates*. PhD thesis, King's College London, 2015.

[23] W. Jamal, *Investigating Phase Synchronisation in EEG signals for Brain Connectivity Analysis*. PhD thesis, 2014.

[24] P. Hagman, O. Sporns, N. Madan, L. Cammoun, R. Pienaar, V. Wedeen, R. Meuli, and J. Thiran, "White matter maturation reshapes structural connectivity in the late developing human brain," *Proccedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 107, no. 44, pp. 19067–19072, 2010.

[25] U. Braun, S. Muldoon, and D. S. Bassett, "On human brain networks in health and disease," *eLS (John Wiley and Sons)*, 2015.

[26] O. Tymofiyeva, C. P. HeB, D. Xu, and A. J. Barkovich, "Structural mri connectome in development: Challenges of the changing brain," *Br. J. Radiol.*, vol. 87, p. no.20140086, July 2014.

[27] E. Baçar, M. Özgören, A. Öniz, C. Schmiedt, and C. Bacsar-Eroglu, "Brain oscillations differentiate the picture of one's own grandmother," *Int. J. Psychophysiol.*, vol. 64, no. 1, pp. 81–90, 2007.

[28] T. Baumgarther, M. Esslen, and L. Jancke, "From emotion perception to emotion experience: emotions evoked by pictures and classical music," *Int. J. Psychophysiol.*, vol. 60, no. 1, pp. 34–43, 2006.

[29] R. Yuvaraj, M. Murugappan, U. Rajendra-Acharya, H. Adeli, N. Ibrahim, and E. Mesquita, "Brain functional connectivity patterns for ememotion state classification in parkinson's disease patients without dementia," *Behav. Brain Res.*, vol. 298, pp. 248–60, 2016.

[30] W. Jamal, S. Das, K. Maharatna, D. Kuyucu, F. Sicca, L. Billeci, F. Apicella, and F. Muratori, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates," *J. Neural Eng.*, vol. 11, p. no. 046019, 2014.

[31] J. E. Hall and A. C. Guyton, *Guyton and Hall textbook of medical physiology.* Saunders/ Elsevier, 13th ed., 2015.

[32] J. Hellier, ed., *The Brain, the Nervous System, and Their Diseases.* ABC CLIO, 2015.

[33] J. Arminy and P. Vuilleumier, eds., *The Cambridge Handbook of Human Affective Neuroscience.* Cambridge University Press, 2013.

[34] D. Purves, *Body and Brain: A Throphic Theory of Neural Connections.* Harvard University Press Cambridge Mass, 1990.

[35] C. Stangor, *Introduction to Psychology.* First World Knowledge, 2011. ISBN: 978-1936126484.

[36] J. W. Hell and M. D. Ehlers, eds., *Structural and Functional Organization of the Synapse.* Springer, 2008.

[37] J. S. Barlow, *The Electroencephalogram: Its Patterns and Origins.* A Bradford Book, 1993.

[38] E. Neidermeyer and F. L. D. Silva, eds., *Electroencephalography: Basic Principles, Clinical Applications and Related Fields.* Lippincott Williams and Wilkins, 1998.

[39] R. Hari and R. Salmelin, "Magnetoencephalography: From squids to neuroscience: Neuroimege 20th anniversary spetial edition," *Neuroimage*, vol. 61, pp. 386–96, june 2012.

[40] B.-J. Casey, M. Davison, and B. Rosen, "Functional magnetic resonance imaging: Basic principles of and application to developmental science," *Developmental Science*, vol. 5, pp. 301–09, August 2002.

[41] E. Amaro and G. J. Barker, "Study design in mri: Basic principles," *Brain Cogn.*, vol. 60, pp. 220–32, January 2006.

[42] P. Jezzard, P. M. Mathews, and S. M. Smith, *Functional Magnetic Response Imaging: An Introduction to Methods.* Oxford Scholarship Online, 2012.

[43] G. S. Freedman, "Radionuclide tomography," *Semin. Nucl. Med.*, vol. 3, pp. 267–83, July 1973.

[44] A. M. Alessio, E. Butterworth, J. H. Cadwell, and J. B. Bassingthwaighte, "Quantitative imaging of coronary blood flow," *Nano Reviews and Experiments*, vol. 1, no. 2010, p. no. 5110, 2010.

[45] C. Eugene and A. Abass, *The Physics of PET/CT: a clinical guide.* Thieme Medical Publishers Inc., 2nd ed., April 2009.

[46] S. I. Ziegler, "Positron emission tomography: Principles, technology, and recent developments," *Nucl. Phys. A*, vol. 752, pp. 679C–87C, April 2005.

[47] L. Livieratos, *Basic Principles of SPECT and PET Imaging (Radionuclide and Hybrid Bone Imaging).* Springer Berlin Heidelberg, 2012.

[48] M. W. Groch and W. D. Erwin, "Spect in the year 2000: Basic principles," *J. Nucl. Med. Technol.*, vol. 28, no. 4, pp. 233–44, 2000.

[49] H. Shibasaki, "Human brain mapping: Hemodynamic response ans electrophysiology," *Clin. Neurophysiol.*, vol. 119, no. 4, pp. 731–43, 2008.

[50] I. Fogelman, G. Gnanasegaran, and H. van der Wall, eds., *Radionuclide and Hybrid Bone Imaging.* Springer-Verlag Berlin, 1 ed., 2012.

