Méthodes d'analyse et de débruitage multicanaux à partir d'ondelettes pour améliorer la détection de potentiels évoqués sans moyennage : application aux interfaces cerveau-ordinateur

Carolina Véronica Saavedra Ruiz

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Méthodes d'analyse et de débruitage multicanaux à partir d'ondelettes pour améliorer la détection de potentiels évoqués sans moyennage

application aux interfaces cerveau-ordinateur

THÈSE

présentée et soutenue publiquement le 14 Novembre 2013

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Carolina Verónica Saavedra Ruiz

Composition du jury

Rapporteurs :	Théodore Papadopoulo, Chargé de recherche, Inria Sophia Antipolis Alain Rakotomamonjy, Professeur, Université de Rouen
Examinateurs :	Anne Boyer, Professeur, Université de Lorraine François Cabestaing, Professeur, Université Lille 1 Bernard Girau, Professeur, Université de Lorraine Laurent Bougrain, Maître de Conférences, Université de Lorraine

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Résumé

Une interface cerveau-ordinateur (en anglais BCI) est un système de contrôle ou de communication qui traduit l'activité cérébrale en commandes pour un logiciel ou d'autres dispositifs comme un système d'écriture ou des prothèses motorisées. Initialement développées pour aider les personnes avec des déficiences motrices importantes, l'utilisation des BCIs se généralise en tant que nouveau moyen d'interaction, en particulier dans le domaine des jeux video. Un des phénomènes physiologiques permettant cette interaction est le potentiel évoqué cognitif P300. Il correspond à une déflexion positive observable par électroencéphalographie autour de 300 ms après la présentation d'un stimulus rare et attendu. Le potentiel P300 présente plusieurs intérêts : i) il permet de proposer un plus grand nombre de commandes que d'autres paradigmes ; ii) aucun entraînement de la part de l'utilisateur n'est nécessaire ; iii) cette réaction est observable chez la quasi totalité des personnes. Ces réactions cérébrales sont cependant difficiles à observer parce que le matériel utilisé pour enregistrer les signaux sur le scalp se trouve éloigné des sources et enregistre également du bruit et des artéfacts qui viennent s'ajouter à la réformation produite par le crâne. Le P300 a une amplitude environ dix fois inférieure au reste de l'activité électroencéphalographique. En conséquence, il est nécessaire d'utiliser des techniques poussées de prétraitement et de classification pour le détecter. Habituellement, pour améliorer le rapport signal sur bruit des potentiels évoqués, les stimulations sont répétées afin de moyenner les réponses enregistrées. Le P300 étant synchronisé sur l'apparition du stimulus, à la différence du reste de l'activité cérébrale, il augmente lors du moyennage alors que le reste de l'activité s'atténue. Les performances de reconnaissance du système s'améliorent mais l'interaction est ralentie. Deux autres problèmes liés au moyennage sont d'une part l'étalement de la moyenne à cause d'une faible mais réelle variation de la latence des réponses et d'autre part parfois l'apparition de faux P300. Dans cette thèse, de nouvelles techniques basées sur la théorie des ondelettes sont développées pour améliorer la performance des systèmes BCI basés sur la détection des P300 en utilisant des mesures de similarité entre les signaux. Une technique présentée dans cette thèse débruite les signaux en considérant simultanément, et non de manière indépendante, la phase des signaux. Cette méthode a été comparée aux techniques classiques de débruitage par seuillage des coefficients d'ondelettes sur une tâche de détection des P300 sans moyennage. Nous avons également étendu cette approche à l'étude des phases et des amplitudes pour obtenir une mesure de corrélation entre les réponses contenant et ne contenant pas de P300. Cette mesure permet d'étudier la localisation temporelle du P300 pour chaque utilisateur dans le but de sélectionner automatiquement la fenêtre temporelle à étudier et de faciliter la détection. Une fenêtre spécifique à chaque canal peut être obtenue à partir des moyennes faites sur les réponses contenant et ne contenant pas de P300 ou une fenêtre commune à tous les canaux peut être obtenue en utilisant les moyennes faites sur tous les canaux. La combinaison de nos techniques de débruitage et de sélection automatique de la fenêtre d'étude a été comparée à l'algorithme xDAWN. Les algorithmes présentés dans cette thèse ont été évalués sur une base de données provenant du laboratoire de neuromagerie de l'université métropolitaine autonome du Mexique et contenant 22 sujets sains. Une deuxième base de données provenant du groupe de traitement de signaux multimedia de l'EPFL et contenant des patients a été utilisée pour valider les performances des algorithmes en utilisant les paramètres sélectionnés pour la première base de données. Les résultats montrent que les méthodes présentées dans cette thèse améliorent significativement la détection des P300 sans moyennage par rapport aux méthodes conventionnelles.

Mots-clés: Interfaces cerveau-ordinateur, potentiels évoqués, analyse temps-fréquence, détection sans moyennage.

Abstract

Brain-Computer Interfaces (BCI) are control and communication systems which were initially developed to assist people with motor disabilities. The main idea behind BCI is to translate the brain activity into commands for a computer application or other devices, such as a spelling system or a robotic limb. Nowadays, BCI systems are becoming very popular because of its potential to be applied in other domains, where the target audience correspond to healthy people, as for example in computer games.

A popular non-invasive technique to record brain signals is the electroencephalography (EEG), because it is easy to use, safe for users and it is not bulky or expensive, compared to other recordings techniques. The EEG records the voltage fluctuations of the brain (brain waves) triggered by different stimuli or mental states. Event-Related potentials (ERPs) are one of the most used signals for BCI that can be recorded using EEG.

Despite the BCI popularity, it is generally difficult to work with brain signals, because the hardware used to record them also records noise and artifacts (*i.e.* body signals from the subject or environmental signals), and because the brain components amplitudes are very small compared to the whole ongoing EEG activity. Moreover, ERP responses have different latencies and amplitudes for each subject (or even for the same subject), causing the use of large time-windows including irrelevant information to perform the analysis. As a consequence it is necessary to use intelligent techniques to recognize ERPs. For these reasons, a commonly used technique was conceived to improve the signal-to-noise ratio of brain signals : averaging several single-trial brain signals obtained under the same conditions. This technique reduces the noise and artifacts, naturally improving the system's performance, at the cost of slowing the whole process due to the extra repetitions required to function. Another problem related with the average of repetitions occurs when there are latency jitter or phase artifacts between single trials, which can cause the elimination of transient characteristics.

In this thesis, new techniques based on wavelet theory are developed to improve the performance of ERP-based BCI systems using similarity measures between signals. One technique presented in this thesis denoises the signals in the wavelet domain simultaneously for all channels, *i.e.* not independently, using the phase information provided by all signals. This method has been compared to classic thresholding techniques based on wavelets on single-trials ERP detection. We have also extended this approach combining the phase and the amplitude information provided by the signals to obtain a correlation between target and non-target signals. This measure allows to localize automatically the ERP in time, for each user, in order to reduce the time-window to analyze, to remove useless features and make the analysis easier. A specific time-window selection can be done per channel using target and non-target averages, or the same time-window can be selected for all channels using target and non-target grand averages. The combination of our two algorithms for denoising the signals and select the time-window has been compared to xDAWN which is an efficient algorithm to enhance and detect ERPs.

The algorithms presented in this thesis are evaluated and compared on single-trial detection using a large dataset for the P300 speller containing 22 first-time users provided by the neuroimaging laboratory of the Metropolitan autonomous university of Mexico. A second database including patients provided by the multimedia signal processing group of EPFL has been blindly used for validation using the parameters selected for the first database. The results show that the generated models exhibit significant improvements with respect to conventional methods.

Keywords: Brain-Computer Interfaces, Event-Related Potential, Signal Analysis, single-trial detection.

Remerciements

My thesis was possible only by a strong institutional support from several entities. First, I have to acknowledge the Chilean government through CONICYT and the Embassy of France in Chile, for granting me the scholarship to pursue the PhD degree here in France. Also, to the University of Lorraine, LORIA and INRIA institutions, which gave me a place to work and the opportunity to join a community of people that enjoy sharing knowledge and always willing to help when I needed.

I want to thank to my advisor, Laurent, for his kindness, humbleness and sympathy. Despite your busy agenda you always found the time to speak with me, even if you were almost always running to meet me or meet someone else. I want you to know I really admire your ability to be calm and confident, even when everything around seems to be chaos. Thank you for all the knowledge you gave me and for been honest when you didn't have an answer. I appreciate the effort you did to study new topics, and to contact an expert when it was needed. For the same reason I also want to thank to Radu Ranta, for all the hours he spent helping me, and Oscar Yañez, who kindly read a chapter of my thesis and helped me to improve it. Also, I want to thank Bernard for all the administrative support and his advices to improve the thesis slides. Furthermore, I want to acknowledge the jury for accepting to correct my thesis and attending to my thesis defense, specially to Alain Rakotomamonjy and Théodore Papadopoulo who spent a lot of time reading and understanding my work.

After four years living in Nancy I got used to the weather, the lack of light in winter, the strange french habits and so many things that would have been impossible to bear without the extraordinary people I found here. I want to thank my friends for all the affection they have given me: it has been difficult to begin saying goodbye to you and I will miss you all. Victor, Mauricio, Daniela, Isabel, Carlos(s), Gabriela, Tomas, Aurore, Nicole, Giorgos, Laure, Arnaud, Maxime, Olivier(s), and so on. I want to specially acknowledge to Jano Yazbeck, who had the courage to hear a thesis rehearsal the night before my thesis defense to help me with my french.

I want to thank my family, for always supported me in my adventures and for their unconditional faith in me. Thank you for teach me how to be a better person and for all the words of encouragement these last months. Finally I want to thank Mauricio: without you in my life these last months would have been a misery for me. Thank you for discussing with me about my thesis when I needed it, even if sometimes you didn't understand what I was talking about. Thanks for stay with me each and every night I worked late, thanks for your support, patience, advice and for sharing with me my successes and failures. I am very grateful to have you in my life, undoubtedly you are the best person I know and I am proud to share my life with you. For that reason I dedicate this thesis to you, because without you it wouldn't been possible.

Contents

List of Figures

List of Tables

Glossary

ALS : Amyotrophic Lateral Sclerosis

BCI : Brain-Computer Interface

CNS : Central Nervous System
CSP : Common Spatial Patterns
CWT : Continuous Wavelet Transform

dMRI : Diffusion Magnetic Resonance Imaging
DWT : Discrete Wavelet Transform

ECoG : Electrocorticography
EEG : Electroencephalography
ERD : Event-Related Desynchronization
ERP : Event-Related Potentials
ERS : Event-Related Synchronization

FLD : Fisher's Linear Discriminant
fMRI : functional Magnetic Resonance Imaging
fNIR : functional Near Infrared
FT : Fourier Transform

ICA : Independent Component Analysis
iEEG : intracranial Electroencephalography
ISI : Interstimulus Interval

LDA : Linear Discriminant Analysis
LFP : Local Field Potentials

MEG : Magnetoencephalography
METS : Multi-channel EEG Thresholding by Similarity
MRL : Mean resultant length

PCA : Principal Component Analysis
PET : Positron Emission Tomography

SCP : Slow Cortical Potentials
SEWS-1 : Semblance-based ERP Window Selection by Channel
SEWS-2 : Semblance-based ERP Window Selection for all Channels
SMR : Sensorimotor Rhythms
SNR : Signal-to-Noise Ratio

SPECT : Single Photon Emission Computed Tomography
SSVEP : Steady-state Visual Evoked Potential
STFT : Short Time Fourier Transform
SURE : Stein Unbiased Risk Estimator
SV : Support Vector
SVM : Support Vector Machine
SWLDA : Stepwise Linear Discriminant Analysis

TcD : Transcranial Doppler Ultrasound
TVEP : Transient Visual Evoked Potential

VEP : Visual Evoked Potentials

WT : Wavelet Transform

Introduction

In the last few decades, the interest in the human brain has increased exponentially, mainly because impressive advances in brain imaging have been developed. New acquisition techniques (*e.g.* the magnetic resonance imaging and its functional and diffusion extensions), new sensors (*e.g.* the multielectrodes array) and new signal processing methods (*e.g.* the wavelet theory) have improved the study of the brain functions and consequently allowed a deeper understanding of the brain. Indeed, the brain activity generates a large amount of information, from the molecular level up to the behavioral level, which can be recorded, represented and processed to extract relevant characteristics for decision making. However, it is practically impossible to analyze directly these data due to the highly complex nature of the information. For this reason, while efficient algorithms have been developed to manage and analyze the information stored in brain signals, there is still an acute need to improve their reliability and performance.

Since the first electroencephalographic (EEG) recordings were made by Hans Berger [Berger 1929], scientists have been interested in understanding the brain activity, but also in interacting with it or using it for communication and control. For example, in 1963, José Manuel Rodriguez Delgado implanted an electrode in the caudate nucleus of a bull to stop it when delivering a current [Delgado 1967]. In 1973, Jacques Vidal presented the seminal idea about brain-computer interfaces (BCI), showing the feasibility of using brain signals in man-machine communication [Vidal 1973]. To accomplish this goal, the *Brain Computer Interface* project was created, using the following BCI definition: "The Brain Computer Interface system is geared to the use of both the spontaneous EEG and the specific evoked responses triggered by time-dependent (visual) sensory stimulation under various conditions". The development of BCI has exponentially increased since this early approach, attracting more and more researchers every day. Nowadays, the BCI research covers a variety of works related to the recording technique, the physiological phenomenon, the feedback, the final user.... Nevertheless, all these BCI works can be described by one of the most used definitions presented in the first BCI international meeting: "A brain–computer interface is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles" [Wolpaw et al. 2000]. It is a worldwide spread research, involving different disciplines, such as biology, medicine and engineering. Two of the most well-known projects in neuroscience integrate a BCI part: the *Blue Brain*[1] project, based in Switzerland, which aims to build a virtual brain, and *the human brain*[2] project, a joint project between the European union and the U.S.A., where the goal is to build an information and communication technology infrastructure for neuroscience in order to understand and emulate brain capabilities. A lot of international projects have been entirely focused on BCI: MAIA[3], Brain-

1. `http://bluebrain.epfl.ch`
2. `http://www.humanbrainproject.eu`
3. `http://cnbi.epfl.ch/page-34069-en.html`

Gate [4], BrainGate2 [5], TOBI [6], Presencia [7], BackHome [8], BrainAble [9], BrIndiSys [10], Decoder [11], VERE [12], etc. In France also, projects contributed to the domain: OpenViBE [13], CoAdapt [14], ROBIK, and OpenViBE2 among others.

As mentioned before, the technologies used for recording the physiological phenomena dealt by the BCI research and their associated medical cares have rapidly improved over the years. More specifically, among the available technologies, non-invasive EEG have had important advances, such as the use of pre-amplifiers to avoid environmental noise and cable movements, dry electrodes designed to be used for long periods of time and wireless systems that improve the mobility of users. EEG is a commonly used method to record in real-time brain activity non-invasively, because it is easy to use, affordable and safe, being the most popular recording system used nowadays in BCI applications. The most common phenomena recorded in EEG are: visual evoked potentials (VEP), sensorimotor rhythms (SMR), slow cortical potentials (SCP) and event-related potentials (ERP). In this thesis, we focus on ERPs, that consist in triggering a specific brain wave through a stimulus. ERPs are well known and popular in BCI applications, because they are time-locked with sensory or cognitive events. In addition, they allow a large number of commands without tedious training for the user and without the need of any particular skill. Finally, ERPs can be easily triggered using the oddball paradigm, which consists in a set composed of rare and common stimulus in an unbalanced proportion. An ERP signal is triggered in response to a rare stimulus.

Current BCI applications based on ERPs are mainly focused on patients for communication or entertainment purposes. In order to be efficient in performance, one important element is to record the response of the same stimulus several times. These repetitions are averaged to improve the ERP detection by increasing the signal-to-noise ratio (SNR). This technique is also used to study ERPs, being able to detect their shapes, peaks and latencies. The applications that use the average of responses to improve ERP detection are very useful for people with disabilities caused by accidents or illness, such as spinal cord injury or amyotrophic lateral sclerosis (ALS). However, they need to use a short interstimulus interval (ISI) to have a higher information transfer rate and to present several times each stimulus, causing in some cases fatigue as a result of the permanent concentration state through long periods of time. This is a major disadvantage in their use by healthy people in applications developed for the general public (*e.g.* computer games) because they require a long time to perform a specific action.

The improvement of the information transfer rate depends on the ability to reduce the number of repetitions. Obviously, the ideal case, that is single-trial detection, is to relate one stimulus to only one brain response. Nowadays the performance in single trial detection is not sufficient to obtain reliable BCI systems, but they are sufficiently attractive to encourage researchers to develop methods dealing with single-trial responses. The use of single-trial responses will provide better conditions for patients in terms of comfort, because the dynamics between the system and the patient will be increased, avoiding onerous conditions and improving the communication

4. `http://www.braingate.com`
5. `http://www.braingate2.org`
6. `http://www.tobi-project.org`
7. `http://www0.cs.ucl.ac.uk/research/vr/Projects/Presencia/index.html`
8. `http://www.backhome-fp7.eu`
9. `http://www.brainable.org`
10. `http://www.brindisys.it`
11. `www.decoderproject.eu`
12. `http://www.vereproject.eu`
13. `http://openvibe.inria.fr`
14. `https://twiki-sop.inria.fr/twiki/bin/view/Projets/Athena/CoAdapt/WebHome`

speed. In order to allow the use of single-trial responses in BCI, signal processing techniques are required to deal with the artifacts and noise embedded in the recordings and the inter- and intra-subject variability. Also the ability to reduce the number of trials will have an impact on applications in other areas for other publics.

Improving the single-trial detection is of course useful for disabled people, but a stronger impact may be expected for healthy people. Indeed, the today's challenge for BCI is to develop plug-and-play applications that can be easily used outside the laboratory. For example, signal acquisition techniques must be improved to be easily deployed for end users, allowing a permanent home-safe use without regular maintenance, and achieving good performance in different environments with high information transfer rates. Despite the fact that the development of BCI applications and its improvement are still in their early steps, others domains are becoming interested in BCI, incorporating EEG technology in applications that are not related to patients nor medicine. One of the most relevant examples in this context is the game industry (see e.g. the Openvibe2 project). A new conference about games based on BCI technology, NeuroGaming [15], occurred in 2013. A survey about BCI games can be found in [Nijholt 2009]. The expansion of BCI to others domains opens the challenge of improving the information transfer rate in order to be attractive for healthy people, which implies that the ideal scenario is to use the fewest number of electrodes (comfort) and a single-trial brain signal detection. We strongly believe that wavelet theory can help to reach this goal.

Wavelet theory (WT) is a well-known technique widely used in areas like astrophysics, seismology and computer graphics. It was created to fulfill the need to analyze high and low frequencies of signals with discontinuities and abrupt changes. The wavelet transform is based on the superposition of functions providing a multi-resolution analysis. The basic function is called *mother wavelet*, where contracted versions of it are used to analyze high frequencies, and dilated versions are used to analyze low frequencies, covering the whole signal spectrum. The result is the analysis at a desired level of detail of the given signal. For example, if we listen classical music, it is possible to distinguish many different musical notes. The analysis using wavelets will give us the main frequencies, when they occur and their duration. Therefore, if the musical notes are played in a different order, wavelets are capable to distinguish the changes, revealing discontinuities and trends. The main advantage of wavelets is that they allow to study the signal jointly in the time and the frequency domains. This is of great interest in the study of brain rhythms in EEG signals, encouraging its use in the BCI domain to discern useful information from the whole signal dynamics. Nevertheless, wavelets have been used only to study single channels in the BCI domain, *i.e.* the signal recorded by one electrode individually without exploring the shared information between signals.

The objective of this thesis is to improve the methodologies used in BCI to enhance the performance of these systems without the need of averaging ERP responses, *i.e.* using single-trial detection. In order to do this, we propose to exploit the information of all the recorded channels as a whole, and not individually as it is usually done in current techniques, using similarity measures for denoising. This is based on the hypothesis that multiple channels are recorded at the same time and on the knowledge that brain signals spread through the scalp instantaneously. Consequently, it is reasonable to exploit the joint information of phase and amplitude. Our study is focused on ERP responses, and specifically on the P300 wave triggered by the oddball paradigm. Its use is widely spread because of the advantages encountered, such as, its high amplitude compared to other ERP events, its time locking characteristic, which allows to

15. `http://www.neurogamingconf.com/`

focus the study of the signal only to a temporal window after the stimulus, and because the P300 can be efficiently produced by almost everyone. The most famous BCI system based on the P300 is the speller presented by Farwell and Donchin in 1988 [Farwell and Donchin 1988], which has been used as the basic model to develop different applications based on the same methodology.

The methods presented in this thesis are based on wavelet theory, specially on correlation measures that are currently used in geology to analyze geophysical data. The use of these measures will allow i) the development of a thresholding technique based on the correlations found between recorded channels ii) and the automatic selection of a temporal window after the stimulus. These methods were developed to improve the performance in single-trial detection. Even though this thesis is focused on ERP detection, the methods presented in this manuscript can be used for other brain signals, since the used model is general enough to be applied to other domains.

This thesis is divided in six chapters. Following this short introduction, Chapter 1 presents a background on brain anatomy and brain functions, different brain signals that can be identified in brain recordings and the existing methods available to actually record them. It should be pointed out that this thesis focuses mainly on electroencephalographic signals, because they offer a high temporal resolution and can be easily recorded using technology that has no contraindication for the user.

In Chapter 2, Brain-Computer Interfaces are presented, starting with the correct design (architecture) of BCI applications, its taxonomy, the physiological phenomena used to develop BCIs, specially focused on P300 and the challenges and problems encountered in its development, whether associated with the user, the paradigm used or the signal processing.

Chapters 3 and 4 describe formal methodologies used in BCI systems for feature extraction and feature classification, such as the Wavelets Theory and the Support Vector Machines respectively. These methodologies are used in the design of the algorithms and experiments introduced in this thesis. These new algorithms are explained in Chapter 5, which are based on the signals correlations in the wavelet domain, in order to be able to study the frequency and time domains jointly. Finally, in Chapter 6,

The algorithms presented in this thesis are evaluated and compared on single-trial detection using a large dataset for the P300 speller containing 22 first-time users provided by the neuroimaging laboratory of the Metropolitan autonomous university of Mexico. A second database including patients provided by the multimedia signal processing group of EPFL has been blindly used for validation using the parameters selected for the first database. The results show that the generated models exhibit significant improvements with respect to conventional methods.

Finally, Chapter 7 presents the conclusion obtained for the presented techniques and the future work proposed to continue in this research direction.

Chapter 1

Brain Signals

Contents

Understanding the structure and the functions of the human brain has been a goal for scientists for many years. At the beginning of such research, it was impossible to perform studies without opening the skull, hindering the progress of the field. Then, a variety of new technologies were developed to study the brain (even for normal subjects) without the need for surgery. This allowed the study of the brain meanwhile performing activities, such as, listening, dreaming or moving body parts, being able to identify which groups of neurons became active at that instant allowing the improvement of brain models. Furthermore, the knowledge gained from the study of brain signals have also enriched the understanding and treatment of mental illness, improving the quality of life of people affected by these diseases.

Nowadays, the brain signals are also used to replace damaged neural pathways in individuals through the development of brain-computer interfaces to control external devices that would be impossible to control otherwise. Even if we are still far from being capable to completely understand what happens in the brain, great efforts are being made to accomplish this goal.

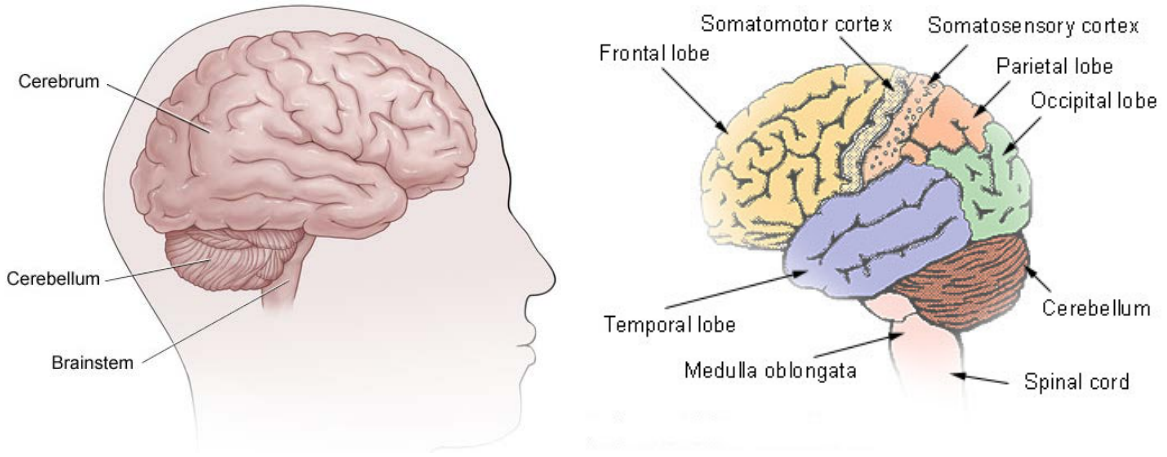
The methods presented in this thesis aim at improving the analysis of brain signals, which can be used specially as a communication channel to send commands to an electronic device. Thus, in this chapter, some basic information about the brain anatomy and the different brain imaging techniques to record its activity are presented. Particularly, we focus on the electroencephalography (EEG), one of the most used techniques to record the brain activity non-invasively.

1.1 The Brain and the Body

The brain is a part of the *central nervous system* (CNS) organ that allows us to analyze what is happening in our environment and interact with it. The brain coordinates body movements, regulates homeostatic functions such as blood pressure, body temperature and heart rate, and also controls our thoughts and feelings.

1.1.1 Brain Structures

The brain is composed of three main parts (see Figure 1.1(a)): the cerebellum, the brainstem and the cerebrum (including the cortex), this last one been the largest part of the human brain. The cortex is composed by the right cerebral hemisphere and the left cerebral hemisphere. The hemispheres are divided by a longitudinal fissure and connected by the corpus callosum. Each hemisphere is subdivided into four lobes by other less deep fissures (see Figure 1.1(b)). The central fissure (or central sulcus) separates the parietal lobe from the frontal lobe, and the primary motor cortex from the primary somatosensory cortex. The lateral fissure separates the frontal and parietal lobes from the temporal lobe, and finally, the parieto-occipital sulcus establishes the boundary between the parietal and occipital lobes.



(a) Major brain parts (OSU Wexner Medical Center) (b) Brain lobes (Alexandra Almonacid E. (CC))

Figure 1.1: Brain Components: According to the performed activity, brain signals are generated in different brain lobes. Please note that the primary motor cortex is labeled as somatomotor cortex.

Broadly speaking, the right hemisphere controls the muscles located in the left side of the body, and the left hemisphere controls the right side of the body. The communication between both hemispheres is made through the corpus callosum, which sends motor and sensory information across both hemispheres. Despite the symmetry between the left and right hemispheres and the fact that for each person the brain works differently, there is a consensus in the functions of each lobe in the brain.

1.1.2 Brain Lobes

The main functions associated to each brain lobe are described below:

- **Frontal Lobe**: Is associated with motor development, problem solving, language, reasoning, planning, memory and is the core for personality and emotions. The primary motor cortex, located in this lobe, is associated to the motor activity, for planning and executing movements.

- **Parietal Lobe**: Plays an important role in the perception of tactile stimuli, because it is able to respond to sensory inputs, such as temperature, pressure and pain. It is also associated with movement coordination and language.

- **Occipital Lobe**: Is responsible for the perception and processing of visual stimuli, also related to spatial recognition.

- **Temporal Lobe**: Plays an important role in the perception of auditory and olfactory stimuli. It is also associated with coordination and balance, emotions, memory and face recognition.

Thus an event is jointly processed by several brain areas located in different lobes.

1.1.3 Neurons

The communication between and within brain regions is made through the electrical activity generated in the nervous system by billions of connected neurons. The *neuron* is the structural unit of the nervous system, differing from normal cells by its inability to reproduce.

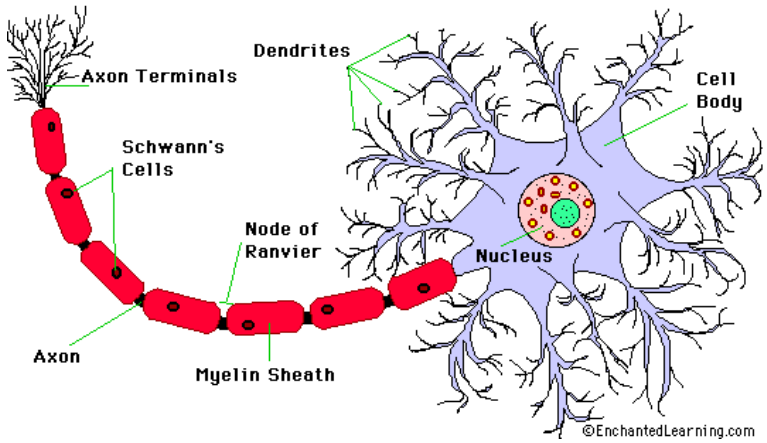


Figure 1.2: Biological Neuron: The diagram shows a single neuron. Dendrites attached to the cell body or soma (in blue) are responsible for receiving the electrical impulses from other neurons. The axon (in red) sends electrical impulses from the cell body to other neurons. The synapse (axon terminals) triggers these electrical impulses between two neurons.

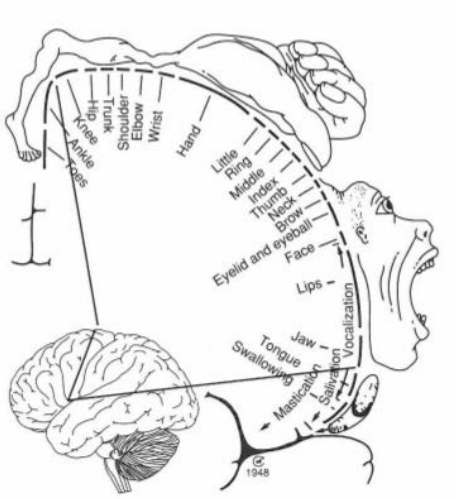
Figure 1.2 shows a basic neuron's representation and its parts consisting of a cell body named *soma*, extensions named *dendrites* and an *axon*, which carries the impulses to and from the cell body respectively. Neurons communicate each other passing electrical currents: this is called *synapse*. The communication takes place using neurotransmitters: chemical substances secreted by the neuron to excite or inhibit an electrical impulse that can propagate to other neurons. The *myelin* is generated by the *Schwann's cells* and is a layer of lipid and protein substances around

axons to increase the speed of the transmission of impulses. Myelin gaps are called *Nodes of Ranvier*.

1.1.4 Peripheral Nervous System

The elements of the central nervous system, *i.e.* the brain and the spinal cord, are connected to the rest of the body through the peripheral nervous system (PNS) to receive and send messages to coordinate body functions. The PNS consists on all the nerves that leave the CNS and branch off to reach all parts of the body. These nerves are the result of the union of fibers located in the brain or in the spinal cord, and depending on this, they can be mainly classify in two types: *Cranial Nerves* and *Spinal Nerves*. There are 12 pairs of cranial nerves emerging directly from the brain. They transmit motor and sensory information to the CNS from organs located mainly in the head and neck of the body. There are 31 pairs of spinal nerves, which emerge from the spinal cord and extend among the vertebrae. They control most of the muscles and glands.

The *primary motor cortex* is a fundamental part of the motor cortex in the frontal lobe. It is responsible for sending information needed to perform voluntary movements and receiving sensory information from other brain regions. The received sensory information is used as feedback to plan and execute movements. Furthermore, it contains a broad representation of the body anatomy, see Figure 1.3, where different body parts are partially controlled by overlapping cortex regions.



[Wolpaw and Wolpaw 2012, p. 15]

Figure 1.3: The motor homunculus: point-by-point correspondence of a body area with a specific area of the primary motor cortex.

1.2 Brain Imaging

Since the early 1990s, a variety of brain imaging techniques allowed to study brain processes in healthy subjects and to relate them with the body behavior. Brain imaging is usually divided in two categories: *structural imaging*, to visualize brain structures and study abnormalities that may affect the normal operation of the brain (*e.g.* tumors), and *functional imaging*, to understand the

functioning of the brain measuring the activity in brain regions while it performs a certain task. It is also used to diagnose diseases, such as epileptic seizures. These brain imaging categories are often used simultaneously to better understand the phenomenon.

To study the brain, it is necessary to record the activity produced by it. Currently, there are several systems to record these activities. They can be classified i) by the location used to capture the signal (invasive and non-invasive), a complete review is available in [Hochberg and Donoghue 2006], and ii) by the type of sensors used to record the brain signals. These sensors can detect electro-physiological, magnetic or metabolic signals.

1.2.1 Invasive Recording Methods

Invasive recordings are usually used in humans for evaluation reasons in patients suffering epilepsy or parkinson disease among others. The electrodes are implanted temporarily (5-14 days [Lal et al. 2005]) to evaluate the possibility of neurosurgical treatment for the extreme cases where patients are resistant to pharmacological treatment [Nair et al. 2008]. The main goal of this technique is to identify the brain tissue responsible for the seizures; to do this, the electrodes are implanted directly on or in the cortex through a surgical procedure, covering small regions of the brain, providing high temporal and spatial resolution. In general, BCI research using invasive recordings methods are based mostly on animals, since humans patients implanted with electrodes are available generally due to therapeutic approaches as described above. Invasive recordings can be subdivided into two major kinds of recordings according to the type of electrode used: i) electrocorticogram and ii) intracortical recording, both are described below.

1.2.1.1 Electrocorticography

The electrocorticography (ECoG) method, also called intracranial EEG (iEEG), works by placing an array of electrodes (on a subdural grid) in direct contact with the surface of the cortex, without penetrating the brain, see Figure 1.4(a). This is achieved through a craniotomy to remove part of the skull and expose the brain. ECoG electrodes are usually arranged in a matrix, e.g. a 8 x 8 matrix. The size of the matrix depends on the region that needs to be covered. The advantage of the signals obtained using ECoG is the accuracy and sensitivity of the recording because of the high spatial resolution, implying that the signal-to-noise ratio is also high due to its greater proximity to the neuronal activity. The problems with ECoG are all the complications related to surgery and the recovery time. In general ECoG is feasible only when people have to be implanted with electrodes to treat another condition, such as intractable drugs epilepsy, decreasing the possibilities for research. In addition ECoG signals are recorded from the surrounding of the electrodes making impossible to rearrange the electrodes without surgery. For further information refer to [Lal et al. 2005].

1.2.1.2 Intracortical Recordings

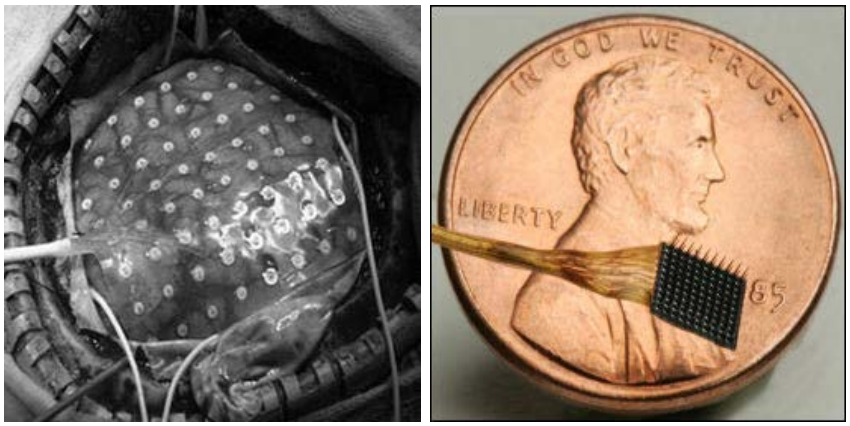
Microelectrodes allow intracellular or extracellular recordings depending on where they are placed, either within or close to the cell membrane of the neuron. Usually, BCI studies based on this type of recordings are limited to animals. In human patients, a pioneering study was performed to drive a cursor across a computer screen for communication purposes [Kennedy et al. 2000]. This study was achieved using a neurotrophic electrode (or cone electrode) specially created to record neural signals [Kennedy et al. 1992]. In the last years, a new brain implant system part of the BrainGate system was developed [Patil 2008]. BrainGate consists of a sensor, a decoder and an external device. The sensor is a 2mm-by-2mm array of 100 electrodes implanted

in the brain to record signals. The decoder translates the brain signals into commands for an external device. This device could be a computer or a robotic limb, for example. A recent study about intracortical recordings in humans beings can be found in [Mukamel and Fried 2012]. The signals recorded with these systems correspond to action potentials named spikes or local field potentials.

- **Neuronal Action Potentials (Spikes)**: Correspond to short electrical pulses emitted by single neurons when they send information down the axon. They can be recorded by placing an electrode very close to the cell body (soma) or the axon. This is a *single-units recording*. When the electrode is able to record more than one neuron the term used is *multi-unit recording*. If multiple neurons are recorded, the process for identifying the number of neurons around the electrode is called spike sorting and it is suitable when neurons have well defined spike characteristics [Lewicki 1998]. A "spike train" or "temporal coding" corresponds to spikes generated in series from a neuron spontaneously, or in response to a external stimulus.

- **Local Field Potentials**: Local Field Potentials (LFP) correspond to the electrical potential generated by the activity of neuronal ensembles using an extracellular electrode. Neurons in the neighborhood generate overlapping spikes causing a LFP amplitude larger than the spikes amplitude, because it is the summed activity information belonging to single neurons, which shows the synchronized activity between them. This is useful to understand the behavior of a set of neurons to a specific stimulus [Liu and Newsome 2006].

A comparison between ECoG and intracortical recordings is shown in Figure 1.4.



(a) ECoG sensors (b) Intracortical multielectrode array

Figure 1.4: Comparison between ECoG and Intracortical Sensors. In the left image (a), a section of the skull is exposed to place an array of electrodes (ECoG) on the cortex. In the right image (b), the intracortical electrode array is inserted in the cortex tissue. A coin is used to grasp the scale of the array.

1.2.2 Non-invasive Recording Methods

The non-invasive technologies to record brain activity are popular because they do not represent any risk for the user, such as, hemorrhages, infections or bioincompatibility compared to invasive methods, avoiding long stages at the hospital for recuperation and the trauma induced by surgery. These technologies are called non-invasive because they do not penetrate the skin physically, meaning that they are painless and safe for users. The main uses of non-invasive techniques is the study of the brain and their use in BCI research. Some of these alternatives include functional magnetic resonance Imaging (fMRI) , diffusion MRI (dMRI) and functional near infrared (fNIR) technology based on blood oxygenation, positron emission tomography (PET) and single photon emission computed tomography (SPECT) technologies based on nuclear medicine, Transcranial Doppler Ultrasound (TcD) , based on ultrasound techniques and finally magnetoencephalography (MEG) and electroencephalography (EEG) based on magnetic fields and brain electrical activity. These methods are described below, pointing the advantages and disadvantages of each one.

1.2.2.1 Functional Magnetic Resonance Imaging

It was discovered that changes in blood flow and blood oxygenation in the brain are closely linked to neural activity [Roy and Sherrington 1890]. From this knowledge, functional Magnetic Resonance Imaging (fMRI) was created to measure magnetic properties of the blood flow to detect active brain regions by the decrease in deoxyhemoglobin. Basically, when neurons go active the blood flow carries glucose and oxygen. Oxygen is transported by hemoglobin in the form of oxyhemoglobin, which has magnetic properties. These cells quickly degrade the local oxygen and increase the level of deoxyhemoglobin (hemoglobin without oxygen). The advantage of this technique is its high spatial resolution (1 mm^2), but it has a low temporal resolution (5-8 seconds) because the inflow of blood is not instantaneous. Also, fMRI is bulky and expensive because it needs superconducting magnets (Figure 1.5(a)).

1.2.2.2 Diffusion Magnetic Resonance Imaging

Diffusion Magnetic Resonance Imaging (dMRI) is a recent imaging technique introduced in the eighties [Bihan and Breton 1985]. It measures the motion or diffusion of water molecules, mainly water, in biological tissues. This allows the study of the brain connectivity, revealing microscopic details about the tissue architecture or brain abnormalities. This technique is currently used to study neurological disorders, especially for the detection of ischemic stroke caused by the loss of blood supply. The main advantage of dMRI compared to MRI is that using a set of images it is possible to compute an new image called *Apparent Diffusion Constant* (ADC), which reflects how the water protons spread out. For more details about dMRI see [Mori and Barker 1999].

1.2.2.3 Functional Near Infrared

Functional Near Infrared (fNIR) is an emerging technology similar to fMRI, but it eliminates many of its drawbacks being more affordable, portable and wearable, such as the kokoro gatari device [Jackson et al. 2013]. fNIR measures the absorption of near infrared light (NIR) in the brain to monitor changes in tissue oxygenation, having good spatial resolution (1 cm^2) and low temporal resolution when determining changes in oxygenation tissue (5-8 seconds). fNIR is a portable system because the sensor is like a "headband" placed in the prefrontal cortex allowing

users comfort. Despite its advantages, fMRI has better spatial resolution, and the ability to obtain an image of the entire brain [Bunce et al. 2006].

1.2.2.4 Positron Emission Tomography

Positron Emission Tomography (PET) is based on nuclear medicine. A short-lived radioactive tracer isotope is injected into the blood system. The tracer concentrates in tissues of interest in approximately one hour. Then the subject is placed into a scanner to obtain an image with a spatial resolution of few millimeters due to the gamma photons emitted. The gamma photons are generated by the collision of positrons with electrons. This technique is considered as non-invasive even if a radioactive substance is used, implying that is not possible to use it regularly in the same person. Also, it is very expensive and specialized personal is needed. More details about PET are available in [Ter-Pogossian 1982].

1.2.2.5 Single Photon Emission Computed Tomography

Single Photon Emission Computed Tomography (SPECT) is very similar to PET but instead of generating the gamma photons through a radioactive isotope emitting positrons, SPECT uses directly a radioisotope to emit photons. The images are obtained using a gamma-camera that rotates around the region of interest in the brain. SPECT has a lower spatial resolution (1 cm^2) than PET, which worsens visualization of deep structures, but it is cheaper and the tracer used (radioisotope) has a longer half-live than PET tracers. For further details about SPECT see [Holman and Tumeh 1990].

1.2.2.6 Transcranial Doppler Ultrasound

Transcranial Doppler Ultrasound (TcD) is a technique that uses ultrasound to measure the changes in cerebral flow velocity caused by neural activity, being useful to evaluate patients with a brain hemorrhage or a brain death. The spatial resolution of this technique can be affected by the thickness of the skull; in order to avoid to block the ultrasound wave through the skull, the temporal brain region is used to perform the recordings, specifically through the eyes, because the bones are thinner there. One of the great strengths of TcD is that it is robust to environmental noise being appealing for its possible use for BCI. More details about TcD can be found in [Myrden et al. 2011].

1.2.2.7 Magnetoencephalography

Magnetoencephalography (MEG) measures the magnetic fields generated by the electrical activity produced in the brain [Cohen 1972]. This technique allows to obtain maps of brain activity with a high spatial (millimeters) and temporal (milliseconds) resolution and with relatively low signal-to-noise ratio (SNR), but the equipment is bulky and expensive (Figure 1.5(b)). Also it is necessary to use a magnetically shielded room to reduce artifacts generated by electrical currents or ferromagnetic objects (such as an elevator). Despite its drawbacks MEG can be a useful tool for brain-computer interfaces. A study investigating the utility of magnetoencephalography in BCI can be found in [Mellinger et al. 2007].

1.2.2.8 Electroencephalography

Electroencephalography (EEG) is the most common method to record brain signals in humans because it is safe and easy to use (Figure 1.5(c)), it is affordable and because it has a high

temporal resolution (in the order of milliseconds). EEG can be roughly defined as the recording of electrical activity in different locations of the head. These recordings are made using electrodes positioned on the scalp with a conductive gel or paste. The electrodes can be of two types, "passive" or "active". Passive electrodes are metal disk connected to an amplifier by a cable. Active electrodes have inbuilt a preamplifier inside to make the electrode less sensitive to environmental noise and cable movements. Both types of electrodes need gel to operate, which does not allow their use for prolonged periods of time because the gel dries causing the system to stop functioning. For this reason, dry electrodes have been developed [Taheri et al. 1994]. Furthermore, wireless EEG systems have been developed in order to allow complete mobility for users. Despite a low spatial resolution (at best 1-2 cm) and a low signal-to-noise ratio, this solution presents many advantages for a wide use. In Section 1.3, EEG will be explained in more details. More information is available in [Niedermeyer and da Silva 2005].

(a) fMRI system (b) MEG system (c) EEG system

Figure 1.5: Non-invasive Brain Imaging examples: fMRI and MEG systems are bulky and expensive to maintain compared to EEG.

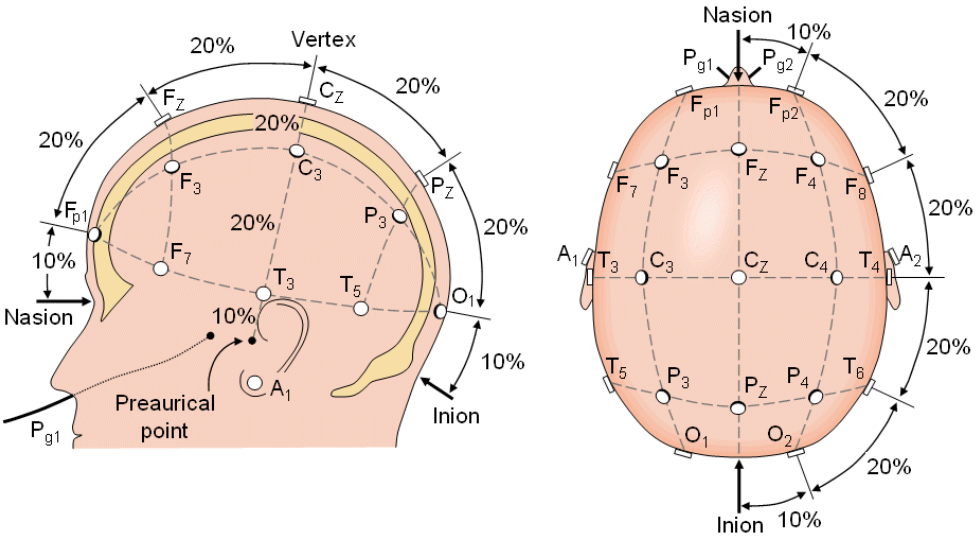
1.3 EEG Recordings

This thesis is focused on electroencephalography as a method to record brain signals, because it is safe and portable, allowing patients to use the system outside the laboratory at a reasonable price. It is also important to realize that invasive methods do not necessarily provide better information that non-invasive methods: they provide different information and the choice between them should be made taking into account the patients needs and the application to be developed [Wolpaw and McFarland 2004]. Besides, there is no concrete evidence that invasive methods have advantages in terms of information transfer rates [Tangermann et al. 2009; Wolpaw 2007], which leads to conclude that the same application made with invasive and non-invasive methods will achieve similar results.

There are different protocols to record brain signals using electroencephalography. The most popular protocols are explained below in addition to the brain signals that can be found in EEG.

1.3.1 10-20 EEG System

In the early recordings of the brain activity using EEG, there were no rules concerning the layout of the electrodes on the head. This situation led to the creation of a common layout to standardize the recordings, been the first of them —yet still used nowadays— the 10-20 system [Jasper 1958]. The name of the system comes from the inter-electrode distance, which corresponds to the 10% or 20% of the distance measured as a function of the skull size (see Figure 1.6). This protocol specifies the location of 21 electrodes on the surface of the scalp, ensuring that the electrodes are placed on the same areas, independently of the skull size. To do it so, two references on the head are needed: the first one is the **nasion**, corresponding to the delve at the top of the nose in line with the eyes. The second one is the **inion**, which is the most prominent point of the back of the head in the occipital bone. Taking into account the front-back distance generated by the nasion and inion points, and the right-left distance using as points the ears (A1-A2 in Figure 1.6) it is possible to find the electrode located in the intersection of the two lines (Cz). From this point, the 10% or 20% of one of the two distances is measured to place the next electrode. As an example, to choose a location for the electrodes C3 and C4, it is necessary to measure to the left and right the 20% of the right-left distance of the skull.