[51] K. Uludag, D. J. Dubowitz, and R. B. Buxton, *Clinical MRI*, ch. Basic principles of functional MRI, pp. 249–87. San Diego: Elsevier, 2005.

[52] S. Clare, *Functional MRI: Methods and Applications.* PhD thesis, 1997.

[53] M. A. Lindquist, "The statistical analysis of fmri data," *Statistical Science*, vol. 23, pp. 439–464, 2008.

[54] R. W. Hill, G. A. Wyse, and M. Anderson, *Animal physiology.* Sinauer Associates, 3 ed., 2012.

[55] G. H. Glover, "Overview of functinal magnetic resonance imaging," *Neurosurg. Clin. N. Am.*, vol. 22, no. 2, pp. 133–139, 2012.

[56] C. M. C. (CMC), "Basics of meg (magnetoencephalography)."

[57] S. P. Singh, "Magnetoencephalography: Basic principles," *Ann Indian Acad Neurol*, vol. 17, pp. Supp S1: 107–12, March 2014.

[58] 4-DNeuroimaging, "Basic principles of magnetoencephalography," http://web.mit.edu/kitmitmeg/whatis.html.

[59] A.-A. of Neurology, *Model Coverage Policy: Magnetoencephalography (MEG) Model Policy*, May 2009.

[60] Codman and Shurtleff, *CODMAN ICP MONITORING SYSTEM: quick set-up guide*. Codman (Jonhson and Jonshon), i-99-000 ed., 2001.

[61] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker," *Nature*, vol. 485, pp. 233–36, May 2012.

[62] D. Keene, S. Whiting, and E. Ventureya, "Electrocorticography," *Epileptic. Disord.*, vol. 2, pp. 57–64, March 2000.

[63] E. C. Leuthardt, K. J.Miller, G. Schalk, R. P. Rao, and J. G. Ojemann, "Electrocorticography-based brain computer interface -the seattle experience," in *IEEE Transactions on neural systems and rehabilitation engineering* (IEEE, ed.), vol. 14, pp. 194–98, June 2006.

[64] R. Carton, "The electric currents of the brain," *Br. Med. J.*, p. 278, August 1875.

[65] H. Berger, "Uber das elektrenkephalogramm des menschen (on the human electroencephalogram)," *Archiv f. Psychiatrie u. Nervenkrankheiten*, vol. 87, pp. 5427–70, 1929.

[66] W. O. Tatum, S. R. Benbadis, D. R. Nordli, P. W. Kaplan, F. W. Drislane, N. Gaspard, and L. J. Hirsch, *Handbook of EEG interpretation*. Demos Medical Publishing LLC., second ed., 2014. ISBN: 9781620700167.

[67] g.Tec: medical engineering, "www.gtec.at."

[68] A. Tyagi, S. Semwal, and G. Shah, "A review of eeg sensors used for data acquisition," in *National Conference on Future Aspects of Artificial intelligence in Industrial Automation (NCFAAILA)*, International Journal of Computer Application, 2012.

[69] Y. M. Chi, T.-P. Jung, and G. Cauwenberghs, "Dry-contact and non-contact biopotental electrodes methodological review," *IEEE-J-RBME*, vol. 3, pp. 106–19, September 2010.

[70] J. Malivuo and R. Plonsey, eds., *Biolectromagnetism*. Oxford University Press, 1995.

[71] Y. M. Chi and G. Cauwenberghs, "Wireless non-contact eeg/ecg electrodes for body sensors networks," in *2010 International Conference on Body Sensors Networks (BSN)*, pp. 297–301, IEEE, 2010.

[72] G. Buzsáki, *Rhythms of the brain.* Oxford University Press, 2006.

[73] M. Rohalova, P. Sykacek, M. Koska, and G. Dorffner, "Detection of the eeg artifacts by means of the (extended) kalman filter," *Measurement Science Review*, vol. 1, no. 1, pp. 59–62, 2001.

[74] T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, "Signal quality of simultaneously recorded invasive and non-invasive eeg," *NeuroImage*, 2009.

[75] O. Vaisanen and J. Malmivuo, "Improving th esnr of eeg generated by deep sources wih weighted multielectrode leads," *J. Physiol. (Paris)*, vol. 103, pp. 306–314, 2009.

[76] G. Repovs, "Dealing with noise in eeg recording and data anal," *Infor. Med. Slov.*, vol. 15, no. 1, pp. 18–25, 2010.

[77] S. J. Luck and E. S. Kappenman, *The Oxford Handbook of Event-Related Potential Components.* Oxford Library of Psychology, December 2011. ISBN: 9780195374148.

[78] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors (Basel)*, vol. 12, pp. 1211–79, January 2012.

[79] C. S. Herrann, S. Rach, J. Vosskuhl, and D. Struber, "Time-frequency analysis of event-related potentials: a brief tutorial," *Brain Topogr.*, vol. 27, pp. 438–50, November 2013.

[80] L. Farwell and E. Donchin, "Talking off the top of your head: Towards a mental prosthesis utilizing event-brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–23, 1988.

[81] . East Tennessee State University, http://www.etsu.edu/cas/psychology/bcilab/.

[82] L. Trejo, R. Rosipal, and B. Mathews, "Brain-computer interfaces for 1-d and 2-d cursor control: designs using volitional control of the eeg spectrum or steady-state visual evoked potentials," vol. 14, pp. 225–29, June 2006.

[83] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, "Brain computer interfaces based on the steady-state visual-evoked response," *IEEE Trans. Rehabil. Eng.*, vol. 8, pp. 211–14, June 2000.

[84] A. Ikeda, K. Terada, N. Mikuni, R. C. Burgess, Y. Comair, W. Taki, T. Hamano, J. Kimura, H. O.Lüders, and H. Shibasaki, "Subdural recording of ictal dc shifts in neocortical seizures in humans," *Epilepsia*, vol. 37, pp. 662–74, July 1996.

[85] U. Strehl, U. Leins, G. Goth, C. Klinger, T. Hinterberger, and N. Birbaumer, "Self-regulation of slow cortical potentials: a new treatment for children with attention deficit/ hyperactivity disorder," *Pediatrics*, vol. 118, pp. e1531–40, November 2006.

[86] A. Kübler, N. Neumann, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. Birbaumer, "Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication," *Arch. Phys. Med. Rehabil.*, vol. 82, pp. 1533–39, November 2001.

[87] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An eeg-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, pp. 252–59, March 1991.

[88] S. Lemm, K.-R. Muller, and G. Curio, "A generalized framework for quantifying the dynamics of eeg event-related desynchronization," *PLoS Comput. Biol.*, vol. 5, August 2009.

[89] E. V. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain-computer interfaces.," *Int. J. Psychophysiol.*, vol. 84, pp. 86–94, 2012.

[90] J. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Brain-actuated interaction," *Artif. Intell.*, vol. 159, no. 1-2, pp. 241–59, 2004.

[91] F. Galan, M. Nuttin, E. Lew, P. Ferrez, G. Vanacker, J. Philips, and J. Millan, "A brain-actuated wheelchair: ashynchronous and non-invasive brain-computer interfaces for continous control of robots," *Clin. Neurophysiol.*, vol. 119, pp. 2159–69, 2008.

[92] E. V. Friedrich, R. Scherer, and C. Neuper, "Stability of event-related (de-) synchronization during brain-computer interface-relevant tasks," *Clin. Neurophysiol.*, vol. 124, pp. 61–69, JAN 2013.

[93] Y. Zhang, Q. Zhao, J. Jing, X. Wang, and A. Cichocki, "A novel bci based on erp components sensitive to configural processing of human faces.," *J. Neural Eng.*, vol. 9, no. 2, p. no.026018, 2012.

[94] L. Chen, J. Jin, I. Daly, Y. Zhang, X. Wang, and A. Cichocki, "Exploring combinations of different color and facial expression stimuli for gaze-independent bcis," *Front. Comput. Neurosci.*, vol. 10, no. JAN, p. no. 5, 2016.

[95] P. Ekman, "Strong evidence for universals in facial expressions: a reply to russell's mistaken critique.," *Psychol. Bull.*, vol. 115, no. 2, pp. 268–87, 1994.

[96] J. Fell and N. Axmacher, "The role of phase synchronization in memory processes," *Nat. Rev. Neurosci.*, vol. 12, pp. 105–18, February 2011.

[97] J. Adey, D. Walter, and C. Hendrix, "Computer techniques in correlation and spectral analysis of cerebral slow wave during discriminative behaviour.," *Exp. Neurol.*, vol. 3, pp. 501–24, 1961.

[98] J. Barlow and M. Brazier, "A note on a correlator for electroencephalographic work," *Electroencephalogr. Clin. Neurophysiol.*, vol. 6, no. 2, pp. 321–25, 1954.

[99] A. Gevins, B. Cutillo, S. Bressler, N. Morgan, R. White, J. Illes, and D. Greer, "Event-related covariances during a bimanual visuomotor tasks. ii. preparation and feedback.," *Electroencephalogr. Clin. Neurophysiol.*, vol. 74, no. 2, pp. 147–60, 1989.

[100] M.-N. Livanov, *Spatial Organization of Cerebral Processes*. I.P.S.T., October 1977.

[101] C. Babiloni, F. Babiloni, F. Carducci, F. Cincotti, C. Vecchio, S. Rossi, C. Miniussi, and P. Rossini, "Functional frontoparietal connectivity during short-term memory as revealed by high-resolution eeg coherence analysis," *Behav. Neurosci.*, vol. 118, pp. 687–97, August 2004.

[102] M. Bota, H. Dong, and L. Swanson, "Combining collation and annotation effort toward completion of the rat and mouse connectomes in bams," *Front. Neuroinf.*, vol. 6, no. 2, 2012.

[103] B. Horwitz, "The elusive concept of brain connectivity," *Neuroimage*, vol. 19, pp. 466–70, February 2003.

[104] P. Roelfsema, A. K. Engel, P. König, and W. Singer, "Visuomotor integration is associated with zero time-lag synchronization among cortical areas," *Nature*, vol. 385, pp. 157–61, January 1997.

[105] F. Varela, J. Lachacux, E. Rodriguez, and J. Martinerie, "The brainweb: Phase syncrhonization and large-scale integration," *Nat. Rev. Neurosci.*, vol. 2, pp. 229–39, April 2001.

[106] K. Friston, "Functional and effective connectivity: A review," *Brain Connect.*, vol. 1, no. 1, pp. 13–36, 2011.

[107] K. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Mapp.*, vol. 2, pp. 56–78, 1994.

[108] A. Aersten and H. Preissl, "Dynamics of activity and connectivity in physiological neuronal networks," *63rd W.E. Haraeus Seminar on Nonlinear Dynamics and Neural Networks.*, vol. 2, pp. 281–301, 1990.