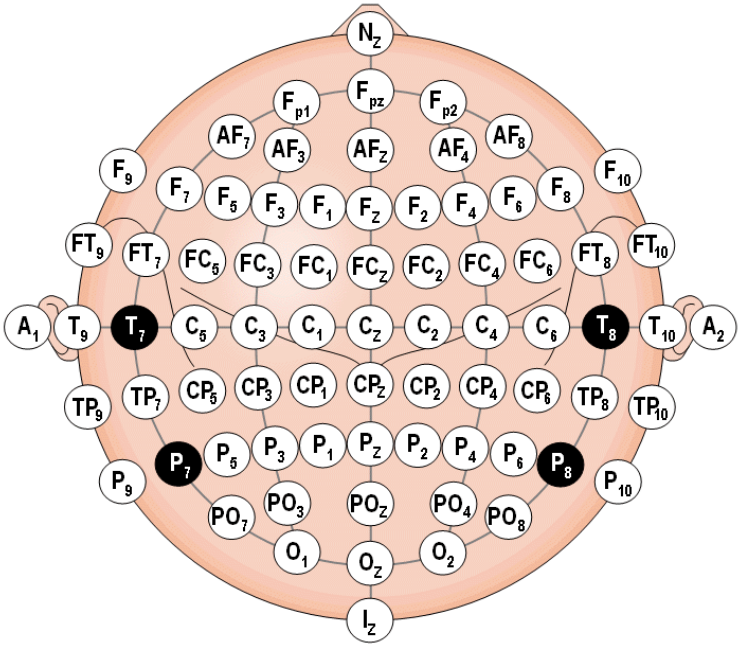
[Malmivuo and Robert Plonsey 1995]

Figure 1.6: Illustration of the international 10–20 EEG system. The left (resp. right) image shows the lateral (resp. top) view of the EEG electrodes layout. The nasion and inion reference points are illustrated in the images, as well as the distances percentages between electrodes locations.

The electrodes numbers are given by the cerebral hemisphere: odd numbers for the left one, even numbers for the right one and the letter "z" (meaning zero to avoid confusion with the letter "O") for the electrodes placed in between. Also the letters used to identify the electrodes correspond to locations on the head, such as, frontal (F), temporal (T), parietal (P), occipital (O), central (C) and combinations of them to indicate intermediate locations.

Furthermore, an extension to the original 10-20 system was proposed in [Chatrian et al. 1985] to increase the number of electrodes to 74 by placing electrodes at every 10% intermediate. This extended 10-20 system is known as the "10% system" or "10-10 system" and has been accepted by the International Federation of Societies for Electroencephalography and Clinical

Neurophysiology [Nuwer et al. 1998]. Nonetheless, the official 10-20 system now includes these positions and it is called the *extended 10-20 system*, being the most used EEG system up today (see Figure 1.7 for a better localization of the electrodes).



[Malmivuo and Robert Plonsey 1995]

Figure 1.7: Extended 10-20 EEG system. In this image the 74 EEG electrodes are shown. The electrodes arrangement is made considering 10% intermediate distances. Electrodes in black correspond to those with different names in the 10-20 system.

1.3.2 Brain Rhythms

The first recording of the electrical brain activity in human beings was made in 1924 by Hans Berger. Berger coined the term "electroencephalogram" in his paper publish in 1929 [Berger 1929]. His findings prompted other scientists to study the brain, identifying different brain oscillations present in EEG recordings. These rhythms or waves are the basis to understand how the brain works and open the possibility to use these kind of signals to develop applications that are not related to the normal brain pathway. These rhythms are divided by frequency bands as follows:

- **Delta Rhythm** (0.1-3.5 Hz): Is a normal brain wave in healthy subjects' EEG recordings. It appears mainly during a deep sleep and rarely can be experienced while awake. The delta rhythm can be found in mental states like dreamless sleep, trance and deep hypnosis. It is affected by drugs, depression, anxiety, alcohol and others circumstances or compounds that can affect directly or indirectly the sleep.

- **Theta Rhythm** (4-7.5 Hz): Detectable in the parietal-temporal area, it is present in REM sleep (sleep with rapid eye movements), drowsiness and in the transition from sleep to waking. It is also triggered by emotional stress, alcohol ingestion and fatigue [Matsuoka

1990]. The theta rhythm can be characterize in two types: the *hippocampal theta rhythm* observed in the hippocampus of mammals (*e.g.* cats) and the *cortical theta rhythms* usually recorded on the human scalp.

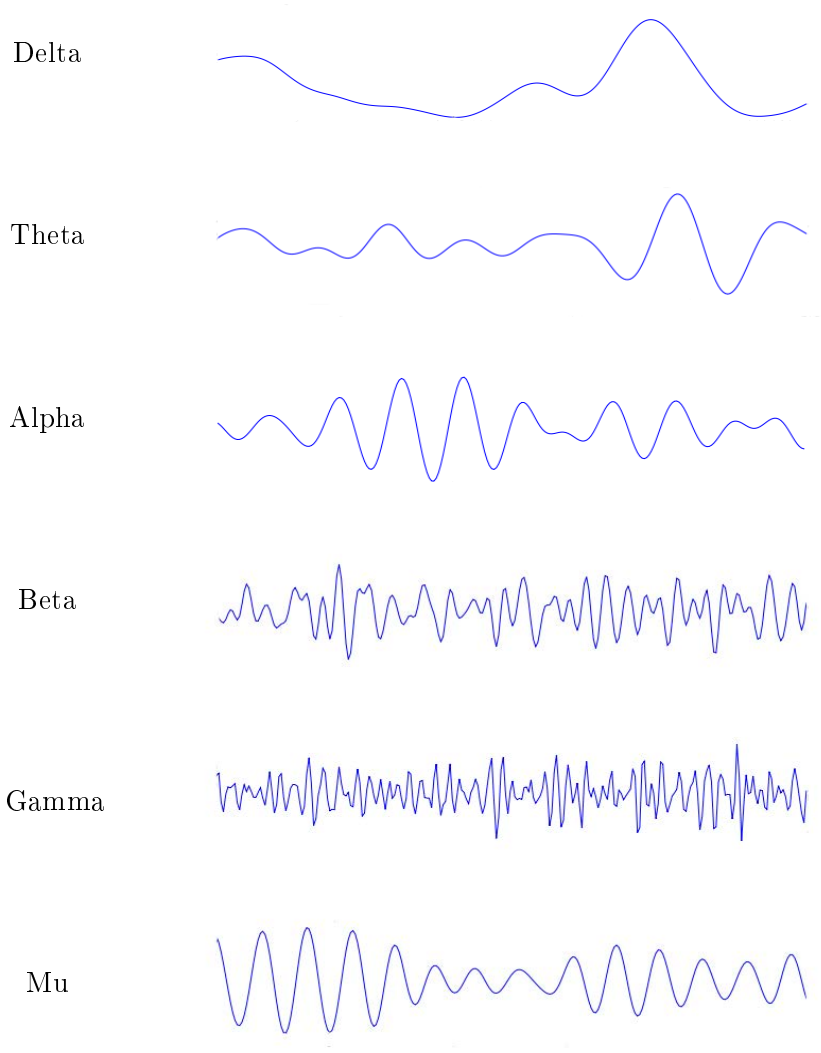
- **Alpha Rhythm** (8-13 Hz): It was defined in [Chatrian et al. 1974] by the terminology committee of the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) as a rhythm "occurring during wakefulness over the posterior region of the head, generally with higher voltage over the occipital areas". This definition is the most widely accepted by the research community. The alpha rhythm is best seen when the subject has the eyes closed in a state of physical relaxation over the occipital, parietal and posterior temporal regions.

- **Beta Rhythm** (14-30 Hz): It is detectable over the parietal and frontal lobes in almost every healthy adult. It is recorded when the person is awake and in full mental activity (alert), over the motor cortex. The beta rhythm is blocked in presence of muscle contractions or tactile stimulation (sensorimotor activity) and it is frequently mixed with a mu rhythm.

- **Gamma Rhythm** (Above 30 Hz): This rhythm is typically associated to perception and consciousness. This assumption is not fully proven by scientists, some of them even say that gamma waves could be artifacts of electromyographic activity [Whitham et al. 2007].

- **Mu Rhythm** (8-13 Hz): It is very similar to alpha rhythm, but it has different topography and is triggered by a different stimuli. The mu activity decreases during movements or preparation of movements and increases with relaxation. These phenomena are called "Event-Related Desynchronization" (ERD) [Niedermeyer and da Silva 2005] and "Event-Related Synchronization" (ERS) in [Pfurtscheller 1992].

Rhythms presented above are shown in Figure 1.8. They correspond to the main rhythms that can be found in EEG recordings. Its is also possible to find ultra-slow and ultra-fast activity in these kind of recordings, such as omega rhythms or rho rhythms. More information about these rhythms and others is available in [Niedermeyer and da Silva 2005].



Images by Hugo Gamboa (CC)

Figure 1.8: Main EEG Rhythms. The electrical activity generated by neurons can be divided in different brain rhythms. This can be useful because different brain states or body functions can be associated to these rhythms, allowing to isolate them for further analysis.

Chapter 2

Brain-Computer Interfaces

Contents

Brain-Computer interfaces (BCI) interpret the brain activity to produce commands on a computer or other devices. A BCI therefore allows its user, and especially a person with high mobility impairment, to interact with its environment only using its brain activity.

The applications based on BCI are mainly focused on improving the quality of life of people who find impossible to properly interact with their environment. Frequently, this situation is caused by degenerative diseases, brainstem strokes, cervical spinal cord injury, muscular dystrophy, amputation or others. The spectrum of applications includes systems for patients with locked-in syndrome (mainly for communication), for patients with limited motor ability (for rehabilitation purposes and control), and even video games, virtual reality and alert systems, etc. for a broad public.

Since the first brain computer interface project [Vidal 1973], many advances have been made in the field, developing different research directions. The current BCI types mainly differ on the type of recordings, which have been presented in the previous chapter, and on the physiological

phenomena used, such as evoked potentials or sensorimotor rhythms (presented below in this chapter).

Despite the progress achieved in recent years [Wolpaw and Wolpaw 2012], BCI systems are commonly not reliable enough to be used out of the laboratories. In fact, these systems do not perform well in all environments, they need frequent maintenance and the information rates are not very high. For these reasons, signal acquisition and translation algorithms are constantly improved to avoid errors, allowing the development of more robust systems and less tedious tasks for the users. This is achieved through the signal processing task of enhancing the features of brain signals and the improvement of the performance of the classifier that are the focus of this thesis. Obviously, signal processing techniques are not enough to improve the performance if the signals generated by the user are informationless. So, it is also important that the user is able to properly control his brain signals. Hence, new feedback techniques, user motivation techniques and new adaptive BCIs protocols are being developed. These user aspects may decrease the variability between signals, simplifying the pattern recognition task.

This chapter introduces the basic BCI principles, starting from its general architecture and its components. Then, the taxonomy that is currently used to classify BCIs is described. Also, the different physiological phenomena used to control BCI systems are summarized. Finally, we focus on ERP-based BCI and the challenging issues in the development of these kind of systems.

2.1 BCI Architecture

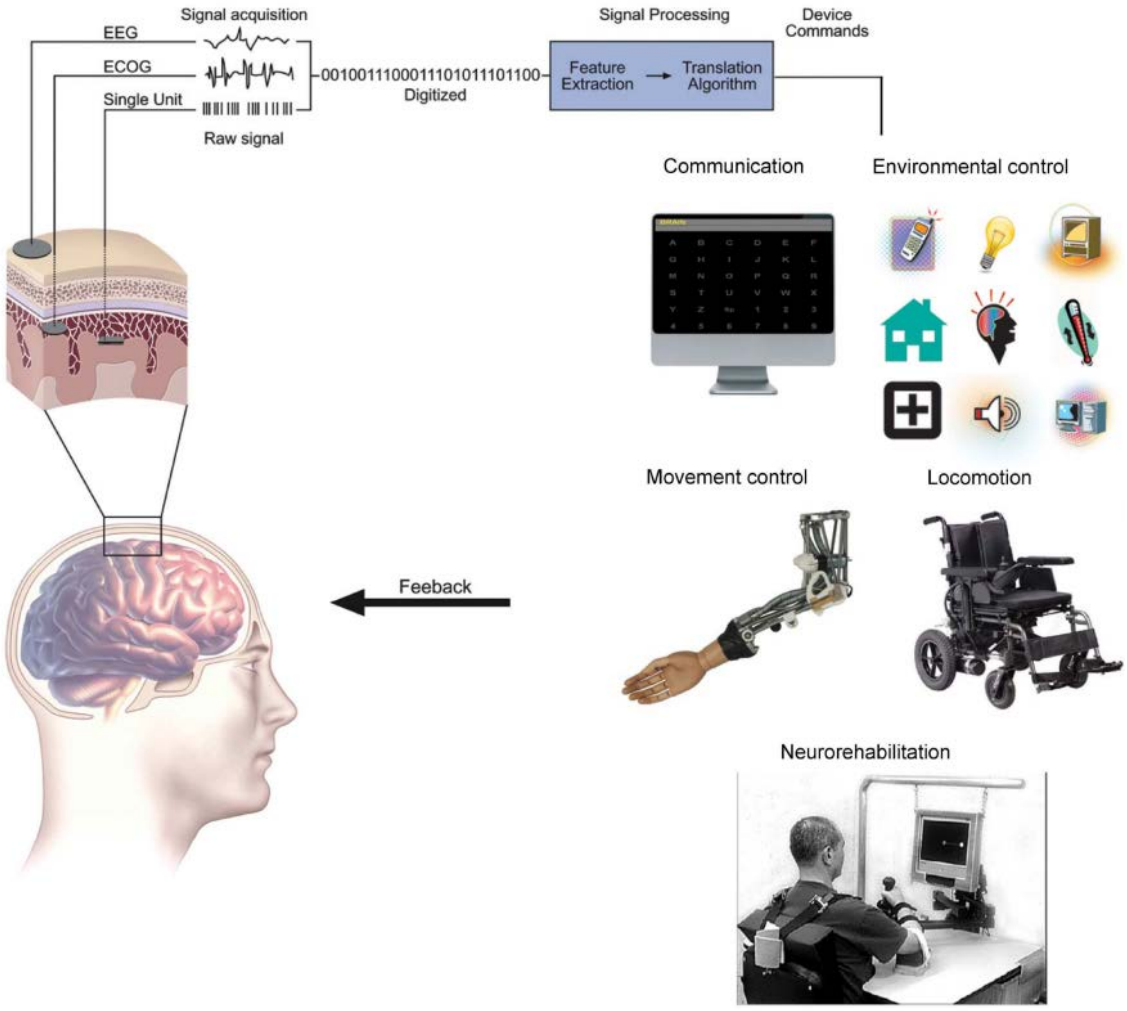
There are several alternatives to design a BCI communication or control system [Mason et al. 2007], but in [Mason and Birch 2003] a general functional model of BCI Systems was presented to be used as a framework to describe them and simplify the comparison between them. There are three main steps that are always present in any BCI design [Wolpaw et al. 2002; Mak and Wolpaw 2009]. These steps correspond to *signal acquisition*, *signal processing* and *device output*. Furthermore, a feedback step can be added to provide information to the subject about the BCI system (brain signals, classifiers outputs, current action). Figure 2.1 was presented in [Mak and Wolpaw 2009], and shows components of a BCI system.

2.1.1 Signal Acquisition

The brain signal is acquired by the recording electrodes, amplified and digitized to be used in the BCI system. First of all, the target type of neural activity must be defined. It can be spontaneous signals generated in the brain (*e.g.* EEG rhythms) or signals generated by an external stimulus, such as a visual or auditory stimulus (*e.g.* flashing letters). After this, it is necessary to choose the best method to record the signal. As shown in section 1.2, these methods can be invasive or non-invasive. Due to ethical reasons, non-invasive recordings are the most used in human beings. In particular, it is difficult for patients with locked-in syndrome to express their consent. All these choices are made depending on the application or the study to be performed.

2.1.2 Signal Processing

Signal processing is the analysis of complex signals to extract and interpret the useful information contained in it. The signal processing is composed by the *feature extraction* and the *feature translation*. This step is crucial for successfully classify the input data and ensure the proper functioning of the system.



[Mak and Wolpaw 2009]

Figure 2.1: BCI architecture: i) the signal acquisition step, in this case either EEG, ECOG or single unit recordings ii) the signal processing step comprising the feature extraction and the translation algorithm iii) the device action for communication, environmental or movement control, locomotion or neurorehabilitation and finally, iv) the feedback step, noted by the black arrow.

2.1.2.1 Feature Extraction

The idea of feature extraction is to process the recorded digitized raw signal to extract "features" that should contain relevant information needed by the classifier. This step may contain a feature selection because all features are not equally significant, and using too many features can dramatically increase the complexity of the classifier and even degrade the performance. The main goal is to find a reduced representation of the input signal to improve the classification accuracy or to increase the processing speed. The recorded brain activity can be analyzed in the time domain, in the frequency domain or in both domains to observe changes in amplitude, voltage, or others. The output of this stage is referred as a feature vector.

2.1.2.2 Feature Translation (classification)

Feature translation actually translates the feature vector into commands to control a device. The main idea is to classify the feature vector among different classes. Usually, the output of this stage is a binary choice (two-class problem). It is also possible to distinguish an answer from many options (multi-class problem). However, the larger the number of classes, the worser the system performance. The success of the translation depends on the capacity of the subject to trigger or control brain signals, the quality of the feature extraction and the accuracy of classifier. The classifier parameters can be obtained by training on a set of supervised examples.

2.1.3 Device Action/Output

Every class in the previous stage is associated to a control action. These actions are used to operate different devices. The number of devices that can be controlled through BCI is unlimited, but nowadays the most typical BCI applications use devices such as computers for communication or gear control, wheelchair for mobility, or robotic limbs to be used in rehabilitation or to accomplish motion actions. There are even some applications for people unable to move their arms, which send signals to the nerves without passing through the spinal cord, allowing to the subject to use their own arm.

2.1.4 Feedback

In some cases, information from the system can be used to improve it or to detect execution errors [Chadwick et al. 2011; Hwang et al. 2009]. This step is not always used in BCI applications, but it is specially useful in on-line sessions, because errors in the operation of the system are critical and informing the user about the progress of the task is important. Nowadays, researchers are studying the possibility of using feedback as a tool to teach and develop the BCI skill in users [Lotte et al. 2013], in order to correctly perform the desired mental commands. This is based on the assumption that, if the recordings are not reliable, the identification of the mental commands may not be properly identified even by advanced signal processing techniques.

2.2 Taxonomy of BCI

In the literature, there are several taxonomies for Brain-Computer Interfaces. This section discusses the most important categories:

- How the brain signals are induced: *Dependently* or *Independently*.
- The nature of the brain signals: *Endogenous* or *Exogenous*.
- According to the processing mode of the data: *Synchronous* or *Asynchronous*
- By the combination of brain signals: *hybrid* BCI.

2.2.1 Dependent and Independent BCI

BCI can be classified into two categories according to the process that generates the brain signals [Wolpaw et al. 2002]. If the use of the peripheral system (even partially) is needed, as for example eye movements to choose a letter to be communicate, the BCI is called *Dependent*, because the triggered brain signals depends on the eye gaze shifting. This implies that this kind of system needs some muscles to operate properly. *Independent* BCI operate with signals that are completely independent of muscles or sensory activities. Using the same example than before,

instead of eye movements to choose a letter to be spelled it would be necessary to "think" in the letter to trigger a brain signal to reveal it. This kind of system is particularly useful when all muscular activity is lost, for example for people with locked-in syndrome.

2.2.2 Endogenous and Exogenous

As shown in section 2.1.1 the brain signals used in BCI can be spontaneous or triggered by an external stimulus. This lead us to a new classification for BCI [Wolpaw et al. 2000]. *Endogenous* brain activity is voluntarily generated by the subject, without the need of an external stimulus, for example imaging some body movements. *Exogenous* brain activity is evoked by an external stimuli, either visual, auditory or somatosensory, specifically to create patterns that are easily detectable. Usually these systems are based on the choice between several options (see for example [Picton 1992; Huettel and McCarthy 2004]). The advantage of exogenous BCI is that it practically does not require user training, contrary to endogenous BCI, yet the last one gives more freedom when developing applications. As an example, an endogenous BCI should allow a subject to move a cursor in any direction in a two dimensional space, while an exogenous BCI may allow a subject to make a choice between different symbols presented on a screen. Using all categories presented so far, endogenous BCI can only be independent and exogenous BCI can be either independent or dependent. This condition is determined by the application. A pilot study comparing endogenous and exogenous BCI for oddball paradigm is presented in [Marchetti et al. 2012].

2.2.3 Synchronous and Asynchronous BCI

BCI can also be classified according to the timing used to send messages or commands to the system, *i.e.* as *Asynchronous* or *Synchronous* BCI. The first paradigm (also called self-paced paradigm) allows the subject to send information freely. For example, in motor imagery tasks (*i.e.* when the user mentally simulates an action) there is no knowledge about when the mental process occurs [Nooh et al. 2011]. In synchronous or cue-based BCI, the mental task must be executed in a predefined time window to send messages or commands according to external cues. The time windows are predefined by the system and are preceded by a stimulus [Boostani et al. 2007]. The advantages of synchronous BCI systems compared to asynchronous systems are a simpler design and performance evaluation, because it is not necessary to evaluate the brain signal at each time. Also they allow movements (artifacts) when the signals are not analyzed. Even if they are a less natural way to interact for the subject, the great majority of current BCI systems are synchronous [Tsui et al. 2009].

2.2.4 Hybrid BCI

An hybrid BCI system is composed either by the combination of different EEG brain signals (*e.g.* event-related potentials and visual-evoked potentials, both explained later in this chapter), the combination of different BCI systems (*e.g.* motor imagery and visual attention) or the combination of EEG signals with others biosignals (*e.g.* EMG). The basic idea is to apply ensemble methods to the BCI domain to improve the communication between the subject and its environment. Hybrid BCIs have been used to improve performance [Leeb et al. 2011], to develop switches/selectors (*e.g.* on/off) and to increase the degrees of freedom in controlling devices [Pfurtscheller et al. 2010].

2.3 Physiological Phenomena

The main goal of BCI is to interpret intentions through monitoring brain signals. For this purpose, several varieties of neurological phenomena are used to decode these intentions and serve to control BCI systems, enabling the development of applications for users. These neurological phenomena are characterized by a specific latency, amplitude, frequency and voltage which are related to the stimulus (external or internal) that caused it. They can be separated in two categories. The first one is called *exogenous neurological phenomena*, being the visual evoked potentials the best known in this group. The second category consists in *endogenous neurological phenomena* such as sensorimotor rhythms, slow cortical potentials and P300 event-related potentials. All these phenomena are described below.

2.3.1 Visual Evoked Potentials

Visual Evoked Potentials (VEPs) are oscillatory waves in the electroencephalogram (EEG), located in the occipital part of the cortex, in response to the stimulation of the retina receptors. This visual stimulus is modulated at a certain frequency or rate. Usually, it corresponds to flashing lights (*e.g.* LEDs, Xe-lights, fluorescent lights), single graphics (*e.g.* arrow, rectangle, square) that appear and disappear from a screen (onset/offset) and pattern reversal characterized by multiple choices that are alternated in a computer screen (*e.g.* black and white checkerboard) [Odom et al. 2004]. To record VEPs, a number of electrodes or sensors are placed in the occipital region and the subject is seated in front of a screen to stare the stimulus repeatedly presented. VEPs are used to clinically test the functional status of the visual system, because it only depends on the stimulus in the retina. These potentials are extensively used in BCIs systems [Wang et al. 2008], because they allow the subject to select among several commands depending on the application, a user training is not needed and the information transfer rate is high. According to the frequency of the stimulation VEPs can be classified as follows.

1. **Transient VEPs (TVEP)** When the twinkling frequency of the stimuli is below 6 Hz, the VEPs are known as TVEPs. This kind of VEP are rarely used in BCI because it takes too long to be detected, but maybe they can be useful for subjects who experience headaches, dizziness or drowsiness, due to looking at high-speed blinking visual stimuli [Yoshimura and Itakura 2011].

2. **Steady-state VEPs (SSVEP)** The only difference with TVEPs is that the stimulus changes at frequencies above 6 Hz. SSVEP amplitude is different for dissimilar stimulation frequencies or different subjects, obtaining a maximum average amplitude around a stimulation frequency of 15 Hz [Pastor et al. 2003]. SSVEPs are very popular to develop BCI system, because they do not required a user training, despite the fact that the subject has to watch steadily the stimuli in the screen. The SSVEP-based BCI performance is influenced by the nature of the visual stimuli in terms of information transfer rate. According to [Zhu et al. 2010], the most popular stimulations used in BCI for generate VEPs are checkerboards, rectangles and LEDs. The best is the LED stimulation with 42 bits/min in terms of transfer rates. According to the specific stimulus sequence modulation SSVEP can be classified into three categories [Bin et al. 2009] as follows:

   (a) **Time modulated VEP (t-VEP)**: All targets are flashed at the same frequency, but the flash sequences are mutually independent, meaning that there is no overlapping

(in time) or synchronization between them. Usually to obtain a better classification rates, repetitions of the same target are averaged to enhanced the response.

(b) **Frequency modulated (f-VEP)**: All targets are flashed at different frequencies, therefore classification (target identification) is usually made through a power spectral analysis. One important advantage of f-VEP is its high information transfers rate, 30-60 bits/min.

(c) **Pseudo-random code modulated (c-VEP)**: The flash sequences are determined by a pseudo-random binary sequence. The target identification is generally made computing the correlation between a template obtained during the training stage and the data being tested. c-VEP provide the highest communication speed of the three categories ($>$ 100 bits/min), but a user training is required.

Despite the advantages offered by these kind of systems, some authors describe them as muscle dependent, because the muscles of the eyes are needed for selective attention. For this reason, VEPs are not suitable for patients in advanced states of ALS or with uncontrollable eye or neck movements.

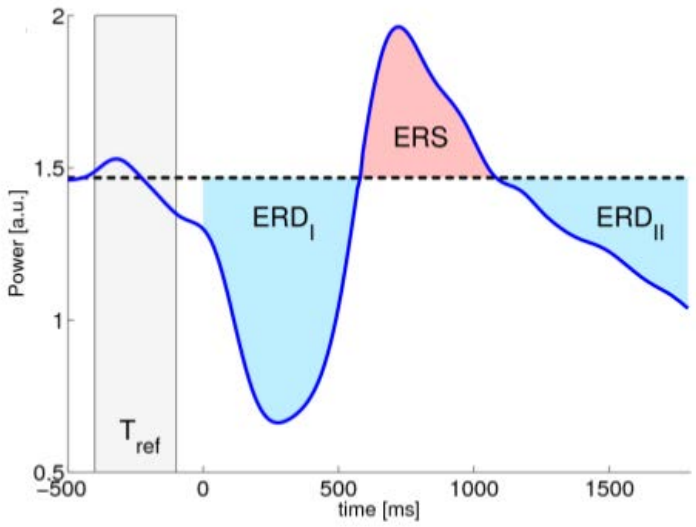
2.3.2 Sensorimotor Rhythms

Sensorimotor Rhythms (SMR) comprises mu and beta rhythms presented in section 1.3.2. The amplitude of these rhythms increases when the brain sensory-motor areas are inactive, while the amplitude decreases when they are active. Essentially, SMR are related to any motor task, without requiring any actual movement [Pfurtscheller et al. 1997], making possible the design of endogenous BCIs. To use these rhythms to control BCIs, a user training is required [Hwang et al. 2009]. This training can be done by allowing the subject to monitor the real-time brain activation maps (quantified in time and space) on his cortex. Nevertheless, a large number of people have trouble controlling these signals. For further information about the relationship between beta and mu rhythms see [Krusienski et al. 2007].

Sensorimotor rhythms can be divided into two different phenomena:

1. **Event-Related Desynchronization (ERD)**: Reflects a decrease of oscillatory brain activity caused by a decrease in the synchronization of neurons in response to an event [Lemm et al. 2009]. For example, the beta and the mu rhythms decrease in presence of activities in motor or sensory areas when these areas are activated. A study in humans demonstrates that mu rhythms present an ERD over the cortical hand area during discrete finger movements [Pfurtscheller and Neuper 1994].

2. **Event-Related Synchronization (ERS)**: The term was first used by Pfurtscheller in 1992 [Pfurtscheller 1992] and corresponds to an increase in the brain activity (synchronization of neurons) caused by active brain areas. For example, in the study presented before [Pfurtscheller and Neuper 1994], the mu rhythms presented an ERS over the hand area with foot movements. Also it is possible to found beta rhythms ERS in the first second after finishing an intentional movement, while mu rhythms still exhibit a ERD pattern [Pfurtscheller and da Silva 1999].



[Lemm et al. 2009]

Figure 2.2: Sensorimotor Rhythms: Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS). The deviation of the event-related dynamics (blue solid line) is measured from a constant baseline that corresponds to the average alpha power, computed during relaxation in the reference period T_{ref} (black dashed line). Blue regions correspond to event-related de-synchronization periods and the red region corresponds to an event-related synchronization.

A comprehensive review of event-related synchronization can be found in [Pfurtscheller et al. 1996].

2.3.3 Slow Cortical Potentials

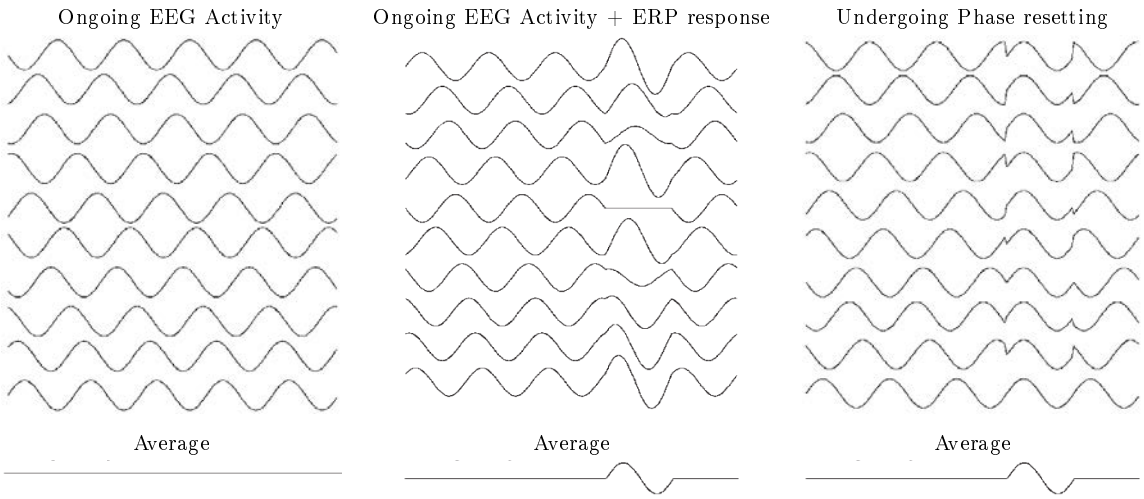
Slow cortical potentials (SCP) are defined in [Birbaumer 1999] as "negative or positive polarizations of the electroencephalogram or magnetic field changes in the magnetoencephalogram (MEG) that last from 300 ms to several seconds". SCPs are associated with changes in the cortical activity, which enable healthy or paralyzed subjects to control these signals, after training, to command BCIs devices [Kübler et al. 1998]. Generally, the shifts in these low-frequency components are used for communication purposes, for example to move a cursor in a computer screen to select targets. The self-regulation of SCPs is usually learned by paralyzed patients using a "thought-translation device" [Birbaumer et al. 2003], through visual-auditory feedback and positive reinforcement of SCPs. After the training, the system can be used for spelling purposes by patients with severe motor disabilities. The principal advantage of SCPs is that they are present in the brain activity of each person, being extensively tested in patients with ALS [Iversen et al. 2008; Kübler et al. 1999], obtaining accuracy rates between 65 and 85 percent. The problem with SCP is the relatively slow information transfer rates (5-12 bits/min).

2.3.4 Event-Related Potentials

Event-Related Potentials (ERPs) are electrical changes in electroencephalographic recordings that are time-locked with sensory or cognitive events. The first recording in humans were published in 1939 [Davis et al. 1939]. From there, the researchers focused on investigating the cognitive processes of the brain, until the P3 wave was discovered by Sutton in 1965 [Sutton et al.

1965] revolutionizing the study of ERPs. The ERPs responses are different for each person in latency and amplitude. They can even change for the same person according to his/her mental state, for example by tiredness or sleepiness.

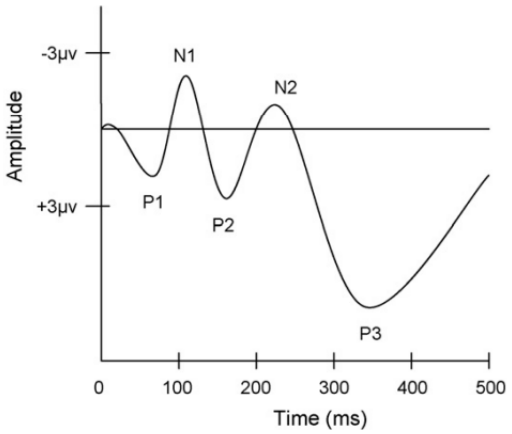
ERPs are very useful to develop BCIs, but it is still unclear why the ERPs are generated. Two hypothesis are the most suitable. The first one establishes that ERPs are generated due to an event-related activation of neural assemblies with the stimulus (evoked response) [Fell et al. 2004], meaning that the ERP is added to the ongoing EEG. The second hypothesis states that ERPs are generated by "phase resetting", meaning that there is an enhanced alignment of the EEG activity, caused by a stimulus, resetting the phase of the ongoing oscillatory activity [Klimesch et al. 2007], which explains why the ERP response is not visible in single trials (see Figure 2.3).



[Sauseng et al. 2007]

Figure 2.3: Hypothetical models of Event-Related Potentials: the ongoing single-trial EEG signals is presented in the left side, the evoked response model is shown in the center and the phase-resetting model is in the right side.

ERPs are not visible in single trials in the EEG because they reflect the ongoing brain activity (neural oscillations) of many processes, making their identification difficult. However, ERPs become more visible averaging many responses to the same stimulus, decreasing the interference of activities not time-locked to the stimulus and highlighting the ERP response. Despite all, scientists were able to identify the ERP waveform components (see Figure 2.4). These components are peaks whose names are designated by their polarity — Negative (N) or positive (P) — and their approximate latency in ms. ERPs can be evoked by visual, auditory or somatosensory stimuli. They have been used to study visual awareness [Koivisto and Revonsuo 2010], attention [Rauss et al. 2011], perception of color and form [Rentzeperis et al. 2012], facial expression processing [Luo et al. 2010] and affective processing [Olofsson et al. 2008] proving that ERPs are a useful tool to track the temporal brain dynamics and understand the neural processing [Saavedra and Bougrain 2012]. Follows a brief description of some of the most studied visual ERPs:



Wikipedia/Event-related potential (CC)

Figure 2.4: Event-Related Potentials. The main ERP components are shown. Their names are designated by their latencies and their polarities, negative for N1 and N2 and positive for P1, P2 and P3.

- **C1** is the first major visual evoked potential (VEP) component generated in the primary visual area (striate cortex; area 17) with an onset around 50 ms and a peak around 90 ms [Russo et al. 2002; Luck 2005; Rauss et al. 2011]. The C1 component is sensitive in amplitude and polarity to the spatial location of the stimulus in the visual field. C1 has a negative polarity when the stimulus is in the upper visual field and a positive one when is presented in the lower visual field [Rauss et al. 2011]. For this reason, the standard nomenclature (N or P) about polarity is not applied to C1. This component is usually used to study attention and learning processes [Rauss et al. 2011] and is sensitive to contrast [Luck 2005].

- **P1**, also called **P100**, is the first positive component with an onset within 60-80 ms and a peak normally observed around 100-130 ms in the lateral occipital cortex [Luck 2005; Koivisto and Revonsuo 2010; Luo et al. 2010]. The early phase of P1 overlaps with C1, making difficult to identify its onset in time. It is a mandatory response elicited by a visual stimulus (exogenous response) because they are influenced by external stimulus, such as luminance [Luck 2005]. P1's amplitude is larger for unpleasant target stimuli than pleasant ones, showing a relation between the P1 component and the affective processing [Olofsson et al. 2008]. A correlation between P1 and face expression has also been found [Luo et al. 2010].

- **N1** is usually called **visual N100** to differentiate it from its auditory version. This potential responds to manipulations of attention. This component has a peak around 100-150 ms in the parietal cortex when influenced by spatial attention (attended vs. unattended locations), and a latter peak around 150-200 ms in the lateral occipital cortex sensitive to a discrimination process [Luck 2005]. N1 is also sensitive to lexical processing, having an increased amplitude in response to pleasant words during silent reading [Kissler et al. 2009].

- **N2** (or N200) can be decomposed into two VEP subcomponents: the N2b which is evoked during conscious stimulus attention (visual discrimination tasks) and the N2c which arises during classification tasks [Patel and Azzam 2005]. Furthermore, some stimuli may evoke a N2pc ("pc" stands for "posterior–contralateral" cortex), produced when the attention is

focus on one object while ignoring others [Koivisto and Revonsuo 2010].

2.3.4.1 P300 Event-Related Potential

In 1965, Sutton and Braren [Sutton et al. 1965] reported a large positive component that peaked around 300 milliseconds after the stimulus onset. This happens when the attended stimulus could not be predicted by the subject. The stimulus used was visual or auditory, and the peak was much smaller when the sequence containing the stimulus was predictable. This component was labeled P3 or P300, because it is a positive deflection with a peak around 300 ms after the onset of the target stimulus. The delay depends on the experimental conditions, the subject, the electrode positions, and others. The P300 event-related brain potential can be separated in two subcomponents, labeled P3a (220-280 msec.) and P3b (310-380 msec.) [Squires et al. 1975]. The P300 and the P3b are equivalent. Instead, the P3a is an additional positive component with a smaller amplitude that occurs earlier than P3b and it is produced by user's attention on changes occurring in the environment [Näätänen 1992] and by the evaluation of the stimulus presented [Polich 2007]. Generally, in order to study the P3a wave, a variation of the oddball paradigm is employed using three stimulus, where the extra stimulus is a "distractor" (or deviant stimulus) that does not require a response from the subject. In [Polich 2007], a P3a and P3b integrative theory is proposed, suggesting that P3a is produced by the frontal lobe as the result of early attention process and the P3b is the result of the signal transmitted to the temporal/parietal lobes. Polich argues that every P300 is composed by a P3a and a P3b, but the communication between the frontal and temporal/parietal lobes affects their morphology depending on task conditions and the stimulus.

As the P300 wave is elicited when the target stimulus is infrequent compared to the standard stimulus, the simplest task to study P300 is using the "Oddball Paradigm" developed to evoke activation in brain regions to generate a P300 component [Picton 1992; Huettel and McCarthy 2004]. This strategy consists in presenting two types of stimuli, one of them being common and the other occasional (or rare). The user's task is to count the rare stimulus among the most common stimuli. The stimuli are typically presented visually or auditorily (usually only one sensory modality is used at a time). Although, this paradigm is very efficient to stimulate the subject, there is no clarity about the cognitive process to assess P300 responses. The P300 response is influenced by internal subject conditions, such as fatigue, temperature, sleep, disease, anxiety, etc. Also, it is influenced by external conditions such as drugs, caffeine, alcohol, etc. [Polich and Kok 1995; Nijboer et al. 2010]. These factors affect the amplitude and latency of the P300 potential, being in some cases useful for improving the implementation of applications based on them. In general, P300 components are impossible to visualize in single trial because they are ten times smaller than the ongoing EEG activity. For these reason, the process is repeated and the responses are averaged to make ERPs visible. In one hand, the responses associated with target stimuli are averaged and in the other the responses associated with non-target stimuli are averaged. Figure 2.5 shows the average difference in amplitude between target and non-target signals.

ERP based applications are usually developed using a repetition of a random sequence (called trial) including target and non-targets stimuli, to enhances the difference between responses of each type and to reduce noise contained in recorded EEG signal by averaging responses. The number of trials that are averaged varies from study to study, but a research concerning the number of trials and the different ways of averaging can be found in [Cohen and Polich 1997]. Averaging enhanced the P300 wave increasing the SNR, but reduces the speed of information

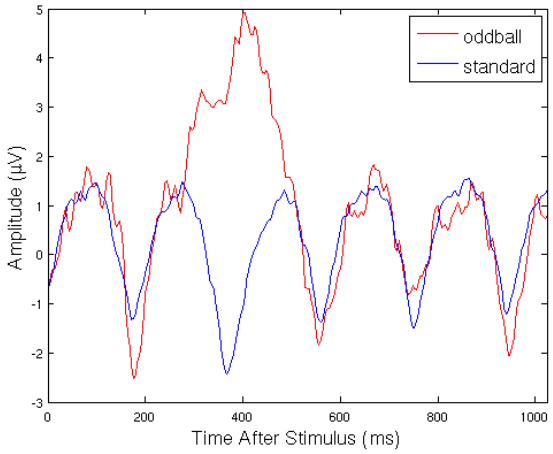


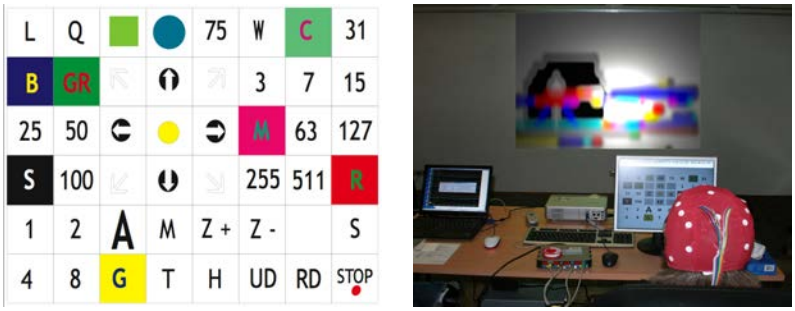
Figure 2.5: Oddball and Standard Grand Averages. The image shows the difference between the averaged signals for oddball (in red) and standard (in blue) stimuli. The grand averages were computed averaging all the (P300) responses corresponding to the oddball stimuli and all the (non-P300) responses corresponding to the standard stimuli. The P300 corresponds to the difference between signals (200-500 ms). The signals correspond to subject LAC of the database [Ledesma-Ramirez et al. 2010] presented in section A.2.1.

transfer rate causing fatigue in some patients, because the task becomes very repetitive. An important issue is to reduce the number of trials to obtain high communication bit-rates and provide a more user friendly system.

2.4 BCI based on P300 Event-Related Potentials

P300s are very popular to develop BCI applications. The main reason is that a classifier can be trained only in a few minutes to identify P300s, which are present in most people including patients. Also, the P300 is time-locked to the stimulus onset, making them ideal to spell characters for communication purposes, since a large number commands can be implemented. Although the P300-based BCI task requires a cognitive process, it does not cause much fatigue and responses are relatively stable between sessions [Sellers and Donchin 2006], allowing daily use for communication and control [Vaughan et al. 2006]. Another advantage is that these systems are suitable for patients in advanced stages of degenerative diseases such as ALS or locked-in syndrome, because they are not significantly affected by the subject gazes. Finally, perhaps the most important reason of using this type of potentials is because almost everybody is able to elicit a P300 wave [Guger et al. 2009].

Almost all applications in BCI based on ERP are implemented using visual stimuli corresponding to a matrix of letters, numbers, or other symbols, being the speller presented by Farwell in 1988 [Farwell and Donchin 1988] the first one and the most used. Similar systems have been developed as an internet browser [Mugler et al. 2008] or the Brain Painting application developed for ALS patient entertainment [Münssinger et al. 2010] (see Figure 2.6).



[Münssinger et al. 2010]

Figure 2.6: Brain Painting. In the left image, the selection matrix of the P300-Brain painting. In the right image, the P300-Brain painting being used by a user.

Currently, others applications for healthy users based on P300 have been developed, for example lie detectors (guilty knowledge test) [Abootalebi et al. 2006, 2009], virtual reality applications such as driving [Bayliss and Ballard 2000], a game controller [Finke et al. 2009] or a music composer [Hamadicharef et al. 2010]. Also, remote control applications have been developed, such as a brain-controlled helicopter by puzzlebox [16], between others.

2.4.1 P300 Speller

The P300 speller is one of the best-known and powerful BCI system based on the non-invasive EEG measured from the subject's scalp. This application allows the user to communicate through spelling words, writing text choosing one letter at time from a flashing matrix. When a highlighted row or column contains the desired letter by the user, a P300 wave is generated in the brain. This P300 wave can be detected to translate the user's intention. The P300 Speller has potentially a strong impact for patients with motor disabilities because it requires no user training, allows a large number of commands and it has a reasonable speed due to the oddball paradigm.

The P300 speller implementation was originally made by Farwell [Farwell and Donchin 1988] in 1988. Later on Donchin described the paradigm in [Donchin et al. 2000], which is considered as the standard for this type of systems. The original speller is composed by 36 characters arranged in a 6x6-matrix (see Figure 2.7). Thus, to spell one character it is necessary to randomly flash 6 columns and 6 rows (12 flashes), while the user pays attention to the desired letter. When the letter is highlighted, a P300 is generated by the brain.

16. `http://brainstorms.puzzlebox.info`

Figure 2.7: A 6x6 P300 speller. The application highlights in a random order columns and rows. The user looks at the desired letter. When the letter is highlighted, a P300 is generated. Detecting this specific transient event twice to identify the right column and the right row, it is possible to know which letter is requested.

The 12 flashes set (6 rows and 6 columns) correspond to a sequence (or a trial). At the end of this sequence, it is possible to classify the responses to identify the row and the column, thus the letter, that triggered a P300 each when they have been highlighted. The user communicates letter by letter to write a word. A letter spelled by the user consists of a particular row and a particular column of the matrix, implying that only 2 stimuli are relevant among all, and therefore these two flashes are considered by the user as rare stimuli (16,7% of the intensifications) making possible the oddball paradigm. The period between the onset of a flash and the onset of the next flash is called the interstimulus interval (ISI) and serves to regulate the speed of the intensifications. For example, an ISI of 62.5 ms corresponds to a faster flashing compared to an ISI of 312.5 ms. For more information about the influence of the ISI in the P300 speller see [Nam et al. 2009]. The subject sat between 50 cm and 80 cm from the displaying matrix and its task is to maintain a mental count of the number of occurrences of flashes comprising the desired letter to maintain attention.

The sequence is repeated several times in a random order to average responses induced by each row/column with the aim to reduce the noise and bring out the potentials to make easier the classification task and improve the accuracy of the system. Usually, the trials repetition is called intensification sequence, meaning that one intensification sequence is composed by 12 flashes. The number of sequences used may vary between 1 and 15 or more when it is really difficult to find the potential in the user's signals. Figure 2.8 shows the improvement in the letter detection using different number of sequences. The accuracy approaches 100% using a large number of sequences. One intensification sequence corresponds to a single-trial detection. The algorithms proposed in this thesis aim at improving this accuracy.

Modifications to the P300 speller paradigm has been proposed with the aim to improve the system for the user seeking higher recognition rate. For example, flashing characters instead of rows and columns [Guan et al. 2004], changing the matrix colors [Takano et al. 2009], changing the size of the matrix [Sellers et al. 2006] or superimposing famous faces in the flashing matrix [Kaufmann et al. 2011]. Also some comparisons concerning the P300 classifications methods exists [Krusienski et al. 2006].