[109] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg," *Comput. Biol. Med.*, vol. 41, pp. 1110–7, July 2011.

[110] C. Granger, "Investigating causal relations by econometric models and corss-spectral methods," *Econometrica*, vol. 37, pp. 424–38, August 1969.

[111] A. Seth, A. Barrett, and L. Barnet, "Granger causality analysis in neuroscience and neuroimaging," *The Journal of Neuroscience*, vol. 35, pp. 3293–97, February 2015.

[112] K. Sameshima and L. A. Baccala, eds., *Methods in brain connectivity inference through multivariate time series analysis.* CRC Press, 2014.

[113] B. L. P. Cheung, B. Riedner, G. Tononi, and B. D. V. Veen, "Estimation of cortical connectivity from eeg using state-space models," *IEEE Tran*, 2010.

[114] K. Friston, "Causal modemodel and brain connectivity in functional magnetic resonance imaging," *PLoS Biology*, vol. 7, no. 2, p. e10033, 2009.

[115] A. Omidvarnia, M. Mesbah, J. O'Toole, P. Coldiltz, and B. Boashash, "Analysis of the time-varying cortical neural connectivity in the newborn eeg: Atime-frequency approach," in *International Worshop on Systems, Signal Processing and their Applications, WoSSPA*, pp. 179–82, IEEE, IEEE, 2011.

[116] M. Gärtner, V. Brodbeck, H. Laufs, and G. Schneider, "A stochastic model for eeg microstate sequence analysis," *Neuroimage*, vol. 104, pp. 199–208, 2015.

[117] D. Lehmann, H. Ozaki, and I. Pal, "Eeg alpha map series: brain micro-states by space-oriented adaptative segmentation.," *Electroencephalogr Clin Neurophysiol.*, vol. 67, no. 3, pp. 271–88, 1987.

[118] A. Khanna, A. Pascual-Leone, C. Michel, and F. Farzan, "Microstates in resting-state eeg: Current status and future directions," *Neurosci. Biobehav. Rev.*, vol. 49, pp. 105–13, 2015.

[119] D. Lehmann, "Deviant microstates ('atoms of thought') in brain electric field sequences of acute schizophrenics," *European Psychiatry*, vol. 13, no. Suplement 4, pp. 197s–98s, 1998.

[120] S. Cacioppo, S. Balogh, and J. Cacioppo, "Implicit attention to negative social, in contrast to nonsocial, words in the stroop task differs between individuals high and low in loneliness: Evidence from event-related brain microsates," *Cortex*, vol. 70, pp. 213–33, 2015.

[121] C. Andreou, P. Faber, G. Leicht, D. Schoettle, N. Polomac, I. Hanganu-Opatz, D. Lehmann, and C. Mulert, "Resting-state connectivity in the prodromal phase of schizophrenia: Insights from eeg microstates," *Schizophr. Res.*, vol. 152, no. 2-3, pp. 513–20, 2014.

[122] K. Nishida, Y. Morishima, M. Yoshimura, T. Isotani, S. Irisawa, K. Jann, T. Dierks, W. Strik, T. Kinoshita, and T. Koenig, "Eeg microsates associ-ated with salience and frontoparietal networks in frontotemporal demen-tia, schizophrenia and alzheimer's disease," *Clin. Neurophysiol.*, vol. 124, no. 6, pp. 1106–14, 2013.

[123] J. Britz, L. Hernandez, T. Ro, and C. Michel, "Eeg-microstate depend-ent emergence of perceptual awareness," *Front. Behav. Neurosci.*, vol. 8, no. May, p. 163, 2014.

[124] M. Angelini, M. Calbi, A. Ferrari, B. Sbriscia-Fioretti, M. Franca, and V. Gallese, "Motor inhibition during overt and covert actions: An elec-trical neuroimaging study," *PLoS One*, vol. 10, no. 5, 2015.

[125] J. E. P. J. Minguillon, M. Coscia, R. Leeb, J. Millán, D. V. D. Ville, and S. Micera, "Modular organization of reaching and grasping movements

investigated using eeg microstates," in *36th IEEE EMBC*, IEEE EMBS, IEEE, August 2014.

[126] A. Khanna, A. Pascual-Leone, and F. Farzan, "Reliability of resting-state microstate features in electroencephalography," *PLoS One*, vol. 9, p. no.e114163, December 2014.

[127] R. Srinivasan, W. Winter, J. Ding, and P. Nunez, "Eeg and meg coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics," *J. Neurosci. Methods*, vol. 166, pp. 41–52, October 2007.

[128] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*. Oxford University Press, 1995. ISBN-10:0195057287.

[129] M.-C. Ho, T.-C. Chen, C.-F. Huang, Y. Cheng-Hsieh, C. Jhih-Ming, H. Ray-Ying, H. Hsing-Chung, and L. Chia-Ju, "Detect ad patients by using eeg coherence analysis," *Journal of Medical Engineering*, vol. 2014, p. no. 236734, 2014.

[130] N. Polomac, G. Leicht, G. Nolte, C. Andreou, T. Scheider, S. Steinmann, K. Engel, and C. Mulert, "Generators and connectivity of the early auditory evoked gamma band response," *Brain Topogr.*, vol. 28, no. 6, pp. 865–78, 2015.

[131] J.-S. George, J. Strunk, R. Mak-McCully, M. Houser, H. Poizner, and A.-R. Aron, "Dopaminergic therapy in parkinson's disease decreases cortical beta band coherence in the resting state and increases cortical beta band power during executive control," *Neuroimage: Clinical*, vol. 3, pp. 261–70, 2013.

[132] M. Ten-Caat, M. Lorist, E. Bezdan, J. Roerdink, and N. Maurits, "High-density eegcoherenceanalysisusing functionalunits applied to mental fatigue," *J. Neurosci. Methods*, vol. 171, no. 2, pp. 271–78, 2008.

[133] Z. Weidong and L. Yingyuan, "Eeg real-time feedback based on stft and coherence analysis," in *Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (P. of the 23rd Annual International Conference of the IEEE, ed.), vol. 2, pp. 1869–1871 vol.2, IEEE, 2001.

[134] W. Jamal, S. Das, K. Maharatna, I. Pan, and D. Kuyucu, "Brain connectivity analysis from eeg signals using stable phase-synchronized states

during face perception tasks," *Physica A: Statistical Mechanics and its Applications*, vol. 434, pp. 273–95, 2015.

[135] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE Transactions of Biomedical Engineering*, vol. 53, no. 2, pp. 266–72, 2006.

[136] T. Bassani and J. Nievola, "Pattern recognition for brain-computer interface on disabled subjects using a wavelet transformation," in *CIBCB'08.*, pp. 180–86, IEEE, 2008.

[137] V. Sakkalis, M. Zervakis, and S. Micheloyannis, "Significant eeg features involved in mathematical reasing: Evidence from wavelet analysis," *Brain Topogr.*, vol. 19, no. 1-2, pp. 53–60, 2006.

[138] Z. Sankari and H. Adeli, "Probabilistic neural networks for diagnosis of alzheimer's disease using conventional and wavelet coherence," *J. Neurosci. Methods*, vol. 197, no. 1, pp. 165–70, 2011.

[139] J. Lachaux, E. Rodriguez, J. Martinerie, and F. Varela, "Measuring phase syncrhony in brain signals," *Hum. Brain Mapp.*, vol. 8, pp. 194–208, November 1999.

[140] E. Gallego-Jutglà, J. Solé-Casals, F.-B. Vialatte, M. Elgendi, A. Cichocki, and J. Dauwels, "A hybrid feature selection approach for the early diagnosis of alzheimer's disease.," *J. Neural Eng.*, vol. 12, no. 1, p. no. 016018, 2015.

[141] J. Dauwels, F. Vialatte, T. Musha, and A. Cichock, "A comparative study of synchrony measures for the early diagnosis of alzheimer's disease based on eeg," *Neuroimage*, vol. 49, pp. 668–93, June 2010.

[142] A. Westdrop, "Volume conduction effects on correlation analysis of eeg data," in *Proceedings of the 12th Southern Biomedical Engineering Conference*, pp. 150–52, IEEE Conference Publications, 1993.

[143] S. van den Broek, F. Reinders, M. Donderwinkel, and M. Peters, "Volumen conduction effect in eeg and meg," *Electroencephalogry and Clinical Neurophysiology*, vol. 106, no. 6, pp. 522–34, 1998.

[144] P. L. Nunez and R. Srinivasan, *Electrical fields of the brain: the neurophysics of EEG.* Oxford University Press, 1981. ISBN: 9780195050387.

[145] M. Hassan, P. Benquet, A. Biraben, C. Berrou, O. Dufor, and F. Wendling, "Dynamic reorganization of functional brain networks during picture naming," *Cortex*, vol. 73, pp. 276–88, 2015.

[146] A. Kabbara, M. Khalil, W. El-Falou, H. Eid, and M. Hassan, "Functional brain connectivity as a new feature for p300 speller," *PLoS One*, vol. 11, no. 1, p. no. e0146282, 2016.

[147] G. Nolte, A. Ziehe, V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K.-R. Muller, "Robustly estimating the flow direction of information in complex physical systems," *Phys. Rev. Lett.*, vol. 100, no. 23, p. no. 234101, 2008.

[148] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186–98, 2009.

[149] G. Ball, P. Aljabar, S. Zebari, A. Edwards, and S. Counsell, "Rich organization of the newborn human brain," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 20, pp. 7456–61, 2014.

[150] D. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, and E. Bullmore, "Adaptive reconfiguration of fractal small-world human brain functional networks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 51, pp. 19518–23, 2006.

[151] C.-Y. Lo, P.-N. Wang, K.-H. Chou, J. Wang, Y. He, and C.-P. Lin, "Diffusion tensor tractography reveals abnormal topological organization in structural cortical networks in alzheimer's disease.," *J. Neurosci.*, vol. 30, pp. 16876–85, December 2010.

[152] L. Wang, C. Zhu, Y. Zang, Q. Cao, H. Zhang, Q. Zhong, and Y. Wang, "Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder," *Hum. Brain Mapp.*, vol. 30, no. 2, pp. 638–49, 2009.

[153] A. Pandit, E. Robinson, P. Aljabar, G. Ball, I. Gousias, Z. Wang, J.-V. Hajnal, D. Rueckert, S. Counsell, G. Montana, and A. Edwards, "Whole-brain mapping of structural connectivity in infants reveals altered connection strength associated with growth and preterm birth," *Cereb. Cortex*, vol. 24, pp. 2324–33, September 2014.