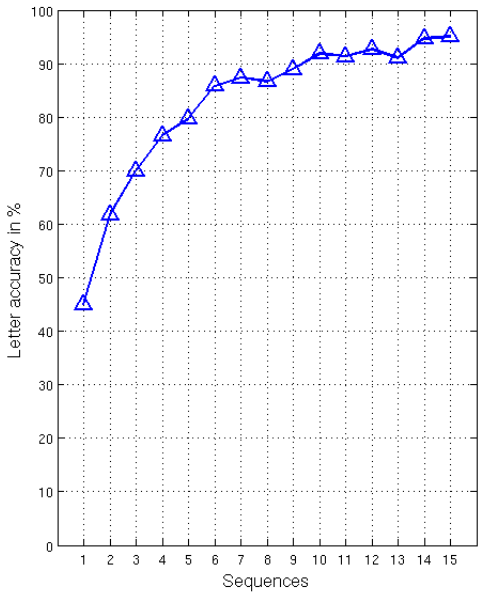


Figure 2.8: Performance curve for 1 to 15 intensification sequences.

2.5 Challenging Issues

Despite all the advantages presented by the P300-based systems, there are several challenging issues that should be addressed to improve P300 based BCI systems. These issues are related with the system ergonomics, *i.e.* its design and implementation or related with signal processing issues. This thesis is focus in the issues presented in the signal processing step, but a quick review is presented below.

2.5.1 Ergonomic Issues

There exists several problems when data is collected from subjects. Some of these problems are related to the subject's mental state as fatigue or distraction. Others are related with the application implementation and design as described below.

2.5.1.1 Adjacency Problem

The adjacency problem corresponds to a human error caused when subject is attracted to a non-target flashed item (column/row) that is adjacent to the target letter generating a P300 response [Fazel-Rezai 2007]. This issue affects the generation of P300 responses, but this mistakes can be removed through average, yet maybe becoming a serious problem if working with single-trial detection. A solution to the adjacency problem is proposed in [Townsend et al. 2010], which changes the pattern of flashed items by disassociating the rows and columns in order to avoid adjacent items to be flashed in the same flash group.

2.5.1.2 Crowding Effect

In BCI spellers the crowding effect is a phenomenon that occurs when a target letter is surrounded by similar ones [van den Berg et al. 2007], causing the inability to discriminate letter features and contours, making it difficult to appropriately respond to the stimulus. The crowding effect is usually caused by font size, font saturation and font tonality. In [Treder and Blankertz 2010], the authors claims that the P300 speller matrix is affected by this problem because it is difficult to pay attention to letters located in the matrix periphery. For this reason several matrix modifications have been proposed, such as superimposing familiar images in the flashing matrix [Kaufmann et al. 2011], changing the size of the matrix and the time between stimulus (flashes), as in [Sellers et al. 2006], or changing visually the matrix to see how it influences the results, as in [Salvaris and Sepulveda 2009]. A study about the effect in visual modifications in the P300 speller, such as, changes in symbols' dimensions, distance between symbols and colors can be found in [Salvaris and Sepulveda 2009]. Although some of these methods have shown some improvements, it is unclear if the effect produced is durable over time, because subjects can get used to these methodologies losing the effect produced at the beginning.

2.5.1.3 Double Flash Problem

The Double flash problem occurs when two targets item are flashed consecutively, causing errors in the subject's response. This is only possible when a column and a row containing the target are consecutively flashed, causing two different phenomena: (1) the second flash is not distinguished by the subject, therefore no P300 response is elicited, and (2) the second flash is distinguished by the subject producing a P300 response that is temporarily overlapped with the first P300 response. The same solution proposed for the adjacency problem in [Townsend et al. 2010] is valid for the double flash problem.

2.5.2 Signal Processing Issues

This thesis is focused on problems found in the P300 feature extraction and translation steps. These problems are challenging essentially because EEG signals are not stationary, explaining the large variability in responses for a given user. Also there is a great variability between different subjects, making difficult to obtain a general model for multiple users. Finally, the low signal-to-noise ratio of the signal makes difficult the feature extraction and classification. All this issues mainly affect the robustness of systems based on ERP's. The main problems encountered when working with this type of signals are explained, as follows.

2.5.2.1 Artifacts and Noise

The recorded signal is a mixture between useful brain signals and useless signals. Noise and artifacts in EEG signals are produced by different phenomena. The noise is generated by the electrical environment, mainly the power lines, and the recording system such as changes in electrodes impedance, voltage bursts, etc. Artifacts are produced by physiological perturbations from the subject's body, such as eye blinks, eye movements, head movements, heart beats, etc. An example of artifacts is shown in Figure 2.9. In both cases, the detection of artifacts is difficult and though sometimes are made by hand by experts, making it slow and useless for the purposes required. For this reason, automated techniques for detection are needed.

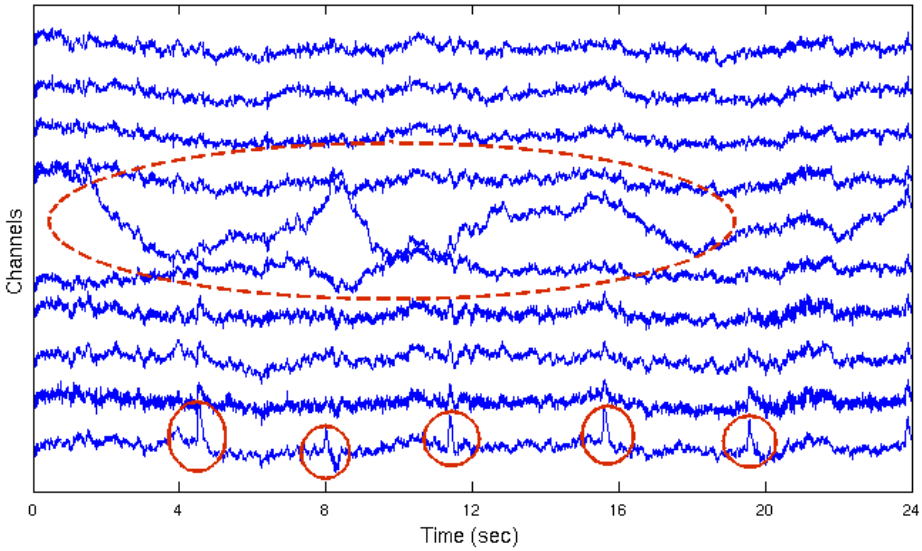


Figure 2.9: EEG Artifacts: an electrode's shift and heart beats can respectively affect one or several signals. The signals correspond to subject PGA of the database [Ledesma-Ramirez et al. 2010] presented in section A.2.1.

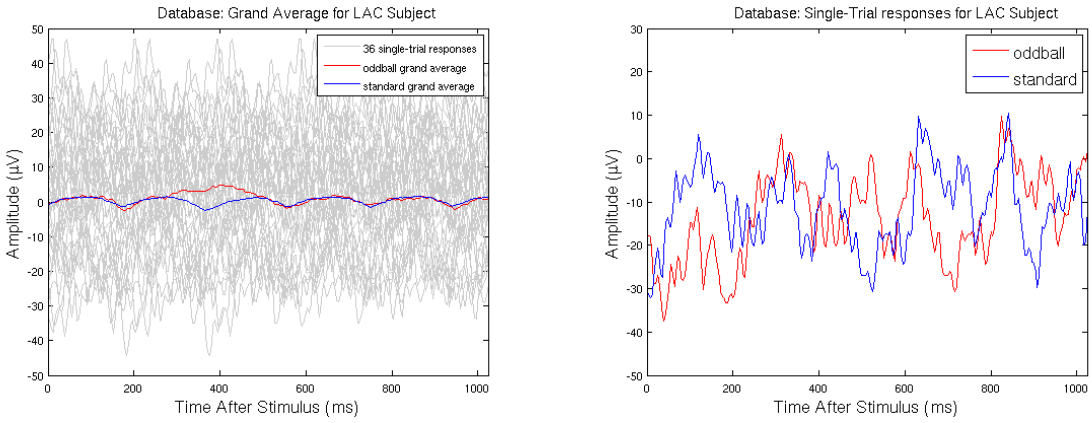


Figure 2.10: Single-trial responses: The left figure shows the grand averages compared to single-trail responses. The right figure shows the comparison between two single-trial responses. The signals correspond to subject LAC of the database [Ledesma-Ramirez et al. 2010] presented in section A.2.1.

2.5.2.2 Amplitude

In addition to the presence of noise and artifacts in the signal, the amplitude of the P300 wave is usually too small to isolate it from the ongoing EEG activity. For this reason the sequence averaging technique is commonly used. The problem with this technique is that it is possible to remove the P300 component if there is jitter or changes in phase in the signal. Also the speed of the application is reduced by the need of multiple recording of the same event to average. The difference between single-trial responses and its average is shown in Figure 2.10. It is clear that

the difference between a P300 and a non-P300 responses is more difficult to make in a single trial than between their averages. It is also important to consider that ERPs vary in amplitude depending on the cognitive tasks, such as selective attention [Gibbons and Stahl 2007], which can undermine the performance of the system.

2.5.2.3 Latency Jitter

P300 waves are triggered about 300 ms after the stimulus. This latency depends on the subject's mental state, such as fluctuations in the state of attention produced by external stimulus that do not belong to the application, which may cause different latencies in the responses. It is also possible to observe a difference in latency between different sessions of the same subject. This may require a constant system recalibration. It was demonstrated that there is a trial-to-trial variability in cortical ERPs in [Truccolo et al. 2002]. This variation in single-trial latencies can negatively impact the performance, since the average can remove the P300 wave as shown in Figure 2.11. Some techniques to analyze the latency jitter in ERPs can be found in [Thornton 2008] and [Woody 1967].

To overcome this problem shift-invariant techniques are needed to deal with non-stationarity and variations between sessions. However, the generalization of these solutions are difficult because signals are highly variable between subjects.

2.5.2.4 Phase artifacts

Single-trials responses are not "phase-locked" or desynchronized, *i.e.* the signals are not synchronized with respect to time. The phase changes may produce a fake ERP response by averaging random signals, giving the impression that there was a response to a stimulus when in fact there is no stimulus. This effect is created by different signals (unrelated to ERPs), as shown in Figure 2.12.

This thesis is focused on overcoming issues related to signal processing as presented above. Some of these problems can be overcome using single-trial responses to remove problems related to the average and to improve the communication bit-rates. But the denoising process has to be very efficient.

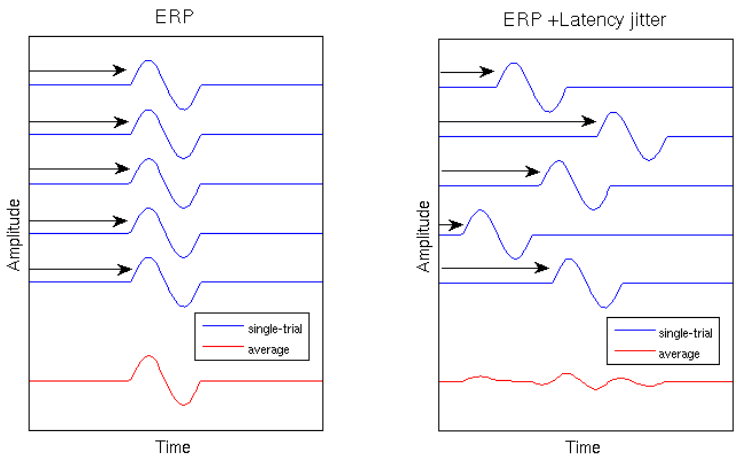


Figure 2.11: In the left, when ERPs responses have the same latency, the average reflects the ERP. In the right, when ERPs responses have an additional jitter, the average produces a flatter curve.

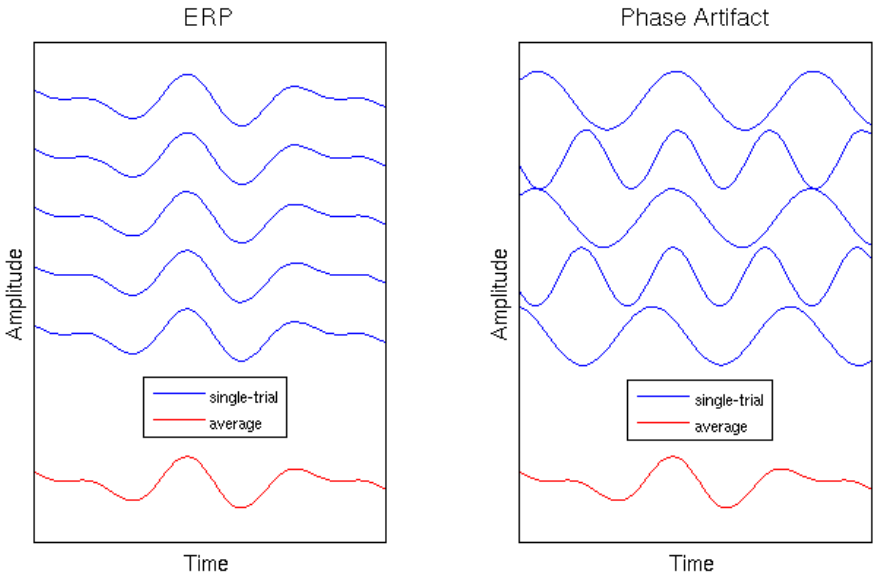


Figure 2.12: In the left, the average corresponds to an ERP while in the right, a fake ERP response is produced by the average of unrelated signals.

Chapter 3

Spatial Filters and Time-Frequency Techniques

Contents

Feature extraction is an essential step in a large number of problems, which amounts to find a suitable representation of input data to simplify classification or pattern detection. Specifically, in the BCI area, EEG brain signals are high dimensional noisy data making difficult to interpret the pattern produced by the user. Therefore, the use of feature extraction techniques is required to improve the input data representation as much as possible, without removing the key information needed for classification. The main goal of this step is to extract the useful information from the original data to allow efficient learning rules to build a predictive model that enables controlling BCI systems. It is important to note that some methods do not affect the data dimensionality, such as methods that improve the signal-to-noise ratio, whereas others actually reduce or enlarge the dimensionality to achieve a greater success of learning machine methods.

Occasionally in BCI systems, artifact detection is performed by a human expert during the offline analysis, removing undesirable characteristics of the signal (such as ocular artifacts), yet making the process slow, laborious and difficult to implement outside the laboratory. This is mainly because the expert cannot do online artifact removal, rendering it rather useless for improving the quality of life of disabled people.

For these reasons several techniques have been used to pre-process EEG signals automatically, with very popular ones being the spatial filters and time-frequency techniques discussed below.

A more extensive survey of signal processing algorithms in BCI can be found in [Bashashati et al. 2007].

3.1 Spatial Filters

Spatial filters have been widely used in EEG for dimensionality reduction, for enhancing ERPs, and for removing certain artifacts, because they have the ability to combine the signals recorded by channels to generate a small set of features which are commonly linear combinations of the original signals. Spatial filters can be divided into two groups: those using the class information of the data and those that do not use this information.

3.1.1 Without Class Information

In this group, we find spatial filters that only use geometrical relationships to find the spatial filters. For example, the *common-average reference* (CAR) computes the common average of all channels at each time point. This average is then subtracted from the channels of interest, reducing those components that are present across all channels. Another spatial filter based on geometrical relationships is the *surface Laplacian spatial filter*, which subtracts to a channel the average of channels located at a fixed radial distance, removing correlated activity between the channels and enhancing the activity that is not shared by them.

In addition, some techniques rely on data to find the spatial filters, which is clearly more complex but can lead to better results. Among all supervised spatial filter techniques, the best known are the *Principal Component Analysis* (PCA), which projects the input data into a lower dimensional space, and the *Independent Component Analysis* (ICA), which separates linearly independent sources. These techniques have been used with success in BCI, but still they pose some limitations. In fact, their main problem is that the BCI task information is not considered, with the risk to discard high-correlated components with the task.

3.1.2 Using Class Information

In this group, the data labels are used to determine the best spatial filters. The basic idea is to obtain, for example, the maximal variance of one class while minimizing the variance of the other class, as in the case of *Common Spatial Patterns* (CSP) . This technique is usually applied in BCI for EEG signals recorded during motor imagery [Ramoser et al. 2000]. The drawback of CSP is that it ignores the short-time temporal characteristics of the data, such as the phase relationships between channels and frequency bands, making it unsuitable for ERP detection. Another very popular technique is *Fisher's Linear Discriminant Analysis* developed by Ronald Fisher in 1936 [Fisher 1936]. The Fisher's criterion is used in order to find a linear combination of features which separates the data of two classes. This is done by seeking a projection that brings together the data of the same class, while maximizing the distance between classes. The Fisher's Linear Discriminant Analysis is commonly used for classification, and therefore, it will be explained in more detail in the next chapter.

Another spatial filter that have been extensively used for ERP classification is the *xDAWN*, presented in [Rivet et al. 2009] as a technique to enhance P300 evoked potentials detection using spatial filters. The main feature of this technique is the dimensionality reduction made to identify ERPs. In the paper, the raw EEG signal is denoted as a matrix $\mathbf{X} \in \mathbb{R}^{N_t \times N_s}$, where the element $\mathbf{X}_{t,j}$ is the recorded signal $x_j(t)$, corresponding to channel j at time t, considering N_s as the number of channels recorded and N_t as the number of time samples. The signal $\mathbf{X}$ can be seen

as the set of ERPs, for each channel, represented by $\mathbf{A} \in \mathbb{R}^{N_e \times N_s}$, where Ne is the temporal duration of the ERP (chosen by the technician, ex. 800 ms), allowing to designate $a_j(t)$ as the ERP signal presented in channel j at time t, and the artifacted ongoing activity $\mathbf{N}$. Using these definitions, the model for EEG brain signal is:

$$\mathbf{X} = \mathbf{DA} + \mathbf{N} \tag{3.1}$$

where the product $\mathbf{DA}$ is the response matrix obtained for the target stimulus for each channel, $\mathbf{D} \in \mathbb{R}^{N_t x N_e}$ is a Toeplitz matrix whose first column is defined as $\mathbf{D}_{\tau_k,1} = 1$, where τ_k is the stimulus onset of the k^{th} target stimulus with all the others elements equal to zero, and $\mathbf{N}$ corresponds to the noise. The authors decided to estimate the matrix $\mathbf{A}$ through the least squares method, obtaining the following solution:

$$\hat{\mathbf{A}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{X} \tag{3.2}$$

As the main idea is to enhance the responses using spatial filters, the model in 3.1 can be written as:

$$\mathbf{XU} = \mathbf{DAU} + \mathbf{NU} \tag{3.3}$$

where $\mathbf{U} \in \mathbb{R}^{N_s \times N_f}$ is the spatial filters matrix, in which u_i correspond to the i^{th} column. The matrix $\mathbf{U}$ can be found using several techniques, but to take into account the noise $\mathbf{N}$ it was proposed to maximize the signal to signal plus noise ratio, shown below:

$$\hat{\mathbf{U}} = \arg\max_{\{\mathbf{U}\}} \frac{Tr(\mathbf{U}^T\hat{\mathbf{A}}^T\mathbf{D}^T\mathbf{D}\hat{\mathbf{A}}\mathbf{U})}{Tr(\mathbf{U}^T\mathbf{X}^T\mathbf{X}\mathbf{U})} \tag{3.4}$$

where $Tr(\cdot)$ refers to the trace operator. Replacing $\mathbf{X}$ and $\mathbf{D}$ by their respective QR factorizations $\mathbf{X} = \mathbf{Q_X}\mathbf{R_X}$ and $\mathbf{D} = \mathbf{Q_D}\mathbf{R_D}$ ($\mathbf{Q_X}$ and $\mathbf{Q_D}$ are orthogonal matrices and $\mathbf{R_X}$ and $\mathbf{R_D}$ upper triangular matrices) and matrix $\hat{\mathbf{A}}$ by Equation 3.2, it is possible to obtain the following equivalent equation:

$$\hat{\mathbf{V}} = \arg\max_{\{\mathbf{V}\}} \frac{Tr(\mathbf{V}^T{\mathbf{Q_X}}^T\mathbf{Q_D}{\mathbf{Q_D}}^T\mathbf{Q_X}\mathbf{V})}{Tr(\mathbf{V}^T\mathbf{V})} \tag{3.5}$$

where, $\mathbf{V} = \mathbf{R_X}\mathbf{U}$. $\hat{\mathbf{V}}$ is expressed as a Rayleigh quotient, whose solution is the concatenation of N_f eigenvectors associated to the N_f largest eigenvalues of matrix ${\mathbf{Q_X}}^T\mathbf{Q_D}{\mathbf{Q_D}}^T\mathbf{Q_X}$. These vectors can be computed from the single value decomposition (SVD) of ${\mathbf{Q_D}}^T\mathbf{Q_X}$:

$${\mathbf{Q_D}}^T\mathbf{Q_X} = \Phi\Lambda\Psi^T \tag{3.6}$$

where $\Lambda \in \mathbb{R}^{N_S \times N_S}$ is the diagonal matrix of singular values sorted in descending order ($1 \geq \Lambda_{1,1} \geq \cdots \geq \Lambda_{N_S,N_S} \geq 0$), $\Phi \in \mathbb{R}^{N_e \times N_S}$ and $\Psi \in \mathbb{R}^{N_S \times N_S}$ are two orthogonal matrices. Splitting these matrices into signal and noise subspaces

$$\Phi = [\Phi_S, \Phi_n],$$
$$\Lambda = \begin{bmatrix} \Lambda_S & 0 \\ 0 & \Lambda_N \end{bmatrix},$$
$$\Psi = [\Psi_S, \Psi_n],$$

leads to $\hat{\mathbf{V}} = \Psi_S$, obtaining as solution for Equation 3.4 $\hat{\mathbf{U}} = {\mathbf{R_X}}^{-1}\Psi_S$. Therefore, the final model can be written as:

$$\mathbf{X} = \mathbf{D}\mathbf{A}'\mathbf{W}^T + \mathbf{N}' \tag{3.7}$$

This model inspired the name algorithm called xDAWN, where $\mathbf{A}' = {\mathbf{R_D}}^{-1}\Phi_S\lambda_S$, $\mathbf{W} = {\mathbf{R_X}}^T\Psi_S$ and $\mathbf{N}' = \mathbf{N} + \mathbf{D}{\mathbf{R_D}}^{-1}\Phi_n\Lambda_n\Psi_n^T\mathbf{R_X}$. The enhanced signal can be obtained by computing:

$$\hat{\mathbf{S}} = \mathbf{X}\hat{\mathbf{U}} = \mathbf{D}\mathbf{A}' + \mathbf{N}{\mathbf{R_X}}^{-1}\Psi_S \tag{3.8}$$

Finally, the subspace is found choosing I columns of $\hat{\mathbf{S}}$. The algorithm presented in [Rivet et al. 2009] is shown below:

Algorithm 3.1 The xDAWN algorithm

Compute QR factorization of $\mathbf{X} \Rightarrow \mathbf{X} = \mathbf{Q_X}\mathbf{R_X}$
Compute QR factorization of $\mathbf{D} \Rightarrow \mathbf{D} = \mathbf{Q_D}\mathbf{R_D}$
Compute SVD of ${\mathbf{Q_D}}^T\mathbf{Q_X} \Rightarrow {\mathbf{Q_D}}^T\mathbf{Q_X} = \Phi\Lambda\Psi^T$
Select the I couples of singular vectors (Φ_i, Ψ_i) associated with the I largest singular values λ_i
Finally $\forall\, i,\ 1 \leq i \leq I,\ (\hat{\mathbf{u}}_i, \mathbf{a}'_i) = (\mathbf{R_X}^{-1}\psi_i, {\mathbf{R_D}}^{-1}\phi_i\lambda_i)$
Estimate sources: $\forall \quad 1 \leq i \leq I, \hat{s}_i(t) = \hat{\mathbf{u}}_i^T\mathbf{x}(t)$

This technique has been widely accepted by the BCI community, being integrated into one of the most famous software for BCI, Openvibe [Renard et al. 2010]. For more information about the xDAWN algorithm see [Rivet et al. 2011]. All these methods are based on the temporal domain, proposing linear combinations of the recorded signals. But it is also useful to investigate the frequency domain. Several standard methods will be presented in the following section.

3.2 Time-Frequency techniques

Signal processing techniques are selected according to certain signal characteristics. There are two different ways to represent a signal. A time domain representation gives information about the maximal time resolution, meaning the intensity of the signal at time t, but no frequency information. A frequency domain representation, on the other hand, gives accurate frequency information, but is not capable of providing any time information. The problem with these representations is that they are mutually exclusive.

The Fourier Transform (FT), developed by Jean Baptiste Fourier in 1807, is a very popular technique to study the frequency spectrum of a signal, being very efficient when the signal is composed by stationary components. The idea behind the technique is that a signal can be represented as a linear superposition of sines and cosines characterized by their frequency f,

which allows to obtain the signal frequency spectrum. If $x(t)$ is a stationary signal, its Fourier transform is defined as the inner product of $x(t)$ and the complex sinusoidal mother functions $e^{-i\omega t}$, where $\omega = 2\pi f$, as follows:

$$\mathcal{F}_x(\omega) = \langle x(t), e^{-i\omega t} \rangle = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \tag{3.9}$$

According to the Fourier inversion theorem, the signal $x(t)$ can be recovered from its Fourier transform as follows:

$$x(t) = \int_{-\infty}^{\infty} \mathcal{F}_x(\omega) e^{i\omega t} d\omega \tag{3.10}$$

Despite its popularity, the FT has two disadvantages. The first one is that the Fourier spectrum is affected by the Gibb's phenomenon [Bocher 1906], if the treated signal is not stationary. The second disadvantage is the lack of time information about the evolution of frequencies. These reasons led to the modification of the FT by Gabor in 1946: a new method was developed to obtain a time-frequency representation, called Short Time Fourier Transform (STFT) or Gabor Transform.

The Short Time Fourier Transform (STFT) [Gabor 1946] allows to analyze a small signal section multiplying the signal $x(t)$ by a *window function* $g(t)$ that slides along the time axis before determining the frequency spectrum. The result is a two-dimensional representation of the signal (in time and frequency). STFT is mainly the same as FT, but it use a windowed signals to make the analysis. In Figure 3.1, the different representations for the same signal are shown. At the top, the representation of the signal in time $x(t)$, where it is possible to observe the signal amplitude at each time. The frequency representation $f(Hz)$ is in the left side of the figure, and shows that the signal is composed by two different frequencies, 10 Hz and 40 Hz, but there is no time information. Finally, the STFT, in the center, shows the frequencies present in the signal in addition to the time when they occur.

The STFT representation can be expressed mathematically as:

$$STFT_x(b, \omega) = \langle x(t), g(t-b) e^{-i\omega t} \rangle = \int_{-\infty}^{\infty} x(t) g(t-b) e^{-i\omega t} dt \tag{3.11}$$

where b indicates the position where the analysis is performed *i.e.* where the window is centered. The STFT provides information about the time and frequencies of a signal under the assumption that the signal's section (defined by the window) being studied is stationary. In some cases, it is difficult to find an appropriate window size where the signal is stationary. Reducing the size of the window is not a solution because an increase in the temporal resolution decreases the frequency resolution, turning impossible to obtain precise signal information in both time and frequency domains. This lack of information can be explained considering the *uncertainty principle*, that states the uncertainty relation between the time and frequency, restricting the maximum joint time-frequency resolution that can be achieved.

3.2.1 The Uncertainty Principle

The uncertainty principle states that is impossible to know at the same time two different physical magnitudes such as the position and velocity of a particle. In other words, the more

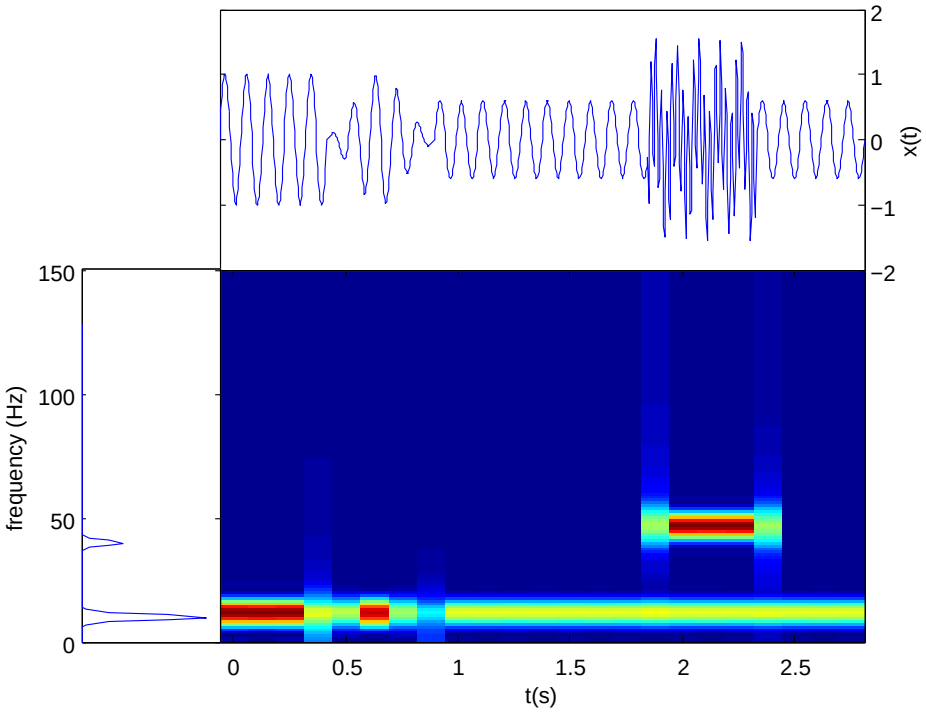


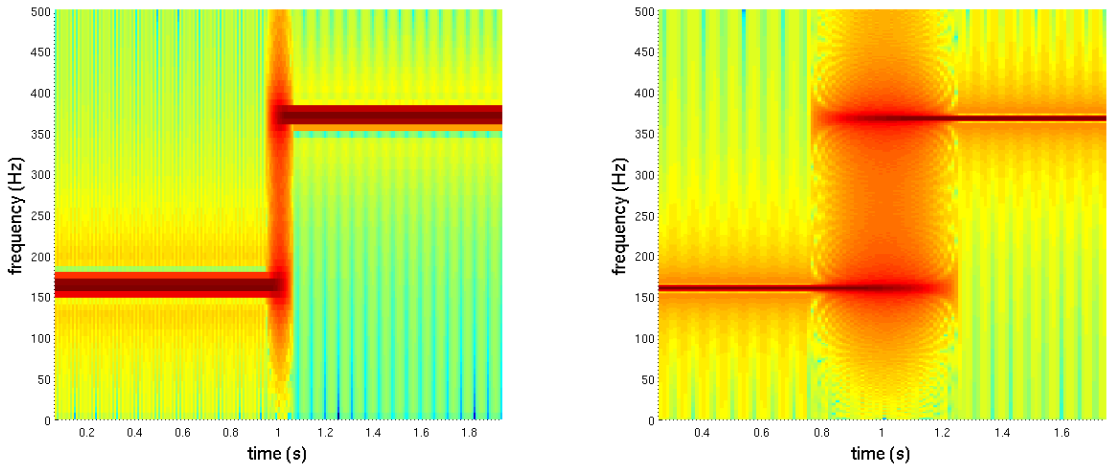
Figure 3.1: Comparison between the time, the frequency and the time-frequency planes. The image at the top shows signal $x(t)$ in the time domain. The image on the left side of the figure shows the frequency $f(Hz)$ plane of signal x and the color image in the center shows the time-frequency representation of signal x using the Short Time Frequency Transform.

accurately the velocity of a particle is known, the less accurately its position can be known and *vice versa*. The uncertainty principle is present in the STFT due to the windowing technique. The principle states that it is impossible to know the frequency and time of an event with an arbitrary precision. Therefore, the signal's representation depends on the width of the selected window $g(t)$, allowing a good frequency resolution and poor time resolution if the window is wide, or a good time resolution and a poor frequency resolution if a narrow window is chosen. For example, in Figure 3.2, two STFT decompositions using different windows for a same signal are shown. Figure 3.2(a) uses a 128 ms window, which allows to identify that the frequency changes in the signal approximately around 1 second, but it is difficult to determine the precise frequencies present in the signal. On the other hand, Figure 3.2(b) uses a 512 ms window, which allows a better frequency precision, but it is difficult to distinguish when the frequency change occurred. Finally, it is possible to infer that the bigger the window in time, the better the frequency resolution and *vice versa*.

In mathematical terms the uncertainty principle must satisfy the following inequality:

$$\Delta t \Delta f \geq \frac{1}{4\pi}$$

where Δt is the time uncertainty (time duration) and Δf is the frequency uncertainty (frequency bandwidth). Taking into account that $\omega = 2\pi f$, the inequality can be also written as $\Delta t \Delta \omega \geq \frac{1}{2}$. Basically this inequality implies that it is not possible to find a temporal width and a frequential width both arbitrarily small.



(a) (b)

Figure 3.2: Time vs Frequency Resolution: the specific precision of each axis is determined by the size of the window. In 3.2(a), a window size of 128 ms has been used. It is possible to identify precisely the time when the frequency changes, but not precisely the frequency values. In 3.2(b), a window size of 512 ms has been used. The time axis is blurred, but it is possible to identify the frequencies in the signal.

The main disadvantage of STFT is that the width of the window is fixed for the entire signal analysis, implying a restricted frequency and time resolution. In addition, the STFT is not able to give information if the windowed signal being analyzed is not stationary, allowing only a limited use in biological signals since most of them are not stationary. This lead to the development of the Wavelet theory to be able to represent non stationary signals in the time-frequency domain mapping in a time-scale representation, where the scale is related to the frequency.

3.2.2 Wavelet Transforms

Wavelet Transform (WT) is a relatively new mathematical tool introduced by Grossmann and Morlet in 1984 to represent a signal using an infinite group of basis functions with finite energy called Wavelets [Mallat 1989]. The term "wavelet" refers to the oscillatory nature and finite length of these functions. Wavelets became popular very fast because its principles can be applied in a large number of different areas, such as biology, geology, electrical engineering, signal processing, etc. attracting a lot of scientists to the wavelet domain.

In brain signals processing, the wavelets theory has been used in several studies [Quiroga and Garcia 2003; Yong et al. 2005; Gupta et al. 2009] because EEG data are not stationary (several rhythms vary at the same time), and contain a lot of artifacts and noise, making them difficult to analyze manually by scientists. These studies focus on the automatic analysis of brain rhythms, evoked potentials and other brain signals. As STFT, wavelets convolve the signal to be analyzed by a window function, but unlike STFT the window sizes used in wavelet are adapted to the analyzed frequencies. Higher frequencies are analyzed using narrow windows to obtain a good time resolution, and lower frequencies are analyzed using wide windows to obtain a good frequency resolution. Due to this basic idea, the multiresolution concept (MRA) has emerged, which as the

name implies, is a signal analysis such that each frequency component is studied with a different resolution. This is an advantage compared to the STFT which analyzes all the frequency components using the same resolution. All these characteristics allow the wavelet transform to focus on brain components in a more precise way compared to the STFT. In Figure 3.3, the difference between STFT and wavelets are shown. Wavelet analysis combines the phase and amplitude information at each point into the coefficient, (displaying red dots for positive peaks and blue dots for negative peaks for this particular figure). Besides, the frequency information in the scale space is also shown. In the STFT the phase information is ignored, so it is possible to see only a continuous line for the sinusoid.

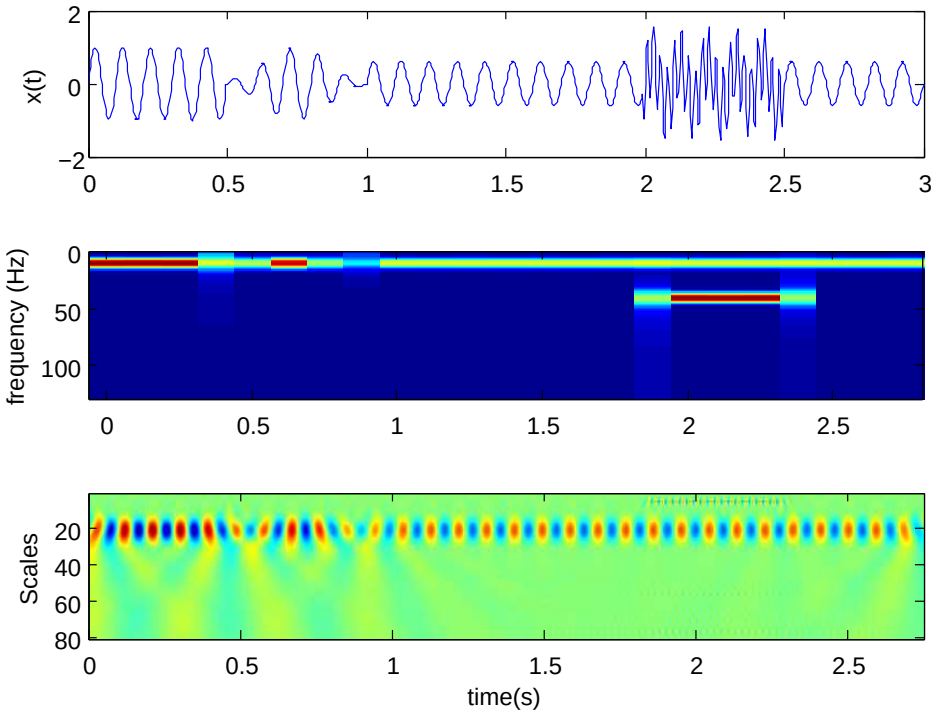


Figure 3.3: Difference between Wavelet and STFT: at the top, the original signal. In the middle the STFT decomposition, which ignores the signal phase information, displaying a continuous line for a sinusoid. At the bottom the wavelet decomposition, which includes phase information displaying red dots for positive peaks and blue dots for negative peaks. Both analysis provide the frequency values, using scales in the case of wavelets.

Wavelet transforms can be divided in *Continuous Wavelet Transform* (CWT) and *Discrete Wavelet Transforms* (DWT), both will be explain later in this section. A signal $x(t)$ is represented in terms of scaled and shifted versions of a *mother wavelet* $\psi(t) \in L^2(\mathbb{R})$, which must satisfy the following *admissibility condition*:

$$C_\psi = \int_0^\infty \frac{|\mathcal{F}_\psi(\omega)|^2}{\omega} d\omega < \infty \tag{3.12}$$

This condition ensures the signal reconstruction by the inverse transform, which also implies that the wavelet must have zero average, *i.e.* it must be oscillatory [Mallat 2008; Daubechies 1992].

$$\int_{-\infty}^{\infty} \psi(t)dt = 0$$

The complexity of the wavelet $\psi(t)$ is given by its *Vanishing moments*, which is the rate decay of the wavelet function, ensuring that the integral converges to zero for a rate decay k. This means that the more vanishing moments the mother wavelet has, the more complex functions can be represented with it. For example, if a wavelet has p vanishing moments, it means that a polynomial of grade $p-1$ can be generated. The number of vanishing moments can be found as follows:

$$\int_{-\infty}^{\infty} t^k \psi(t)dt = 0, \qquad \text{for } 0 \leq k < p \tag{3.13}$$

The order of a wavelet transform is usually given by its number of vanishing moments, for example the family Daubechies has N orders corresponding to N positive integers, which gives us, for example, Daubechies order 4 (Db4) or Daubechies order 8 (Db8). The higher the wavelet transform order is, better the signals approximation will be.

Another important feature of wavelets is the *compact support*, that establishes that the wavelet should have finite duration, meaning that the wavelet equals to zero outside a closed interval, allowing less complex calculations.

The *wavelet family* is generated using dilated and translated versions of the mother wavelet. These functions are called *child wavelets* or just *wavelets*.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \tag{3.14}$$

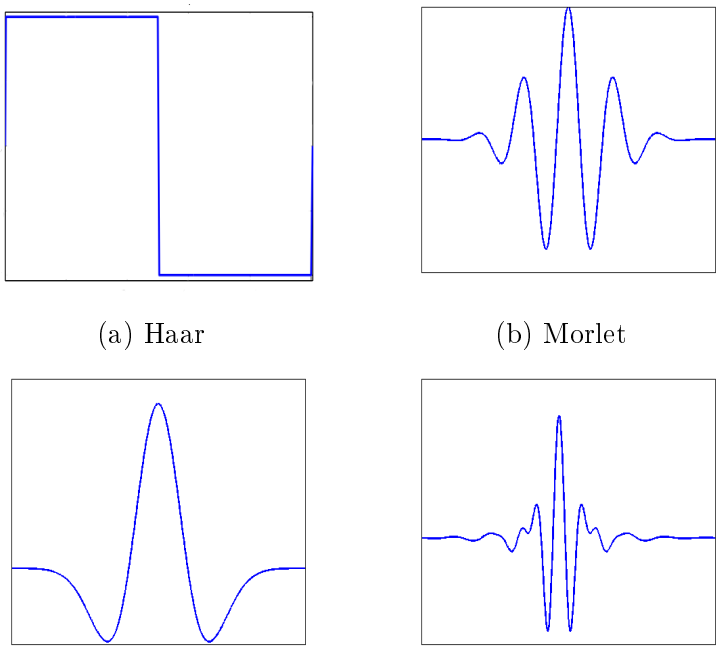
where $a \in \mathbb{R}^+$ is the scale parameter, whose change alters the spectrum (wavelet central frequency) and the size of the time-frequency window. The parameter $b \in \mathbb{R}$ corresponds to the translation along the time axis, specifying its location in time. A normalization constant $\frac{1}{\sqrt{a}}$ is added to ensure that the energy $E_\psi = \int_{-\infty}^{\infty} |\psi(t)|^2 dt$ is the same at all scales a.

Nowadays, there are several wavelets families that are popular for ERP detection, each one with its specific form. Some of these families are shown in Figure 3.4. However, the choice of a wavelet family waveform depends on the application and the type of signals to be analyzed.

As mentioned before, the theory of wavelet transforms is composed by the *Continuous wavelet transform* (CWT) which uses a continuous wavelet function to do the signal analysis, and the *Discrete wavelet transform* (DWT), which uses filter banks to obtain a multiresolution time-frequency representation.

3.2.2.1 Continuous Wavelet Transform

As previously explained, a mother wavelet (or analyzing function) $\psi(t)$ is needed to perform a signal analysis. This mother wavelet is the equivalent to the window used in the STFT. Its choice



(a) Haar (b) Morlet

(c) Mexican Hat (d) Meyer

Figure 3.4: Some popular Wavelet Mothers. Each wavelet mother in these images corresponds to an analysis waveform or prototype, that provides a basic template to be dilated and translated to obtain the child wavelets waveforms.

is made to analyze particular shapes in signal $x(t)$, because the wavelet coefficients measures the similarity of the mother wavelet shape and the original signal in a current scale and at a specific time position. The CWT of $x(t) \in L^2(\mathbb{R})$ is defined as:

$$W_\psi^x(a,b) = \langle x, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} x(t)\psi_{a,b}^*(t)dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a} \right) dt \tag{3.15}$$

where a and b change continuously, and ψ^* is the complex conjugate of ψ. The CWT coefficients measure the variation of x in a neighborhood of point b, whose size is proportional to a, obtaining a mapping of a one-dimensional signal in a two-dimensional space.

As shown in Equation 3.14, the wavelet analysis representation is made in terms of the scale a and not the frequency f. These two are inversely related by

$$f_a = \frac{f_c}{a f_s} \tag{3.16}$$

where, f_c is the wavelet central frequency, f_s is the signal sampling rate, and f_a is the frequency corresponding to the scale a.

The reconstruction of the original signal is possible due to the *admissibility condition* (3.12) using:

$$x(t) = \frac{1}{C_\psi} \int_0^{\infty} \int_{-\infty}^{\infty} W_\psi^x(a,b) \frac{1}{a^2} \psi_{a,b}(t) da db \tag{3.17}$$

For practical implementation purposes, the continuous wavelet transform is often performed using a summation instead of the integral, where a discrete grid of scales and translations is used. The resolution of the discretization makes possible or impossible the inverse wavelet transform to reconstruct the original signal. It is important not to confuse the discretization of the CWT with the actual DWT, where the scales and translation parameters are discretized to avoid the high redundancy in the signal representation obtained using the CWT.

3.2.2.2 Discrete Wavelet Transform

In wavelet analysis, the discrete form of Equation 3.14 is often used. More precisely, the discrete orthogonal wavelet decomposition is obtained using a discretized scale a and translation b. In the most general case, this is made as follows:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right), \qquad a_0 > 1, b_0 > 0 \tag{3.18}$$

where the scale a and translation b are set through the integers m and n. a_0 and b_0 are integers corresponding to fixed scale and translation step parameters. A widely used discretization for the scale and translation is the dyadic steps, because it is simple enough and allows the construction of orthonormal wavelets, that ensures the signal reconstruction as shown later. The wavelet family functions using dyadic steps is obtained setting $a_0 = 2$ and $b_0 = 1$ respectively in Equation 3.18. By using these parameters for the discrete wavelet transform on a given scale $m + 1$, there are twice less coefficients than on the previous scale m. The new wavelet family can be written as follows:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}}\psi\left(\frac{t - n2^m}{2^m}\right) = 2^{-m/2}\psi(2^{-m}t - n). \tag{3.19}$$

Using this notation it is possible to write the following equation for the discrete wavelet transform:

$$W_\psi^x(m,n) = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt = 2^{-m/2}\int_{-\infty}^{\infty} x(t)\psi^*\left(2^{-m}t - n\right)dt \tag{3.20}$$

The wavelet coefficients obtained with this transformation are also called *Detail coefficients*. The original signal can be reconstructed adding all coefficients obtained through Equation 3.20 as follows:

$$x(t) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} W_\psi^x(m,n)\psi_{m,n}(t). \tag{3.21}$$

For a perfect reconstruction of $x(t)$, the wavelets must be orthogonal to each other and normalized to have a unit energy. As a consequence, there is no redundancy in the wavelet coefficients. To be orthogonal, the wavelet must satisfy the following equation:

$$\int_{-\infty}^{\infty} \psi_{m,n}(t)\psi_{m',n'}(t)dt = \begin{cases} 1 & \text{if } m = m' \text{ and } n = n' \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

When using a finite depth decomposition, the dilatations (thus the scales) are limited, *i.e.* $m \in \{1, 2, \ldots, M\}$. In this case, wavelets are not sufficient to represent an arbitrary signal containing low frequencies. This is because in each scale there are twice less coefficients than in the previous scale. Therefore, a continuous signal $x(t)$ will require the use of infinite scales to study the entire signal spectrum. Therefore, *father wavelets* (also known as scaling functions) are introduced in the decomposition: as the mother wavelets, they have a unit norm, but they do not have a zero mean. For a given family of scaled and translated mother wavelets, a family of translated and scaling functions $\phi_{m,n}$ is used to ensure that the entire signal spectrum is analyzed [Mallat 1989]. Father wavelets are defined as follows:

$$\phi_{m,n} = 2^{-m/2}\phi(2^{-m}t - n) \tag{3.23}$$

The wavelet coefficients obtained through father wavelets are called *Approximation coefficients* and are computed as:

$$W_{\phi}^{x}(m,n) = \int_{-\infty}^{\infty} x(t)\phi_{m,n}(t)dt \tag{3.24}$$

Using these notations, a discretized signal $x(t)$ can be written as:

$$x(t) = \sum_{m,n} W_{\psi}^{x}(m,n)\psi_{m,n}(t) + \sum_{n} W_{\phi}^{x}(M,n)\phi_{M,n}(t), \tag{3.25}$$

where the left term corresponds to the sum of the detail coefficients through all scales and the right term corresponds to the approximation coefficients on the last scale M. The approximations coefficients correspond to the high-scale, low-frequency components of the signal. These coefficients contain the most important part of the signal: the part with the "essence" of the signal. The detail coefficients are the low-scale, high-frequency components, *i.e.* all the small features that makes the signal "unique".

From Equation 3.25, it is possible to infer that the sum of the approximation coefficients $x_m(t)$ and the detail coefficients $d_m(t)$ for the same scale m, gives as a result the signal approximation at the previous scale:

$$x_{m-1}(t) = x_m(t) + d_m(t) \tag{3.26}$$

Until now, the signal $x(t)$ has been considered as continuous. Obviously this is not the case in digital data, making the possible number of scales to be used limited, depending exclusively on the size of the signal. If the signal length is $N = 2^M$, there will be at most M scales, where $0 < m \leq M, M \in \mathbb{N}^*$. The *fast wavelet transform* algorithm by Mallat in 1988 [Mallat 1989] was developed based on Equation 3.26. This algorithm implements the DWT using filters, and starts computing the approximation and detail coefficients for the first level (scale) until the last one. Figure 3.5 shows the decomposition procedure for signal $\mathrm{x_0^N}$ of size N. The decomposition is performed using a low-pass filter to obtain the approximation coefficients $\mathrm{x_1^{n/2}}$ of the first level and a high-pass filter to obtain the detail coefficients $\mathrm{d_1^{n/2}}$ of the same level. According to the Nyquist's theorem, half of the samples can be eliminated. Therefore the obtained coefficients are downsampled by two discarding every other sample in order to preserve the number of points in the input data. Then, the process in repeated using the approximation coefficients as new input,

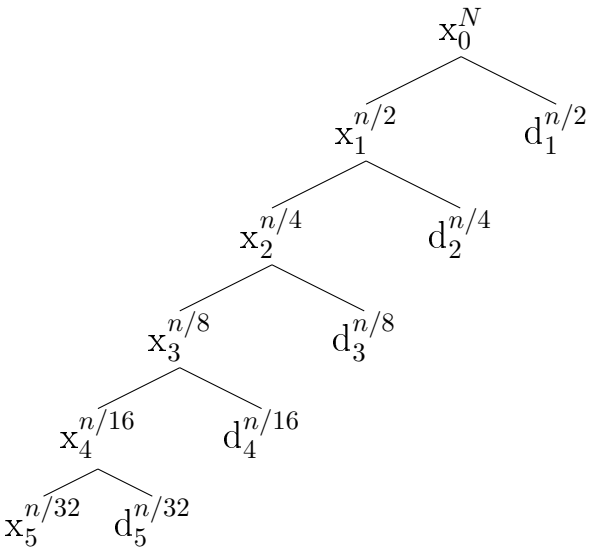


Figure 3.5: Discrete Wavelet Decomposition over five levels. The decomposition of signal x_0^N, where N is the signal's size and 0 the level, is performed by filtering and downsampling in order to obtain the approximation $x_1^{n/2}$ and detail $d_1^{n/2}$ coefficients. The process is repeated using approximation coefficients until a desired level of decomposition.

in this case $x_1^{n/2}$, for the next level. The process ends when the desired decomposition level is reached. In the example used in Figure 3.5, five levels are used.

In order to reconstruct the original signal, the inverse process is applied. The approximation and detail coefficients of the last decomposition level are upsampled by two, inserting zeros at odd-indexed elements. Then they are filtered by reconstruction filters associated with those used in the decomposition process. Finally, the approximation and detail coefficients are summed to obtain the approximation coefficients of the previous level, because $x_4 = x_5 + d_5$.

3.2.3 Semblance Analysis

Wavelet analysis is also useful for bivariate analysis, making possible the study of two different signals to discover the relationship between them. The cross-wavelet analysis allows to find mutual characteristics between signals using the available information in the wavelet transform. To perform this type of analysis, at first, it is necessary to extend some concepts used in Fourier analysis for wavelets analysis.

The wavelet power spectrum is defined for wavelets as the autocorrelation of a signal $x(t)$ in the wavelet domain:

$$W_\psi^{xx} = W_\psi^x W_\psi^{x*} \tag{3.27}$$

where, W_ψ^{x*} is the complex conjugate of W_ψ^x. This concept can be extended to compare two different signals $x(t)$ and $y(t)$. The cross-wavelet spectrum [Torrence and Compo 1998] can be defined as the expectation value of the product of both wavelet decomposition W_ψ^x and W_ψ^{y*}:

$$W_\psi^{xy} = W_\psi^x W_\psi^{y*} \tag{3.28}$$

W_ψ^{xy} is a complex value and can be decomposed into phase $\theta = tan^{-1}(\Im(W_\psi^{xy})/\Re(W_\psi^{xy}))$ and amplitude $|W_\psi^x W_\psi^{y*}|$.

Semblance analysis [Cooper and Cowan 2008] was introduced to compare two given signals $x(t)$ and $y(t)$ based on phase correlations between their wavelet decompositions W_ψ^x and W_ψ^y. This measure is based on the cross-wavelet transform and is defined as:

$$S = cos^n(\theta), \qquad S \in [-1, 1] \tag{3.29}$$

where n is an odd integer greater than zero. The reason that n should be odd is to preserve the sign of the cosine. Besides, the use of large numbers for n produces a sharp semblance response as demonstrated in [Cooper and Cowan 2008].

The values of S correspond to the correlation between the two signals. $S = 1$ means that signals are fully correlated, $S = 0$ means that signals are uncorrelated, *i.e.* it is not possible to conclude that there is a correlation between them. Conversely, $S = -1$ implies that signals are fully inversely correlated. As Equation 3.29 does not consider amplitude information, it is possible to combine the phase information of S and the amplitude information as follows:

$$D = cos^n(\theta)|W_\psi^x W_\psi^{y*}| \tag{3.30}$$

This measure can be useful if the signal's amplitudes are important to be analyzed.

3.2.4 Multi-Semblance Analysis

Nevertheless, the semblance measure is only useful to compare two signals. For this reason the concept was extended to compare more than two signals at the same time.

This new semblance measure is called **Mean resultant length** (MRL) and it was presented by Cooper in [Cooper 2009] based on circular statistics [Davis 1986, p. 319], considering that the beginning and the end of the phases coincide ($0° = 360°$). The phases can be treated as vectors, because the angles denote a direction. It is possible to find the mean direction of phases for all the signals by connecting all the vectors. If the mean direction is divided by the number of vectors involved, the Mean Resultant Length is obtained. The MRL can be calculated according to the number N of signals treated, at each scale a and time t:

$$MRL(a,t) = \frac{\sqrt{(\sum_{i=1}^{N} \Re(W_\psi^i(a,t)))^2 + (\sum_{i=1}^{N} \Im(W_\psi^i(a,t)))^2}}{\sum_{i=1}^{N} |W_\psi^i(a,t)|}. \tag{3.31}$$

For more than two signals, the inversely correlated concept does not apply, that is verified by the MRL values ranging from 0 for uncorrelated signals to 1 for correlated signals.

These measures are used in the geology field to analyze geologic sequences of data through statistical procedures [Davis 1986].

3.2.5 Wavelet Thresholding Methods

Discrete wavelet transform has been successfully applied in different domains, such as medicine, astronomy, geology, informatics to remove noise from non-stationary signals, since they do not

require any particular hypothesis about the nature of the signal under study, using thresholding methods. The first thresholding method applied to the wavelets coefficients was introduced by Donoho in [Donoho and Johnstone 1994, 1995], allowing to analyze the noise of a signal at each scale, adapting the threshold accordingly. This technique is of great interest for the study of highly noisy biological signals. Some papers where this technique is applied for ERP detection are [Quiroga and Garcia 2003; Yong et al. 2005; Gupta et al. 2009].

The fundamental hypothesis of wavelet denoising is that wavelets are correlated with the informative signal and not correlated with the noise, which globally means that large coefficients correspond to signal and small coefficients correspond to noise. Therefore, noise canceling can be performed by thresholding: only large coefficients will then be used to reconstruct the informative signal (see [Antoniadis et al. 2001; Antoniadis 2007] for a review of denoising methods).

Let $z(t) = x(t) + n(t)$ be a noisy discretized signal, where $z(t)$ is the given discrete-time recorded signal of size N, $x(t)$ is the noise-free unknown version of $z(t)$ and $n(t)$ is the noise. The goal is to estimate the signal $\hat{x}(t)$ from the noisy data $z(t)$ with a small mean square error (MSE) between the original data $x(t)$ and its denoised version $\hat{x}(t)$. According to Equation 3.25 the orthogonal wavelet decomposition of $z(t)$ can be written as:

$$z(t) = \sum_{m,n} W_{\psi}^{z}(m,n)\ \psi_{m,n}(t) + \sum_{n} W_{\phi}^{z}(M,n)\ \phi_{M,n}(t) \tag{3.32}$$

where m is the scale, n the position, ψ the wavelet, ϕ the scaling function and M the analysis depth [Mallat 2008]. The approximation coefficients correspond to W_{ϕ}^{z} and the detail coefficients to W_{ψ}^{z}, which can be considered as the noisy version of the wavelet coefficients of the noise-free signal $x(t)$. As mentioned above, small coefficients of the detail are considered as noise. So the coefficients W_{ψ}^{z} are modified by setting to zero all coefficients below a certain threshold τ. Generally, the approximation coefficients W_{ϕ}^{z} remain unchanged, because they represent important components (low-frequencies) in the signal.

The new wavelet detail coefficients after thresholding are obtained as follows:

$$\hat{W}_{\psi}^{x}(m,n) = \delta_{\tau}(W_{\psi}^{z}(m,n))\ W_{\psi}^{z}(m,n), \tag{3.33}$$

where $\delta_{\tau}(W_{\psi}^{z}(m,n))$ is a thresholding function applied on the measured signal coefficients. From Equation 3.32, an estimate of the denoised signal $\hat{x}(t)$ can be obtained by the inverse wavelet transform of the coefficients vector $\hat{W}_{\psi}^{x}(m,n)$.

The *wavelet thresholding* usually refers to the use of the *hard thresholding* rule, where, as explained before, all coefficients below the threshold τ are set to zero.

$$\delta_{\tau}(W_{\psi}^{z}(m,n)) = \max\left(0, \operatorname{sign}(|W_{\psi}^{z}(m,n)| - \tau)\right), \tag{3.34}$$

On the other hand, *wavelet shrinkage* usually refers to the use of the *soft thresholding* rule, that shrinks the coefficients with high amplitude towards zero.

$$\delta_{\tau}(W_{\psi}^{z}(m,n)) = \max\left(0, \frac{|W_{\psi}^{z}(m,n)| - \tau}{|W_{\psi}^{z}(m,n)|}\right) \tag{3.35}$$

In Figure 3.6, the difference between the hard and soft thresholding rules are shown. The threshold was set in 0.4 (red-dash line). The hard threshold rule causes a discontinuity in the signal because it keeps or sets to zero a coefficient. On the other hand, the soft threshold rule shrinkage the large coefficients, avoiding discontinuities. Clearly, these thresholding techniques have advantages and disadvantages, which will not be discussed in this thesis. For more details about thresholding rules, see [Fan 1997].

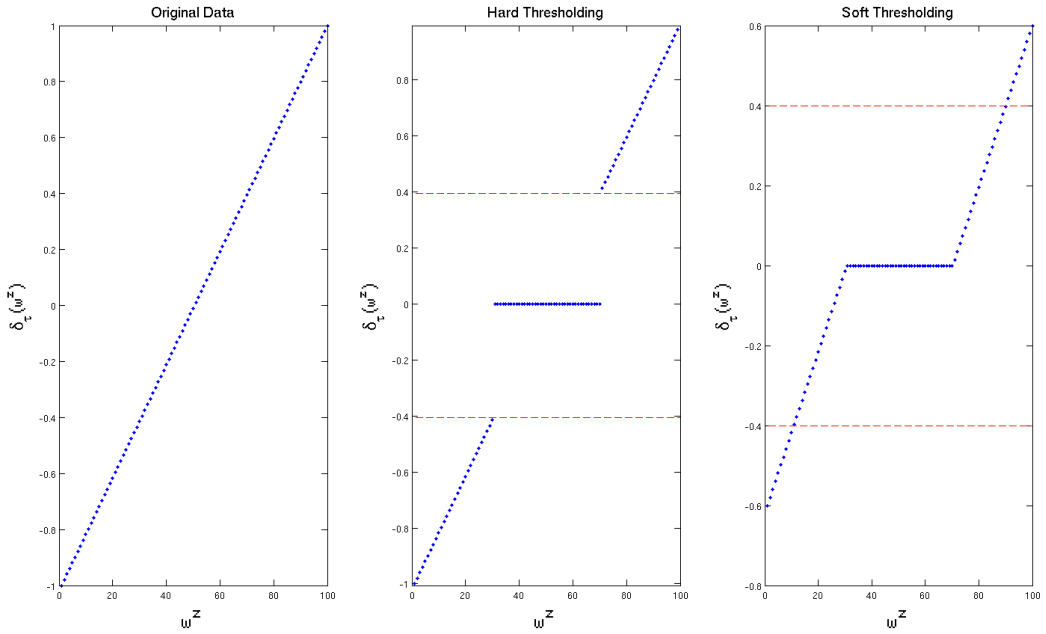


Figure 3.6: Threshold selection rules. The original signal is presented on the left. The effect of the hard thresholding rule is presented in the middle, that sets to zero coefficients below the threshold, causing discontinuities in the signal. And the effect of the soft threshold rule is shown on the right side, that shrinks large coefficients avoiding discontinuities.

3.2.6 Classic Thresholds

There are different methods to select the appropriate threshold to filter coefficients, the best known are the universal, sure and minimax thresholds:

– **Universal Threshold**

The most widely used is the universal threshold τ_U proposed by Donoho and Johnstone in their algorithm *VisuShrink* [Donoho and Johnstone 1994]. *VisuShrink* is used to achieve *complete asymptotic elimination of the normal Gaussian noise* and it can be shown (using extreme values statistics) that this is achieved by setting $\tau_U = \sqrt{2 \log N}$. This method is very appealing because of its simplicity. However, eliminating all noise can represent a major drawback because this often leads to less precise reconstruction of the signal of interest.

– **Sure Threshold**

A different approach is proposed by the *SureShrink* algorithm [Donoho and Johnstone 1995], that aims at *estimating as precisely as possible the "clean" signal* by minimizing the expected value of the mean square error between the denoised and noise-free signals, known as the *Stein Unbiased Risk Estimator* (SURE). Let us use the vector notation $\mathbf{w}_m^z = (W_\psi^z(m,1), W_\psi^z(m,2), \ldots, W_\psi^z(m,P))^T = (w_1, w_2, \ldots, w_P)^T$, where P is the last detail coefficient of scale m. The problem is to find an estimator $\hat{\mu}$ of the mean vector $\mu = (\mu_1, \mu_2, \ldots, \mu_P)^T$ from $\mathbf{w}_m^z$ with the minimum risk, *i.e.* estimating the loss $||\mu - \hat{\mu}||_2$ that, according to stein, can be written as:

$$\mathrm{SURE}(\tau_S; \mathbf{w}_m^z) = P + \sum_{i=1}^{P} [\min(|w_i|, \tau_S)]^2 - 2 \cdot |\{i : |w_i| < \tau_S\}| \tag{3.36}$$

This risk is minimized by exhaustively searching the optimal threshold τ_S according to the coefficients. The obtained threshold τ_S (or thresholds, as the method is usually implemented by scale) has as upper bound, the τ_U threshold.

$$\tau_S^* = \underset{\{0<\tau_S\leq\tau_U\}}{\arg\min}\ \mathrm{SURE}(\tau_S; \mathbf{w}_m^z) \tag{3.37}$$

The SURE principle has a drawback when the wavelets coefficients present a significant sparsity, since the information contributed to SURE by the coefficients that are zero is higher than the information contributed by non-zero coefficients. For this reason, the final threshold τ_S is selected as follows:

$$\tau_S(\mathbf{w}_m^z) = \begin{cases} \sqrt{2\log P} & \text{if } \nu_s(\mathbf{w}_m^z) \leq 1 \\ \tau_S^* & \text{otherwise} \end{cases} \tag{3.38}$$

where $\nu_s(\mathbf{w}_m^z)$ is a sparsity measure [Donoho and Johnstone 1995] defined as:

$$\nu_s(\mathbf{w}_m^z) = P^{-1/2} \cdot \frac{\sum_{i=1}^{P}(w_i^2 - 1)}{\log_2^{3/2}(P)} \tag{3.39}$$

Given that m corresponds to the scale, it can be derived that the threshold τ_S is level dependent and should be noted as $\tau_S = \tau_S(m)$.

– **Minimax Threshold**

Another classic threshold τ_M is the minimax one [Donoho and Johnstone 1992, 1998] which is used as a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics in order to design estimators, where the minimax risk bound is defined as:

$$\Lambda_P^* = \inf_{\tau_M} \sup_{\mu} \left\{ \frac{R_{\tau_M}(\mu)}{P^{-1} + \min(\mu^2, 1)} \right\} \tag{3.40}$$

where the threshold τ_M is the one that minimizes it. Under the shrinkage function $\delta_\tau(\cdot)$ it is possible to define the mean, the variance and the risk function $R_{\tau_M}(\mu)$, in L_2, as:

$$\begin{aligned} M_{\tau_M} &= E\{\delta_\tau(\mathbf{w}_m^z)\mathbf{w}_m^z\} \\ V_{\tau_M} &= \mathrm{var}\{\delta_\tau(\mathbf{w}_m^z)\mathbf{w}_m^z\} \\ R_{\tau_M}(\mu) &= E\{(\delta_\tau(\mathbf{w}_m^z)\mathbf{w}_m^z - \mu)^2\} = V_{\tau_M} + (M_{\tau_M} - \mu)^2 \end{aligned}$$

Since the denoised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the one that finds the minimum of the maximum mean square error obtained for the worst function in a given set. As for τ_S, the obtained value will be lower than the universal threshold. Consequently, more information is preserved but the denoised signal has a noisier appearance.

To better grasp how the denoising techniques are working, an example of the wavelet thresholding using a hard thresholding rule, the universal threshold, and Coiflet order 3 as a mother wavelet to decompose the signal in five scales is shown in Figure 3.7. The original signal is contaminated with noise, which can be extracted using wavelet thresholding while maintaining some high frequency coefficients at the end of the denoised signal.

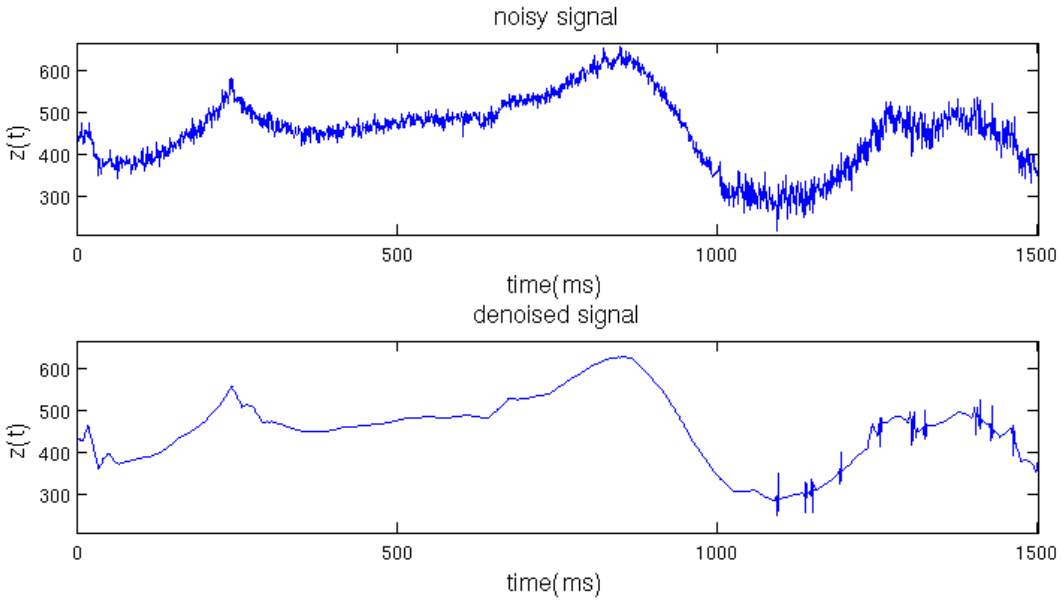


Figure 3.7: The original signal is shown at the top. The denoised version is shown at the bottom, which is obtained using the universal threshold and a hard thresholding rule. The wavelet mother is a Coiflet order 3 using five scales.

The detail coefficients of each level are shown in the left column of Figure 3.8. The number of elements in the first level (D1) corresponds to half the size of the original signal, and the subsequent levels have half coefficients than the previous one. In the right column of Figure 3.8, the detail coefficients that were not removed are shown.

There are other thresholds in the literature, such as the NeighBlock threshold [Cai and Silverman 2001], that incorporates the neighborhood information of the wavelet coefficients and

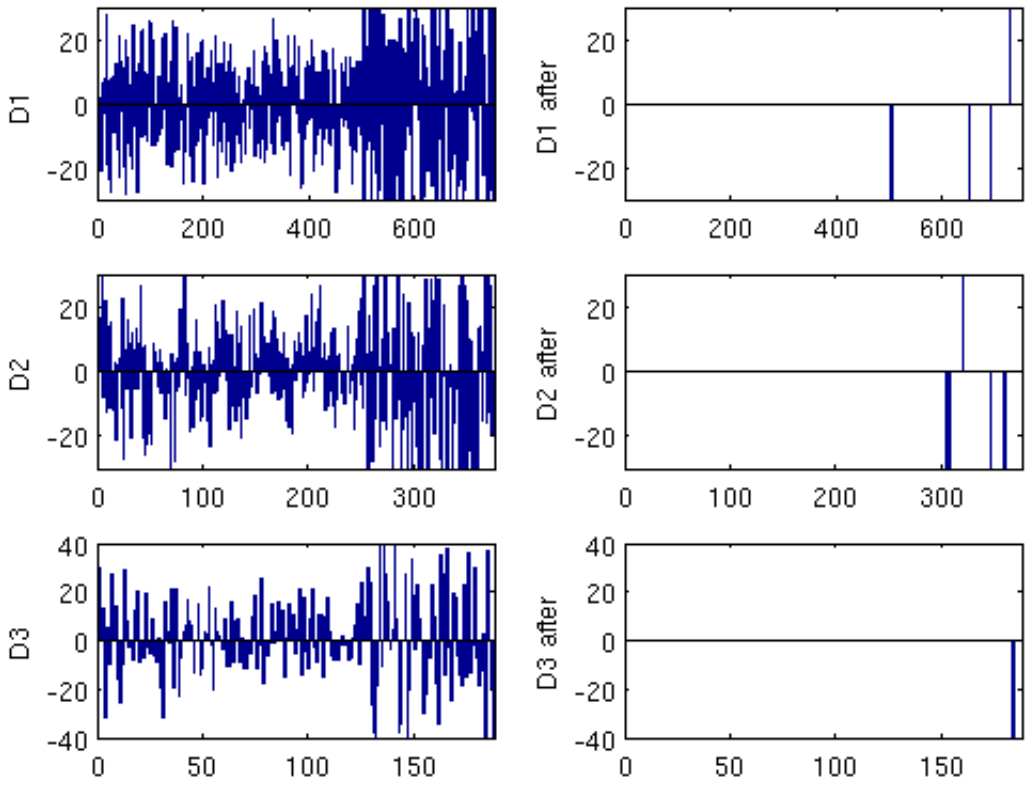


Figure 3.8: Detail coefficients by level before and after the denoising process. On the left column, the original detail coefficients and on the right column, the thresholded coefficients. The universal threshold is used, which is computed for each scale.

the Hysteresis thresholds, that uses a heuristic combination of the classic threshold presented in this chapter [Ranta and Louis-Dorr 2010].

Chapter 4

Classification

Contents

In the previous chapter, feature extraction methods were presented on behalf of their influence on pattern recognition systems. As explained before, the feature extraction process is made in order to obtain relevant information, which is particularly useful for EEG signals due to their high dimensionality, unstable measures and low SNR. The objective of the feature extraction process is to improve the performance or/and speed up the classification process. In BCI, the goal is to translate brain activity into commands for a device, which is achieved by a classification or a regression algorithm, being the classification the most used approach [McFarland et al. 2006].

The classification process can be made using two different types of training: supervised or unsupervised. In this thesis, we focus on supervised learning techniques. The main idea behind these techniques is to use the information provided by a labeled data set (called training set) to generate decision functions which are able to classify new data into different existing classes. The decision function's parameters are adjusted by the samples' information of the training set (features and labels) in order to maximize the number of well-classified elements in the training set, and hopefully to maximize the classification accuracy for future elements. In this chapter, the classifier characteristics and a taxonomy are presented, together with the main classification algorithms used for ERP detection: the stepwise linear discriminant analysis (SWLDA) and support vector machines (SVM).

4.1 Classifier Characteristics

According to [Anderson 2005], the main desirable characteristics in a classifier are:

1. *Accuracy*: This implies to achieve a minimum correctness for the classifier, with the ability to adapt to different subjects and environments without the need of adjusting the system parameters regularly, *i.e.* the system should be reliable.

2. *Fast, Responsive*: Meaning that the classifier's training should take the less computational time as possible, requiring reasonable amount of data, and making decisions quickly to provide a comfortable interaction for the user. Finally, the classifier should have low storage requirements.

3. *Interpretable*: Must be capable to explain why a particular BCI command has be selected, in addition to allow the use of feedback in real-time, to improve the system performance or the error detection.

4. *Practical*: The ultimate goal of most BCI systems is to be used outside of the laboratory (such as home) for disabled patients. For this reason, it is important to develop a system that is easy to setup, with the ability to differentiate between multiple mental states, being suitable and easy to use by the subject.

4.2 Classifier Taxonomy

Besides the desirable characteristics in a classifier, a taxonomy for BCI classifiers is presented in [Lotte et al. 2007], with the main objective of describing the different types of classifiers, which can be separated into the following categories:

- *Generative or informative classifier - Discriminative classifier*: Generative classifiers maximize the data likelihood, obtaining a statistical model of the data, such as, Bayes quadratic. The decision is made using the Bayes rule to calculate the most probable class. In contrast, discriminative classifiers learn directly a function that separates the feature space into different classes, without defining a model of how the data was generated [Ng and Jordan 2001].

- *Static classifier - Dynamic classifier*: On the one hand, static classifiers consider only a batch of data for classification, such as multilayer perceptron. On the other hand, dynamic classifiers build a classification model progressively as the data is presented. This has the advantage of not requiring re-training, but the downside relying on strong assumptions of the generating process of data, such as the Markov assumption in hidden Markov models.

- *Stable classifier - Unstable classifier*: Stable classifiers have the ability of generalizing a problem, meaning that slight changes in the data do not considerably affect the performance, diminishing the need of re-training, such as in linear discriminant analysis [Vapnik 1999]. In contrast, unstable classifiers are highly complex, such as the multilayer perceptron, and their performance can change considerably if there are small changes in the

dataset.

– *Regularized classifier*: Regularization consists in carefully penalize the model to control its complexity, prevent the influence of outliers and overtraining, obtaining a good generalization performance and robustness with respect to outliers. Regularization helps limiting the complexity of the classifier and the raggedness of the decision surface [Müller et al. 2004].

As an example of applying this taxonomy, we can categorize the SVM as a discriminative, static, stable and regularized classifier. A summary list with the properties of well-known classifiers can be found in [Lotte et al. 2007].

Among the different types of classifiers that can be used for BCI, linear classifiers achieve a classification decision based on linear functions, being the most popular algorithms used for BCI the Linear Discriminant Analysis (LDA) and Linear Support Vector Machines (Linear-SVM) [Lotte et al. 2007], which will be immediately presented in the next section. A comparison of classification techniques for the P300 speller responses can be found in [Krusienski et al. 2006].

4.3 Linear Generative Models for Classification

A linear classifier makes a decision based on linear combinations of the variables [Bishop 2006, chap. 4]. It creates hyperplanes as boundaries between classes. These classifiers are simple and often robust, which is important when working with noisy data.

Linear Discriminant Analysis (LDA) is a very popular linear pattern classification technique especially in BCI field. More precisely, articles mention several close-related techniques such as LDA, Fisher's Linear Discriminant (FLD) and Stepwise Linear Discriminant Analysis (SWLDA).

Let us present their specificities to avoid confusion. Even if these techniques are intrinsically available for a multi-class problem, we will only describe these methods for a two-class problem which is the kind of problem we are dealing with in this manuscript *i.e.* absence/presence of ERP in EEG.

4.3.1 Linear Discriminant Analysis

Assuming that the elements of a class follows a normal distribution and share a common covariance matrix Σ between classes, we have a multivariate normal probabilistic model. We also assume that variables have been centered to simplify the formulas. Let us apply the Bayes rule which is a linear rule that decides to assign a pattern $\mathbf{x}$ to the class with the highest posterior probability [Ripley 1995, p. 21, 36 and 92]:

$$\log p(k|\mathbf{x}) = -0.5(\mathbf{x} - \mu_k)^T \Sigma^{-1} (\mathbf{x} - \mu_k) + \log p(k) + \delta$$

The first term of the sum is the square of the Mahalanobis distance from the pattern to the class mean that takes into account the covariance matrix, $p(k)$ is the prior probability of class k and δ is a constant that corresponds to a sum of terms not specific to a class. Expanding the first term and removing terms non specific to each class again, we look for the following largest value:

$$\mathbf{x}^T \Sigma^{-1} \mu_k - 0.5 \mu_k^T \Sigma^{-1} \mu_k + \log p(k)$$

For a two-class problem, we can observe the difference between the value of each class:

$$\mathbf{x}^T \Sigma^{-1} (\mu_1 - \mu_2) - 0.5 \mu_1^T \Sigma^{-1} \mu_1 + 0.5 \mu_2^T \Sigma^{-1} \mu_2 + \log(p(k=1)/p(k=2))$$

Let $\mathbf{w}$ be a weight vector such as $\mathbf{w} \propto \Sigma^{-1}(\mu_1 - \mu_2)$. When $\mathbf{x}^T\mathbf{w}$ is greater than a threshold, $\mathbf{x}$ belongs to class 1 (ω_1), otherwise to class 2 (ω_2).

Thus, a linear discriminant function is found to assign a pattern $\mathbf{x}$ to one of the existing two classes according to the value of the projection of $\mathbf{x}$ onto $\mathbf{w}$. The distance to the separating hyperplane perpendicular to $\mathbf{w}$ can be represented as the linear function:

$$y(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + \mathrm{b}, \tag{4.1}$$

where, $\mathbf{x}$ is the input vector to classify, $\mathbf{w}$ is the weight vector perpendicular to the hyperplane and b is the bias which translates the hyperplane from the origin. $\mathbf{w}^{\mathrm{T}}\mathbf{x}$ is the dot product between two vectors.

Σ, μ_k and $p(k)$ are approximated by the data. If we do not make the homoscedasticity assumption, we consider a specific class covariance Σ_k for each class. Then, the classifier is no longer linear and is referred to as the quadratic discriminant analysis (QDA).

4.3.2 Fisher's Linear Discriminant

The advantages of the Fisher's Linear Discriminant (FLD) compared to LDA are that it does not assume a normal distribution neither a common covariance matrix. The idea is to seek a close projection for data belonging to the same class but far from data belonging to the other classes, which means maximizing the square difference between the means, divided by the sum of the variance of the individual classes. This is called the Fisher's criterion and can be written as the following ratio:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}, \tag{4.2}$$

where $\mathbf{w}$ is the projection vector that is a normal vector of the discriminant hyperplane, S_B is the between-class covariance matrix and S_W is the within-class covariance matrix in the input space. This matrices are computed as follows:

$$\mathbf{S}_B = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^{\mathrm{T}}$$

$$\mathbf{S}_W = \sum_{i\in\omega_1}(\mathbf{x}_i - \mathbf{m}_1)(\mathbf{x}_i - \mathbf{m}_1)^{\mathrm{T}} + \sum_{i\in\omega_2}(\mathbf{x}_i - \mathbf{m}_2)(\mathbf{x}_i - \mathbf{m}_2)^{\mathrm{T}}$$

where $\mathbf{m_1}$ and $\mathbf{m_2}$ are the mean for class ω_1 that contains N_1 elements and class ω_2 that contains N_2 elements in the input space.

$$\mathbf{m_1} = \frac{1}{\mathrm{N}_1}\sum_{i\in\omega_1}\mathbf{x}_i \qquad \text{and} \qquad \mathbf{m_2} = \frac{1}{\mathrm{N}_2}\sum_{i\in\omega_2}\mathbf{x}_i.$$

The ratio in Equation 4.2 can be maximized by keeping the denominator constant and maximizing the numerator, *i.e.* the problem can be expressed as:

$$\text{maximize}_{\{\mathbf{w}\}}\mathbf{w}^T\mathbf{S}_B\mathbf{w}$$

$$\text{subject to} \qquad \mathbf{w}^T\mathbf{S}_W\mathbf{w} = \delta,$$

where δ is a constant. Using Lagrange multipliers it is possible to define the Lagrangian as:

$$L = \mathbf{w}^T \mathbf{S}_B \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{S}_W \mathbf{w} - \delta), \tag{4.3}$$

which can be maximized with respect to both $\mathbf{w}$ and λ, differentiating with respect to w and setting the gradient $\nabla_w L$ to zero:

$$L = \mathbf{w}^T(\mathbf{S}_B - \lambda \mathbf{S}_W)\mathbf{w} + \lambda\delta$$

$$\nabla_w L = 2(\mathbf{S}_B - \lambda \mathbf{S}_W)\mathbf{w} = 0 \qquad \text{or}$$

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}.$$

If $\mathbf{S}_W$ has a full rank, *i.e.* its inverse exists, then $\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = \lambda\mathbf{w}$. The direction of $\mathbf{S}_B\mathbf{w}$ is in the same direction as $(\mathbf{m}_2 - \mathbf{m}_1)$. Considering that the magnitude λ of $\mathbf{w}$ is not important, we can obtain $\mathbf{w}^*$ as:

$$\mathbf{w}^* = S_W^{-1}(\mathbf{m}_2 - \mathbf{m}_1) \tag{4.4}$$

Once the perpendicular vector $\mathbf{w}^*$ to the hyperplane is found, a threshold θ can be fixed in order to define a classifier. For our problem with two classes, a new example $\mathbf{x}$ is assigned to the ω_1 if $y(\mathbf{x}) \geq \theta$ and to ω_2 otherwise.

Thus, for a two-class problem, FLD obtains the same result as LDA using the Fisher's criterion that does not make assumptions about a common covariance matrix and a normal distribution of classes.

For an arbitrary number C of classes, a natural extension of FLD is to use multiple discriminant analysis [Duda et al. 2001, pg. 121]. The idea is to produce a transformation into a space of dimension at most of $\mathrm{C} - 1$, where the features are ordered in terms of importance for discrimination, which can be can be grouped in a matrix, so $\mathbf{y} = \mathbf{W}^{\mathrm{T}}\mathbf{x}$.

The idea in multiple discriminant analysis is to find a subspace with bases W that maximizes:

$$J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}}, \tag{4.5}$$

where $\mathbf{S}_W$ and $\mathbf{S}_B$ are the generalization of the within- and between-class covariances. whose solution is obtained by the generalized eigenvalue problem,

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i, \tag{4.6}$$

where each $\mathbf{w}_i$ is a generalized eigenvector. In practice the generalized eigenvalues are found as the roots of the polynomial $|\mathbf{S}_B - \lambda_i \mathbf{S}_W| = 0$ and then for each eigenvalue λ_i the generalized eigenvector can be found through $(\mathbf{S}_B - \lambda_i \mathbf{S}_W)\mathbf{w}_i = 0$.

The transformation obtained does not provide a discriminative rule. For example, a multi-class classification scheme could be used to directly detect row and column codes in a P300

speller. There are several techniques to build a classifier, where the most common ones are the *one-against-all* elements not labeled as the target class are labeled as non target. In the *One-against-one* approach, which use $C(C-1)/2$ classifiers, each classifier discriminates between two existing classes. The *Discriminant functions* approach, which define C linear discriminant functions and assign $\mathbf{x}$ to the class whose discriminant function is the largest value, between others. The one against all method has a disadvantage compared with one against one method in the training stage. Training is slower, because it is made using all data for each FLD model, instead of using the data belonging to two classes. On the contrary, the one against all method is faster to classify new items, because there are less models to evaluate.

Unfortunately there are drawbacks using FLD when $J(\mathbf{w}) = 0$, wich means data have the same means, or if there are too much overlap between classes.

4.3.3 Stepwise Linear Discriminant Analysis

Emanuel Donchin has used, for a long time, the Stepwise Linear Discriminant Analysis (SWLDA) to discriminate ERP [Squires and Donchin 1976], which is an extension of the previous methods that makes variable selection [Draper and Smith 1981, p. 213]. This method has become very popular since it has been implemented in BCI2000 [17] [Schalk et al. 2004], the first general-purpose system for brain-computer interface research.

SWLDA compares the efficacy of the current model with another model constructed by adding (or removing, according to the variant) a variable, *i.e.* a dimension of the data. Different variants exist. A *forward stepwise selection* adds a new variable at each step to the previously chosen subset until a stopping criterion is met. A *backward elimination* is when all variables are considered at the beginning and the useless variables are eliminated step by step, finishing when is not longer possible to improve the model or when a stopping criterion is met. F-statistics are used to compare two models and decide which one is the best *i.e.* if the variable is added/removed or not. One basic stopping criterion halts the process when adding (resp. removing) any remaining (resp. existing) variable does not improve the performances. Stepwise methods do not guarantee to find the best possible model, because neither forward selection nor backward elimination consider all the possible subsets of variables. It is a suboptimal search. If a variable is considered important for the model; it is added or eliminated (depending on the method). Yet, this variable is not re-evaluated with the arrival of new variables into the model. For this reason a new method called *stepwise regression* was developed, which rechecks at each step the importance of all previously included variables. This meaning that it uses both previously presented methods. The SWLDA algorithm used by BCI2000 for the P300 speller is based on a combination of forward and backward stepwise regression. Algorithm 4.1 describes this method according to the matlab code and the C++ code of BCI2000 since there is no complete information available in [Krusienski et al. 2008] [18] [19].

17. `http://www.schalklab.org/research/bci2000`
18. `http://www.bci2000.org/wiki/index.php/User_Reference:P300Classifier`
19. Statistic toolbox 7, userguide, stepwisefit function p. 20.1711

Algorithm 4.1 The SWLDA algorithm

Input: A training data set $\chi = \{(\mathbf{x}_1, \omega_1), (\mathbf{x}_2, \omega_2), \ldots, (\mathbf{x}_N, \omega_N)\}$ where $\omega_i \in \{0, 1\}$.
Output: The SWLDA model

for all variables **do**
  Compute the correlation between the variable and the target class.
end for
Fit a linear model with the variable most highly correlated to the class by performing a linear regression. This model will be referenced as the current model.
repeat
  for all unused variables **do**
    Fit a model augmented by the unused variable by performing a linear regression
    Compute the F-test p-value between the current model and the augmented model.
  end for
  if the min F-test p-value of the variable is less than 0.1 **then**
    Accept the augmented model as the new current model.
    repeat
      for all variables in the model (except the last added or removed variable) **do**
        Fit a model by performing a linear regression removing the variable
        Compute the F-test p-value between the current model and the reduced model.
      end for
      if the max F-test p-value of the variable is greater than 0.15 **then**
        Accept the reduced model as the new model
      end if
    until no variable are included in the model or no variable has been removed
  end if
until The pre-defined maximum number of variables to used in the model has been reached or no variable has been added

The stopping criterion defines the variables subset size. Several criteria have been studied to be used as stooping criterion [Hocking 1976; Bendel and Afifi 1977]. Some of the most common stopping rules are the F-test, the coefficient of determination R^2 and the residual mean square. A disadvantage concerning the stopping criterion is to find only a part of the best model, since the stopping criterion is met before that all important variables are found.

In summary, two modified versions of LDA are used to classify ERP signals: (1) Fisher's Linear Discriminant (FLD) [Mika et al. 1999], which is based on the Fisher's criterion, defined as the ratio of the variance between the classes to the variance within the classes; and (2) Stepwise linear discriminant analysis (SWLDA) [Draper and Smith 1981], which selects suitable predictor variables to be included in a multiple regression model, that can be used to discriminate responses. A study comparing FLD and SWLDA can be found in [Krusienski et al. 2006], where it is shown that SWLDA is more stable than FLD when there are more features than training samples, because SWLDA constraints the number of features by removing insignificant terms.

4.4 Support Vector Machines

Support Vector Machines (SVM) are supervised learning machines for classification and regression problems, based on the statistical learning theory developed by Vapnik in the seventies [Vapnik 1982]. In the last years, SVM have become very popular mainly due to the use of kernels functions, that make them capable to work directly with non-linear problems. Also, the final model can be expressed using only the support vectors found in the training, and only a few parameters must be chosen.

Support vector machines were designed as classifiers to be used in binary problems, meaning that the items to be classified belong to one of two categories [Cortes and Vapnik 1995]. This binary classification is done through a separating hyperplane, in a high dimensional feature space if necessary, that leaves aside the elements of one class from the elements of the other class. This section presents an introductory SVM review focused on pattern recognition problems.

Formally, if a classification task is performed for a set of N elements $\{(\mathbf{x}_i, y_i) | i \in \mathcal{I} = \{1, \ldots, N\}\}$, where each example has n inputs, *i.e.* $\mathbf{x}_i \in \mathbb{R}^n$ belonging to two different classes $y_i \in \{-1, +1\}$. The separating hyperplane constructed by the SVM and represented by $(\mathbf{w}, \mathrm{b})$ (Equation 4.1) can be equally expressed by all pairs $\{\lambda\mathbf{w}, \lambda\mathrm{b}\}$, since $\lambda \in \mathbb{R}^+$ is a scaling parameter. In order to define a *canonical hyperplane*, which separates data, it must satisfy two constraints for all $i \in \mathcal{I}$:

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq +1 \qquad \text{when } y_i = +1 \tag{4.7}$$

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \qquad \text{when } y_i = -1 \tag{4.8}$$

such that, the elements from one class are in one side of the hyperplane and the elements of the other class are in the opposite side and at least one example on both sides has a distance of 1 to the hyperplane. The constraints 4.7 and 4.8 can be rewritten in one constraint as:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1, \quad \forall i \in \mathcal{I} \tag{4.9}$$

Then, the classification function is defined as:

$$f_{\mathbf{w},b}(\mathbf{x}) = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x} + b). \tag{4.10}$$

The SVM maximizes the margin between the classes in the space of characteristics [Boser et al. 1992], if the nearest examples to the hyperplane have a distance of ± 1 (depending on the class) then the geometric distance between a data point $\mathbf{x}_i$ and the hyperplane is:

$$d\left((\mathbf{w}, b), \mathbf{x_i}\right) = \frac{y_i(\mathbf{x}_i \cdot \mathbf{w} + b)}{\parallel \mathbf{w} \parallel} \geq \frac{1}{\parallel \mathbf{w} \parallel}, \tag{4.11}$$

implying that the margin of the hyperplane ρ can be defined as twice the distance from the hyperplane to the nearest example (or examples), *i.e.* $\rho = 2/ \parallel \mathbf{w} \parallel$. As in any supervised machine learning technique, the training stage is conducted to find the classification function parameters $\mathbf{w}$ and b, which in this case maximize the margin width.

4.4.1 Margin Maximization

If the problem is linearly separable, there are an infinite number of hyperplanes that solve the problem, as shown in Figure 4.1(a), but only one maximizes the margin *i.e.* the distance between hyperplanes $\mathcal{H}_1 : \mathbf{w} \cdot \mathbf{x} + b = 1$ and $\mathcal{H}_2 : \mathbf{w} \cdot \mathbf{x} + b = -1$, as shown in Figure 4.1(b). The examples that lie on hyperplanes $\mathcal{H}_1$ and $\mathcal{H}_2$ are called support vectors (SVs).

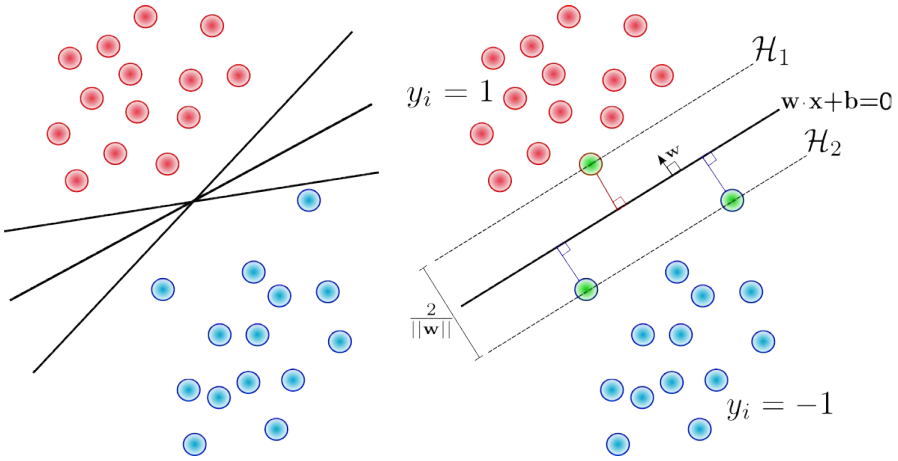


Figure 4.1: Margin Maximization: For a binary class problem, (a) several possibilities exist for the separating hyperplane that solves the problem. (b) The hyperplane that maximizes the margin between classes, together with the elements involved in the mathematical formulation of the hyperplane are drawn. The support vectors are located on the margin.

In order to maximize the margin distance $\rho = 2/ \parallel \mathbf{w} \parallel$, it is possible to minimize $\parallel \mathbf{w} \parallel$. For this reason the formulation to find the hyperplane can be defined as:

$$\begin{aligned} J(\mathbf{w}) : \textbf{minimize} \quad & \parallel \mathbf{w} \parallel \\ \text{subject to} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \qquad \forall i \in \mathcal{I}, \end{aligned} \tag{4.12}$$

Since this problem is difficult to solve, because of the square root in the norm, it is possible to reformulate the problem changing $\parallel \mathbf{w} \parallel$ by $\frac{1}{2} \parallel \mathbf{w} \parallel^2$, which is easier to solve and has the same minimizer as 4.12:

$$J(\mathbf{w}) : \textbf{minimize} \quad \frac{1}{2} \parallel \mathbf{w} \parallel^2 \tag{4.13}$$

$$\text{subject to} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \qquad \forall i \in \mathcal{I}. \tag{4.14}$$

This formulation ensures that each example is correctly classified, since the problem is considered linearly separable. In practice, the problem is often not linearly separable and even if it is, a better margin is achieved if the classifier can misclassify some examples. In order to allow errors, the constraint 4.14 is replaced by:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \qquad \forall i \in \mathcal{I}, \tag{4.15}$$

where the variables ξ_i are called *slack variables* and they allow an example to be in the margin if $0 \leq \xi_i \leq 1$ or misclassified if $\xi_i > 1$. To avoid a large number of misclassified data, the penalty term $P \sum_{i \in \mathcal{I}} \xi_i$ is added to $J(\mathbf{w})$, which gives us the following optimization problem:

$$J(\mathbf{w}) : \textbf{minimize} \quad \frac{1}{2} \parallel \mathbf{w} \parallel^2 + P \sum_{i \in \mathcal{I}} \xi_i \tag{4.16}$$
$$\text{subject to} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \quad \forall i \in \mathcal{I}.$$

The constant P corresponds to the compromise between maximizing the margin and minimizing the amount of slack.
There are two different alternatives to solve the problem. The first one is to solve directly the *primal* optimization problem [Chapelle 2007] (*i.e.* directly addressing $J(\mathbf{w})$). However, what is usually done is to use the technique of Lagrange multipliers and solve the *dual* optimization problem (dual formulation) [Schölkopf and Smola 2002], which is expressed in terms of α_i as follows:

$$\textbf{maximize}_{\{\alpha\}} D(\alpha) = \sum_{i \in \mathcal{I}} \alpha_i - \frac{1}{2} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j), \tag{4.17}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} \alpha_i y_i = 0, \tag{4.18}$$
$$0 \leq \alpha_i \leq P, \qquad \forall i \in \mathcal{I} \tag{4.19}$$

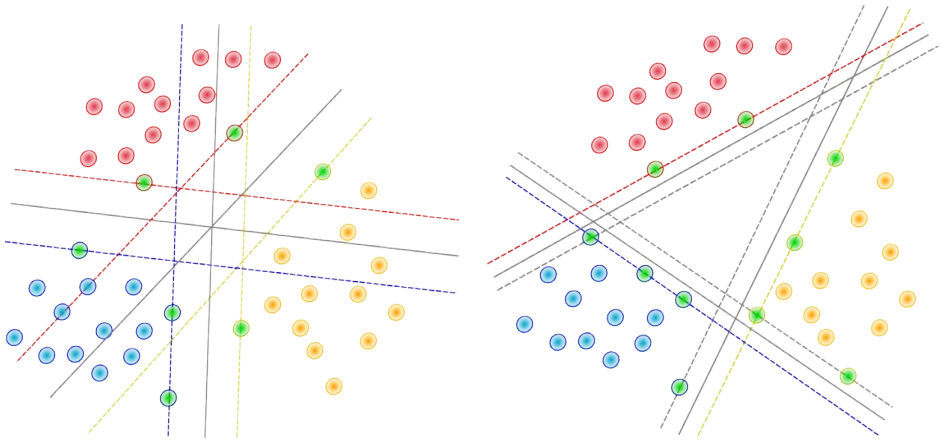
The dual problem allows to solve the primal one because the dimension of the dual problem is finite (and equal to the number of examples) whereas the primal problem has infinite dimension. In fact, the discriminant function 4.10 can be written as:

$$y(x) = \sum_{i \in \mathcal{I}} \alpha_i y_i \mathbf{x}_i + b. \tag{4.20}$$

Only the support vectors $\mathbf{x}_i$ verify the condition $\alpha_i > 0$, and they define the limits of the margin and the separating hyperplane. In Figure 4.1(b), the SVs are represented as green disks.