[154] W. Jamal, S. Das, K. Maharatna, F. Apicella, G. Chronaki, F. Sicca, D. Cohen, and F. Muratori, "On the existence of synchrostates in multichannel eeg signals during face-perception tasks," *Phys. Eng. Express*, vol. 1, no. 1, p. no. 15002, 2015.

[155] F. Guo, B. Hong, X. Gao, and S. Gao, "A brain-computer interface using motion-onset visual evoked potential.," *J Neural Eng*, vol. 5, pp. 477–85, 2008.

[156] B. Hong, F. Guo, T. Liu, X. Gao, and S. Gao, "N200-speller using motion-onset visual response," *Clin. Neurophysiol.*, vol. 120, no. 9, pp. 1658–66, 2009.

[157] C. Babiloni, F. Vecchio, P. Buffo, M. Buttiglione, G. Cibelli, and P. M. Rossini, "Cortical responses to consciousness of schematic emotional facial expressions: A high-resolution eeg study," *Hum. Brain Mapp.*, vol. 31, no. 10, pp. 1556–69, 2010.

[158] A. Bachiller, J. Poza, C. Gomez, V. Molina, V. Suazo, and R. Hornero, "A comparative study of event-related coupling patterns during an auditory oddball task in schizophrenia," *J. Neural Eng.*, vol. 12, p. no. 016007, February 2015.

[159] A. Medl, D. Flotzinger, and G. Pfurts, "Hilbert-transform based predictions of hand movements from eeg measures," in *14th Annual International Conference of the IEEE EMBS*, IEEE, 1992.

[160] P. Tass, M. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H. Freund, "Detection of n:m phase locking from noisy data: Application to magnetoencephalography," *Phys. Rev. Lett.*, vol. 81, no. 15, pp. 3291–94, 1998.

[161] T. Peng, A.B.Rowley, P. Ainslie, M. Poulin, and S. Payne, "Wavelet phase synchronization analysis of cerebral blood flow autoregulation," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 57, pp. 960–68, 2010.

[162] S. Bozhokin and I. Suslova, "Wavelet-based analysis of spectral rearrangements of eeg patterns and of non-stationary correlations," *Physica A*, vol. 421, pp. 151–60, 2015.

[163] X. Li, X. Yao, J. Fox, and J. G. Jefferys, "Interaction dynamics of neuronal oscillations analysed using wavelet transforms," *Journal of Neurosciene Methods*, vol. 160, pp. 178–185, 2007.

[164] D. Li, X. Li, D. Cui, and Z. Li, "Phase synchronization with harmonic wavelet transform with application to neuronal populations," *Neurocomputing*, vol. 74, no. 2011, pp. 3389–3403, 2011.

[165] P. S. Addison, *The Illustrated Wavelet Transform Handbook*. IOP Publishing Ltd, July 2002.

[166] G. Lance and W. Williams, "A general theory of classificatory sorting strategies ii. clustering systems.," *The Computer Journal*, vol. 9, no. 4, pp. 373–380, 1967.

[167] M. Kyan, P. Muneesawang, K. Jarrah, and L. Guan, *Unsupervised Learning: A Dynamic Approach*. John Wiley and Sons, 2014.

[168] A. Singh, A. Yadav, and A. Rana, "K-means with three different distance metrics," *International Journal of Computer Applications*, vol. 67, no. 10, pp. 13–17, 2013.

[169] C. Aggarwal and C. Reddy, *Data clustering: algorithms and applications*. Data Mining and Knowledge Discovery Series (Book 31), Chapman and Hall/ CRC, 2014. ISBN: 9781466558212.

[170] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to pattern recognition*. Elsevier, 2010.

[171] C. Spironelli, A. Angrilli, A. Calorego, and L. Stegagno, "Delta eeg band as a marker of left hypofrontality for language in schizophrenia patients," *Schizophr. Bull.*, vol. 37, no. 4, pp. 757–767, 2011.

[172] W. Jamal, S. Das, K. Maharatna, D. Kuyucu, F. Sicca, L. Billeci, F. Apicella, and F. Muratori, "Using brain connectivity measure of eeg synchrostates for discriminating typical and autism spectrum disorder," *International IEEE/EMBS Conference on Neural Engineering-NER*, pp. 1402–05, 2013.

[173] M. Latka, M. Turalska, M. Glaubic-Latka, W. Kolodziej, D. Latka, and B. J. West, "Phase dynamics in cerebral autoregulation.," *American journal of physiology. Heart and circulatory physiology*, vol. 289, no. 5, pp. H2272–9, 2005.

[174] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–69, 2010.

[175] M. Kaiser, "A tutorial in connectome analysis: Topological and spatial features of brain networks," *NeuroImage*, vol. 57, no. 3, pp. 892–907, 2011.

[176] A. Fornito, A. Zalesky, and E. T. Bullmore, *Fundamentals of Brain Network An.* Elsevier, 2016.

[177] D. Watts and S. Strogatz, "Collective dynamics of 'small-word' networks.," *Nature*, vol. 393, pp. 440–42, 1998.

[178] S. Milgram, "The small-world problem," *Psychology Today*, vol. 2, no. 1, pp. 61–67, 1967.

[179] D. S. Bassett and E. T. Bullmore, "Small-world brain network revisited," *The Neuroscientist*, no. 1073858416667720, 2016.