The original SVM models are not directly applicable to classification problems with more than two classes, for this reason several approaches have been proposed. As presented in the FLD's section, *one-against-all* and *one-against-one* approaches can be used with SVM to obtain a multi-class support vector machine based on binary SVMs. Figure 4.2 shows the difference in the final model using these two methods in a problem with C = 3.

There are others approaches derived from the "one-against-..." methods, such as, the error-correcting codes [Dietterich and Bakiri 1995], employed as a distributed output representation or the Directed Acyclic Graph SVM (DAGSVM) [Platt et al. 2000], which is based on the one-against-one method, but generates a decision tree with $C(C-1)/2$ decision nodes. The final class decision is derived from the final leaf node obtained through an evaluation path.



(a) (b)

Figure 4.2: Multi-class SVM: The multi-class problem is decomposed into binary problems. Image (a) presents the *one-against-one* method, which uses binary SVMs to separate each pair of classes. Image (b) presents the *one-against-all* method, which uses one SVM per class to separate it from the others. The margin colors indicate the class associated to the hyperplane and the support vectors are shown as green disks.

It is also possible to modify the SVM structure to obtain an intrinsic multi-class SVM, in order to use only one classifier to classify more than two classes simultaneously. Thus, it allows to train only one classifier instead of several ones. But the training algorithm is specific for a certain SVM design, making their use more complex. Some examples of these type of algorithms can be found in [Crammer and Singer 2001; Lee et al. 2004].

Some comparison of multi-class vector machines methods can be found in [Hsu and Lin 2002; Duan and Keerthi 2005; Guermeur 2012].

4.5 Linear Classifiers using Kernel functions

In some cases, the items are not linearly separable and a non-linear classifier provides a better solution to this problem. It is possible to extend linear classifiers to generate non-linear classifiers without explicitly computing a mapping of the data into a high dimensional feature space, which can be separated linearly. These classifiers are obtained applying the called "Kernel Trick" [Aizerman et al. 1964], where the mapping is made by replacing the dot product by a kernel function $\mathcal{K} = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$, where $\varphi(\cdot)$ is a transformation that translates the input data into a high dimensional space. The advantage of this method is that it is only needed to define $\mathcal{K}$, without projecting the data into the new high dimensional space (see Figure 4.3). The most common kernel functions are linear, polynomial and radial basis, defined as:

- Linear: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$
- Polynomial: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \cdot \mathbf{x}_i^{\mathrm{T}} \cdot \mathbf{x}_j + r)^d, \quad \gamma > 0$
- Radial basis function: $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma > 0,$

where γ, r and d are parameters.

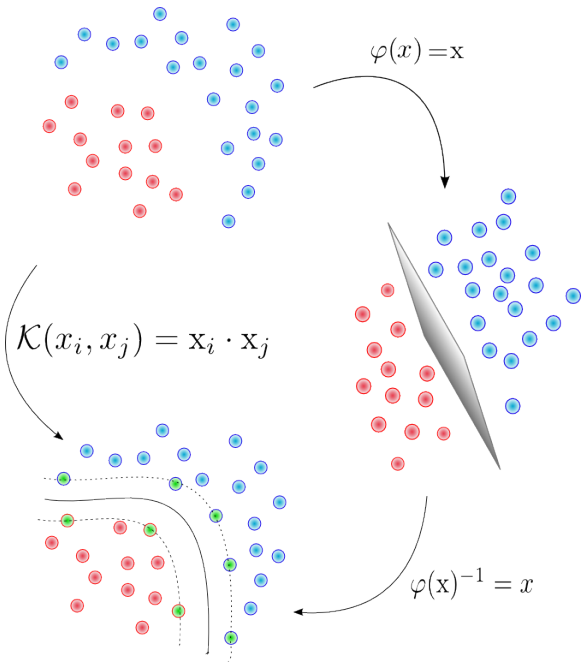


Figure 4.3: Kernel Trick: Not linearly separable data. Data are projected into a high dimensional space, where they are linearly separable.

4.5.1 Support Vector Machines and Kernels

To use the kernel trick in SVMs, it is only necessary to replace the inner product $\mathbf{x}_i \cdot \mathbf{x}_j$ by the kernel function $\mathcal{K}(x_i, x_j)$ in the dual problem described by Equation 4.19:

$$\textbf{maximize}_{\{\alpha\}} D(\alpha) = \sum_{i \in \mathcal{I}} \alpha_i - \frac{1}{2} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I}} y_i y_j \alpha_i \alpha_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \tag{4.21}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} \alpha_i y_i = 0, \tag{4.22}$$

$$\alpha_i \geq 0, \qquad \forall i \in \mathcal{I} \tag{4.23}$$

4.5.2 Fisher's Linear Discriminant and Kernels

The kernel trick can also be used in FLD, allowing the use of non-linear mappings. The FLD can be extended modifying the $J(\mathbf{w})$ function by reformulating the problem in terms of dot products as follows:

$$J(\alpha) = \frac{\alpha^{\mathrm{T}} \mathbf{M} \alpha}{\alpha^{\mathrm{T}} \mathbf{N} \alpha}, \tag{4.24}$$

where, the relationship between $\mathbf{w}$ and α_i is given by:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i). \tag{4.25}$$

and $\mathbf{M}$ and $\mathbf{N}$ are defined as follows:

$$\mathbf{M} = (\mathbf{M}_2 - \mathbf{M}_1)(\mathbf{M}_2 - \mathbf{M}_1)^{\mathrm{T}}, \qquad (\mathbf{M}_c)_i = \frac{1}{N_c} \sum_{\mathbf{x}_j \in \omega_c} \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j), \qquad \forall i \in \mathcal{I} \tag{4.26}$$

$$\mathbf{N} = \mathbf{N}_1 + \mathbf{N}_2, \qquad \mathbf{N}_c = \mathbf{K}_c(\mathbf{I} - \mathbf{1}_{N_c})\mathbf{K}_c^{\mathrm{T}}. \tag{4.27}$$

c corresponds to the index of the class. The matrix $\mathbf{K}_c$ is defined for indexes i and j as $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{I}$ is the identity matrix and $\mathbf{1}_{N_c}$ the matrix with all entries equal to $1/N_c$. For more details see [Mika et al. 1999].

Despite the fact there are many other classification techniques, the ones presented in this chapter are robust techniques widely used in BCI systems. Thus, the stepwise linear discriminant analysis (SWLDA) and support vector machines (SVM) with a linear kernel will be used as classifiers in our processing chain. The next chapter will present our contributions on signal processing.

Chapter 5

Wavelet-based Semblance methods for enhanced ERPs detection

Contents

Studying the waveform of event-related potentials (ERP) can help to understand brain's functions and detect pathologies (see Section 2.3.4). Also, the mere detection of the presence or absence of the ERP is a useful way of interaction in brain-computer interfaces. In these systems, the stimuli are usually repeated in order to average the EEG responses and enhance the ERP to make easier the detection. The idea is to reduce the ERP variability and increase its amplitude, which obviously improves the certainty about the target BCI command. However, these repetitions can become tedious and tiring for the user, besides the evident decrease of the system's speed caused by this averaging technique.

Several studies have shown that it is possible, yet difficult, to distinguish ERPs in single trials from the EEG background [Quiroga and Garcia 2003; Yong et al. 2005; Gupta et al. 2009; Saavedra et al. 2011; Saavedra and Bougrain 2010]. To successfully reduce the number of repetitions, different signal-analysis techniques for EEG can be applied, each one useful in their own merit. In fact, this choice depends on the unanswered questions in which researchers are currently interested. It is important to accept that there is no bulletproof technique to perform EEG analysis, but these techniques, as a whole, are helping to unify the understanding of the brain. This open a new research line to better understand the brain and to improve the performance of current ERP-based BCI systems, such as spellers and control devices.

Through this chapter, novel methods to denoise, localize and isolate ERPs are presented. They are based on the correlations between the recorded channels, to improve ERP detection and, as a consequence, to improve current systems based on this type of signals. More precisely,

these methods are based on the similarity measures developed for wavelet theory, mainly because it is possible to study the temporal and spectral information of a signal simultaneously, meaning that signals will be studied in two domains: the time domain and the frequency domain. Although the methods presented in this thesis are focused on ERPs, these methods can be applied to other brain signals or even to other research areas with similar requirements.

5.1 Studying ERP responses based on Wavelet Semblance Measures

In this section, we will explain how the wavelet semblance measures are helpful in the study of ERPs. We will illustrate the capabilities of the similarity measures presented in Sections 3.2.3 and 3.2.4, with the help of simple examples.

5.1.1 Semblance Measure

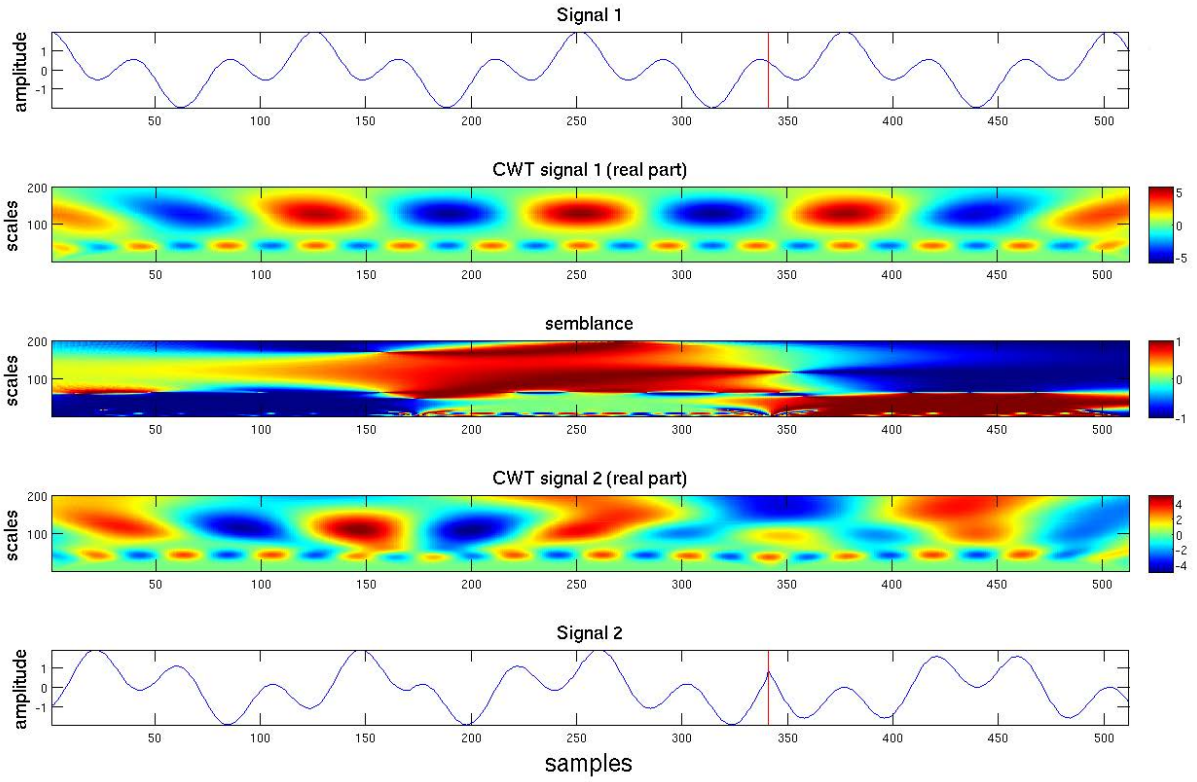


Figure 5.1: Semblance of two synthetic signals. Each signal is shown along with the real part of its CWT. At the center position, the semblance between the two signals is shown. S ranges from -1 to 1, where $S = -1$ means signals are inversely correlated, $S = 0$ uncorrelated and $S = 1$ correlated. The semblance was computed using $n = 1$ (see Section 3.2.3). The red line divides the signals in two parts with different semblances.

The first example uses the semblance measure (based on CWT) for two synthetic signals,

allowing to visualize their relationship using heat maps (see Figure 5.1). The signals, their respective wavelet transforms and the semblance measure are shown. Each signal is composed by two cosines waves, as it can be seen in the repetitive patterns of the CWTs transforms. On the left side of the red line, the semblance of the cosine component with the larger scale increases as we approach to the red line, meaning that the signals become more similar as the time increases. This can be verified by looking to the actual signals. In contrast, the cosine component with the smaller scale starts inversely correlated and becomes uncorrelated near the red line.

In the right side of the red line, the semblance image shows that the cosine with the larger scale is inversely correlated, since the semblance is low in this part of the image, implying a difference of 180 degrees in their phase. To the contrary, the cosine with the smaller scale is fully correlated in both signals, because the semblance exhibits high values for this portion of the image. It is important to note that the limitations of the semblance measure are the same than for wavelets transform, *i.e.* that at short scales the semblance is more clear or sharp than at long scales. To conclude, the semblance measure is a new technique that allows to study the correlation between two signals in the time-frequency domain.

5.1.2 MRL Measure

As explained in Section 3.2.4, the MRL measure is the extension of the semblance to multiple signals. Let's consider Figure 5.2, where 30 artificial signals generated according to the phase-resetting theory [20] [Klimesch et al. 2007] are shown along with their MRL. This last one is computed using Equation 3.31 and its coefficients range from 0 (uncorrelated signals) to 1 (correlated signals). The MRL shows that the signals are correlated for almost all scales around time $t = 200$ ms, and that they are uncorrelated before and after ($t = 0$, $t = 400$). Even though this can be easily detected using naked eye in the actual signals, it is difficult to identify the correlation in other points of the time axis (*e.g.* $t = 100$). This implies that it is possible to distinguish single-trial ERPs using the MRL similarity measure.

The next example presents, in Figure 5.3, the averaged event-related responses on 14 different channels from one subject in an auditory task [Makeig et al. 1997], and their MRL coefficients. In this case, the differences are smaller than in the previous example. The highest values of the MRL coefficients are located in the first part of the signals, decreasing its values with the time, and arriving to some zones not correlated at different times and scales. Clearly, in this example signals are highly correlated, because they are the result of the average of single trials, which as stated before removes noise and allows the ERP visualization using naked eye. It is possible to see that the auditory ERP is located (around the 100 ms) and the similarity is stronger in this part of the image than in the rest of it. The signals are correlated not only for the ERP but for most of the MRL time because channels are recording mostly the same information, namely the EEG background and maybe some artifacts.

Through these examples, we have shown that the semblance and MRL measures enrich the analysis of EEG and ERP signals, allowing to study the correlation between the stationary and non-stationary data recorded by different channels. The rest of this chapter will explore some of the possible applications of this theory to ERP detection and BCI systems.

5.1.3 D Measure

According to recent literature, there is no consensus on the most appropriate pre-processing to improve ERP detection in BCI systems. Research publications refer to various low-pass filters,

20. http://www.cs.bris.ac.uk/ rafal/phasereset/

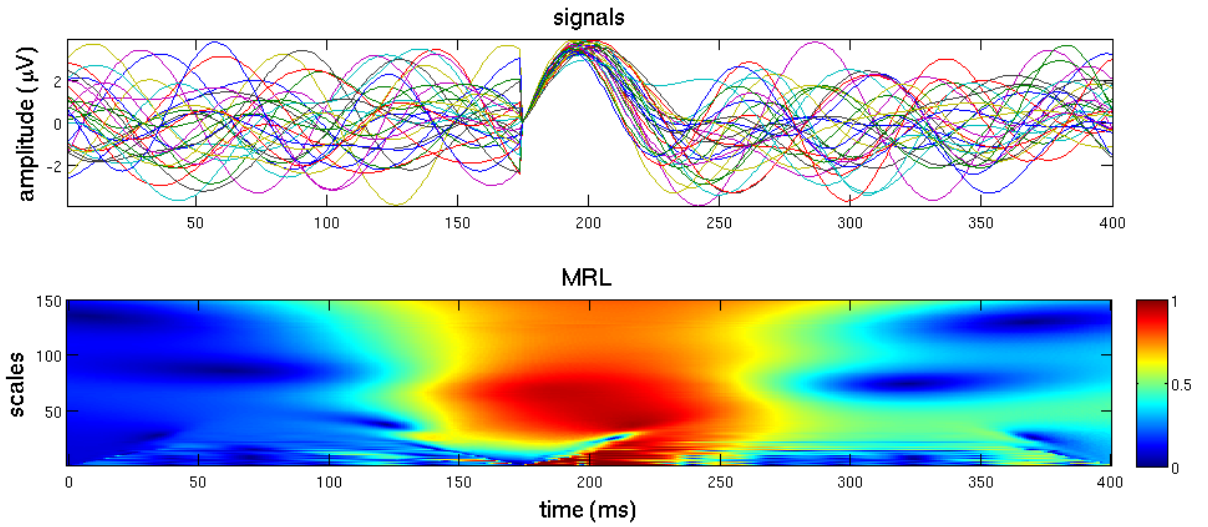


Figure 5.2: MRL of 30 artificial signals corresponding to the phase-resetting theory to generate ERPs. MRL coefficients range from 0 (uncorrelated) to 1 (correlated).

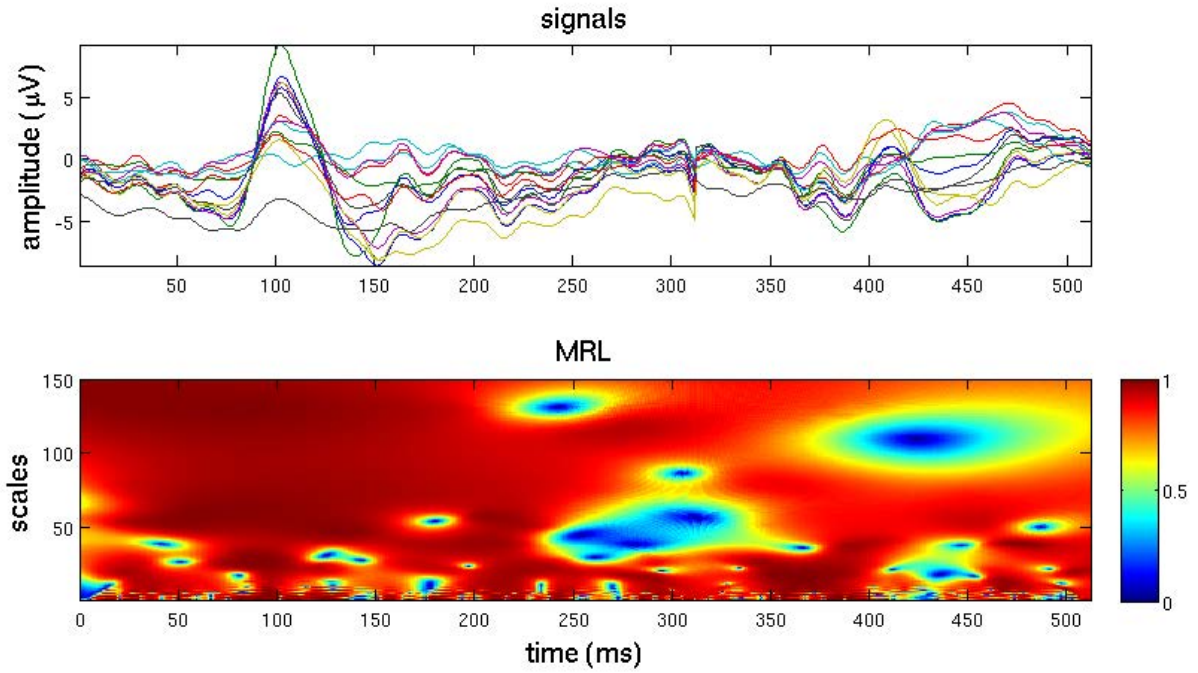


Figure 5.3: The top image corresponds to 14 channels of the averaged event-related responses for one subject in an auditory task. The bottom image shows the MRL coefficients ranging from 0 (uncorrelated) to 1 (correlated).

high-pass filters, baseline, subsampling or feature selection.

Let us compare the different pre-processing stages used by the community, for the P300 odd-

ball paradigm such as used in the speller by Farwell and Donchin [Farwell and Donchin 1988]. As explained before, the main problem faced when detecting a P300 is that the signal-to-noise ratio is so low that it is necessary to apply preprocessing techniques to increase the amplitude of the P300 and reduce the EEG background activity. Pre-processing is a well known and important step in such systems, even when averaging is applied. But practices still diverge since Farwell and Donchin proposed to use a band-pass filter from 0.02 to 35 Hz [Bougrain et al. 2012].

Recently, 27 articles on P300 detection were published in the proceedings of the BCI conference held in Graz in September 2011. These studies present a set of cutoff frequencies (see Figure 5.5). On one side, we observe that no specific value is clearly used for a high-pass filter, even though a 0.1 value is the most commonly selected. On the other side, a 30 Hz low-pass filter seems to be the most commonly used value, but the spread is also important. In this thesis a study is performed using a database with 22 healthy subjects to compare the effect on the letter accuracy (single-trial detection) based on different filters to pre-process input signals. These results are presented in Section 6.2.1.

The influence of the use of filters can be studied using the semblance measures, through the example shown in Figure 5.4. In this example, different low cutoff frequencies (60 Hz, 30 Hz, 20 Hz, 10 Hz and 8 Hz) were applied to the original signals, to then compute the grand averages for the target signals and non-target signals over channels. The dot product D, *i.e.* the semblance between these two signals including the amplitude information, was computed (see Equation 3.30 page 52) and plotted in Figure 5.4. The ERP corresponds to the blue zone in the heat map, since it is present only in one of the two signals. The red zones indicate that the same information is present in both signals, which can make difficult the identification of the ERP. Using strong cutoff frequencies allows to remove correlated components making easier the identification of the ERP, as using a 20 Hz filter. For a 10 Hz filter, the ERP is less defined, but there is no components that may hinder its identification. The difficulty is to predict the compromise between removing useless components and ERP components, as for example using a 8 Hz filter, where the ERP is very difficult to see.

5.2 Denoising based on Channels Similarity

EEG recordings are usually overlapped with noise and artifacts like muscle activity. Their presence hinders the EEG detection, requiring novel methods to remove them from the underlying true brain signal. One important assumption about noise is that it is supposed to be independent from brain activity. Also it is assumed that brain signals are recorded (almost) instantaneously by the electrodes, implying that each recorded channel is highly correlated with the others. Following this thought, it is possible to conclude that signal components that are not correlated over channels are assumed to be noise or artifacts, and can be removed from the recorded signals.

As shown in Section 3, the fundamental hypothesis of wavelet denoising is that large coefficients correspond to signal and small coefficients correspond to noise. Therefore, noise canceling can be performed by thresholding [Antoniadis 2007]; only large coefficients will then be used to reconstruct the informative signal. The problem with current methods is that they denoise one channel at a time, regardless the information presented in others channels, wasting the information provided by the ensemble (such as, phase and amplitude information). For this reason this thesis proposes the use in the wavelet domain of all channels information, based on their phase angles correlations. For this purpose, the methods introduced in Section 3.2.3 are integrated in

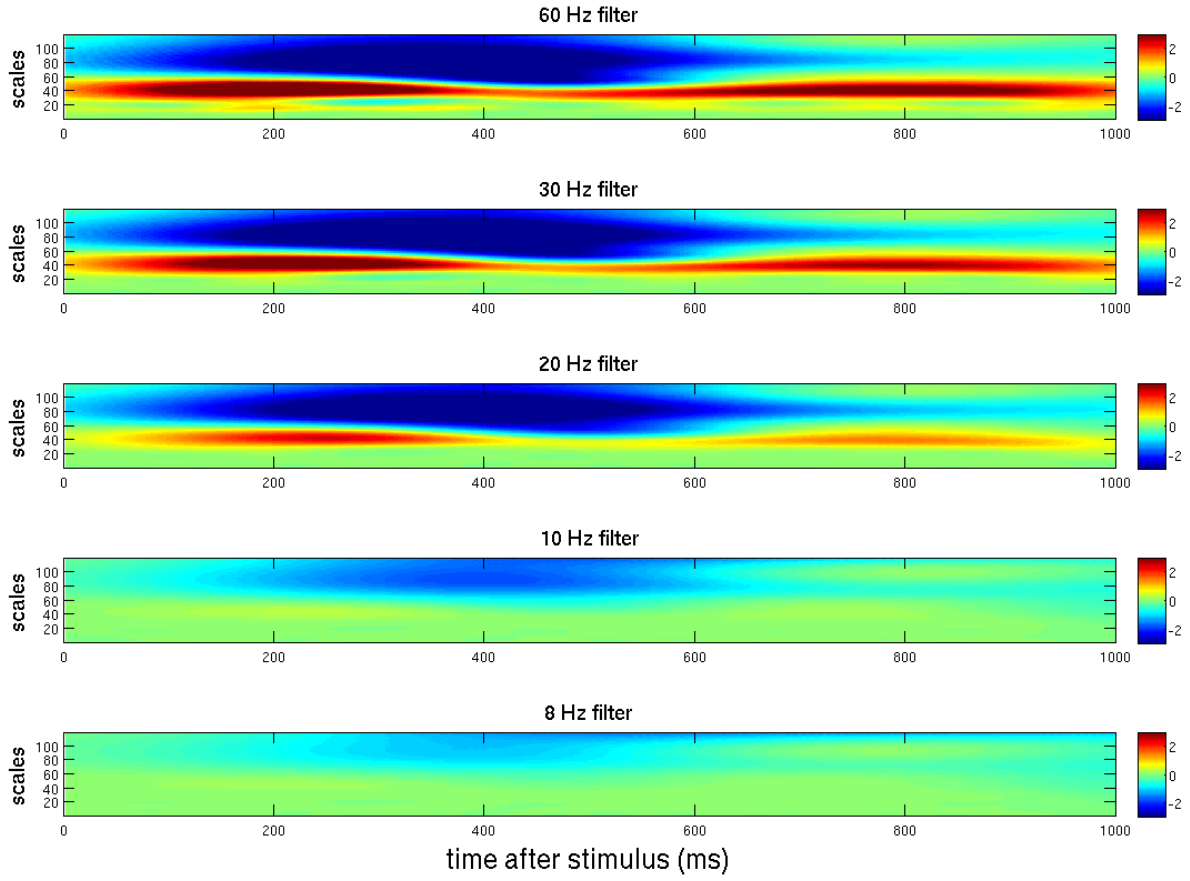


Figure 5.4: Low cutoff Frequencies influence in the ERP detection. The dot product D for the grand averages of target and non-target signals is presented. Each image corresponds to a different low cutoff frequency applied to the originals signals as a pre-processing. The blue (resp. red) zones represent different (resp. the same) information in both signals.

the denoising process.

Formally, let $\mathbf{X}$ be the whole data and $x_c(t)$ be the signal recorded by the c^{th} electrode, $c \in \{1, ..., C\}$, at time t, $t \in \{1, ..., T\}$. The matrix of recorded EEG signals can be defined as $\mathbf{X} \in \mathbb{R}^{T \mathbf{x} C}$. The denoising through thresholding can be done using the Mean Resultant Length coefficients, *i.e.* using coefficients obtained for all channels, instead of the individual channels coefficients. In order to do this, the MRL is computed using the wavelet decomposition of all channels $W_{\psi}^{x_c}$, for a mother wavelet ψ, $\forall c$ through Equation 3.31 page 52. After this, it is possible to set to zero all coefficients that are below the given threshold τ_d and finally to reconstruct the signal using the filtered wavelet coefficients. The MRL computation is made through the combination of the phase angles of the real and imaginary parts of the wavelet decomposition. DWT uses wavelets families that are orthogonal to each other, so the imaginary part can be obtained with the Hilbert transform of the channel [Cooper 2009].

In simple words, we keep those components with high similarity between channels to produce a denoised EEG signal. This novel approach is called *Multi-channel EEG Thresholding by*

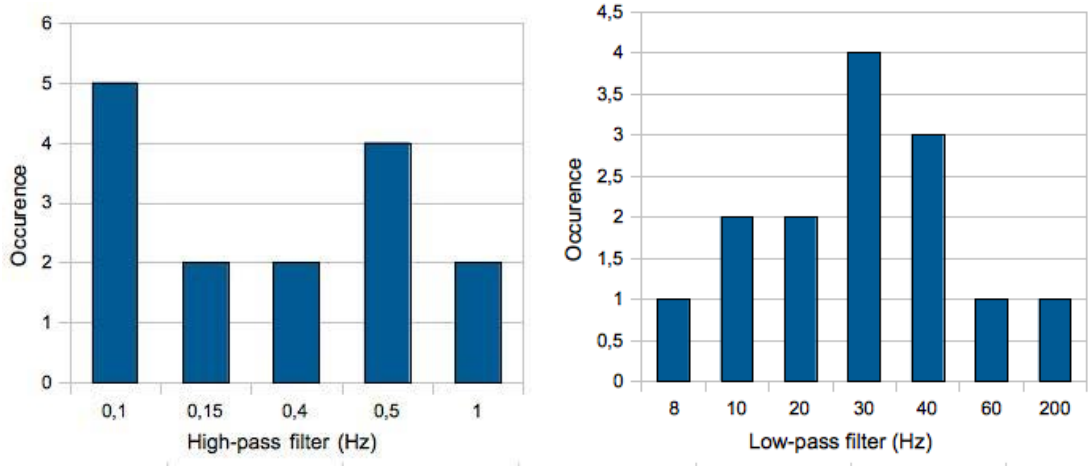


Figure 5.5: Cutoff frequencies commonly used with the oddball paradigm in BCI: the left (resp. right) figure presents the occurrence of different high-pass (resp. low-pass) filters.

Similarity (METS), and the full wavelet denoising process is described in Algorithm 5.1.

Algorithm 5.1 Multi-channel EEG Thresholding by Similarity (METS)

Input: Given the EEG signal matrix $\mathbf{X}$, with C channels and T temporal samples.
Output: The denoised signals $\widetilde{\mathbf{X}}$

Set the correlation threshold τ_d, $0 \leq \tau_d \leq 1$
for $c = 1$ to C **do**
    Compute the Hilbert transform H_c of x_c.
    Compute the DWT $W_\psi^{x_c}$ of signal x_c and $W_\psi^{H_c}$ of H_c using Equation 3.20 page 49.
end for
for $t = 1$ to T **do**
    Compute the MRL(t) using Equation 3.31 page 52
    if $MRL(t) < \tau_d$ **then**
        set to zero $W_\psi^{x_c}$ at time t, $\forall c$
    end if
end for
for $c = 1$ to C **do**
    Reconstruct the signal for channel c using the new $W_\psi^{x_c}$.
end for

An example of our denoised method can be seen in Figure 5.6, showing how a signal is better visualized after a denoising procedure, making the signal less blurred.

The METS algorithm is generic enough to be applied to any multi-channel (EEG) signals. But it is also flexible enough to exploit the specific properties of the analyzed phenomenon through the parameter selection. In fact, one important parameter is the mother wavelet. In practice, the wavelet methods are influenced by the choice of the mother wavelet. Yet unfortunately there is no formal method to choose a mother wavelet. It can be chosen by its resemblance in shape to the underlaying target phenomenon or because of their properties. In ERPs studies, the most common mother wavelet used are the quadratic B-spline [Quiroga and Garcia 2003; Demiralp

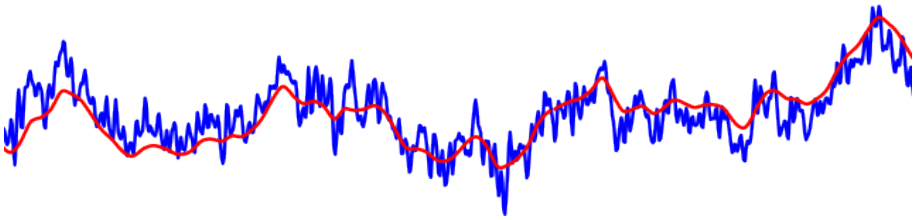


Figure 5.6: Difference between the original signal and its denoised version using METS. The shown segment corresponds to 3 seconds.

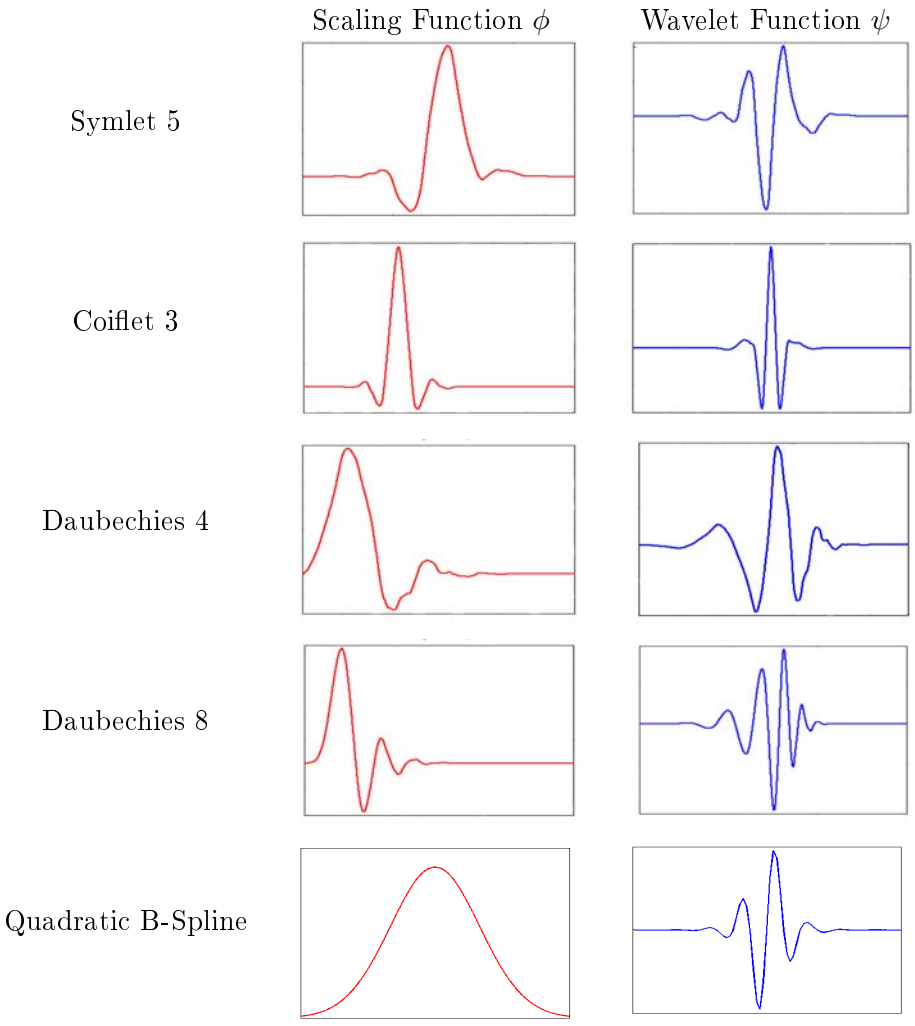


Figure 5.7: Mother wavelets used for the study of ERPs.

et al. 2001; Başar et al. 2001], the Daubechies wavelets Db4 and Db8 [Polikar et al. 2007], the Symlet wavelet [Polikar et al. 2007; Dear and Hart 1999] and Coiflet [Yong et al. 2005]. The wavelets shapes are shown in Figure 5.7. The impact in the selection of the mother wavelet is studied in Section 6.2.3.

5.3 Temporal Window Selection

In the vast majority of BCI applications, the ERP responses are studied during a predefined temporal window starting after the stimulus onset. Usually this window is chosen large enough to be sure to include the ERP components under study. However, the ERP responses have different latencies (and amplitudes) for each subject, causing irrelevant data to be included in the temporal window. This can increase both the difficulty of training a classifier with irrelevant variables and the complexity of detecting the ERP. Figure 5.8 shows the difference in latency for three different subjects from the UAM database [21]. The P300 peaks differs in time and width for the three subjects.

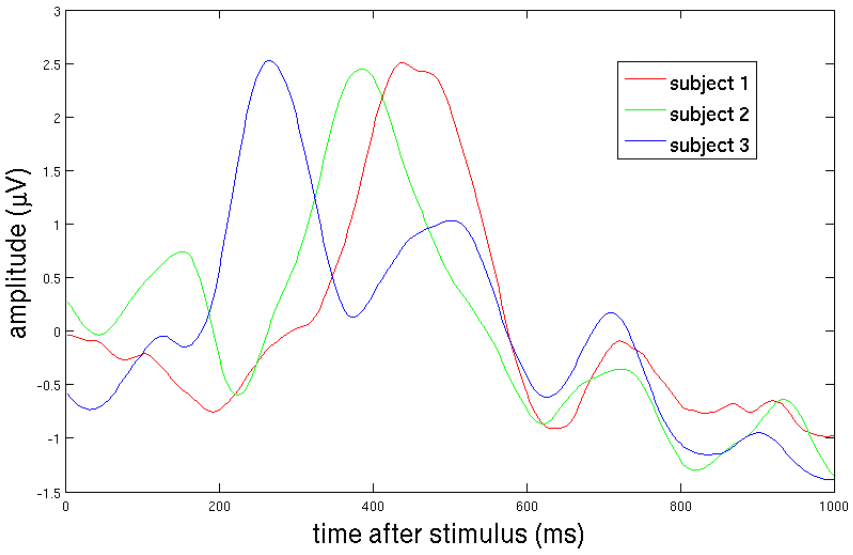


Figure 5.8: The difference in latency for three different subjects from the UAM database (ICE, WFG and XCL) spelling the same letters with a P300 speller is shown.

This problem can be found even for the same subject in different sessions, as shown in Figure 5.9. In this case, the variability of the P300, in average for channel Cz, shows the ERP response shape and latency changes between two sessions.

The straightforward solution is to consider a variable time window, though this poses new challenges. The first one is to correctly detect when the ERP begins and finishes. Also, the low SNR of the recordings impedes the isolated detection in single-trial. To select this temporal window for each subject, we propose to compare the averages of target and non-target signals by channel or the grand averages over channels, through the wavelet-based semblance measures introduced in Section 3. The idea is to have a better expressiveness of both amplitude and phase in order to isolate the ERP wave. The hypothesis is that the ERP is not correlated with the EEG background activity, and this should be reflected in the semblance analysis.

In Figure 5.10, the difference between the grand averages of target signals and non-target signals over channels for a P300 ERP are shown. In this case the initial temporal window is 1 second duration, but it is possible to reduce the time window to where the P300 actually lies, for example approximately between 200 and 600 ms. Even though the P300 name states that the

21. `http://akimpech.izt.uam.mx/p300db/doku.php?id=deposits.en`

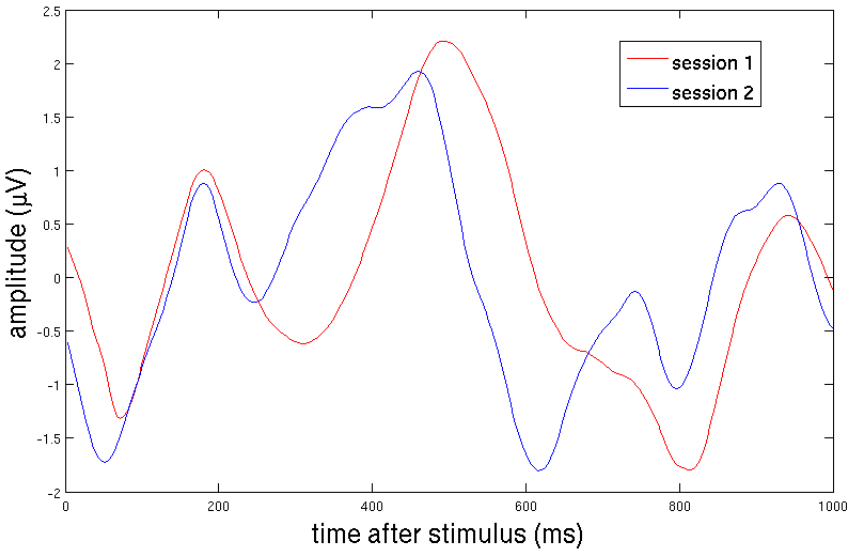


Figure 5.9: The difference between two P300 averages for channel Cz in different sessions for the same subject from the UAM database (GCE) is shown.

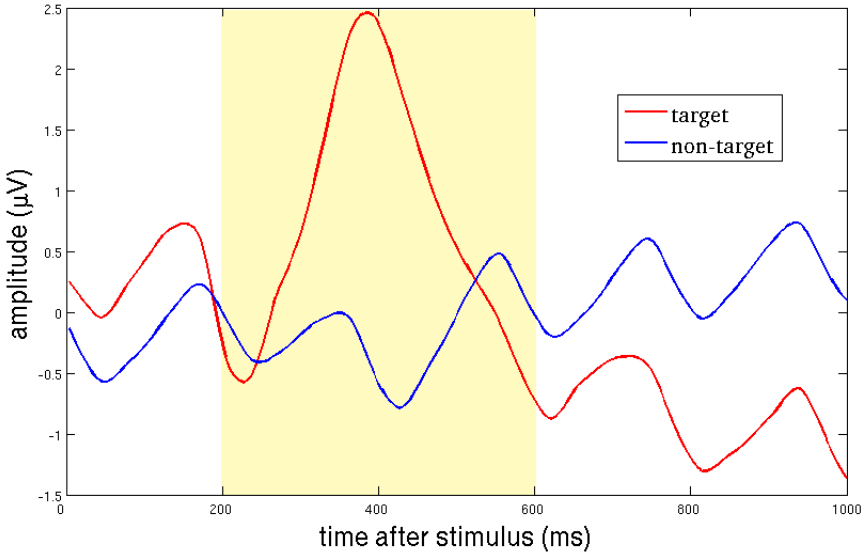
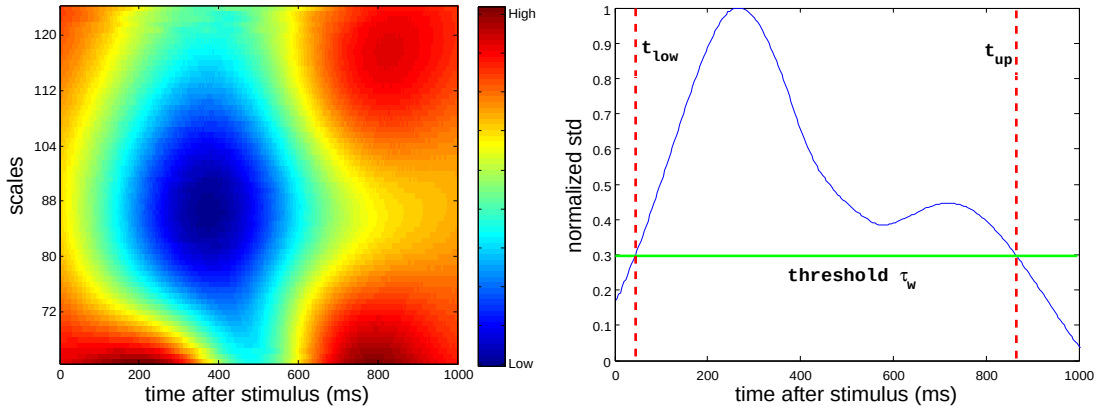


Figure 5.10: Difference between the Grand Averages of target and non-target signals, over channels, in a temporal window of 1 second. The signals were recorded using the P300 speller (see Section 2.4.1). The signals were previously denoised using METS algorithm to better visualize the difference between them, and the channels were normalized and averaged. The P300 average (red) was computed using 480 target responses and the non-target average (blue) was computed using 2400 non-target responses. The yellow band stand out the section were there are more differences between the signals.

ERP is triggered around 300 ms after the stimulus, in this case it starts earlier. The selection of a smaller temporal window allows to maximize the discrimination information between the two classes, by removing features that do not carry useful information.

Our approach consists in computing the correlation between the averages independently for each channel or the grand averages over channels, using the continuous wavelet transform, in order to obtain a continuous trend in time of the correlations by each scale. According to [Kolev et al. 1997], the most significant differences between target and not-target responses, in the frequency domain, are found in Delta and Theta brain rhythms. Consequently, the scales used to analyze the signals averages were selected according to these rhythms.



(a) (b)

Figure 5.11: (a) Dot product D. The colors in the image range from blue to red indicating the similarity between the grand average of targets responses $A^{\mathcal{T}}$ and the grand average of non-target responses $A^{\mathcal{N}}$ based on the amplitude and phase information ; (b) Normalized standard deviation of D over the scales (between 0 and 1).

In Figure 5.11(a), an example of this trend is shown, where it is easy to identify the less correlated components in cold colors and the more correlated components in warmer colors. These last ones correspond to EEG background, yet the correlation is not uniform between scales. It is possible to localize the ERP (in blue) within a thinner time window than the original. However, the ERP is not present in all scales, making harder an automatic detection of this temporal window. Fortunately, the ERP produces a high variability over the scales at a specific time on the dot product D, which leads to analyze the standard deviation for each time point. As the magnitudes of the wavelet coefficients can be different for each subject, we normalize the standard deviation between 0 and 1. At last, we select the temporal window using a predefined threshold as shown in Figure 5.11(b). This technique is completely independent of the previous denoising step introduced in Section 5.2, though better results are expected if both are used together.

Please note that in the example of Figure 5.10, the ERP is specially strong and the EEG background is highly correlated between target and non-target grand averages. Also, the signals were previously denoised using METS to better appreciate the signal's components. In fact, a naive approach such as detecting the average's peak of the target signal and select a temporal window around it may be effective for this subject under these conditions. Unfortunately, this cannot be done in all cases, since the grand averages varies in amplitude, shape and time lo-

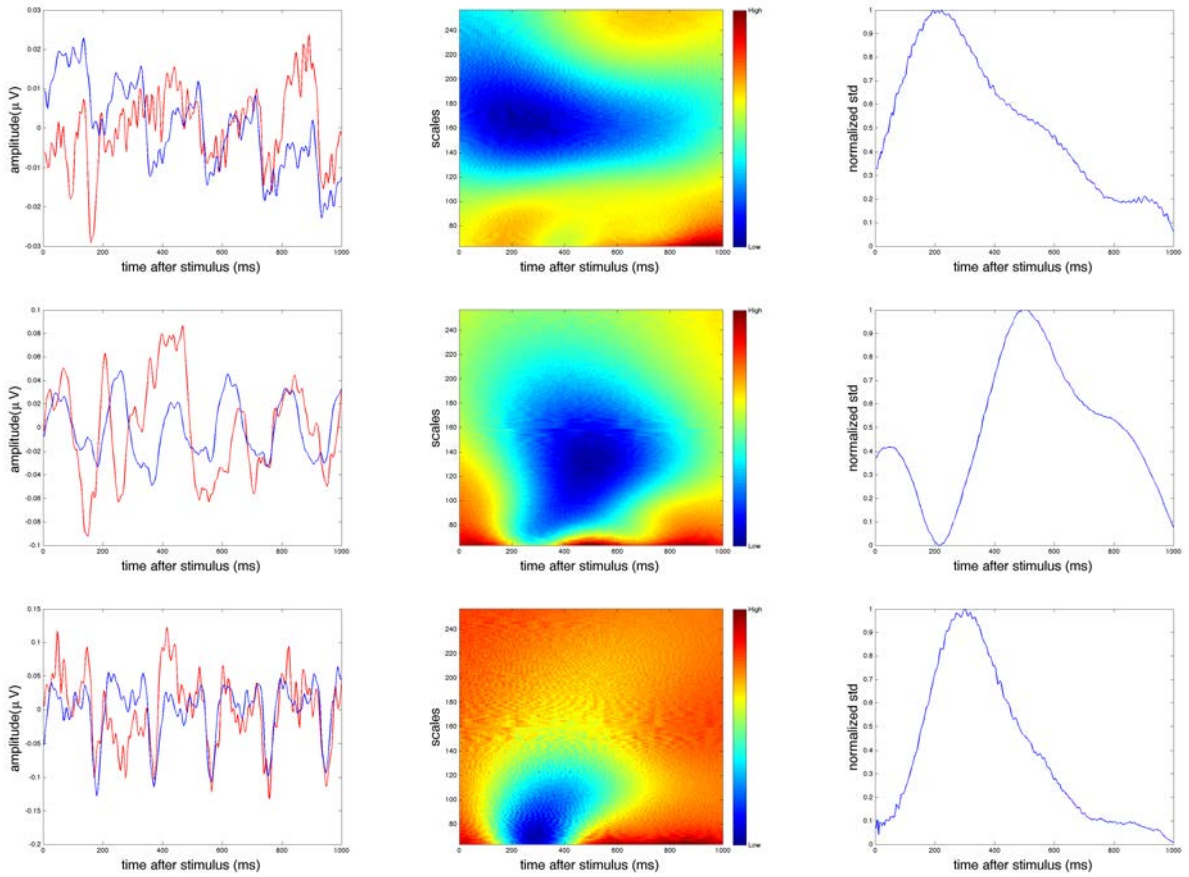


Figure 5.12: Analysis of ERPs in the time-frequency domain using D. The figure shows: in the left column, the target and non-target Grand Averages for three more subjects, recorded under the same conditions used for the subject of figure 5.10; in the center column, the heat map of the computed dot product D for their respective Grand Averages; and in the right column the normalized standard deviations of the scales at each time point. As the ERP creates a high variance in D, it is easier to select a time window in the time-frequency domain.

cation, making difficult to isolate the ERP. For example, in Figure 5.12, three other cases are shown, including the grand averages for target and non-target signals over channels without denoising, their respective dot product D and its normalized standard deviation. Even though some differences between target and non-target signals can be identified in the grand averages, it is hard to analyze them directly using only the time domain. The ERP can be easily identified in the time-frequency domain using the dot product, because, as stated before, the low correlation zones (in blue) usually correspond to the ERP. Moreover, D allows to properly visualize not only the locations, but also the sizes of the ERPs in both time and frequency domains. The normalized standard deviations show that relevant information of when the ERP was evoked can be obtained from the dot product, due to the fact that the ERP is not present in all scales. In summary, Figure 5.12 shows that it is easier to select a time window using the time-frequency domain rather that using only the grand averages at the left of the figure.

Algorithm 5.2 Semblance-based ERP Window Selection by Channel (SEWS-1)

Input: Given the EEG signal matrix $\mathbf{X}$, with C channels and T temporal samples.
Output: The bounds of the temporal window t_c^{low} and t_c^{up} for each C channel.

Set the window threshold τ_w, $0 \leq \tau_w \leq 1$
for $c = 1$ to C **do**
    Compute the Average $A_c^{\mathcal{T}}$ of responses belonging to the target class for channel c.
    Compute the Average $A_c^{\mathcal{N}}$ of responses belonging to the non-target class for channel c.
    Compute the CWT $W_\psi^{A_c^{\mathcal{T}}}$ and $W_\psi^{A_c^{\mathcal{N}}}$ using Equation 3.15 page 48.
    Compute the Semblance using Equation 3.29 page 52.
    Compute the dot product D using Equation 3.30 page 52.
    Compute the standard deviation $\sigma(D)$ of D over the scales and standardize it between 0 and 1.
    The limit t_c^{low} is given by the first t from the left which meets $\sigma(D(t)) > \tau_w$
    The limit t_c^{up} is given by the first t from the right which meets $\sigma(D(t)) > \tau_w$
end for

5.3.1 Semblance-based ERP Window Selection by Channel

Formally, let the signals be denoted by $x_c(t)$, where c corresponds to the channel and $t \in \{1, ..., T\}$, and $\mathcal{M}$ be the set of all the stimulus, where $\mathcal{T}$ correspond to the subset of target stimulus and $\mathcal{N}$ corresponds to the subset of non-target stimulus, *i.e.* $\mathcal{M} = \{\mathcal{T}, \mathcal{N}\}$. Then, the responses to a stimulus in a predefined temporal window, of size t^{up}, can be extracted as follows:

$$r_{i,c}(t) = x_c(s_i + t), \quad t \in \{1, ..., t^{up}\}, \tag{5.1}$$

where s_i corresponds to the stimulus onset i. The average response for each type of stimuli, and for each channel, can be computed as follows:

$$A_c^{\mathcal{T}}(t) = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} r_{i,c}(t) \qquad A_c^{\mathcal{N}}(t) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} r_{i,c}(t) \tag{5.2}$$

where the operator $|\cdot|$ denotes the cardinal number. After obtaining the averages, we compute the continuous wavelet transforms of them $W_\psi^{A_c^{\mathcal{T}}}$ and $W_\psi^{A_c^{\mathcal{N}}}$ to compute the dot product D through equation 3.30 page 52.

In the example of Figure 5.11, the result obtained for the dot product D corresponds to Figure 5.11(a), where cold colors indicate that the signals have a maximum difference, showing where the P300 is located in the spatial space. The normalized standard deviation of D is shown in Figure 5.11(b). Using a threshold τ_w, $0 \leq \tau_w \leq 1$, the new temporal window for the current channel is selected within limits t^{low} and t^{up}, obtained through a sequential search of the first points that exceeds the threshold both from the left and right boundaries. We call this algorithm *Semblance-based ERP Window Selection by Channel* (SEWS-1), and Algorithm 5.2 describes the complete window selection process.

5.3.2 Semblance-based ERP Window Selection over Channels

A variation on this model is to perform the *Semblance-based ERP Window Selection over Channels* (SEWS-2) as summarized in Algorithm 5.3, where the same temporal window is selected

for all channels. To do this, the only difference in the process is to compute the Grand Average through all channels using Equation 5.3.

$$GA^{\mathcal{T}}(t) = \frac{1}{|C|}\sum_{c=1}^{C} A_c^{\mathcal{T}}(t) \qquad GA^{\mathcal{N}}(t) = \frac{1}{|C|}\sum_{c=1}^{C} A_c^{\mathcal{N}}(t) \tag{5.3}$$

Algorithm 5.3 Semblance-based ERP Window Selection over Channels (SEWS-2)

1: **Input**: Given the EEG signal matrix $\mathbf{X}$, with C channels and T temporal samples.
2: **Output**: The margins for the temporal window t^{low} and t^{up}
3:
4: Set the window threshold τ_w, $0 \leq \tau_w \leq 1$
5: Compute the Grand Average $GA^{\mathcal{T}}$ of responses belonging to the target class over channels.
6: Compute the Grand Average $GA^{\mathcal{N}}$ of responses belonging to the non-target class over channels.
7: Compute the *CWT* $W_\psi^{GA^{\mathcal{T}}}$ and $W_\psi^{GA^{\mathcal{N}}}$ using Equation 3.15 page 48.
8: Compute the Semblance using Equation 3.29 page 52.
9: Compute the dot product D using Equation 3.30 page 52.
10: Compute the standard deviation $\sigma(D)$ of D and standardize it between 0 and 1.
11: The limit t^{low} is given by the first t from the left which meets $\sigma(D(t)) > \tau_w$
12: The limit t^{up} is given by the first t from the right which meets $\sigma(D(t)) > \tau_w$

5.4 Discussion

As explained in the above sections, wavelet-based techniques can be applied in BCI research, not only to extract frequency information, but also to extract information in the time domain. In particular the MRL is a natural way to analyze the simultaneous channels recordings as a whole, as Algorithm 5.1 does. This method allows to detect signal components that are not present in all channels, improving the analysis of EEG signals in general. We believe that our method is more robust compared to classic techniques that analyze one channel at a time. Furthermore, if our assumption that all channels are correlated is correct, the use of the same threshold to denoise all channels at the same time is justified. The parameters in our techniques resemble the ones used in classic wavelet thresholding, such as the mother wavelet, the threshold, and the level of decomposition. This means that similar improvements can be applied in our algorithm, such as the automatic selection of the threshold performed by these techniques.

The use of the continuous wavelet transform is a theoretically sound method to analyze the ERP time-window. The core element of the time-window selection algorithms is the dot product D and the techniques presented in this chapter only explore a few possibilities of what can be done using D. In fact, the standard deviation is only one way of exploiting the information carried by D: more elaborated or informed methods could be developed in the future to increase the performance or robustness of the time-window selection. Moreover, the analysis of the similarity of signals based on wavelets opens several research opportunities for EEG applications. For example, the same principle behind the time-window selection can be applied to detect scales that are constant through the whole signal, in order to select the frequency bands that are useful for the problem under study.

Another applications can be to detect channels with anomalous behavior through semblance analysis, like electrodes wrongly adjusted to the head or faulty electrodes, which do not properly record the signal. Finally, these similarity measures can be applied to study and compare the EEG behavior of different subjects for the same application, in order to understand better the physiological responses of the brain, or to develop more robust BCI techniques.

The presented algorithms will be evaluated and compared with some state of the art algorithms in the next chapter.

Chapter 6

Experiments

Contents

This chapter presents the results obtained for several experiments. The first part aims to compare different methodologies used by the BCI scientific community. Then, experiments to evaluate the performance of the new methods introduced in this thesis are presented, contrasting them with the xDAWN algorithm and the state of the art on wavelet thresholding techniques.

6.1 Experimental Setup

The software used for implementing all the algorithms used in these experiments is MATLAB 7.12.0 (R2011a). The only toolbox used outside the MATLAB distribution is the library for Support Vector Machines, LIBSVM, by Chih-Chung Chang and Chih-Jen Lin [Chang and Lin 2001], due to its efficient implementation and straightforward interface with MATLAB.

6.1.1 UAM Database

We use a large P300 database [Ledesma-Ramirez et al. 2010] recorded by the Neuroimaging Laboratory of the Universidad Autónoma Metropolitana (UAM), Mexico, using the P300 speller

[Farwell and Donchin 1988] (described in Section 2.4.1) included in the BCI2000 platform [Schalk et al. 2004]. The dataset contains 22 first-time healthy users, labeled ACS, APM, ASG, ASR, DCM, ELC, FSZ, GCE, ICE, JLD, JLP, JMR, JSC, JST, LAC, LAG, PGA, WFG and XCL, which will be referred as subjects s_1 to s_{22} for simplicity. All subjects have similar characteristics, such as, sleep duration, drugs, age, etc. 10 electrodes were recorded,(Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz, see Figure 6.1), which provide the best features for discrimination [Krusienski et al. 2006, 2008]. The signals were recorded at 256 Hz using a g.tec gUSBamp EEG amplifier, a right ear reference and a right mastoid ground. An 8th order 0.1-60 Hz Chebyshev bandpass filter, and a 60 Hz notch have been used. The stimulus is highlighted for 62.5 ms with an inter-stimuli interval of 125 ms. Parameters are summarized in Table 6.1. The dataset is freely available along with a full description on the database webpage [22].

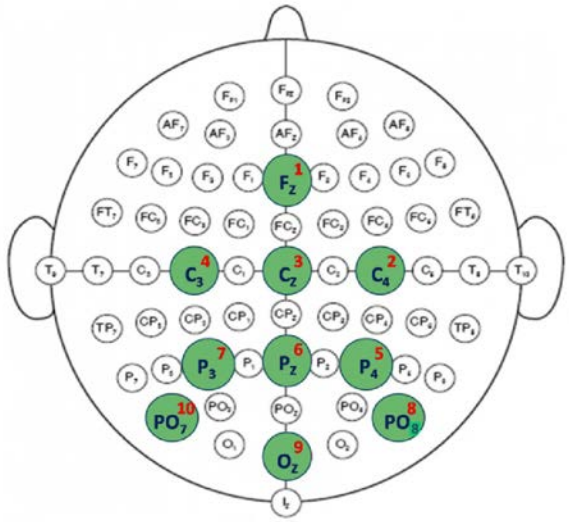


Figure 6.1: Channels included in the UAM database are enlarged and colored in green.

The dataset contains four different sessions:

1. Three copy-spelling runs, corresponding to words CALOR, CARINO and SUSHI.
2. One copy-spelling run (corresponding to word SUSHI) with feedback using a classifier trained on data from session one.
3. Three free-spelling runs (user-selected words, around 15 characters per subject).
4. Variable free-spelling runs reducing the number of repetitions as indicated by the bit-rate analysis.

In order to validate our algorithms, we use session 1 (copy spelling session) to train a classifier and session 3 (free spelling session) to test the generated models. This choice was made because both sessions do not use feedback and because the detection methods have to be robust under different conditions, *i.e.* the algorithms will be study under a multi-session scheme.

22. `http://akimpech.izt.uam.mx/p300db/doku.php?id=deposits.en`

P300 Speller	
Recorded channels	10
Sampling frequency	256
Sequences	15
Band-pass	Chebyshev; Bandwidth: 0.1 Hz - 60 Hz; 8th order
Notch filter	Chebyshev, Frequencies rejected: 58 Hz - 62 Hz; 4th order
PreRunDuration	2 seconds
PostRunDuration	0.5 seconds
PreSequenceDuration	2 seconds
PostSequenceDuration	2 seconds
Stimulus duration	62.5 ms
ISIMinDuration	125 ms
ISIMaxDuration	125 ms
Matrix	6 rows, 6 columns
Window	width: 1680; height: 1050

Table 6.1: Recording Parameters of the UAM Database.

From a single trial perspective, the dataset contains 240 letters by person for session 1, and 250 letters approximately for session 3, because the number of letters varies depending on subject. Considering the 22 subjects, there are 5280 letters in total for session 1 and 5485 letters for session 3. Finally, considering that for the detection of one letter it is necessary to identify two P300s, it follows that there are 10560 P300s and 52800 non-target responses in session 1 and 10970 P300s in session 3 and 54850 non-target responses.

6.1.2 Methodology

The methodology used to assess our algorithms (presented in Section 5) with respect to the state of the art, is a generic approach of the signal processing stage of the BCI architecture (see Section 2.1). We have divided this process in six successive sequentially steps, including the different algorithms explained in the previous chapters:

1. **Normalization**: The recorded signal is normalized within the range [0,1];

2. **Filtering**: The normalized signal is filtered using a fourth-order Butterworth;

3. **Feature Extraction**: The filtered signal is enhanced for classification. We consider the following techniques:

   - The xDAWN algorithm of Section 3.1
   - The classic thresholding techniques of Section 3.2.5
   - The EEG denoising technique based on channels similarity (METS) of Section 5.2
   - The ERP window selection based on the semblance measure (SEWS) of Section 5.3

4. **Channel Concatenation**: For each ERP response, the features obtained for each channel are concatenated all together in a single feature vector, obtaining a single matrix for all data.

5. **Training**: We consider two classifiers from the literature:

- SVM of Section 4.4
- SWLDA of Section 4.3.3

6. **Classification**: The results obtained from the P300 classifier are used to identify the target letter, by selecting the most likely column and row.

Consequently, every experiment will take the raw data as input and will output the letter accuracy obtained on the test data. The basic performance criterion is the letter accuracy percentage, however other measures are used like the P300 accuracy percentage by subject.

Every step has an important impact in the overall performance, and some of them are independent of the novel methods proposed in this thesis. Indeed, the cutoff frequencies (filtering), the classifier and the mother wavelet, are choices that must be fixed before assessing our methods. Hence, next section explore these possibilities because there are no conclusive arguments in the literature to fix them *a priori*.

6.2 Preliminary Studies

For the experiments of this section, the UAM database previously described is used, because it is composed by a large number of subjects, enabling a clear comparison between methodologies. The first experiment is focused on the different combinations of cutoff frequencies of low-pass and high-pass filters used to pre-process brain signals, as shown in Section 5.1.3. The range of filters used is considerably wide, making difficult to select an optimal combination based on previous studies. The second experiment intends to compare the two main classification algorithms used in BCI, the LSVM and SWLDA algorithms. The last experiment of this section aims to compare the influence on the performance of the selected mother wavelet.

6.2.1 Experiment #1: Pre-processing Filtering

To better understand the practical impact of the high-pass and low-pass filter values, we compute the single-trial letter accuracy in the P300 speller using a combination of cutoff frequencies. The accuracy corresponds to the average of single-trial letter accuracy on the test session of the 22 subjects from the database, obtained through training a LSVM classifier for each subject.

Several band-pass filters were used, obtained by combining 2nd-order Butterworth low-pass and high-pass filters. The low cutoff frequencies are 0.1, 0.15, 0.4, 0.5 and 1 Hz and the high cutoff frequencies are 8, 10, 20, 30, 40 and 60 Hz. As shown in Section 5.1.3, scientists also use 200 Hz for the low cutoff frequency; this frequency is not included in our study because recordings were already band-pass filtered under 60 Hz frequency.

As shown in Table 6.2, the best results were found using either a 10 Hz or a 20 Hz low cutoff frequency, since the difference between both results is not significant. Also it is possible to observe that the standard deviation is directly related to the low cutoff frequency. The standard deviation increase its value with the low cutoff frequency. For the high cutoff frequency, the best performance was achieved using 0.1 Hz, but in order to understand the dependency between the low and high cutoff frequencies, we visualize the information using boxplot graphs.

The boxplots in Figure 6.2 are grouped by high cutoff frequencies, where there are four groups (one for each high cutoff frequency) and each group is composed by six boxplot (one for each low cutoff frequency). It is possible to conclude that, independently of the high cutoff frequencies, the best selection for the low cutoff frequency are 10 Hz or 20 Hz, as seen before. It is also possible to deduce that if we increase the low cutoff frequency the accuracy for subjects

High Cutoff Frequencies \ Low Cutoff Frequencies	8	10	20	30	40	60	μ	σ
0.1	52.71	53.23	**53.60**	52.15	50.51	**48.23**	**51.74**	2.03
0.15	52.43	53.07	**53.56**	**52.16**	**50.56**	48.16	51.66	2.00
0.4	52.76	53.38	**53.72**	51.66	50.04	47.44	51.50	2.40
0.5	53.01	**53.74**	53.17	51.42	50.00	47.51	51.48	2.38
1	**53.13**	**53.83**	52.50	50.62	48.95	46.22	50.88	2.89
μ	52.81	**53.45**	53.31	51.60	50.01	47.51		
σ	0.27	0.33	0.5	0.63	0.65	0.81		

Table 6.2: Single-trial letter accuracy averaged over 22 subjects for various combinations of low-pass and high-pass filters, including their respective mean and standard deviation statistics. The best results for the low cutoff frequencies (green) and for the high cutoff frequencies (blue) are highlighted. When the best results coincides for low and high frequencies, only this value is highlighted (black).

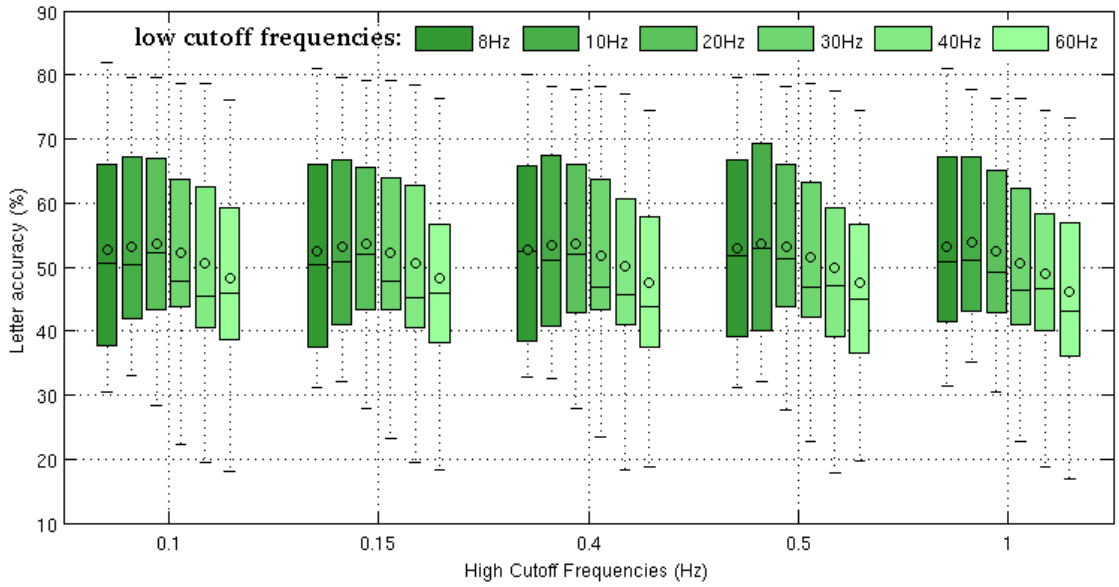


Figure 6.2: Boxplot for the 22 subjects using single-trial responses grouped by high cutoff frequencies. The circle represents the mean and the horizontal line represents the median.

with high performances decreases approximately only by a 5 percent, but for subjects with bad performances the accuracies drop dramatically.

In Figure 6.3, the boxplots are grouped by low cutoff frequencies and each group is composed by the high cutoff frequencies used in the experiment. As seen in Table 6.2, the high cutoff frequency has less influence in the final results. However, it seems that results are more stable for the 0.1 Hz frequency if the low cutoff is increased. Taking into account that the goal of this stage is to improve the system performance using a frequency window as wide as possible to keep as much as information from the original signal, we conclude that a [0.1-20] Hz band-pass filter

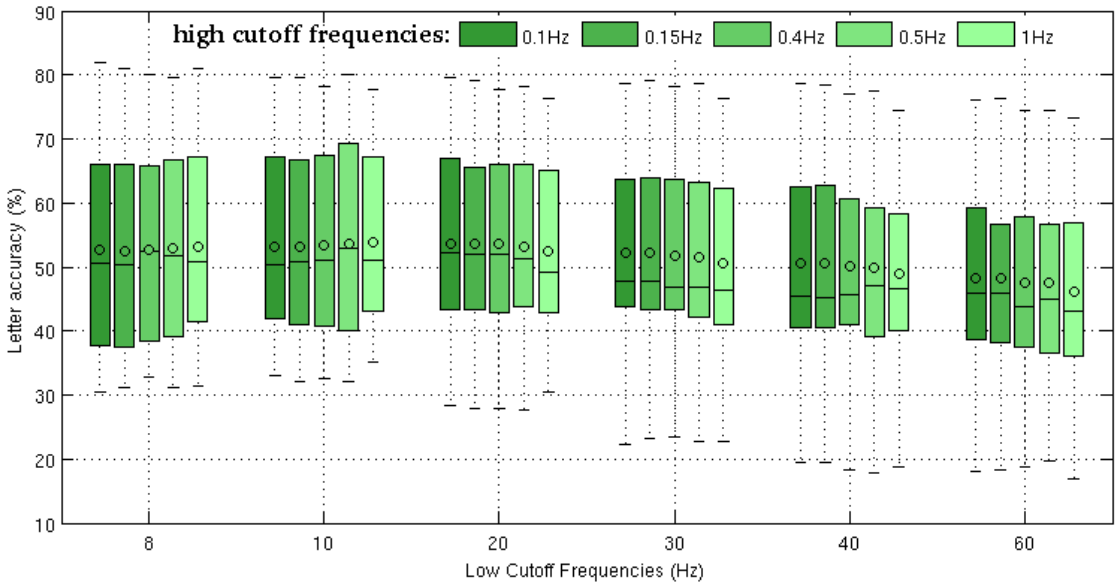


Figure 6.3: Boxplot for the 22 subjects using single-trial responses grouped by low cutoff frequencies. The circle represents the mean and the horizontal line represents the median.

offers a good compromise between filtering and preserving information. For these reasons, in the next experiments, this frequency band will be used for detecting ERPs.

6.2.2 Experiment #2: SWLDA vs LSVM

Two of the most used classifiers (see Chapter 4) are SWLDA, implemented in the BCI2000 system, and LSVM, implemented in the Openvibe software. Although kernelized versions of these classifiers can also be used, the original linear versions are used here because ERP classification is commonly considered as a linear-separable problem.

In this experiment, a comparison between SVM and SWLDA is presented using the same dataset in order to expose the advantages and disadvantages of each method. The parameters used for SWLDA are those proposed in [Krusienski et al. 2006], which are the same ones used in BCI2000. These parameters correspond to the maximum p-value for a variable to be added to the model, which is setted to 0.1 and the maximum p-value for a variable to be removed, setted to 0.15. The SVM uses a linear kernel (LSVM) and the penalty term P was setted to the default value 1. All signals have been filtered between [0.1-20] Hz.

The results in letter accuracy, for the training and test datasets, for single-trial detection are shown in Table 6.3. The algorithm SWLDA is reasonably efficient because it selects the suitable features to discriminate the P300 responses, but the problem using SWLDA algorithm is that there is no certainty of convergence since the parameters may not satisfy the heuristic, causing the absence of a classification model. Two examples of this situation are subjects s_6 and s_{19}, which can not meet the parameters to create a model.

The mean μ and the standard deviation σ of the letter accuracies were computed without subjects s_6 and s_{19}. In terms of single-trial letter accuracy, LSVM performs considerably better than SWLDA. The difference between the training and test results can be explained because

Sbj.	Training Set		Testing Set	
	SWLDA	SVM	SWLDA	SVM
s_1	45.83	80.83	36.67	45.71
s_2	62.08	86.67	64.70	64.71
s_3	60.41	82.08	59.58	55.00
s_4	42.91	53.75	49.52	28.25
s_5	65.83	93.33	65.83	66.94
s_6	-	52.50	-	38.89
s_7	22.5	67.50	20.56	39.44
s_8	60.83	87.08	57.41	69.63
s_9	52.50	78.33	51.90	56.67
s_{10}	40.42	65.00	38.10	45.24
s_{11}	27.50	75.00	26.67	45.96
s_{12}	64.58	82.50	53.89	55.00
s_{13}	46.25	65.42	35.18	43.33
s_{14}	60.00	86.25	36.67	49.58
s_{15}	37.08	62.50	23.81	42.38
s_{16}	70.00	91.67	68.21	75.38
s_{17}	76.25	92.08	75.71	79.52
s_{18}	70.83	90.00	64.06	54.78
s_{19}	-	81.25	-	73.33
s_{20}	24.17	58.75	17.19	35.09
s_{21}	31.67	67.08	26.67	46.67
s_{22}	73.33	93.33	60.56	67.78
μ	51.75	77.96	46.64	53.35
σ	17.10	12.64	17.98	13.75

Table 6.3: Letter accuracy per subject on the training and testing set in single-trial detection for the SWLDA and LSVM classifiers. The mean and standard deviation statistics were computed without considering subjects s_6 and s_{19}, because the standard parameters stop the forward selection from the beginning for SWLDA. The best results for each set are highlighted.

the EEG signals recorded for the training and the test belong to different sessions, with possibly different amplitudes and frequencies distributions due to the variable mental state of the subject. The advantage of LSVM is that there is no need to set free parameters as in other kernels.

Our final process will use the LSVM classifier, not only because it provides in general a better performance in our setup, but also because the same parameters can be used for all subjects, unlike SWLDA.

6.2.3 Experiment #3: Mother Wavelet (Selection)

As stated in the previous chapters, there is no formal method to select a mother wavelet. Therefore, we study the differences in letter accuracy by changing only the mother wavelet being used. This experiment was performed using the classic thresholding technique, based on the soft thresholding rule and the SURE threshold. The level of decomposition was setted to 3. The signals have been filtered and between [0.1-20] Hz. The results for 5 different wavelet mothers (Coiflet order 3, Daubechies of order 4 and 8, Symlet of order 5 and B-spline) are shown in Table 6.4. The results include the number of letters and P300s detected in place of their percentage, since the differences between the wavelet mothers are easier to observe. Here, the P300 accuracy was computed to test the ability of a certain wavelet mother to detect the waveform. The results shows that the wavelet mother selection strongly depends on the subject, what makes us conclude that P300 waveforms are different for each subject. The overall best

performances are for Coiflet of order 3 and B-spline, since several subjects obtained the best results with these wavelet mothers. However, Table 6.4 shows that Daubechies of order 4 and 8 obtained also very competitive results, suggesting that a smart selection of a mother wavelet in this set, does not substantially contribute to improve the P300 recognition. Therefore, for simplicity we will use the Coiflet of order 3 as wavelet mother for the rest of the experiments.

Sbj.	total	total	**Coif3**		**Db4**		**Db8**		**Sym5**		**B-spline**	
	#p3	#let.	#p3	#let.	#p3	#let.	#p3	#let.	#p3	#let.	#p3	#let.
s1	420	210	277	99	276	96	274	99	271	94	269	93
s2	510	255	416	169	408	164	415	168	412	166	411	165
s3	480	240	373	140	368	136	372	139	367	135	375	142
s4	630	315	321	104	322	99	327	109	314	99	329	106
s5	704	352	507	249	509	246	506	242	508	248	510	248
s6	540	270	340	111	333	105	339	108	334	104	336	107
s7	360	180	227	64	228	67	224	63	226	65	232	69
s8	540	270	454	189	454	187	451	189	454	188	457	190
s9	420	210	324	128	323	126	323	126	327	130	324	126
s10	420	210	267	89	271	94	269	90	264	90	270	90
s11	556	278	360	134	351	129	350	129	345	124	355	134
s12	360	180	271	102	271	103	267	98	273	105	270	100
s13	526	263	334	119	332	118	333	118	331	117	332	116
s14	480	240	344	122	349	124	345	121	346	121	350	124
s15	420	210	279	94	281	95	278	94	284	97	275	94
s16	390	195	340	149	342	151	338	147	340	149	339	148
s17	420	210	369	165	371	167	365	164	371	168	371	167
s18	630	315	429	197	424	192	425	188	413	182	426	197
s19	584	292	472	218	472	219	475	223	472	218	470	218
s20	542	271	315	109	307	106	307	111	305	104	309	106
s21	678	339	410	167	411	168	408	166	414	171	411	167
s22	360	180	296	124	296	122	292	120	298	124	296	122
Total	10970	5485	**7725**	**3042**	7699	3014	7683	3012	7669	2999	7717	3029

Table 6.4: Number of successfully detected P300s and successfully detected letters per subject using a SURE denoising with different mother wavelets (single-trial detection). This experiment uses the LSVM classifier for 5 different mother wavelets: Coiflet order 3 (Coif3), Daubechies order 4 and 8 (Db4 and Db8), Symlet order 5 (Sym5) and B-splines. The best results for each subject are highlighted.

6.3 Denoising Based on Channels Correlation

In this section, we assess our proposed denoising algorithm, compared to the state of the art of thresholding techniques based on wavelets presented in Section 3.2.5. As explained in Section 5, one of the main contributions of this thesis is the wavelet-based denoising using the channels correlation through the mean resultant length.

6.3.1 Experiment #4: Threshold Analysis

As shown in Section 3.2.4, the MRL values run from 0 to 1, where 0 represents that signals are not correlated and 1 that they are fully correlated. In principle, the MRL threshold τ_d used to set to zero small coefficients can be selected between 0 and 1 to denoise the recorded channels.

In this experiment, we study the distribution of the MRL coefficients to better understand how the threshold τ_d should be selected.

In Figure 6.4, the histogram of MRL coefficients are shown for only six subjects to illustrate the distributions, since no difference can be really observed between subjects. The MRL coefficients exhibit an exponential distribution. Therefore it is possible to establish a correlation threshold τ_d based on an exponential scale, in order to set to zero all coefficients that are below the given threshold. After this process, we can reconstruct the signal using the filtered wavelet coefficients.

Different threshold values were tested, using the formula $\tau_d = 1 - 10^{-y}, y \in [0, \infty]$ due to the exponential nature of the distribution of the MRL coefficients. The idea is to focus the search of a suitable threshold in the values that produce an effective denoising in a exponential distribution (high values). Indeed, for $y < 1$ ($\tau_d < 0.9$) only high uncorrelated coefficient with respect to the whole distribution are removed.

The threshold was tested for only a few integer values of y using all the subjects in the database. We obtained a fairly good overall result for $y = 3$, where $\tau_d = 0.999$. The details of these experiments are not included in this manuscript, since the objective is not to optimally select the threshold, but to understand which are the ranges where the threshold makes the desired effect. This threshold will be used for the following experiments.

6.3.2 Experiment #5: METS vs Classic Wavelet Methods

The results of our denoising algorithm METS, compared with the results of classic thresholding methods, are shown in Table 6.5. Also, the baseline results using only a band-pass filter [0.1-20] Hz are presented. All thresholding techniques improves the baseline result on average, meaning that these techniques are actually denoising the original signals and not removing relevant information. Nevertheless, these techniques are not perfect since the performance of the subject with the maximum value slightly decreases.

Although the improvements in the results of these techniques may seem small, please recall that the probability of detecting a letter by chance in the P300 speller is only 1/36 (detecting exactly twice 1 ERP among 6 responses), in contrast to the probability of randomly detecting a single ERP of 1/2. This means that there is a high probability of missing a letter that was correctly detected (0.97) due to randomness, and only a small chance of correctly detect a letter (0.03) that was previously misclassified.

Table 6.5 reveals that our algorithm is, in general, capable to remove the noise from the signal's subjects, and it improves the results obtained by the classic thresholding techniques based on wavelets. Even though the results are not substantially greater than the classic techniques, our algorithm performs consistently better. Actually, the standard deviation is less in our approach, meaning that the increase in the mean is global and not due the improvement of subjects with best results. In fact, the maximum value is the same obtained by Minimax and Universal thresholds, while the minimum is improved by METS. This suggests that METS is improving the results of those subjects with original poor performance.

We attribute these improvements to the introduction in the process of the channels correlation and to the fact that channels are processed as a whole. Indeed, the thresholds in classic methods are selected using statistical techniques to infer a suitable threshold for each subject, while for METS we used a fixed threshold of $\tau_d = 0,999$ (*i.e.* $y = 3$) for all subjects. Thus, even though the performance of our algorithm METS is appealing, it could be extended to use in addition a similar automatic threshold selection in order to obtain better results.

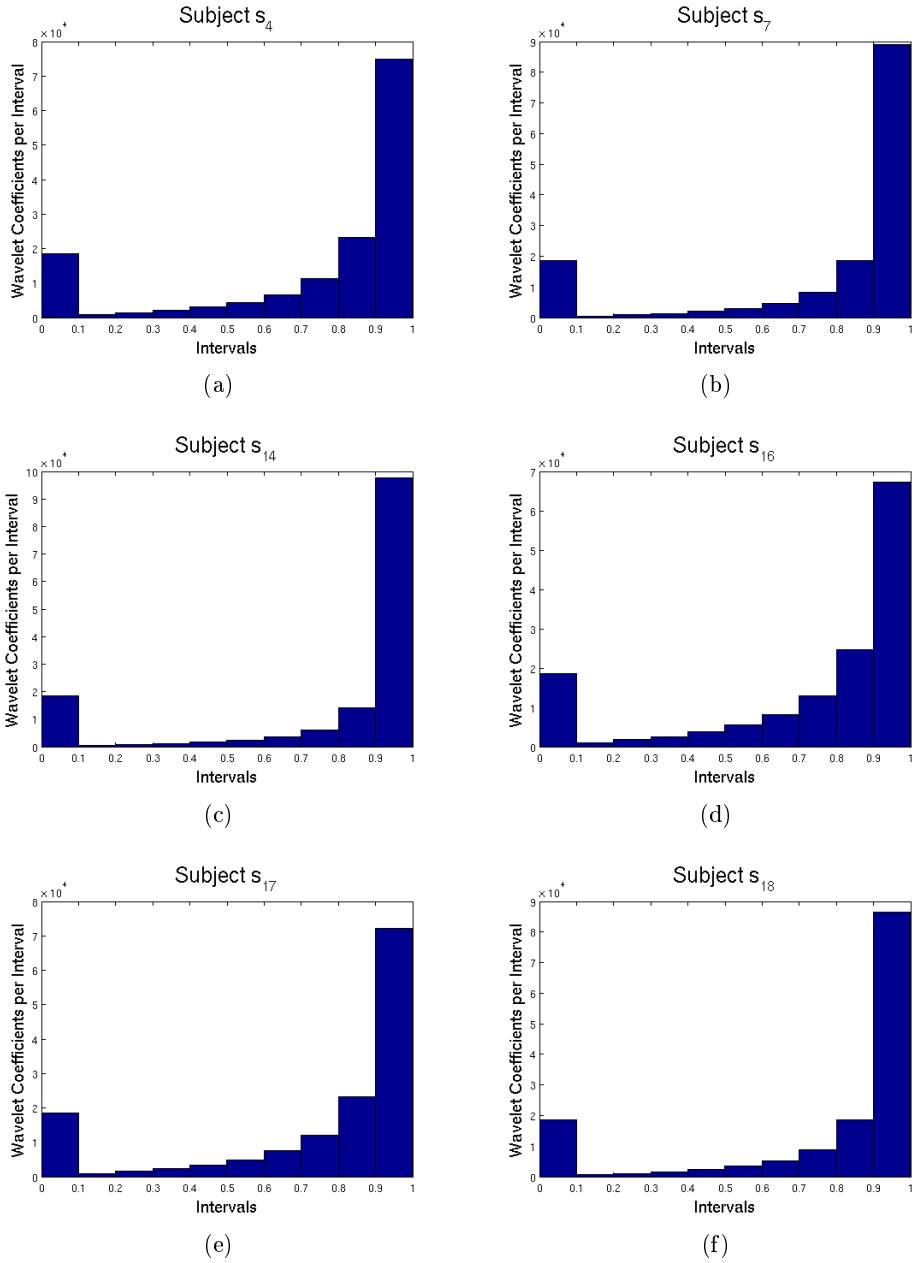


Figure 6.4: Histogram of MRL coefficients between the recorded channels for 6 different subjects. The MRL coefficients exhibit an exponential distribution, so the threshold τ_d will be based on a exponential scale.

6.4 Temporal-Window Selection

As explained in Section 5.3, a temporal window must be set to specify the segment to analyze. This window is commonly selected within a range of [0,1] second, in order to ensure that the P300 is within the recorded window, independently of the user reaction time to the stimulus.

In this section, we test the window selection algorithms presented in Section 5.3 to analyze the impact of a smart selection of a thinner window.

Method	μ	σ	min	max
[0.1-20] Hz Filter	53.60	14.14	28.25	79.52
SURE	54.80	13.90	33.02	78.57
Minimax	55.00	13.93	32.70	79.05
Universal	55.07	13.92	33.02	79.05
METS	55.20	13.69	33.65	79.05

Table 6.5: Letter accuracy for the METS, Minimax, Universal and SURE denoising algorithms over 22 subjects of the UAM database. Descriptive statistics (mean, standard deviation, min and max) on letter accuracy are reported. The best performances are highlighted.

6.4.1 Experiment #6: SEWS-1 and SEWS-2 algorithms

The time-window selection techniques were tested using only as a pre-processing a [0.1-20] Hz filtering, in order to isolate the performance of the time-window selection from the denoising. These algorithms are labeled SEWS-1 for the algorithm using specific windows size for each channel and SEWS-2 for the algorithm using the same window size for all channels (see Section 5.3). Results for different thresholds τ for both algorithms are shown in Tables A.2 and A.3.

Pre-processing	τ_w	μ	σ	min	max	$\mu(t^{low})$(ms)	$\mu(t^{up})$(ms)
[0.1-20] Hz Filter	-	53.60	14.14	28.25	79.52	0	1000
SEWS-1	0.1	53.73	14.70	28.57	80.95	17	951
SEWS-1	0.2	54.10	14.16	32.70	80.48	34	900
SEWS-1	0.3	53.98	13.71	32.06	80.48	56	844
SEWS-1	0.4	54.10	13.28	36.83	80.00	90	781
SEWS-1	0.5	53.16	14.12	32.86	79.05	128	724

Table 6.6: Letter accuracy for the SEWS-1 algorithm without denoising over 22 subjects of the UAM database. Descriptive statistics (mean, standard deviation, min and max) on letter accuracy are reported. Also, the begin (t^{low}) and the end (t^{up}) of the time window for each subject are reported. The best results are highlighted.

The SEWS-1 algorithm was tested using increasing thresholds until the baseline was no longer outperformed. This happens for $\tau_w = 0.5$ where the algorithm is removing in average 40% of the size of the temporal window. For all the other thresholds, the original temporal window of 1 second can be reduced without a performance loss, showing that the information removed from the signal made it difficult to classify. Also, the minimum and maximum accuracy are improved, showing that the technique is flexible enough to deal with different types of subjects. The compromise between reducing the temporal window and keeping relevant information is controlled by the threshold parameter τ_w, yet selecting the threshold is not an easy task. We can observe in Table A.2 that $\tau_w = 0.1$ already improves the results by removing 6% of the signal in average, while $\tau_w = 0.2$ and $\tau_w = 0.4$ offer the best average. As the thresholds increases, the minimum value and variance are reduced, at the cost of a worse performance for the subjects with best results. Actually, the best result obtained for $\tau_w = 0.4$ removes in average the 30% of the signal, which is an aggressive window selection, while the same average performance is obtained by removing only a 16% of the signal.

The results for the SEWS-2 algorithm are summarized in Table A.3, showing that the be-

Pre-processing	τ_w	μ	σ	**min**	**max**	$\mu(t^{low})$(ms)	$\mu(t^{up})$(ms)
[0.1-20] Hz Filter	-	53.60	14.14	28.25	79.52	0	1000
SEWS-2	0.1	54.08	14.60	29.52	80.48	14	960
SEWS-2	0.2	54.29	14.29	32.38	79.52	28	917
SEWS-2	0.3	54.41	14.03	31.43	78.10	49	869
SEWS-2	0.4	52.59	15.21	32.86	79.05	85	827

Table 6.7: Letter accuracy for the SEWS-2 algorithm without denoising over 22 subjects of the UAM database. Descriptive statistics (mean, standard deviation, min and max) on letter accuracy are reported. Also, the average begin (t^{low}) and the average end (t^{up}) of the time windows are reported for each subject. The best results are highlighted.

havior of the temporal-window selection is different than SEWS-1. Here, the algorithm is more conservative (on average), producing wider windows than SEWS-1, because the grand average considers the information of all channels. This means that the contribution of components which increase the standard deviation for a channel, will be most likely present in the standard deviation of the grand averages.

For the same reason, the performances increase as stronger thresholds are used, until it dramatically falls below the baseline when $\tau_w = 0.4$, where the algorithm removes 26% of the signal (on average). This is less than what SEWS-1 can discard (on average) without removing important information. However, the best results of SEWS-2 outperform SEWS-1, because in theory all channels carry similar information, meaning that the selected windows should be alike. As SEWS-2 naturally compensates the presence of excessively narrow or wide windows, most of the subjects may benefit to some degree from the window selection. On the other hand, if the threshold is too high, most of the subjects may lose informative components.

In summary, this experiment shows that reducing the time window based on semblance is an effective technique to automatically locate the ERP response in time, and it seems particularly useful for subject with low performances.

Please remember that these results were obtained without using METS, showing that SEWS can work with noisy signals. However, as denoising is a common practice in signal processing, SEWS must also perform correctly with denoised signals. Next section deepens the analysis of the time-window selection techniques when they are used in conjunction with our denoising algorithm METS.

6.4.2 Experiment #7: METS and SEWSs algorithms

In this experiment, we propose to apply our two main contributions of Section 5, the SEWS time-window selection techniques and the METS denoising algorithm. We denote as METS & SEWS-1 the algorithm that denoises and uses a specific temporal-window size for each channel, and as METS & SEWS-2 the algorithm that denoises and uses the same temporal window for all channels.

Table A.4 shows the impact of using the METS algorithm as a previous step for the time-window selection. The threshold values for the algorithms are $\tau_d = 1 - 10^{-3}$ for METS, $\tau_w = 0.1$ for SEWS-1 and $\tau_w = 0.2$ for SEWS-2 according to the previous study. Other non-aggressive thresholds for SEWS were also tested, producing very similar results to the ones presented here. A paired t-test shows that all algorithms are statistically significant at 1% compared to the [0.1-20] Hz Filter, since the p-values are considerably small.

Pre-processing	μ	σ	min	max	t-test p-value
[0.1-20] Hz Filter	53.60	14.14	28.25	79.52	-
METS	55.20	13.19	33.65	79.05	0.0017
METS & SEWS-1	56.00	13.64	35.56	80	0.0004
METS & SEWS-2	55.91	14.13	34.44	78.97	0.0005

Table 6.8: Letter accuracy for the time-window selection algorithms (SEWS) when using denoising (METS), over 22 subjects of the UAM database. Descriptive statistics (mean, standard deviation, min and max) on letter accuracy are reported, in addition to a paired t-test between the baseline ([0.1-20] Hz Filter) and the algorithms. The best results are highlighted.

Both SEWS algorithms improve the average, minimum and maximum results compared to only using METS, implying that the time-window selection is an effective technique also for denoised signals. In fact, the generated time windows are very similar to the ones presented in the previous section. Comparing with the baseline filter results, the combination of our two independent algorithms significantly improves the ERP detection in single trial, incurring only in a small computational cost compared to the time required for the training and classification phases. In particular, SEWS-1 obtains the best maximum and minimum accuracy, but in average both window selection algorithms are competitive.

More details about the window sizes obtained for each subject by the algorithms can be found in Tables 6.9 and 6.10. We can observe that the windows sizes remain quite large. All of them start between [0-133] ms and end between [762-1000] ms. The removed information usually lies at the end of the original temporal window, beyond the segment where the P300 theoretically lies. Also, for several subjects t^{low} is equal or close to zero, probably because early ERP components are present in the signal besides the P300 response as discussed in Section 5.4.

Subject	$\mu(\mathbf{t^{low}})$	$\mu(\mathbf{t^{up}})$	μ(**size**)	Subject	$\mu(\mathbf{t^{low}})$	$\mu(\mathbf{t^{up}})$	μ(**size**)
s_1	0	961	961	**s_{12}**	0	959	959
s_2	0	926	926	**s_{13}**	0	938	938
s_3	21	967	946	**s_{14}**	4	966	962
s_4	36	980	944	**s_{15}**	50	980	931
s_5	17	941	924	**s_{16}**	7	927	919
s_6	0	967	967	**s_{17}**	0	960	960
s_7	16	948	933	**s_{18}**	6	972	966
s_8	5	887	882	**s_{19}**	22	984	962
s_9	32	984	952	**s_{20}**	20	949	929
s_{10}	61	994	933	**s_{21}**	12	889	877
s_{11}	0	883	883	**s_{22}**	11	963	952

Table 6.9: Average time windows per subject obtained by METS & SEWS-1. The begin, end and size of the windows were averaged over the channels. All the values are reported in milliseconds.

The results of this section show that our algorithm outperforms the baseline scenario, but this is not sufficient to justify a broad use of our methods. In the next section, further experiments are presented to explore the extents of our proposal.

Subject	t^{low}	t^{up}	size	Subject	t^{low}	t^{up}	size
s_1	0	922	922	s_{12}	0	938	938
s_2	0	895	895	s_{13}	0	824	824
s_3	12	891	879	s_{14}	20	949	929
s_4	51	965	914	s_{15}	90	953	863
s_5	0	926	926	s_{16}	0	836	836
s_6	0	941	941	s_{17}	0	930	930
s_7	51	898	847	s_{18}	47	961	914
s_8	0	930	930	s_{19}	43	984	941
s_9	70	969	899	s_{20}	27	910	883
s_{10}	133	1000	867	s_{21}	39	820	781
s_{11}	0	762	762	s_{22}	0	977	977

Table 6.10: Time window per subject obtained by METS & SEWS-2. The begin, end and size of the windows are are reported in milliseconds.