[180] M. Humphries and K. Gurney, "Network 'small-world-ness': a quantitative method for determining canonical network equivalence," *PLoS One*, vol. 3, no. 4, p. no. e0002051, 2008.

[181] M. Kaiser, "Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks," *New Journal of Physics*, vol. 10, p. no.083042, 2008.

[182] D. Simard, L. Nadeau, and H. Kroguer, "Fastest learning in small-world neural networks," *Phys. Lett. A*, vol. 336, no. 1, pp. 8–15, 2005.

[183] M. Hassan, M. Shamas, M. Khalil, W. Falou, and F. Wendling, "Eegnet: An open source tool for analyzing and visualizing m/eeg connectome," *PLoS ONE*, vol. 10, p. 9, 2015.

[184] M. Hassan, O. Dufor, I. Merlet, C. Berrou, and F. Wendling, "Eeg source connectivity analysis: from dense array recordings to brain networks," *PLoS One*, vol. 9, no. 8, p. e105041, 2014.

[185] R. Rosario, P. Cardoso, M. Munoz, P. Montoya, and J. Miranda, "Motif-synchronization: A new method for anlysis of dynamic brain network with eeg," *Physica A*, vol. 439, no. August, pp. 7–19, 2015.

[186] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora, "Small-world behaviour in time-varying graphs," *Physical Review*, vol. 81, p. no.055101(R), 2010.

[187] S. Mehrkanoon, M. Breakspear, and T. W. Boonstra, "Low-dimensional dynamics of resting-state cortical activity," *Brain Topogr.*, vol. 27, no. 3, pp. 338–52, 2014.

[188] C. Chang, Z. Liu, M. C. Chen, X. Liu, and J. H. Duyn, "Eeg correlates of time-varying bold functional connectivity," *NeuroI*, vol. 72, pp. 227–36, 2013.

[189] C. A. Filho, B. Campos, T. Costa, L. Uribe, C. Barreto, R. Attux, and G. Castellano, "Graphs metrics as features for an lda based classifier for motor imagery data," in *3rd Brainn Congress*, vol. April, pp. 12–13, 2016.

[190] C. A. Filho, B. Campos, T. Costa, L. Uribe, C. Barreto, R. Attux, and G. Castellano, "Can graph metrics be used for mi-bcis?," in *3rd Brainn Congress*, p. 4128, 2016.

[191] J. Asensio-Cubero, J. Q. Gan, and R. Palaniappan, "Multiresolution analysis over graphs for a motor imagery based online bci game," *Comput. Biol. Med.*, vol. 68, pp. 21–26, 2016.

[192] J. Pineda, B. Allison, and A. Vankov, "The effects of self-movement, observation, and imagination on mu rhythms and readiness potentials (rp's): Toward a brain-computer interface (bci)," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 8, pp. 219–22, 2000.

[193] G. Pfurtscheller and F. L. da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–57, 1999.

[194] F. van der Heijden, R. P. Duin, D. de Ridder, and D. M. Tax, *Classification, Parameter Estimation and State Estimation*. John Wiley & Sons, 2004.

[195] I. Rojas, G. Joya, and A. Catala, eds., *Advances in Computational Intelligence: Proceedings of the 13th International Work-Conference on Artificial Neural Networks (IWANN). Part 2*. Springer, 2015.

[196] A. Soro, E. Vargiu, G. Armano, and G. Paddeu, eds., *Information Retrieval and Mining in Distributed Environments*. Springer, 2010.

[197] *MathWorks, Statistics and Machine Learning Toolbox: User's Guide (2016)*.

[198] S. Rogers and M. Girolami, *A First Course in Machine Learning.* Machine Learning and Pattern Recognition, Chapman and Hall/ CRC, 2011. ISBN-13: 978-1498738484.

[199] E. Parvinnia, M. Sabeti, M. Jahromi, M. Zolghadri, and R. Boostani, "Classification of eeg signals using adaptive weighted distance nearest neighbor algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 26, no. 1, pp. 1–6, 2014.

[200] A. Khasnobish and S. Bhattacharyya, "K-nearest neighbor classification of left-right limb movement using eeg data," in *International Conference on Systems in Medicine and Biology, ICSMB*, vol. 2010, pp. 126–31, 2010.

[201] T. Rahman, A. Ghosh, M.-H. Shuvo, and M. Rahman, "Mental stress recognition using k-nearest neighbor (knn) classifier on eeg signals," in *International Conference on Materials, Electronics and Information Engineering (ICMEIE)*, 2015.

[202] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *SpringerPlus*, vol. 5, no. 1, p. 1304, 2016.

[203] R. Zhang, P. Huang, L. Guo, Y. Zhang, P. Li, and D. Yao, "Z-score linear discriminant analysis for eeg based brain-computer interfaces," *PLoS ONE*, vol. 8, no. 9, 2013.

[204] Y. Li, J. Long, T. Yu, Z. Yu, C. Wang, H. Zhang, and C. Guan, "An eeg-based bci system for 2-d cursor control by combining mu/beta rhythm and p300 potential," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10 part I, pp. 2495–505, 2010.

[205] P. Herman, G. Prasad, T.-M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 317–26, 2008.

[206] S. M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from eeg signals in brain-computer interface," *Information Sciences*, vol. 178, no. 6, pp. 1629–40, 2008.

[207] Q. Xu, H. Zhou, Y. Wang, and J. Huang, "Fuzzy support vector machine for classification of eeg signals using wavelet-based features," *Med. Eng. Phys.*, vol. 31, no. 7, pp. 858–65, 2009.