6.5 Further Studies

When using a single-trial approach the SNR is very low, making the ERP detection a highly complex problem. Indeed, BCI feature-extraction techniques usually perform well when using average of signals, but they find difficult to assess single-trial signals, specially for subject with difficulties to handle the BCI skill. Moreover, the ERP identification must be performed as fast as possible for real-time applications, which requires employing robust techniques under dimensionality reduction.

In this section, we will show that our techniques outperform the well-known xDAWN spatial filter for single trials and explore the robustness of our techniques using less features. Finally, a new database is introduced to validate our algorithms using the same setup and parameters than in the UAM database.

6.5.1 Experiment #8: Comparison with xDAWN algorithm

The xDAWN algorithm (see Section 3.1.2) plays a similar role as the algorithms proposed in this thesis, yet using a completely different strategy. xDAWN automatically enhances the ERP for classification by using spatial filters, while SEWS automatically identifies the ERP segment to improve classification. On the other hand, xDAWN combines the multi-channel information to put aside useless components, while METS denoises useless components by using the multi-channel information. In the next experiments, we contrast this state-of-the-art approach with our proposal.

In Table 6.11, the results for xDAWN using different numbers of estimated sources are shown. The signal pre-processing was performed according to the specifications in [Rivet et al. 2009]. The only difference with our experimental setup is the filter high cutoff frequency of 1 Hz, instead of 0.1 Hz. Surprisingly enough, the mean results do not improve the baseline of 52.50% (see Table 6.12) obtained with the [1-20] Hz filter, obtaining good results for subjects with high performances and very low results for subjects with low performances, as shown in Figure 6.5. The best result is obtained using 3 sources, which will be used to compare with our algorithms.

Table 6.12 summarizes statistics on the letter accuracy for xDAWN compared to our algorithms. The best result was obtained by our algorithms, showing improvements in mean,

	xDAWN									
Sources	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
μ	44.64	50.24	51.03	50.18	50.42	50.26	50.88	50.52	49.68	49.59
σ	19.72	16.57	15.80	15.96	15.35	14.59	14.97	14.32	14.85	14.76
min	11.43	23.89	24.44	27.22	28.89	27.3	26.98	24.76	22.22	21.27
max	76.67	78.1	80	76.19	75.71	75	74.76	72.67	74.29	72.33

Table 6.11: xDAWN algorithm statistics of the letter accuracy for the 22 subjects in single-trial detection, using different number of sources. The signal pre-processing was performed according to [Rivet et al. 2009] with a band-pass filter of [1-20] Hz. The best results are highlighted.

Pre-processing	μ	σ	**min**	**max**
Filter [1-20] Hz	52.50	13.49	30.48	76.41
xDAWN	51.03	15.80	24.44	80.00
METS & SEWS-1	56.00	13.64	35.56	80
METS & SEWS-2	55.91	14.13	34.44	78.97

Table 6.12: Comparison between METS & SEWSs and xDAWN algorithms. Descriptive statistics (mean, standard deviation, min and max) on letter accuracy are reported. The best results are highlighted.

standard deviation, maximum and minimum values. In Figure 6.5, the results by subject are shown for xDAWN, METS & SEWS-1 and METS & SEWS-2. Our algorithm shows better performances for 16 over 22 subjects. It is possible to notice that the algorithms have similar results when the subjects have a high accuracy rate, which suggests that data is less noisy or that the ERP response is stronger making easier the classification. For subjects with low performances, our algorithms perform consistently better, showing that they are more stable among subjects than xDAWN, allowing a more generalized solution for single-trial detection. Also, our solution seems more useful for subjects with poor performances, either because they are first time users or do not generate a strong P300.

Therefore, we conclude that Wavelet-based Semblance performs better than xDAWN for single-trial detection in presence of artifacts and noise. However, our algorithms are conceived to analyze single trials and xDAWN should have a better performance using averaged responses. For this reason, we compare the behavior of the algorithms using different number of repetitions. The results are shown in Figure 6.6, going from a single trial up to five repetitions. It is possible to see that METS & SEWS-2 performs better than METS & SEWS-1 using repetitions. As expected, METS & SEWS performs better than xDAWN for a few (two and three) repetitions. For four repetitions, the results are very similar and for five repetitions xDAWN outperforms our algorithms.

These results are consistent with the theory behind both algorithms: xDAWN improves with more repetitions because averaged ERPs are easier to enhance due to the increased SNR, while the effect of METS becomes moderate because averaging naturally removes uncorrelated components. Also, a better SNR makes easier the classification and implies that useless features removed by SEWS have less impact on the decision boundaries. Nevertheless, xDAWN and our proposed algorithms are not mutually exclusive, meaning that smart combinations of them could be developed in the future.

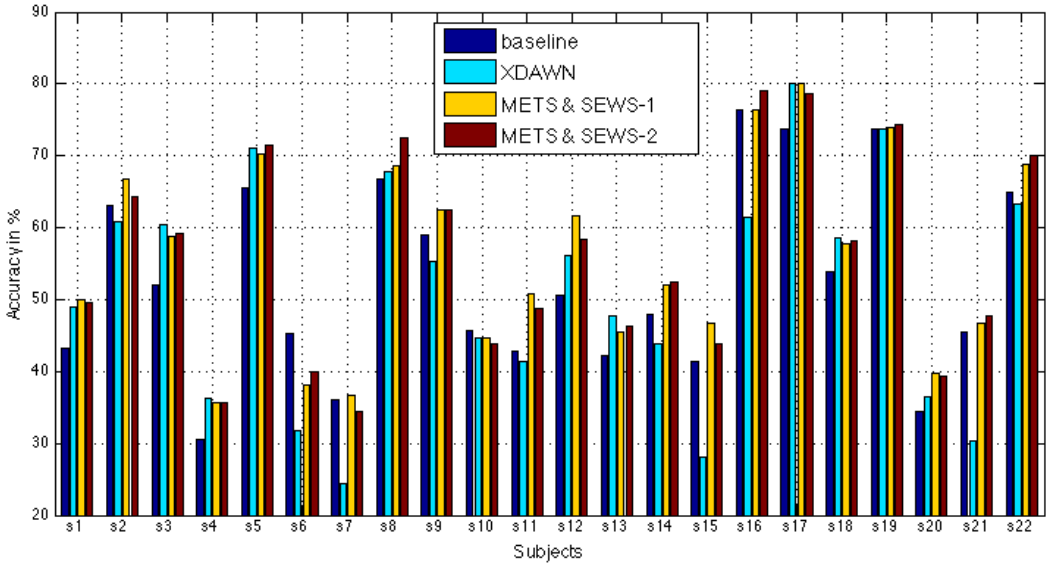


Figure 6.5: Letter accuracy for single-trial detection for the 22 subjects for xDAWN, METS & SEWS-1 and METS & SEWS-2 algorithms and a baseline corresponding to a [1-20] band-pass filter.

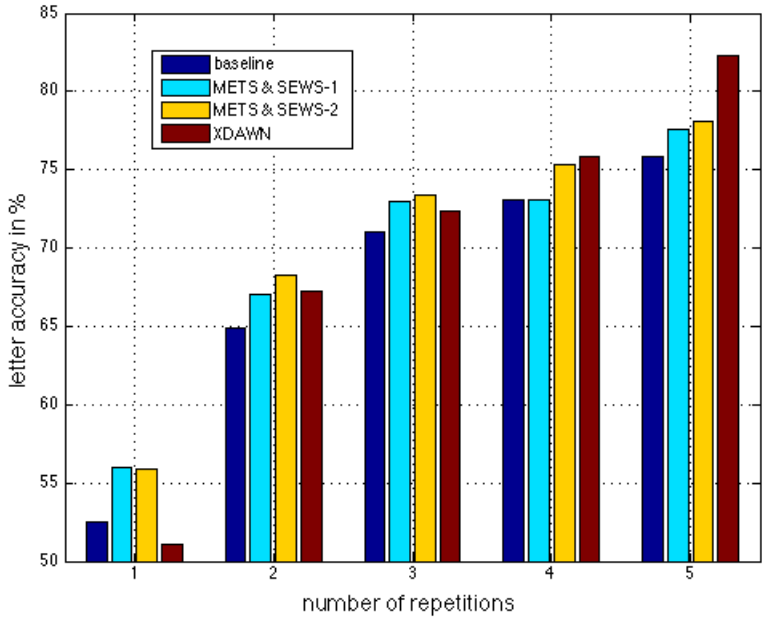


Figure 6.6: Letter accuracy comparison using average of trials for xDAWN, METS & SEWS algorithms and a baseline corresponding to a [1-20] Hz band-pass filter. The parameters used for METS & SEWS algorithms are the same than in Table 6.12.

6.5.2 Experiment #9: Dimensionality Reduction

Dimensionality reduction is useful to i) reduce overtraining and ii) speed up the training and the online classification. Thus, in real-time applications, it could be useful to improve the throughput of BCI systems, making it appropriate to the subjects needs in real life. In this experiment, we reduce the number of features of our algorithms to show that our performance is not severely affected. For doing this, we subsample the output of our algorithms, reducing the number of features by steps of 10%. The initial sampling rate is 256 Hz.

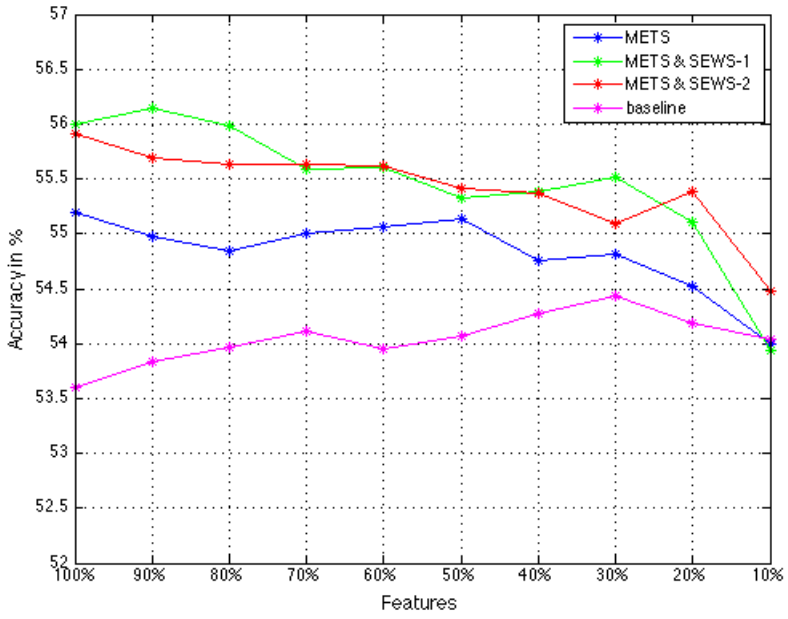


Figure 6.7: Average letter accuracy of METS and SEWS algorithms for the 22 subjects using different percentage of the total features.

In Figure 6.7 the average letter accuracies for METS and SEWS algorithms are presented. As expected, using less features reduces in general the letter accuracy, but the decay is slow: all the results lose around 2% maximum compare to the best value. The baseline improves its results. It can be explained because the subsampling infers a smoothness (due to a low-pass filtering and an interpolation) in the signal that makes easier the classification.

6.5.3 Experiment #10: EPFL Database

We tested our algorithms on a different database to confirm that our approach is generic and do not depend on the type of data of the UAM database. For this purpose, we have selected the EPFL database [Hoffmann et al. 2008], which is a P300 application different to the speller including patients.

This dataset was recorded by the *Ecole Polytechnique Fédérale de Lausanne* using six different images to evoke a P300 response, (see Figure 6.8). The images were individually and randomly flashed for 100 ms with an interstimulus interval of 400 ms. 32 electrodes were recorded (Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz) using a Biosemi Active Two system at a

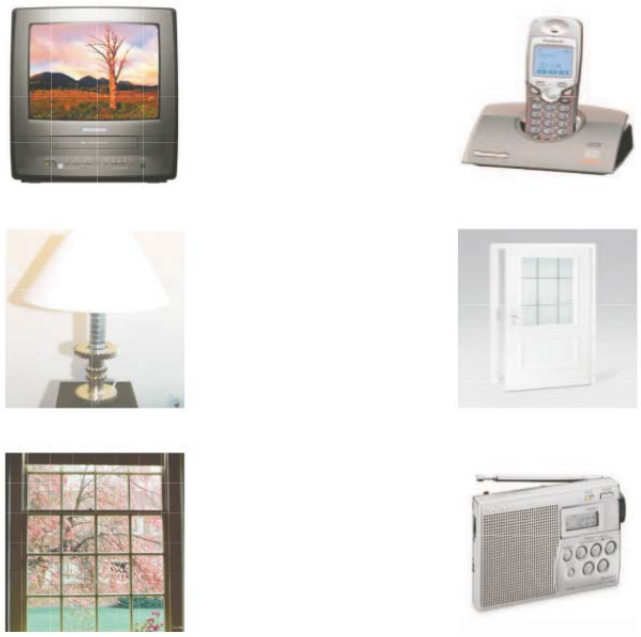


Figure 6.8: EPFL database display matrix. The images were flashed by changing their overall brightness.

sampling rate of 2048 Hz.

The database contains two groups of subjects. In the first group, we found four disabled subjects, each one with a different diagnosis: (1) cerebral palsy, (2) multiple sclerosis, (3) late-stage amyotrophic lateral sclerosis, and (4) traumatic brain and spinal cord injury. The second group is composed by four able-bodied subjects [23].

We used this database blindly, meaning that the parameters of our algorithms and methodology are the same that the ones derived from the UAM database (classifier, mother wavelet, scales and thresholds). We applied the extracting technique and the 8 electrodes configuration (Fz, Cz, Pz, P3, P4, P7, P8 and Oz) presented in [Hoffmann et al. 2008]. The only difference is that our decimation was performed from 2048 Hz to 256 Hz rather than to 32 Hz in [Hoffmann et al. 2008] to have the same sampling rate than the one of the UAM database.

The results for the EPFL database are available in Table 6.13 and show that our algorithms improve the baseline. For this database, the statistical significance is hard to evaluate due to the small number of subjects [24], but METS already improves in average the P300 detection and the combination of METS and SEWS offers a clear improvement for almost all the subjects. In particular, SEWS-2 is more effective for disabled subjects (*i.e.* the first four subjects) and it obtains an impressive improvement from 45.31 % to 62.50 % for the able-bodied subject 8. Also, these results confirm that our techniques are specially useful for subjects with difficulties to handle the BCI skill in single-trial detection.

In summary, this experiment validates our techniques for single-trial ERP detection, not only because it is a different database with a different display matrix and patients, but also because it

23. For further details about subjects see `http://mmspg.epfl.ch/BCI_datasets`.

24. Although a paired t-test is not very meaningful for this database, METS & SEWS-1 was significant at 5% compared to the baseline, and METS & SEWS-2 at 6%.

Subject	Baseline	METS	METS & SEWS-1	METS & SEWS-2
s1	44.53	40.88	42.34	45.26
s2	41.41	49.22	49.22	50.00
s3	58.33	62.88	64.39	64.39
s4	49.21	48.41	52.38	50.00
s5	44.62	46.15	46.92	43.08
s6	48.18	54.01	60.58	55.47
s7	72.93	65.41	70.68	72.93
s8	45.31	53.12	55.47	62.50
μ	50.57	52.51	55.25	55.45
σ	10.36	8.28	9.49	10.36
min	41.41	40.88	42.34	43.08
max	72.93	65.41	70.68	72.93

Table 6.13: P300 detection percentage for the EPFL database using the proposed algorithms for each subject. Descriptive statistics (mean, standard deviation, min and max) on the P300 detection percentage are also reported. The best results are highlighted

uses exactly the same parameters used for the UAM database. Obviously, better performances are expected if the parameters (like the thresholds) are fine-tuned specifically for this database.

Chapter 7

Conclusions

Electroencephalography is the one of the most used brain imaging technique to study the brain activity, and an essential part of neurology and BCI applications. Nonetheless, the information contained in EEG signals are complex and damaged. Indeed, EEG recordings are a mixture of individual sources located at different depths and projected on the scalp. Moreover, even for a repetitive task, the activity of thousand of neurons involved in it has an important variability. Finally, the presence of physiological perturbations (artifacts) and noise disturbs the brain signal, making difficult to isolate the target brain activity. Therefore, specific algorithms have to be developed to extract useful information from these signals. Broadly speaking, the signals should be preprocessed to enhance their characteristics, making easier their analysis, classification and final use by a BCI system.

Nowadays, it is known that EEG contains transient events, named Evoked-Related Potentials (ERP), which are triggered by specific visual, auditory or tactile stimuli. These signals do not require a training to be produced and are sufficiently strongly generated by almost everybody. Nevertheless, these signals are typically one order of magnitude smaller than the EEG background. Thus, a vast majority of BCI applications are based on the average of single-trial responses to the same stimulus in order to amplify the signal-to-noise ratio (SNR). Since attention is an inherent quality of human beings, ERP-based BCI applications usually require only a small cost in terms of cognitive effort, but it is possible that the task becomes mentally difficult when using averaging, because the same stimulus is repeated several times in order to compute the mean of similar responses. Also, this method implicitly assumes that the ERPs have the same latency. This assumption can produce negative effects when it is not true. Although, it is very difficult to directly relate one stimulus to one signal response, it is important to focus the research in single-trial direction to improve brain signals decoding and future applications based on them, such as brain-computer interfaces.

In the last years, the interest in single-trial research has increased, mainly to improve the information transfer rate and to be able to apply BCI in other domains, such as gaming. This research branch has contributed with new algorithms to better analyze and detect single-trial brain signals. Specifically, single-trial signals required to develop pre-processing techniques to deal with the low SNR. However, these techniques usually do not focus in the shared information between recorded channels, which can be particularly useful to identify relevant information when only a single trial response is available. Also, only a small attention has been given to the analysis of within- and between-subjects variability in single trials, which can be used to improve the ERP detection and consequently the study of brain functions.

In the rest of this chapter, we will highlight the general and specific contributions of this

thesis, showing the relevance of our proposal and presenting some of the research lines that this work opens.

7.1 Contributions

In this manuscript, single-trial responses were studied to propose new methods that enhance the information contained in them by denoising them and improving their localization in order to better detect transient events for BCI applications. More precisely, the focus of our research is to study the common information between the recorded channels, emphasizing the useful information contained in time and frequency domains to improve the ERP detection based on single-trial responses [Saavedra and Bougrain 2013a,b].

In this section, we present the main contributions of this thesis with their respective conclusions. The most important contributions are our novel algorithms METS and SEWS. The first one removes artifacts and noise based on the phase correlations among the recorded EEG channels in the wavelet domain. The second one selects automatically a temporal window adapted to the user characteristics in order to remove features in the time domain that do not provide information under our criteria: this temporal-window selection can be performed independently for each channel or jointly for all channels.

7.1.1 Improved Time-frequency Analysis

This thesis has made use of the wavelet transform to analyze the amplitude and phase of the EEG signals components. The expressiveness of this representation enriches the analysis leading to a better identification of the components and allowing interesting applications such as classic wavelet denoising of EEG signals.

We proposed to extend this analysis for EEG using measures of similarity between signals in the wavelet domain, specifically the Semblance and its Mean Resultant Length (MRL) generalization. Measuring the degree of similarity between signals at different scales and time periods can help to identify the variations of various processes and sources. These measures are currently used almost exclusively for geosciences, yet we have shown that they are very interesting tools for the EEG analysis and for the BCI research development.

Specifically, the use of similarity measures between channels or the combination of channels information in general is a promising approach, since channels are recording the same signals from different locations at the same time.

7.1.2 Further Studies of Classic Techniques

The BCI architecture involves several steps, and for each one different techniques can be chosen. The wrong selection of one of them may cause the others to not exhibit their potential. Therefore, it is very important to select a good set of baseline techniques before trying new approaches. This is specially true for single-trial applications, where every step is crucial for achieving a reasonable performance. We have studied the impact of some of the most known techniques used for ERP-based applications, using a large database to assess them.

The band-pass filter has a significant impact in the classification accuracies, playing an important role when setting up the system. The range of band-pass filters used in literature is considerable wide, making difficult the selection without previous information. In particular the low-pass filter seems to influence more the results than the high-pass filter, and the design of a specific filter per subject can improve the classification accuracy. A good compromise for the

UAM database is to choose a [0.1-20] Hz band-pass filter, ensuring that the ERP is not removed from the signal but removing enough undesired components to achieve improved results.

We have tested the two most used classifiers for BCI, namely SWLDA and SVM. It is difficult to get an algorithm that improves accuracy for all subjects of the selected database, due to the differences between subjects brain activities and the noise effect in the dataset. But, SVM clearly outperforms SWLDA, not only in letter accuracy, but also in reliability.

We have studied the impact of different wavelet families for denoising evoked potentials from EEG background activities to improve the detection in a single-trial P300 Speller. These wavelet families are those that are adapted to the ERP waveform. Despite the fact that all these mother wavelets "look similar" to the target P300 wave, and that the performances are more or less competitive, an improvement in terms of stability can be found using Coiflets of order 3. Moreover, since the difference in the results do not substantially improve the ERP detection, we believe that the design of a specific mother wavelet by subject will not produce significant changes.

7.1.3 Multi-channel EEG Thresholding by Similarity

We proposed an alternative wavelet-based denoising technique using the channels correlation in the time-frequency domain. This technique is useful to recognize correlated components between signals, which is particularly useful when the objective is to isolate EEG events that are common to all channels, such as ERPs. Please be aware that the channels under study should be recording approximately the same brain activity. This is a fair constraint, because most BCI applications uses a subset of electrodes that are appropriate to a particular task. An advantage compared with classic wavelet thresholding techniques is that EEG components with small wavelet coefficients, but present in all channels at the same time, will not be removed using our technique.

We empirically demonstrate, using P300-based applications, that the information among channels is useful to remove undesirable components of the signals, improving the letter accuracy for most subjects under study. In some cases, the features extraction for a given stimulus in single trial gave, in the best case, a performance of 80 % of recognized letters using the classic P300 speller, showing that it is possible to work with real data under this methodology. Also, this technique was applied to disabled patients, showing robustness under challenging conditions.

7.1.4 Semblance-based ERP Window Selection

The algorithms presented in this thesis do not need to use a fixed temporal window for subjects, since this window is adjusted according to the individual characteristics of a particular subject. Furthermore, the automatic selection of the temporal window decreases the amount of information to be processed by the classifier.

This technique uses the class information in the time-frequency domain to analyze the differences between classes. The key point of the algorithm is the use of a semblance analysis to study the ERP, which opens several opportunities to develop new algorithms, such as the ones presented in this thesis.

In this manuscript, two different versions of the temporal-window selection were presented, each one corresponding to a different line of thought. The first one (SEWS-1), using a specific temporal window for each channel, is supported by the fact that a single electrode with strong perturbations can disrupt the algorithm performance, for example when an electrode is loose. The second one (SEWS-2), which corresponds to use of the same temporal window for all channels,

is supported by the assumption that the information carried by all electrodes is similar, though similar windows should be expected. As both arguments are strong enough to be worth studying them, both variations were tested.

In the experiments, we empirically demonstrate that both variations improve the target identification preserving their properties under denoised signals. Also, the fact that SEWS-2 benefits of the informations of all channels makes it more stable than SEWS-1. An insightful conclusion drawn from the empirical results is that these techniques are specially useful for subjects with low performances. Our explanation is that, for these subjects, it is difficult for the classifier to find a proper model that distinguishes between classes, because the useless features interfere with the training process. This is why SEWS can rapidly improve the results by removing useless information and facilitating the classification stage. On the contrary, for subjects with high performances, the useless information does not disturb the classifier, which causes that removing useless features have only a minor impact in the results.

Finally, the methods presented in this thesis show more stability among subjects than previous works like xDAWN in single trials, and show a robust behavior under dimensionality reduction.

7.2 Future Work

Despite the advantages of the presented models, further studies are needed to select the best threshold parameters for the algorithms. A possible approach is to select subject-dependent thresholds automatically in order to improve as much as possible the performances, but is still unclear how to develop such algorithms.

A direct future work that emerges from this thesis is the development of new tools based on similarity measures in the wavelet domain. For example, channels selection algorithms can be developed using the MRL measure, in order to detect electrodes with anomalous recordings. For the time-window selection algorithms, other measures (than the variance) can be tested to improve the temporal-window selection. In addition, It is possible to apply the same principle based on channels correlations to, for example, determine a proper band-pass filter for each subject. This can be done by analyzing the variance of the coefficients in the frequency (scales) domain instead of their variance in the temporal domain. Moreover, the proposed analysis could be applied to other steps of a BCI system, such as the automatically detection of a trial that needs to be repeated, or using the similarity information to feed a classifier.

Another opportunity arises from the fact that spatial filters have not been fully explored to use single-trial signals. The algorithms presented in this thesis and the spatial filters are not mutually exclusive, implying that they could be mixed in order to use the best qualities of both techniques for BCI applications.

Also, the methodologies presented here can be used to study the differences between subjects and finally be able to generate a global robust system.

The generality of our methods leads us to think that they can be successfully applied to other BCI problems, such as the study of signals synchronization, or even in other areas, such as finding non-stationary events in astronomical sources (*e.g.* starlight intensity changes due to planet eclipses).

Bibliography

Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *International Journal of Psychophysiology*, 62:309–320.

Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Computer Methods and Programs in Biomedicine*, 94:48–57.

Aizerman, A., Braverman, E. M., and Rozoner, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837.

Anderson, C. (2005). Taxonomy of feature extraction and translation methods for BCI. Online Resource.

Antoniadis, A. (2007). Wavelet methods in statistics: Some recent developements and their applications. *Statistics Surveys*, 1:16–55.

Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, 6(6):1–83.

Başar, E., Schürmann, M., Demiralp, T., Başar-Eroglu, C., and Ademoglu, A. (2001). Event-related oscillations are 'real brain responses' - wavelet analysis and new strategies. *International Journal of Psychophysiology*, 39:91–127.

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E. (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, 4:R32–R57.

Bayliss, J. D. and Ballard, D. H. (2000). A Virtual Reality Testbed for Brain-Computer Interface Research. *IEEE Transaction on Rehabilitation Engineering*, 8:188–190.

Bendel, R. B. and Afifi, A. A. (1977). Comparison of stopping rules in forward "stepwise" regression. *Journal of the American Statistical Association*, 72:46–53.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience*, 87:527–570.

Bihan, D. L. and Breton, E. (1985). Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *C. R. Acad. Sci. (Paris)*, 301(15):1109–1112.

Bin, G., Gao, X., Wang, Y., Hong, B., and Gao, S. (2009). VEP-based Brain-Computer Interfaces: Time, Frequency, and Code modulations. *Comp. Intell. Mag.*, 4(4):22–26.

Birbaumer, N. (1999). Slow Cortical Potentials: Plasticity, Operant Control, and Behavioral Effects. *Neuroscientist*, 5:74–78.

Birbaumer, N., Hinterberger, T., Kübler, A., and Neumann, N. (2003). The thought-translation device (TTD): neurobehavioral mechanisms and clinical outcome. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:120–123.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc.

Bocher, M. (1906). Introduction to the Theory of Fourier's Series. *Annals of Mathematics*, 7(3):81–152.

Boostani, R., Graimann, B., Moradi, M. H., and Pfurtscheller, G. (2007). A comparison approach toward finding the best feature and classifier in cue-based BCI. *Medical and Biological Engineering and Computing*, 45:403–412.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152.

Bougrain, L., Saavedra, C., and Ranta, R. (2012). Finally, what is the best filter for P300 detection? In *Proceedings of the 3rd TOBI Workshop*.

Bunce, S. C., Izzetoglu, M., Izzetoglu, K., Onaral, B., and Pourrezaei, K. (2006). Functional Near-Infrared Spectroscopy. *IEEE Engineering in Medicine and Biology Magazine*, 25:54–62.

Cai, T. and Silverman, B. (2001). Incorporating Information on Neighbouring Coefficients into Wavelet Estimation. *Sankhya: The Indian Journal of Statistics*, 63:127–148.

Chadwick, E. K., Blana, D., Simeral, J. D., Lambrecht, J., Kim, S. P., Cornwell, A. S., Taylor, D. M., Hochberg, L. R., Donoghue, J. P., and Kirsch, R. F. (2011). Continuous neuronal ensemble control of simulated arm reaching by a human with tetraplegia. *Journal of Neural Engineering*, 8(3):034003.

Chang, C. and Lin, C. (2001). LIBSVM: a library for support vector machines.

Chapelle, O. (2007). Training a Support Vector Machine in the Primal. *Neural Comput.*, 19(5):1155–1178.

Chatrian, G. E., Bergamini, L., and Dondey, M. (1974). A glossary of terms most commonly used by clinical electroencephalographers. *Electroencephalographic Clinical Neurophysiology*, 37:538–553.

Chatrian, G. E., Lettich, E., and Nelson, P. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity. *American Journal Of EEG Technology*, 25:83–92.

Cohen, D. (1972). Magnetoencephalography: Detection of the Brain's Electrical Activity with a Superconducting Magnetometer. *Science*, 175:664–666.

Cohen, J. and Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, 25:249–255.

Cooper, G. (2009). Wavelet-based semblance filtering. *Computers & Geosciences*, 35(10):1988–1991.

Cooper, G. and Cowan, D. (2008). Wavelet-based semblance analysis. *Computers & Geosciences*, 34(2):95–102.

Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20:273–297.

Crammer, K. and Singer, Y. (2001). On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines. *Journal of Machine Learning Research*, 2:265–292.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. pub-siam.

Davis, H., Davis, P. A., Loomis, A. L., Hervey, E. N., and Hobart, G. (1939). Electrical reactions of the human brain to auditory stimulation during sleep. *Journal of Neurophysiology*, 2:500–514.

Davis, J. (1986). *Statistics and Data Analysis in Geology*. John Wiley & sons.

Dear, S. P. and Hart, C. B. (1999). Synchronized Cortical Potentials and Wavelet Packets: A Potential Mechanism for Perceptual Binding and Conveying Information. *Brain and Language*, 66:201–231.

Delgado, J. M. (1967). Agression and defense under cerebral radio control. *UCLA Forum Med Sci*, 7:171–193.

Demiralp, T., Ademoglub, A., Istefanopulos, Y., Basar-Eroglu, C., and Basar, E. (2001). Wavelet analysis of oddball P300. *International Journal of Psychophysiology*, 39:221–227.

Dietterich, T. G. and Bakiri, G. (1995). Solving Multiclass Learning Problems via Error-Correcting Output Codes. *J. Artif. Int. Res.*, 2(1):263–286.

Donchin, E., Spencer, K. M., and Wijesinghe, R. (2000). The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface. *IEEE Transaction on Rehabilitation Engineering*, 8:174–179.

Donoho, D. and Johnstone, I. (1992). Minimax Estimation via Wavelet Shrinkage. Technical Report 402, Stanford University.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455.

Donoho, D. and Johnstone, I. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224.

Donoho, D. L. and Johnstone, I. M. (1998). Minimax Estimation via Wavelet Shrinkage. *The Annals of Statistics*, 26(3):879–921.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis*. Wiley.

Duan, K.-B. and Keerthi, S. S. (2005). Which Is the Best Multiclass SVM Method? An Empirical Study. In *Proceedings of the Sixth International Workshop on Multiple Classifier Systems*.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern Classification*. Wiley.

Fan, J. (1997). Comments on "wavelets in statistics: A review" by a. antoniadis. *Journal of the Italian Statistical Society*, 6(2):131–138.

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

Fazel-Rezai, R. (2007). Human Error in P300 Speller Paradigm for Brain-Computer Interface. *The 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France.*, 29:2516–2519.

Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., Schaller, C., Elger, C. E., and Fernández, G. (2004). Neural Bases of Cognitive ERPs: More than Phase Reset. *Journal of Cognitive Neuroscience*, 16:1595–1604.

Finke, A., Lenhardt, A., and Ritter, H. (2009). The MindGame: A P300-based brain-computer interface game. *Neural Networks*, 22:1329–1333.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7:179–188.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the IEE.*, 93:429–457.

Gibbons, H. and Stahl, J. (2007). Response-time corrected averaging of event-related potentials. *Clinical Neurophysiology*, 118:197–208.

Guan, C., Thulasidas, M., and Wu, J. (2004). High performance P300 speller for brain-computer interface. In *IEEE Biomedical Circuits and Systems*.

Guermeur, Y. (2012). A generic model of multi-class support vector machine. *International Journal of Intelligent Information and Database Systems (IJIIDS)*, 6(6):555–577.

Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., Gramatica, F., and Edlinger, G. (2009). How many people are able to control a P300-based brain-computer interface (BCI)? *Neuroscience Letters*, 462:94–98.

Gupta, C. N., Khan, Y. U., Palaniappan, R., and Sepulveda, F. (2009). Wavelet Framework for Improved Target Detection in Oddball Paradigms Using P300 and Gamma Band Analysis. *Biomedical Soft Computing and Human Sciences*, 14(2):61–67.

Hamadicharef, B., Xu, M., and Aditya, S. (2010). Brain-Computer Interface (BCI) based Musical Composition. In *Proceedings of the 2010 International Conference on Cyberworlds*.

Hochberg, L. R. and Donoghue, J. P. (2006). Sensors for Brain-Computer Interfaces. *IEEE Engineering in Medicine and Biology Magazine*, 25(5):32–38.

Hocking, R. R. (1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1):1–49.

Hoffmann, U., Vesin, J.-M., Ebrahimi, T., and Diserens, K. (2008). An efficient P300-based brain-computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125.

Holman, L. and Tumeh, S. S. (1990). Single-Photon Emission Computed Tomography (SPECT) Applications and Potential. *JAMA*, 263(4):561–564.

Hsu, C.-W. and Lin, C.-J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on neural networks*, 13:415–425.

Huettel, S. A. and McCarthy, G. (2004). What is odd in the oddball task?: Prefrontal cortex is activated by dynamic changes in response strategy. *Neuropsychologia*, 42(3):379–386.

Hwang, H., Kwon, K., and Im, C. (2009). Neurofeedback-based motor imagery training for brain-computer interface (BCI). *Journal of Neuroscience Methods*, 179:150–156.

Iversen, I., Ghanayim, N., Kübler, A., Neumann, N., Birbaumer, N., and Kaiser, J. (2008). A brain-computer interface tool to assess cognitive functions in completely paralyzed patients with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119:2214–2223.

Jackson, M. M., Ozawa, K., Kido, K., McClendon, I., and Kerwin, R. (2013). Field Study of an fNIR-Based Brain-Computer Interface for Communication. In *Proceedings of the Fifth International Brain-Computer Interface Meeting*.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10(2):371–375.

Kaufmann, T., Schulz, S., Grünzinger, C., and Kübler, A. (2011). Flashing Characters with Famous Faces Improves ERP-based Brain Computer Interface Performance. *Journal of Neural Engineering*, 8:056016 (10pp).

Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K., and Goldwaithe, J. (2000). Direct Control of a Computer from the Human Central Nervous System. *IEEE Transactions on Rehabilitation Engineering*, 8:198–202.

Kennedy, P. R., Mirra, S. S., and Bakay, R. A. (1992). The cone electrode: ultrastructural studies following long-term recording in rat and monkey cortex. *Neuroscience Letters*, 142:89–94.

Kissler, J., Herbert, C., Winkler, I., and Junghofer, M. (2009). Emotion and attention in visual word processing: an ERP study. *Biol Psychol*, 80(1):75–83.

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., and Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, 31:1003–1016.

Koivisto, M. and Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience & Biobehavioral Reviews*, 34:922–934.

Kolev, V., Demiralp, T., Yordanova, J., Ademoglu, A., and Isoglu-Alkaç, U. (1997). Time-Frequency Analysis Reveals Multiple Functional Components During Oddball P300. *NeuroReport*, 8:2061–2065.

Krusienski, D., Sellers, E., Cabestaing, F., Bayoudh, S., McFarland, D., Vaughan, T., and Wolpaw, J. (2006). A comparison of classification techniques for the P300 speller. *Journal of neural engineering*, 3:299–305.

Krusienski, D., Sellers, E., McFarland, D., Vaughan, T., and Wolpaw, J. (2008). Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167:15–21.

Krusienski, D. J., Schalk, G., McFarland, D. J., and Wolpaw, J. R. (2007). A mu-Rhythm Matched Filter for Continuous Control of a Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering*, 54:273–280.

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

Kübler, A., Kotchoubey, B., Salzmann, H.-P., Ghanayim, N., Perelmouter, J., Hömberg, V., and Birbaumer, N. (1998). Self-regulation of slow cortical potentials in completely paralyzed human patients. *Neuroscience Letters*, 252:171–174.

Lal, T. N., Hinterberger, T., Widman, G., Schroeder, M., Hill, J., Rosenstiel, W., Elger, C., Schölkopf, B., and Birbaumer, N. (2005). Methods Towards Invasive Human Brain Computer Interfaces. *Advances in Neural Information Processing System*, 17:737–744.

Ledesma-Ramirez, C., Bojorges-Valdez, E., Yanez-Suarez, O., Saavedra, C., Bougrain, L., and Gentilleti., G. (2010). An open-Access P300 Speller Database. In *BCI meeting 2010*.

Lee, Y., Lin, Y., and Wahba, G. (2004). Multicategory Support Vector Machines: Theory and Application to the Classification of Microarray Data and Satellite Radiance Data. *Journal of the American Statistical Association*, 99(465):67–81.

Leeb, R., Sagha, H., Chavarriaga, R., and Millán, J. (2011). A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. *Journal of Neural Engineering*, 8:025011.

Lemm, S., Müller, K.-R., and Curio, G. (2009). A Generalized Framework for Quantifying the Dynamics of EEG Event-Related Desynchronization. *PLOS Computational Biology*, 5(8):e1000453.

Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9:53–78.

Liu, J. and Newsome, W. T. (2006). Local Field Potential in Cortical Area MT: Stimulus Tuning and Behavioral Correlations. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26:7779–7790.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13.

Lotte, F., Scherer, R., and Lécuyer, A. (2013). A:8 Teaching the BCI Skill: Feedback and Human Training Approaches. BCI meeting 2013 workshop.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. The MIT press.

Luo, W., Feng, W., He, W., Wang, N.-Y., and Luo, Y.-J. (2010). Three stages of facial expression processing: ERP study with rapid serial visual presentation. *NeuroImage*, 49(2):1857–1867.

Mak, J. N. and Wolpaw, J. R. (2009). Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects. *IEEE Reviews in Biomedical Engineering*, 2:187–199.

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci.*, 94:10979–10984.

Mallat, S. (1989). Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition.

Malmivuo, J. and Robert Plonsey, R. (1995). *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. New York, Oxford University Press.

Marchetti, M., Piccione, F., Silvoni, S., and Priftis, K. (2012). Exogenous and endogenous orienting of visuospatial attention in P300-guided brain computer interfaces: A pilot study on healthy participants. *Clinical Neurophysiology*, 123:774–779.

Mason, S. G., Bashashati, A., Fatourechi, M., Navarro, K. F., and Birch, G. E. (2007). A Comprehensive Survey of Brain Interface Technology Designs. *Annals of Biomedical Engineering*, 35:137–169.

Mason, S. G. and Birch, G. E. (2003). A General Framework for Brain-Computer Interface Design. *EEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:70–85.

Matsuoka, S. (1990). Theta Rhythms: State of Consciousness. *Brain Topography*, 3:203–208.

McFarland, D. J., Anderson, C. W., Müller, K.-R., Schlögl, A., and Krusienski, D. J. (2006). BCI Meeting 2005—Workshop on BCI Signal Processing: Feature Extraction and Translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–137.

Mellinger, J., Schalk, G., Braun, C., Preissl, H., Rosenstiel, W., Birbaumer, N., and Kübler, A. (2007). An MEG-based brain-computer interface (BCI). *NeuroImage*, 36:581–593.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K. R. (1999). Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41–48.

Mori, S. and Barker, P. B. (1999). Diffusion Magnetic Resonance Imaging: Its Principle and Applications. *The Anatomical Record*, 257:102–109.

Mugler, E., Bensch, M., Halder, S., Rosenstiel, W., Bogdan, M., Birbaumer, N., and Kübler, A. (2008). Control of an Internet Browser Using the P300 Event-Related Potential. *International Journal of Bioelectromagnetism*, 10:56–63.

Mukamel, R. and Fried, I. (2012). Human Intracranial Recordings and Cognitive Neuroscience. *Annual Review of Psychology*, 63:511–537.

Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G., and Blankertz, B. (2004). Machine Learning Techniques for Brain-computer Interfaces. *Biomed Tech*, 49(1):11–22.

Münssinger, J. I., Halder, S., Kleih, S. C., Furdea, A., Raco, V., Hösle, A., and Kübler, A. (2010). Brain Painting: First Evaluation of a New Brain-Computer Interface Application with ALS-Patients and Healthy Volunteers. *Front. Neurosci.*, 4:182.

Myrden, A. J. B., Kushki, A., Sejdic, E., Guerguerian, A.-M., and Chau, T. (2011). A Brain-Computer Interface Based on Bilateral Transcranial Doppler Ultrasound. *PLoS ONE*, 6(9):e24170.

Näätänen, R. (1992). *Attention and Brain Function*. Lawrence Erlbaum Associates.

Nair, D. R., Burgess, R., McIntyre, C. C., and Lüders, H. (2008). Chronic subdural electrodes in the management of epilepsy. *Clinical Neurophysiology*, 119:11–28.

Nam, C. S., Jeon, Y., Li, Y., Kim, Y.-J., and Yoon, H.-Y. (2009). Usability of the P300 Speller: Towards a More Sustainable Brain-Computer Interface. *e-Minds: International Journal on Human-Computer Interaction*, 1:111–125.

Ng, A. Y. and Jordan, M. I. (2001). On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes. *NIPS*, pages 841–848.

Niedermeyer, E. and da Silva, F. L. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Philadelphia , Lippincott Williams & Wilkins.

Nijboer, F., Birbaumer, N., and Kübler, A. (2010). The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study. *Frontiers in Neuroprosthetics*, 4.

Nijholt, A. (2009). BCI for Games: A 'State of the Art' Survey. In *Proceedings of the 7th International Conference on Entertainment Computing*, pages 225–228.

Nooh, A. A., Yunus, J., and Daud, S. M. (2011). A Review of Asynchronous Electroencephalogram-based Brain Computer Interface Systems. *International Conference on Biomedical Engineering and Technology IPCBEE*, 11:55–59.

Nuwer, M., Comi, C., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J.-M., Hinrichs, H., Ikeda, A., Luccas, F., and Rappelsburger, P. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and clinical Neurophysiology*, 106:259–261.

Odom, J. V., Bach, M., Barber, C., Brigell, M., Marmor, M. F., Tormene, A. P., Holder, G. E., and Vaegan (2004). Visual evoked potentials standard (2004). *Documenta Ophthalmologica*, 108:115–123.

Olofsson, J. K., Nordin, S., Sequeira, H., and Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biol Psychol*, 77(3):247–265.

Pastor, M. A., Artieda, J., Arbizu, J., Valencia, M., and Masdeu, J. C. (2003). Human Cerebral Activation during Steady-State Visual-Evoked Responses. *The Journal of Neuroscience*, 23:11621–11627.

Patel, S. H. and Azzam, P. N. (2005). Characterization of N200 and P300: Selected Studies of the Event-related Potential. *Int. J. Med. Sci.*, 2(4):147–154.

Patil, S. A. (2008). Brain Gate as an Assistive and Solution Providing Technology for Disabled People. *ICBME Proceedings*, 23:1232–1235.

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalography and clinical Neurophysiology*, 83:62–69.

Pfurtscheller, G., Allison, B. Z., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., Zander, T. O., Mueller-Putz, G., Neuper, C., and Birbaumer, N. (2010). The hybrid BCI. *Front. Neurosci.*, 4(30).

Pfurtscheller, G. and da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110:1842–1857.

Pfurtscheller, G., Jr., A. S., and Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band - an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, 24:39–46.

Pfurtscheller, G. and Neuper, C. (1994). Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man. *Neuroscience Letters*, 174:93–96.

Pfurtscheller, G., Neupera, C., Flotzingera, D., and Pregenzerb, M. (1997). EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and Clinical Neurophysiology*, 103:642–651.

Picton, W. T. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9(4):456–479.

Platt, J. C., Cristianini, N., and Shawe-Taylor, J. (2000). Large Margin DAGs for Multiclass Classification. In *Advances in Neural Information Processing Systems 12*, pages 547–553.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118:2128–2148.

Polich, J. and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, 41:103–146.

Polikar, R., Topalis, A., Green, D., Kounios, J., and Clark, C. M. (2007). Comparative multiresolution wavelet analysis of ERP spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease. *Computers in Biology and Medicine*, 37:542–558.

Quiroga, R. and Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. *Clinical Neurophysiology*, 114(2):376–390.

Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal Spatial Filtering of Single Trial EEG During Imagined Hand Movement. *IEEE Transactions on Rehabilitation Engineering*, 8:441–446.

Ranta, R. and Louis-Dorr, V. (2010). Hysteresis Thresholding: A Graph-Based Wavelet Block Denoising Algorithm. *The Open Signal Processing Journal*, 3:6–12.

Rauss, K., Schwartz, S., and Pourtois, G. (2011). Top-down effects on early visual processing in humans: A predictive coding framework. *Neuroscience & Biobehavioral Reviews*, 35:1237–1253.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., and Lécuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments. *Teleoperators and virtual environments*, 19:35–53.

Rentzeperis, I., Nikolaev, A. R., Kiper, D. C., and Leeuwen, C. V. (2012). Relationship between neural response and adaptation selectivity to form and color: an ERP study. *Front Hum Neurosci.*, 6:89.

Ripley, B. (1995). *Pattern Recognition and Neural Networks*. Cambridge university press.

Rivet, B., Cecotti, H., Souloumiac, A., Maby, E., and Mattout, J. (2011). Theoretical Analysis of XDAWN Algorithm: Application to an Efficient Sensor Selection in a P300 BCI. *19th European Signal Processing Conference*.

Rivet, B., Souloumiac, A., Attina, V., and Gibert, G. (2009). xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain Computer Interface. *IEEE Trans. Biomed. Engineering*, 56(8):2035–2043.

Roy, C. and Sherrington, C. (1890). On the regulation of the blood-supply of the brain. *The Journal of Physiology*, 11(1-2):85–117.

Russo, F. D., Martínez, A., Sereno, M. I., Pitzalis, S., and Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Hum. Brain Mapp*, 15:95–111.

Saavedra, C. and Bougrain, L. (2010). Wavelet denoising for p300 single-trial detection. In *Proceedings of 5th french conference in computational neuroscience*.

Saavedra, C. and Bougrain, L. (2012). Processing stages of visual stimuli and event-related potentials. In *Neurocomp/KEOpS'12 workshop*.

Saavedra, C. and Bougrain, L. (2013a). Denoising and time-window selection using wavelet-based semblance for improving erp detection. In *Proceedings of the Fifth International Brain-Computer Interface Meeting*.

Saavedra, C. and Bougrain, L. (2013b). Wavelt-based semblance for p300 single-trial detection. *International Conference on Bio-Inspired Systems and Signal Processing*.

Saavedra, C., Bougrain, L., and Ranta, R. (2011). Hysteresis thresholding for wavelet denoising applied to p300 single-trial detection. In *INRIA technical report, ISNR INRIA/RR-7723*.

Salvaris, M. and Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *J Neural Eng.*, 6:046011.

Sauseng, P., Klimesch, W., Gruber, W. R., Hanslmaryr, S., Freunberger, R., and Doppelmayr, M. (2007). Are event-related potentials components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience*, 146:1435–1444.

Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., and Wolpaw, J. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043.

Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 1st edition.

Sellers, E. W. and Donchin, E. (2006). A P300-based brain-computer interface: Initial tests by ALS patients. *Clinical Neurophysiology*, 117:538–548.

Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., and Wolpaw, J. R. (2006). A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73:242–252.

Squires, K. and Donchin, E. (1976). Beyond averaging: the use of discriminant functions to recognize event related potentials elicited by single auditory stimuli. *Electroencephalogr Clin. Neurophysiol.*, 41(5):449–59.

Squires, N., Squires, K., and Hillyard, S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin. Neurophysiol.*, 38:387–401.

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, 150:1187–1188.

Taheri, B. A., Knight, R. T., and Smith, R. L. (1994). A dry electrode for EEG recording. *Electroencephalography and Clinical Neurophysiology*, 90:376–383.

Takano, K., Komatsu, T., Hata, N., Nakajima, Y., and Kansaku, K. (2009). Visual stimuli for the P300 brain-computer interface: A comparison of white/gray and green/blue flicker matrices. *Clinical Neurophysiology*, 120:1562–1566.

Tangermann, M., Krauledat, M., Grzeska, K., Sagebaum, M., Blankertz, B., Vidaurre, C., and Müller, K.-R. (2009). Playing pinball with non-invasive BCI. *Advances in Neural Information Processing Systems*, 21:1641–1648.

Ter-Pogossian, M. M. (1982). Positron Emission Tomography (PET). *Journal of Medical Systems*, 6:569–577.

Thornton, A. R. D. (2008). Evaluation of a technique to measure latency jitter in event-related potentials. *Journal of Neuroscience Methods*, 168:248–255.

Torrence, C. and Compo, G. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78.

Townsend, G., LaPallo, B. K., Boulay, C. B., Krusienski, D. J., Frye, G. E., Hauser, C. K., Schwartz, N. E., Vaughan, T. M., Wolpaw, J. R., and Sellers, E. W. (2010). A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clin Neurophysiol*, 7:1109–1120.

Treder, M. S. and Blankertz, B. (2010). (C)overt attention and visual speller design in an ERP-based brain-computer interface. *Behavioral and Brain Functions*, 6(28):1–13.

Truccolo, W. A., Ding, M., Knuth, K. H., Nakamura, R., and Bressler, S. L. (2002). Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity. *Clinical Neurophysiology*, 113:206–226.

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

van den Berg, R., Roerdink, J. B. T. M., and Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision*, 7(2):14.1–11.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics*. Springer-Verlag New York, Inc.

Vapnik, V. (1999). An Overview of Statistical Learning Theory. *IEEE Transaction on Neural Networks*, 10(5):988–999.

Vaughan, T. M., McFarland, D. J., Schalk, G., Sarnacki, W. A., Krusienski, D. J., Sellers, E. W., and Wolpaw, J. R. (2006). The Wadsworth BCI Research and Development Program: At Home With BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14:229–233.

Vidal, J. J. (1973). Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, 2:157–180.

Wang, Y., Gao, X., Hong, B., Jia, C., and Gao, S. (2008). Brain-Computer Interfaces Based on Visual Evoked Potentials. *IEEE Engineering in Medicine and Biology Magazine*, 27:64–71.

Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., Broberg, M., Wallace, A., DeLosAngeles, D., Lillie, P., Hardy, A., Fronsko, R., Pulbrook, A., and Willoughby, J. O. (2007). Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. *Clinical Neurophysiology*, 118:1877–1888.

Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., and Vaughan, T. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791.

Wolpaw, J. R. (2007). Brain-computer interfaces as new brain output pathways. *The Journal of Physiology*, 579:613–619.

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., Donchin, E., Quatrano, L. A., Robinson, C. J., and Vaughan, T. M. (2000). Brain-Computer Interface Technology: A Review of the First International Meeting. *IEEE Transactions on Rehabilitation Engineering*, 8:164–173.

Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. *PNAS.*, 101(51):17849–17854.

Wolpaw, J. R. and Wolpaw, E. W. (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press.

Woody, C. D. (1967). Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. *Medical and Biological Engineering and Computing*, 5:539–554.

Yong, Y., Hurley, N., and Silvestre, G. (2005). Single-trial EEG classification for brain-computer interface using wavelet decomposition. In *European Signal Processing Conference, EUSIPCO 2005*. Citeseer.

Yoshimura, N. and Itakura, N. (2011). Usability of Transient VEPs in BCIs. *Recent Advances in Brain-Computer Interfaces*, 6:119–134.

Zhu, D., Bieger, J., Molina, G. G., and Aarts, R. M. (2010). A survey of stimulation methods used in SSVEP-based BCIs. *Intell. Neuroscience*, 2010.

Annexe A

Résumé étendu

Une interface cerveau-ordinateur (en anglais Brain-Computer Interface ou BCI) est un système de contrôle ou de communication qui traduit l'activité cérébrale en commandes pour un logiciel ou d'autres dispositifs comme des prothèses motorisées. Initialement développées pour aider les personnes présentant des déficiences motrices importantes, les BCIs peuvent aussi avoir des applications pour des personnes en bonne santé en tant que nouveau moyen d'interaction, en particulier dans le domaine des jeux video.

Un des phénomènes physiologiques permettant cette interaction est le potentiel évoqué cognitif P300. Il correspond à une déflexion positive observable par électroencéphalographie autour de 300 ms après la présentation d'un stimulus rare et attendu (voir figure A.1).

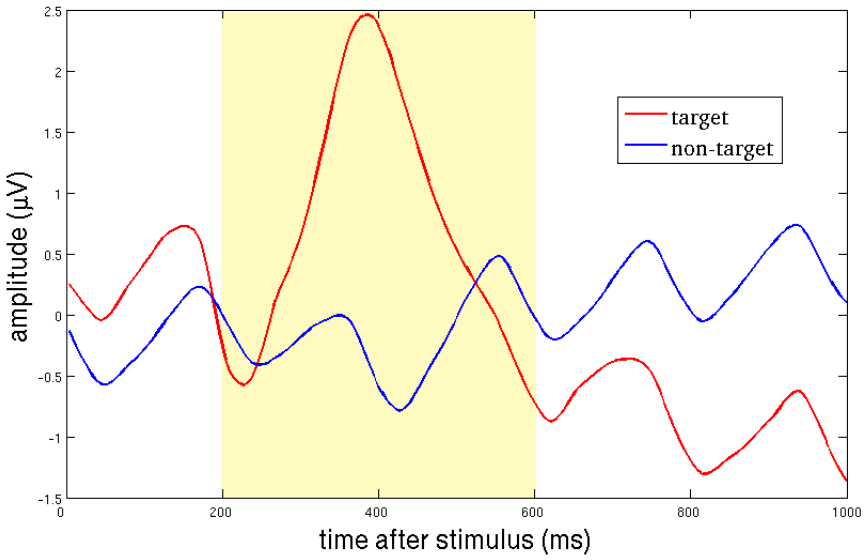


FIGURE A.1 – Moyennes de 480 réponses contenant un P300 (en rouge) et de 2400 réponses n'en contenant pas (en bleu) pour la fenêtre classique [0, 1000] ms. Ici, la différence est manifeste dans la fenêtre [200, 600 ms.

Le potentiel P300 présente plusieurs intérêts : i) il permet de proposer un plus grand nombre de commandes que d'autres phénomènes neuronaux exploitables dans le cadre d'une interface

cerveau-ordinateur comme la variation de l'énergie dans certaines bandes de fréquences ; ii) aucun entraînement de la part de l'utilisateur n'est nécessaire ; iii) cette réaction est observable chez la quasi totalité des personnes y compris les patients.

Ces réactions cérébrales sont cependant difficiles à observer parce que le matériel utilisé pour enregistrer les signaux sur le scalp se situe à une distance importante des sources et enregistre également du bruit provenant du matériel d'enregistrement et de l'environnement électromagnétique, et des artefacts provenant des mouvements musculaires de l'utilisateur. Le P300 a une amplitude environ dix fois inférieure au reste de l'activité électroencéphalographique enregistrée. En conséquence, il est nécessaire d'utiliser des techniques poussées de prétraitement et de discrimination pour reconnaître certaines composantes utiles comme le P300. Habituellement, pour améliorer le rapport signal sur bruit des potentiels évoqués, les stimulations sont répétées afin de moyenner les réponses enregistrées. Le P300 étant synchronisé sur l'apparition du stimulus, à la différence du reste de l'activité cérébrale, il augmente lors du moyennage alors que le reste de l'activité, qui peut être apparentée à une sinusoïde bruitée, s'atténue. Les performances de reconnaissance du système s'améliorent mais l'interaction est ralentie. Deux autres problèmes liés au moyennage sont d'une part l'étalement de la moyenne à cause d'une faible mais réelle variation de la latence des réponses et d'autre part l'apparition éventuelle de faux P300.

Dans cette thèse, de nouvelles techniques basées sur la théorie des ondelettes et utilisant seulement des réponses individuelles sont développées pour améliorer la performance des systèmes BCI basés sur la détection du P300. La solution présentée dans cette thèse débruite le signal dans le domaine des ondelettes en considérant simultanément, et non de manière indépendante, l'information de phase fournie par l'ensemble des canaux enregistrés. Nous avons également étendu cette approche à l'étude des phases et des amplitudes pour sélectionner automatiquement la fenêtre du signal à observer pour chaque utilisateur. Les algorithmes présentés dans cette thèse sont évalués sur deux bases de données l'une contenant 22 sujets novices et l'autre contenant 8 sujets dont 4 patients. Les résultats montrent que les modèles présentés dans cette thèse améliorent significativement la détection des P300 sans moyennage par rapport aux méthodes conventionnelles.

A.1 Méthodes multicanaux d'analyse et de décorrélation à base d'ondelettes

L'étude des potentiels évoqués (PE) permet de mieux comprendre le fonctionnement du cerveau. Habituellement, cette étude est basée sur la réponse moyenne de plusieurs essais (réalisés dans les mêmes conditions) afin de diminuer la variabilité entre les essais. Cependant, outre l'évidente augmentation du temps d'acquisition pour obtenir une moyenne, la répétition du protocole est ennuyeuse pour l'utilisateur, ce qui a pour effet de diminuer l'amplitude du PE. Plusieurs études ont montré qu'il est possible de détecter des PE sans moyennage [Quiroga et Garcia, 2003 ; Yong et al. 2005 ; Gupta et al. 2009]. Cette possibilité permettrait d'améliorer la performance des BCI basées sur les PE, telles que le système d'écriture de Farwell et Donchin [Farwell and Donchin 1988].

Cette thèse présente de nouvelles méthodes pour localiser et isoler les PE sans moyennage. Ces méthodes se basent sur les similarités qui existent entre les canaux enregistrés afin d'améliorer le débruitage et la sélection de la fenêtre d'étude. Elles permettent donc de mieux détecter les PE sans moyennage et, par conséquent, dans le cas des BCI basées sur le PE, d'accélérer l'interaction. Ces méthodes sont basées sur la théorie des ondelettes car elles permettent d'étudier les signaux

dans les domaines temporel et fréquentiel en même temps.

A.1.1 Méthode multicanaux de débruitage

Les enregistrements électroencéphalographiques (EEG) contiennent généralement du bruit provenant des lignes électriques et des artefacts tels que l'activité des muscles du visage. Leur présence dégrade l'analyse des signaux EEG y compris la détection des PE nécessitant de nouvelles méthodes pour retrouver les signaux propre à l'activité cérébrale. Une hypothèse importante dans le traitement du bruit est qu'il est supposé indépendant de l'activité cérébrale. En outre, il est supposé que les signaux cérébraux sont enregistrés (presque) instantanément par les électrodes, ce qui implique que chaque canal enregistré est fortement corrélé avec les autres. Sous cette condition, il est possible de conclure que les éléments des signaux qui ne sont pas corrélés sur les canaux sont supposés être du bruit ou des artefacts, et peuvent être retirés de l'enregistrement.

Du point de vue de la théorie des ondelettes [Mallat 2008], l'hypothèse fondamentale de débruitage est que les grands coefficients obtenus par ces méthodes d'analyse correspondent aux signaux et que les petits coefficients correspondent au bruit. Par conséquent, l'annulation de bruit peut être effectuée par seuillage [Antoniadis 2007] ; seuls les grands coefficients seront utilisés pour construire le signal débruité. Le problème avec les méthodes actuelles est qu'elles débruitent chaque canal indépendamment des autres, perdant au passage les informations fournies par l'ensemble des canaux (comme la phase et l'amplitude). Pour cette raison, dans cette thèse, nous considérons, dans le domaine des ondelettes, l'information disponible dans tous les canaux, en particulier les angles de phase.

Soit $\mathbf{X} \in \mathbb{R}^{T\mathbf{x}C}$ l'ensemble des données et $x_c(t)$ le signal enregistré par la $c^{\text{ème}}$ électrode, $c \in \{1, \ldots, C\}$, au temps t, $t \in \{1, \ldots, T\}$. Le débruitage que nous utilisons s'obtient par seuillage et s'appuie sur les coefficients de la *Mean Resultant Length* (MRL) [Cooper 2009], *i.e.* à partir des coefficients obtenus pour tous les canaux, et non des coefficients des canaux individuels. Pour ce faire, la MRL est calculée en utilisant une décomposition par ondelettes de tous les canaux $W_\Psi^{x_c}$, avec Ψ l'ondelette mère, pour un instant t et une échelle a donnés, grâce à l'équation suivante :

$$MRL(a,t) = \frac{\sqrt{(\sum_{i=1}^{N} \Re(W_\Psi^i(a,t)))^2 + (\sum_{i=1}^{N} \Im(W_\Psi^i(a,t)))^2}}{\sum_{i=1}^{N} |W_\Psi^i(a,t)|}. \tag{A.1}$$

Une valeur proche de 0 indique que les signaux sont peu similaires alors qu'une valeur proche de 1 indique qu'ils le sont fortement.

Le calcul de la MRL se fait via la combinaison des angles de phase des parties réelle et imaginaire de la décomposition en ondelettes. La transformée en ondelettes discrète utilise des familles d'ondelettes qui sont orthogonales les unes par rapport aux autres, de sorte que la partie imaginaire peut être obtenue avec la transformée de Hilbert du canal. Ensuite, il est possible de mettre à zéro tous les coefficients inférieurs à un seuil fixé τ_d et finalement de reconstruire le signal à l'aide des coefficients d'ondelettes filtrés. L'algorithme A.1 montre le processus complet de débruitage à l'aide de la MRL.

A.1.2 Sélection de la fenêtre temporelle

Dans la grande majorité des applications BCI basées sur la détection du P300, la réponse au stimulus est étudiée dans une fenêtre temporelle débutant à l'apparition de celui-ci et pendant

Algorithm A.1 Seuillage multicanaux de signaux EEG par similarité (METS, Multichannel Thresholding EEG Signal)

Données : Soit $\mathbf{X}$ la matrice des signaux EEG, avec C canaux et T échantillons.
Résultat : Le signal débruité $\widetilde{\mathbf{X}}$

Choisir le seuil de corrélation τ_d, $0 \leq \tau_d \leq 1$
pour chaque canal c de 1 à C **faire**
    Calculer la transformée de Hilbert H_c du signal x_c.
    Calculer la transformée en ondelettes discrète $W_\psi^{x_c}$ de x_c et $W_\psi^{H_c}$ de H_c avec l'Equation 3.20.
fin de pour
pour t de 1 à T **faire**
    Calculer MRL(t) avec l'Equation A.1
    si MRL(t) $< \tau_d$ **then**
        mettre à zéro $W_\psi^{x_c}$ au temps t, $\forall c$
    fin de si
fin de pour
pour c de 1 à C **faire**
    Reconstruire le signal du canal c avec $W_\psi^{x_c}$.
fin de pour

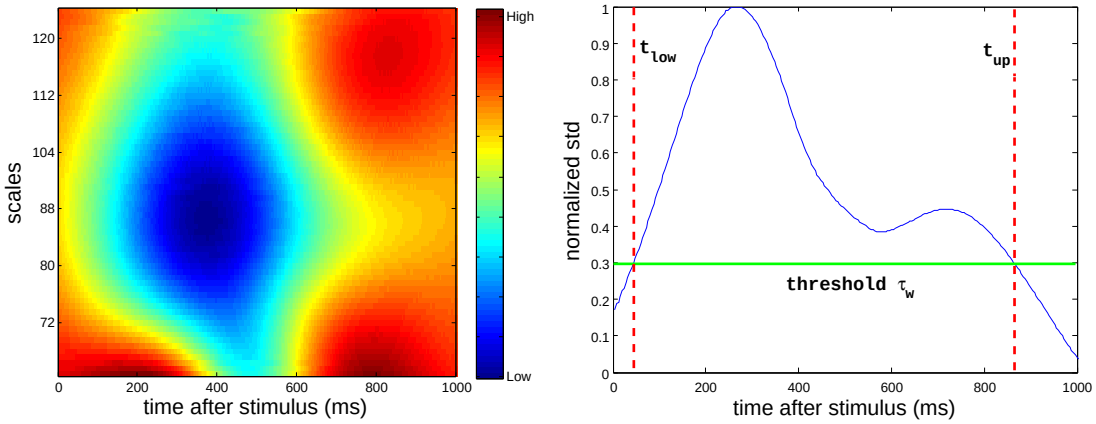
une durée d'une seconde pour être sûr de contenir le P300 et d'autres composantes utiles pour l'analyse. Cependant, les réponses ont des latences et des amplitudes différentes pour chaque sujet. Une fenêtre trop large inclut donc des données non pertinentes qui peuvent rendre l'apprentissage plus difficile et diminuer les performances du classifieur.

Il est donc utile de déterminer une fenêtre temporelle adaptée à chaque sujet. Pour cela, il faudrait déjà être capable de dire à quoi correspond le début et la fin d'un P300. De plus, d'autres composantes précédant le P300 sont également utiles pour détecter une réaction du sujet. Le faible rapport signal sur bruit empêche généralement cette identification si l'on observe une réponse unique. Pour y parvenir, nous proposons de comparer la moyenne des réponses contenant un P300 avec celle des autres réponses en utilisant les mesures de similarité basées sur les ondelettes que nous avons définies dans la section 3.

La figure A.1 montre la différence entre les réponses moyennées contenant un P300 ou pas. Par défaut, la fenêtre temporelle d'analyse utilisée est [0, 1000] ms mais il est possible de la réduire à la seule zone où le P300 réside. Par exemple, ici, la zone [200, 600] ms suffirait. On constate aussi que bien que la désignation P300 suppose un pic de réponse 300 ms après le stimulus, ici, il se situe plus tôt. La comparaison des moyennes permet de préciser la localisation de la fenêtre d'étude afin d'éliminer l'information non pertinente.

Dans cette thèse, nous proposons d'analyser et de comparer les moyennes dans le domaine des ondelettes. L'idée est d'avoir une meilleure expressivité à la fois de l'amplitude et de la phase pour isoler le P300 et les autres composantes utiles. L'hypothèse faite est que les éléments corrélés dans les deux signaux moyennés correspondent à une activité EEG de fond tandis que les éléments non corrélés correspondent aux composantes du PE. Ces similarités peuvent être quantifiées à l'aide d'une mesure dans le domaine des ondelettes.

Soient $x_c(t)$ le signal enregistré par la $c^{\text{ème}}$ électrode, $c \in \{1, \ldots, C\}$, au temps t, $t \in \{1, \ldots, T\}$, $\mathcal{T}$ l'ensemble des signaux appartenant à la classe cible (target) c'est-à-dire contenant un P300 et $\mathcal{N}$ l'ensemble des signaux n'appartenant pas à la classe cible (non-target). Les réponses



(a) (b)

FIGURE A.2 – (a) Valeurs de la mesure de similarité D obtenues avec l'équation 3.30. Les couleurs indiquent la similarité entre $GA_{\mathcal{T}}$ et $GA_{\mathcal{N}}$ en fonction de l'amplitude et de la phase. Une forte similarité apparaît en rouge et une faible similarité apparaît en bleu ; (b) Moyenne sur les échelles normalisée (entre 0 et 1) de D.

à un stimulus dans une fenêtre temporelle prédéfinie, de taille t^{up}, peuvent être extraites comme suit :

$$r_{i,c}(t) = x_c(s_i + t), \quad t \in \{1, ..., t^{up}\}, \tag{A.2}$$

où s_i correspond au début du stimulus i. La moyenne A de chaque classe $(\mathcal{T}, \mathcal{N})$ et chaque canal e peut être calculée par :

$$A_c^{\mathcal{T}}(t) = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} r_{i,c}(t) \qquad A_c^{\mathcal{N}}(t) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} r_{i,c}(t) \tag{A.3}$$

où l'opérateur $|\cdot|$ correspond au cardinal de l'ensemble.

Suite au calcul des As, nous calculons leurs transformées en ondelettes continues $W_\psi^{A_c^{\mathcal{T}}}$ et $W_\psi^{A_c^{\mathcal{N}}}$ pour obtenir le produit D avec l'équation 3.30.

Le résultat obtenu pour D est illustré par la figure A.2(a), où les couleurs froides indiquent une différence maximale des signaux des deux classes et par conséquent la localisation du P300 dans le temps. La variation dans le temps de l'écart-type normalisé de D est illustré dans la figure A.2(b)). En considérant le seuil τ_w, $0 \leq \tau_w \leq 1$, la nouvelle fenêtre temporelle, pour le canal actuel, est sélectionnée entre les bornes t^{low} et t^{up}. Ces bornes correspondent aux instants pour lesquels le seuil est dépassé et elles sont obtenues à l'aide d'une recherche séquentielle en partant des extrémités droite et gauche de la fenêtre. Finalement, nous obtenons pour chaque canal une sélection automatique de la fenêtre temporelle dans laquelle effectuer la discrimination. Nous appelons cet algorithme *Semblance-based ERP*[25] *Window Selection by Channel (SEWS-1) (voir l'algorithme A.2 pour une description complète du processus de la sélection de la fenêtre d'étude.*

25. Event-Related Potential *i.e.* Potentiel Evoqué

Algorithm A.2 Algorithme de sélection automatique de la fenêtre d'étude spécifique à chaque canal.

Données : Soit $\mathbf{X}$ la matrice des signaux EEG, avec C canaux et T échantillons
Résultats : Les bornes t_c^{low} et t_c^{up} pour chacun des C canaux.

Choisir le seuil de corrélation τ_d, $0 \leq \tau_d \leq 1$
pour $c = 1$ à C **faire**
    Calculer le signal moyen de la classe cible $A_c^{\mathcal{T}}$ *i.e.* des réponses contenant un P300.
    Calculer le signal moyen de la classe non cible $A_c^{\mathcal{N}}$ *i.e.* des réponses ne contenant pas un P300.
    Calculer les transformées en ondelettes continues $W_\psi^{A_c^{\mathcal{T}}}$ et $W_\psi^{A_c^{\mathcal{N}}}$ avec l'équation 3.15.
    Calculer la similarité avec l'équation $S = cos^n(\theta)$.
    Calculer le produit scalaire D avec l'équation $D = cos^n(\theta)|W_\Psi^x W_\Psi^{y*}|$.
    Calculer l'écart-type $\sigma(D)$ de D et le normaliser entre 0 et 1.
    La borne t_c^{low} correspond à l'instant t à partir du début de la fenêtre pour lequel $\sigma(D(t)) > \tau_w$
    La borne t_c^{up} correspond à l'instant t à partir de la fin de la fenêtre pour lequel $\sigma(D(t)) > \tau_w$
fin de pour

Une variation de cette méthode consiste à prendre la même fenêtre temporelle pour tous les canaux. Dans ce cas, la seule différence dans le processus ci-dessus est de calculer le signal moyen de chaque classe tous canaux confondus avec l'équation suivante (voir algorithme A.3).

$$GA^{\mathcal{T}}(t) = \frac{1}{|C|}\sum_{c=1}^{C} A_c^{\mathcal{T}}(t) \qquad GA^{\mathcal{N}}(t) = \frac{1}{|C|}\sum_{c=1}^{C} A_c^{\mathcal{N}}(t) \tag{A.4}$$

Algorithm A.3 Algorithme de sélection automatique de la fenêtre d'étude commune à tous les canaux.

Données : Soit $\mathbf{X}$ la matrice des signaux EEG, avec C canaux et T échantillons
Résultats : Les bornes t^{low} et t^{up} communes à tous les canaux.

Choisir le seuil de corrélation τ_d, $0 \leq \tau_d \leq 1$
Calculer le signal moyen de la classe cible $GA^{\mathcal{T}}$ *i.e.* des réponses contenant un P300 tous canaux confondus.
Calculer le signal moyen de la classe non cible $GA^{\mathcal{N}}$ *i.e.* des réponses ne contenant pas un P300 tous canaux confondus.
Calculer les transformées en ondelettes continues $W_\psi^{GA^{\mathcal{T}}}$ et $W_\psi^{GA^{\mathcal{N}}}$ avec l'équation 3.15.
Calculer la similarité avec l'équation $S = cos^n(\theta)$.
Calculer le produit scalaire D avec l'équation $D = cos^n(\theta)|W_\Psi^x W_\Psi^{y*}|$.
Calculer l'écart-type $\sigma(D)$ de D et le normaliser entre 0 et 1.
La borne t^{low} correspond à l'instant t à partir du début de la fenêtre pour lequel $\sigma(D(t)) > \tau_w$
La borne t^{up} correspond à l'instant t à partir de la fin de la fenêtre pour lequel $\sigma(D(t)) > \tau_w$

Selon [Kolev et al. 1997], les différences les plus significatives entre les réponses "cible" et "non cibles", dans le domaine fréquentiel, se trouvent dans les rythmes Delta et Theta du cerveau.

Par conséquent, les échelles utilisées pour analyser les signaux moyennés ont été sélectionnées en fonction de ces rythmes.

Les paramètres de nos algorithmes sont comparables à ceux utilisés dans les méthodes classiques de débruitage à base d'ondelettes utilisant un seuil. Cela signifie que des améliorations similaires peuvent être appliquées à nos algorithmes comme par exemple une sélection automatique du seuil. Ces méthodes sont indépendantes de l'étape de débruitage introduite à la section A.1.1. Cependant, leur combinaison devrait s'avérer plus efficace que leur utilisation séparée.

A.2 Expérimentations

Comme nous l'avons présenté dans la section A.1, les contributions principales de cette thèse concernent un débruitage à partir de la théorie des ondelettes basé sur la similarité des canaux via la MRL, et la sélection automatique de la fenêtre d'analyse à partir de la similarité.

A.2.1 La base de données

Une base de données correspondant au système d'écriture de Farwell et Donchin, *i.e.* basé sur la détection du P300 [Farwell and Donchin 1988], a été utilisée pour évaluer nos contributions. Cette base de données a été enregistrée par le laboratoire de neuroimagerie de l'université autonome métropolitaine de Mexico (UAM) [Ledesma-Ramirez et al. 2010] avec le logiciel BCI2000 [Schalk et al. 2004]. La base de données contient 22 primo-utilisateurs sains, identifiés par les acronymes ACS, APM, ASG, ASR, DCM, ELC, FSZ, GCE, ICE, JLD, JLP, JMR, JSC, JST, LAC, LAG, PGA, WFG et XCL, qui seront désignés comme les sujets s_1 à s_{22} par simplicité. 10 electrodes ont été enregistrées (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz), lesquelles présentent le plus d'intérêt pour une discrimination des P300 [Krusienski et al. 2008]. Les signaux ont été enregistrés avec une fréquence d'acquisition de 256 Hz avec l'amplificateur gUSBamp de la société g.tec. La référence a été prise à l'oreille droite et la masse sur la mastoïde droite. Un filtre de Chebyshev passe-bande 0.1-60 Hz d'ordre 8 et un filtre coupe-bande à 60 Hz ont été utilisés. Le stimulus a été flashé pendant 62,5 ms et l'intervalle inter-stimuli était de 125 ms. La base de données est disponible gratuitement en ligne avec une description complète[26]. Le système d'écriture de Farwell et Donchin utilise une matrice 6x6 de lettres et de chiffres. Les colonnes et les lignes sont mises en surbrillance dans un ordre aléatoire. La bonne reconnaissance d'une lettre (ou d'un chiffre) nécessite de détecter le P300 généré après la mise en surbrillance de la ligne et de la colonne qui contiennent le symbole souhaité. C'est donc cette double reconnaissance qui est mesurée dans les expérimentations suivantes.

A.2.2 Débruitage basé sur la similarité des canaux

Nous avons évalué notre algorithme de débruitage par rapport aux techniques standard à seuil basées sur les ondelettes.

Les résultats de notre algorithme de débruitage, comparés à ceux des méthodes standard de débruitage à base d'ondelettes, sont présentés dans le tableau A.1. Le tableau montre que notre algorithme est, en général, capable de supprimer du bruit plus efficacement que les méthodes standard. Bien que les résultats de notre algorithme ne sont pas largement meilleurs, ils le sont de manière significative. Rappelons que l'intérêt principal de notre algorithme est qu'il prend en considération la similarité entre tous les canaux.

26. http ://akimpech.izt.uam.mx/p300db/doku.php ?id=deposits.en

Method	μ	σ	min	max
SURE	54.80	13.90	33.02	78.57
Minimax	55.00	13.93	32.70	79.05
Universal	55.07	13.92	33.02	79.05
METS	55.20	13.69	33.65	79.05

TABLE A.1 – Pourcentage moyen de reconnaissance des lettres sans moyennage des réponses pour les 22 sujets.

A.2.2.1 Efficacité de la sélection automatique de la fenêtre sans débruitage

Nous avons présenté dans la section A.1.2 qu'une fenêtre temporelle devait être spécifiée pour étudier la présence ou non d'un P300. Cette fenêtre correspond généralement à l'intervalle [0,1] s, quel que soit le temps de réaction de l'utilisateur au stimulus. Nous avons évalué les effets de notre algorithme de sélection automatique des bornes de la fenêtre.

L'algorithme pour la sélection d'une fenêtre spécifique à chaque canal est noté $SEWS-1$ et l'algorithme pour la sélection d'une fenêtre commune à tous les canaux est noté $SEWS-2$.

A.2.2.2 Fenêtrage multicanaux ou par canal

Les techniques de sélection de fenêtres temporelles ont été testées en appliquant uniquement une étape de pré-traitement correspondant à un filtrage [0.1-20] Hz. Les résultats pour différents seuils τ pour les deux algorithmes sont présentés dans les tableaux A.2 et A.3.

Prétraitement	τ_w	μ	σ	min	max	$\mu(t^{low})$(ms)	$\mu(t^{up})$(ms)
[0.1-20] Hz Filter	-	53.60	14.14	28.25	79.52	0	1000
SEWS-1	0.1	53.73	14.70	28.57	80.95	17	951
SEWS-1	0.2	54.10	14.16	32.70	80.48	34	900
SEWS-1	0.3	53.98	13.71	32.06	80.48	56	844
SEWS-1	0.4	54.10	13.28	36.83	80.00	90	781
SEWS-1	0.5	53.16	14.12	32.86	79.05	128	724

TABLE A.2 – Taux de reconnaissance des lettres obtenus, sans moyennage et sans débruitage préalable, par l'algorithme SEWS-1 pour les 22 sujets de la base de l'UAM. Des statistiques descriptives (moyenne, écart-type, minimum, maximum) des taux de reconnaissance des lettres sont indiquées. De plus, les bornes (t^{low}) et (t^{up}) de la fenêtre temporelle pour différents seuils sont aussi présentées. Les meilleurs résultats sont mis en surbrillance.

L'algorithme SEWS-1 a été testé en augmentant la valeur du seuil jusqu'à ce que les performances soient inférieures à celles d'un simple filtrage. Cela se produit pour $\tau_w = 0.5$ où l'algorithme supprime en moyenne 40% de la taille de la fenêtre temporelle. Pour tous les autres seuils, la fenêtre temporelle originale de une seconde peut être réduite sans perte de performance, ce qui montre que cette partie du signal complique la reconnaissance et qu'il est préférable qu'elle soit supprimée. De plus, les taux de reconnaissance minimal et maximal parmi les sujets sont améliorés, ce qui montre que la technique est assez souple pour s'adapter aux différents types de sujets. Le compromis entre la réduction de la fenêtre temporelle et la conservation de l'information pertinente est contrôlée par le paramètre de seuil τ_w, mais la sélection du seuil n'est pas une

tâche facile. Nous pouvons observer dans le tableau A.2 que $\tau_w = 0.1$ améliore déjà les résultats en supprimant 6% du signal en moyenne, tandis que $\tau_w = 0.2$ et $\tau_w = 0.4$ offrent un meilleur taux de reconnaissance. Lorsque les seuils augmentent, la valeur minimale et la variance sont réduites, au prix d'une performance moins bonne pour les sujets ayant les meilleurs résultats. En fait, ici, les meilleurs résultats correspondent à une élimination en moyenne de 16% et 30% du signal, soit dans le dernier cas à une sélection assez forte.

Prétraitement	τ_w	μ	σ	**min**	**max**	$\mu(t^{low})$(ms)	$\mu(t^{up})$(ms)
[0.1-20] Hz Filter	-	53.60	14.14	28.25	79.52	0	1000
SEWS-2	0.1	54.08	14.60	29.52	80.48	14	960
SEWS-2	0.2	54.29	14.29	32.38	79.52	28	917
SEWS-2	0.3	54.41	14.03	31.43	78.10	49	869
SEWS-2	0.4	52.59	15.21	32.86	79.05	85	827

TABLE A.3 – Taux de reconnaissance des lettres obtenus, sans moyennage et sans débruitage préalable, par l'algorithme SEWS-2 pour les 22 sujets de la base de l'UAM. Des statistiques descriptives (moyenne, écart-type, minimum, maximum) des taux de reconnaissance des lettres sont indiquées. De plus, les bornes (t^{low}) et (t^{up}) de la fenêtre temporelle pour chaque seuil sont aussi présentées. Les meilleurs résultats sont mis en surbrillance.

Les résultats de l'algorithme SEWS-2 sont résumés dans le tableau A.3. Ils montrent que le comportement de l'algorithme de la sélection fenêtre temporelle commune à tous les canaux est différent de SEWS-1. Ici, l'algorithme est plus conservateur (en moyenne). Les fenêtres sont plus larges que SEWS-1 parce que l'information utile sur chaque canal est considérée pour sélectionner la fenêtre. Cela signifie que les instants qui augmentent l'écart-type pour un canal augmentent aussi l'écart-type pour de la moyenne des canaux et seront conservés.

Pour la même raison, les performances augmentent quand on utilise des seuils plus forts jusqu'à ce qu'ils suppriment trop d'information et fassent tomber les performances en dessous de celles d'un filtrage *i.e.* quand $\tau_w = 0.4$ et que 26% du signal sont supprimés (en moyenne). Ce pourcentage est (en moyenne) inférieur à celui de SEWS-1 pour un même seuil. SEWS-2 conserve plus informations. Cependant, les résultats de SEWS-2 sont meilleurs que ceux de SEWS-1 car généralement tous les canaux transportent des informations similaires, ce qui signifie que les fenêtres sélectionnées devraient être similaires. SEWS-2 compense naturellement la présence de fenêtres trop étroites ou larges, *i.e.* que la plupart des sujets peuvent bénéficier d'une sélection de la fenêtre plus robuste. Par contre, si le seuil est trop élevé, la plupart des sujets peuvent perdre des informations utiles.

En résumé, cette expérience montre que la réduction de la fenêtre temporelle par rapport à une fenêtre originale de [0,1] s est une technique efficace pour localiser automatiquement le PE dans le temps. Cette sélection semble particulièrement utile pour le sujet avec de faibles performances.

Nous rappelons que ces résultats ont été obtenus sans utiliser METS, montrant que SEWS peut travailler avec des signaux bruités. Cependant, SEWS devrait également fonctionner correctement avec les signaux débruités. La section suivante approfondit l'analyse de la sélection de la fenêtre temporelle lorsque la sélection est combinée avec notre algorithme de débruitage METS.

A.2.2.3 Efficacité de la sélection automatique de la fenêtre avec débruitage

Dans cette expérience, nous proposons d'appliquer nos deux principales contributions de la section A.1, *i.e.* l'algorithme de débruitage METS et les techniques SEWS-1 et SEWS-2 de sélection de la fenêtre temporelle. Notons METS & SEWS-1 l'algorithme qui débruite le signal et sélectionne une fenêtre spécifique à chaque canal, et METS & SEWS-2 l'algorithme qui débruite et sélectionne une même fenêtre temporelle pour tous les canaux.

Prétraitement	μ	σ	**min**	**max**	seuil de signification
[0.1-20] Hz Filter	53.60	14.14	28.25	79.52	-
METS	55.20	13.19	33.65	79.05	0.0017
METS & SEWS-1	56.00	13.64	35.56	80	0.0004
METS & SEWS-2	55.91	14.13	34.44	78.97	0.0005

TABLE A.4 – Pourcentage de reconnaissance des lettres par les algorithmes SEWS-1 et SEWS-2 sans moyennage et avec un débruitage assuré par l'algorithme METS, pour les 22 sujets de la base de l'UAM. Des statistiques descriptives (moyenne, écart-type, minimum, maximum) des taux de reconnaissance des lettres sont indiquées aux côtés du seuil de signification d'un test de Student apparié par rapport à un filtre [0.1-20] Hz. Les meilleurs résultats sont mis en surbrillance.

Les deux algorithmes SEWS améliorent la moyenne, le minimum et le maximum des résultats comparativement à l'utilisation seule de l'algorithme de débruitage METS, ce qui implique que l'application de notre sélection temporelle est efficace également pour les signaux débruitées. En comparant ces résultats avec ceux d'un filtrage, la combinaison de nos deux algorithmes améliore la détection des PE sans moyennage de manière significative, SEWS-1 obtenant des résultats légèrement meilleurs.

A.2.3 Comparaison avec l'algorithme xDAWN

L'algorithme XDAWN [Rivet et al. 2009] est une méthode de débruitage non basée sur les ondelettes efficace pour améliorer la détection du P300. Nous avons procédé à une comparaison de ses performances sujet par sujet avec notre propre algorithme de débruitage.

Au regard de la figure A.3, nous constatons que notre méthode présente un meilleur pourcentage de reconnaissance des lettres sans moyennage pour 16 des 22 sujets. Les performances sont très similaires quand le pourcentage de reconnaissance est élevé, ce que nous interprétons comme le fait que les signaux sont relativement propres et que la technique de débruitage n'a pas d'incidence majeure. Par contre, nos méthodes obtiennent de meilleures performances dans les cas où le taux de reconnaissance du xDAWN est plus faible *i.e.* en présence d'artefacts ou de fort bruit.

En conséquence, nous concluons que nos algorithmes améliorent les performances du xDAWN dans le cas d'une détection sans moyennage en présence d'artefacts et de bruit. Cependant, nos algorithmes sont conçus pour analyser des réponses sans moyennage et xDAWN devrait avoir de meilleures performances en utilisant des réponses moyennées. Pour cette raison, nous comparons le comportement des différents algorithmes en utilisant un nombre différent de répétitions. La figure A.4 présente les résultats obtenus pour une à cinq répétitions. Il est possible de voir que METS & SEWS-2 fonctionnent mieux que METS & SEWS-1 si les réponses sont moyennées. Comme prévu, METS & SEWS sont plus performants que xDAWN lorsque le nombre de réponses moyennées est faible (un, deux ou trois). A partir de quatre répétitions, les résultats sont très

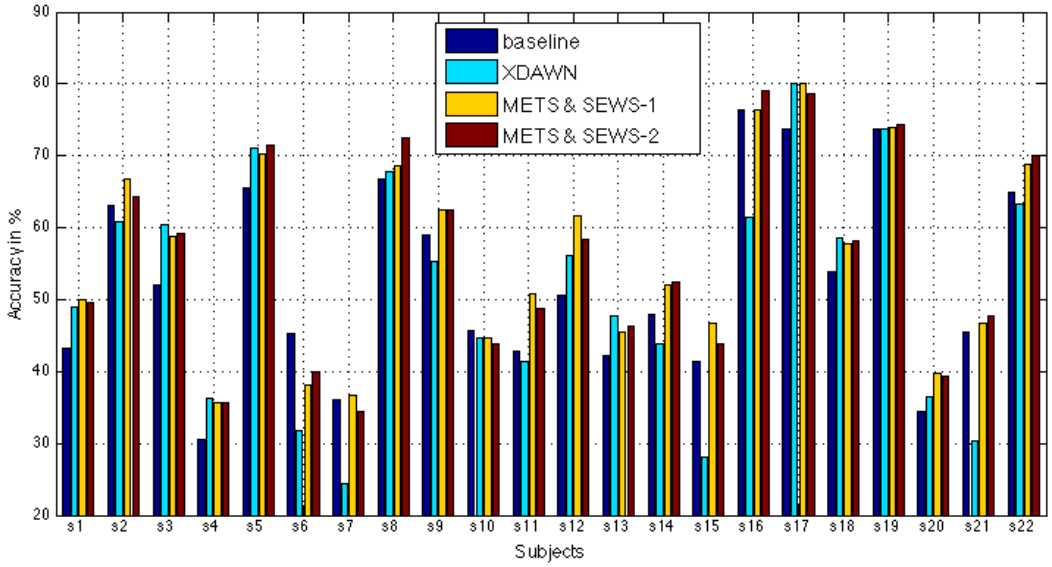


FIGURE A.3 – Comparaison du pourcentage de lettres reconnues sujet par sujet par les algorithmes METS & SEWS-1, METS & SEWS-2 et xDAWN.

similaires et pour cinq répétitions xDAWN surpasse nos algorithmes.

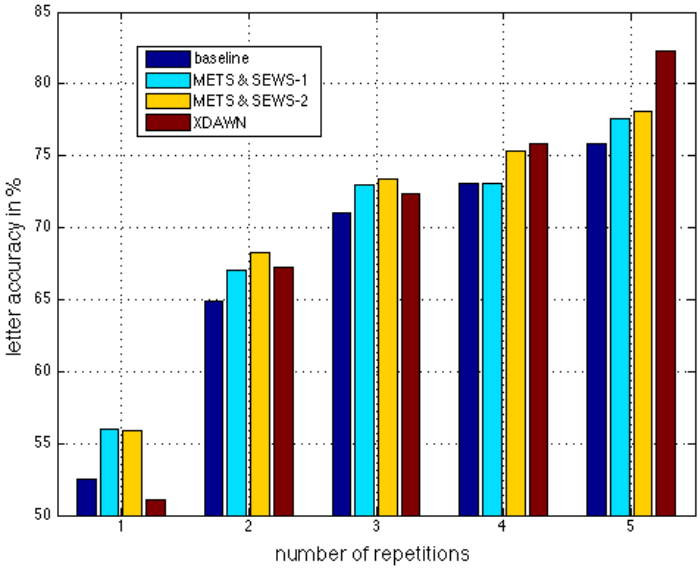


FIGURE A.4 – Comparaison du taux de reconnaissance de lettres obtenu à partir de réponses moyennées pour les algorithmes xDAWN, METS & SEWS et un filtrage passe-bande [1-20] Hz.

A.2.4 Base de Données EPFL

Nous avons testé nos algorithmes sur une autre base de données pour confirmer que notre approche est générique et adaptée aux patients. Cette base de données enregistrée par l'*Ecole Polytechnique Fédérale de Lausanne*, utilise six images (télévision, téléphone, lampe, porte, fenêtre, radio) pour évoquer une réponse P300 [Hoffmann et al. 2008]. Ces images sont organisées dans une matrice 3x2 et sont flashées individuellement et de manière aléatoire pendant 100 ms toutes les 400 ms. 32 électrodes ont été enregistrées (FP1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, FP2, Fz, Cz) à une fréquence d'échantillonnage de 2048 Hz.

La base de données contient deux groupes de sujets. Le premier groupe contient quatre patients, chacun avec un handicap différent : infirmité motrice cérébrale, sclérose en plaques, sclérose latérale amyotrophique avec un traumatisme cérébrale, et lésion de la moelle épinière. Le deuxième groupe est composé de quatre sujets valides [27].

Nous avons utilisé cette base de données comme base de test, ce qui signifie que les paramètres de nos algorithmes et la méthodologie sont les mêmes que ceux retenus après évaluation pour la base de données de l'UAM (type de classificateur, mère ondelettes, échelles et seuils). Nous avons utilisé la même technique d'extraction de caractéristiques et la configuration de 8 électrodes (Fz, Cz, Pz, P3, P4, P7, P8 et Oz) présentées dans [Hoffmann et al. 2008]. La seule différence est que nous avons utilisé un sous-échantillonage de 2048 Hz à 256 Hz au lieu de 32 Hz afin d'avoir le même échantillonnage que celui de la base de données de l'UAM.

Sujet	Filtre [0,20] Hz	METS	METS & SEWS-1	METS & SEWS-2
s1	44.53	40.88	42.34	45.26
s2	41.41	49.22	49.22	50.00
s3	58.33	62.88	64.39	64.39
s4	49.21	48.41	52.38	50.00
s5	44.62	46.15	46.92	43.08
s6	48.18	54.01	60.58	55.47
s7	72.93	65.41	70.68	72.93
s8	45.31	53.12	55.47	62.50
μ	50.57	52.51	55.25	55.45
σ	10.36	8.28	9.49	10.36
min	41.41	40.88	42.34	43.08
max	72.93	65.41	70.68	72.93

TABLE A.5 – Taux de reconnaissance des lettres obtenus, sans moyennage, par les algorithmes METS, METS & SEWS-1 et METS & SEWS-2 pour les 8 sujets de la base de l'EPFL. Des statistiques descriptives (moyenne, écart-type, minimum, maximum) des taux de reconnaissance des lettres sont indiquées. Les meilleurs résultats sont mis en surbrillance.

Les résultats de nos algorithmes sur la base de données de l'EPFL sont disponibles dans le tableau A.5 et montrent qu'ils améliorent les résultats. Pour cette base de données, la signification statistique est difficile à évaluer en raison du petit nombre de sujets [28], mais METS améliore déjà en moyenne la détection des P300 et la combinaison de METS et SEWS offre une nette

27. Pour plus de détails sur des sujets voir `http://mmspg.epfl.ch/BCI_datasets`.
28. Un test de Student apparié indique que METS & SEWS-1 est significativement meilleur à 5% par rapport au filtre, et METS & SEWS-2 au 6%.

amélioration pour presque tous les sujets. En particulier, SEWS-2 est plus efficace pour les sujets handicapés (*i.e.* les quatre premiers sujets) et il obtient une impressionnante amélioration de 45,31% à 62,50% pour un sujet normal. En outre, ces résultats confirment que nos techniques sont particulièrement utiles pour sujets dont les réponses sont plus difficilement exploitables.

A.2.5 Conclusion

Les interfaces cerveau-ordinateur basées sur la détection des potentiels évoqués cognitifs permettent un grand nombre de commandes sans apprentissage spécifique de l'utilisateur. Mais cette détection nécessite généralement de répéter la présentation du stimulus déclencheur du potentiel évoqué pour, par moyennage des réponses, mieux faire apparaître celui-ci. Ces répétitions réduisent le nombre de commandes possibles par minute et lassent l'utilisateur. Dans cette thèse, de nouvelles techniques basées sur la théorie des ondelettes sont proposées pour débruiter et isoler les potentiels évoqués sans moyennage. Une méthode présentée dans ce manuscrit permet de débruiter le signal dans le domaine des ondelettes en considérant simultanément, et non de manière indépendante, l'information de phase fournie par l'ensemble des canaux enregistrés. Nous avons également étendu cette approche à l'étude des phases et des amplitudes pour sélectionner automatiquement la fenêtre du signal à observer pour chaque utilisateur. Les algorithmes présentés dans cette thèse ont été évalués sur deux bases de données contenant respectivement 22 et 8 sujets. En outre, nos techniques d'amélioration de la détection des potentiels évoqués sans moyennage ont été appliquées à la deuxième base de données qui inclut des patients avec les paramêtres sélectionnés pour la première base montrant ainsi que nos algorithmes sont génériques et utiles pour des bases de données différentes. Les résultats montrent que les modèles présentés dans cette thèse améliorent significativement la détection des P300 sans moyennage par rapport aux méthodes conventionnelles en particulier lorsque ces dernières donnent de faibles résultats.