[208] A. Celisse and S. Robin, "Non-parametric density estimation by exact leave-p-out cross-valdiation," *Comput. Stat. Data Anal.*, vol. 52, pp. 2350–2368, 2008.

[209] D. Drehmer and G. Morris, "Cross-validation with small samples: An algorithm for computing gollob's estimator," *Educational and Psychological Measurement*, vol. 41, pp. 195–200, 1981.

[210] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: practical techniques management.* Morgan Kaufmann Publishers, 3rd. ed., 2011.

[211] D. H. Krishna, I. Pasha, and T. S. Savithri, "Classification of eeg motor imagery multi class singals based on cross correlation," *Procedia Comput. Sci.*, vol. 85, pp. 490–495, 2016.

[212] Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, "Classification of motor imagery eeg singals with support vector machines and particle swarm optimization," *Computational and Mathematical Methods in Medicine*, vol. 2016, p. ID. 4941235, 2016.

[213] S. Siuly and Y. Li, "Improving the separability of motor imagery eeg signals using a cross correlation-based least square support vector machine for brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, pp. 526–538, July 2012.

[214] S.-H. Park and S.-G. Lee, "Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern," *IEEE Sens. J.*, vol. 17, no. 10, pp. 2977–2983, 2017.

[215] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, April 2014.

[216] E. Goeleven, R. D. Raedt, L. Leyman, and B. Verschuere, "The karolinska directed emotional faces: a validation study," *Cognition and emotion*, vol. 22, no. 6, pp. 1094–118, 2008.

[217] D. Lundqvist, A. Flykt, and A. Ohmar, "The karolinska directed emotional faces (fdef)," *Stockholm: Department of Neurosciences Karolinska Hospital*, 1998.

[218] E. Dong, C. Li, S. Du, A. N. Belkacem, and C. Chen, "Classification of multi-class motor imagery with a novel hierarchical svm algorithm for brain-computer interfaces," *Med. Biol. Eng.*, pp. 1–10, 2017.

[219] Y. Hashimoto and J. Ushiba, "Eeg-based classification of imaginary left and right foot movements using beta rebound," *Clin. Neurophysiol.*, vol. 124, pp. 2153–2160, 2013.

[220] E. Baçar, "A review of gamma oscillations in healthy subjects and in cognitive impairment," *Int. J. Psychophysiol.*, vol. 90, pp. 99–117, Nov 2013.

[221] C. Wilke, G. Worrel, and B. He, "Graph analysis of epileptogenic networks in human partial epilepsy," *Epilepsia*, vol. 52, pp. 84–93, January 2011.

[222] Y. Punsawad, S. Ngmrussameewong, and Y. Wongsawat, "On the development of bci and its neurofeedback training system for assistive communication device in persons with severe disability," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–4, IEEE, 2016.

[223] H. Ji, J. Li, R. Lu, R. Gu, L. Cao, and X. Gong, "Eeg classification for hybrid brain-computer interface using a tensor based multiclass multimodal analysis scheme," *Computational Intelligence and Neuroscience*, p. no.1732836, 2016.

[224] A. C. Ramos, R. G. Hernandez, and M. Vellasco, "Feature selection methods applied to motor imagery task classification," in *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, IEEE, 2016.

[225] BBCI, "Bci competition." http://bbci.de/competition/.

[226] S. Siuly and Y. Li, "Improving the seprability of motor imagery eeg signals using a cross correlation-based least square support vector machine for brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, 2012.

[227] A. Zarei and F. Ghassemi, "Spatial filter bank based on probabilistic common spatial pattern in motor imagery bci systems," in *23rd Iranian Conference on Biomedical Engineering (ICBME)*, IEEE, 2016.

[228] H. Lu, K. Plataniotis, and A. Venetsanopoulus, "Regularized common spatial patterns with generic learning for eeg signal classification," in *Proc. EMBC*, pp. 441–446, IEEE EMBS, 2009.

[229] J. S. Kirar and R. Agrawal, "Composite kernel support vector machine based performance enhancement of brain computer interface in conjuntion with spatial filter," *Biomed. Signal Process. Control*, vol. 33, no. 151-160, 2017.

[230] S. Seifzadeh and M. Rezaei, "Fast and efficient four-class motor imagery electroencephalography singal analysis using common spatial pattern-ridge regression algorithm for the purpose of brain-computer interface," *J. Med. Sign. Sence.*, vol. 7, pp. 80–85, 2017.

[231] S.-L. Wu, Y.-T. Liu, T.-Y. Hsieh, Y.-Y. Lin, C.-Y. Chen, Chun-Hsiang, and C.-T. Lin, "Fuzzy integral with particle swarm optimization for a motor-imagery-based brain-computer interface," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 21–27, 2017.

[232] T. Qian, W. Zhou, Z. Ling, S. Gao, H. Liu, and B. Hong, "Fast presurgical functional mapping using task-related intracranial high gamma activity," *J. Neurosurg.*, vol. 119, pp. 26–36, July 2013.

[233] K. Pearson, "Notes on regression and inheritance in the case of two parents," *Proc. Roy. Soc. London*, vol. 58, pp. 240–42, June 1895.

[234] D. T. Bundy, M. Pahwa, N. Szrama, and E. C. Leuthardt, "Deconding three-dimensional reaching movements using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 13, no. 2, p. 026021, 2016.