New signal processing and machine learning methods for EEG data analysis of patients with Alzheimer's disease

PhD Thesis by

Esteve Gallego Jutglà

Directors:
Dr. Jordi Solé Casals
Dr. François Benoît Vialatte

Doctoral programme:
Digital and Information Technology

UVIC
UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

Escola de Doctorat de la Universitat de Vic - Universitat Central de Catalunya, December 2014

"A smooth sea never made a skillful sailor"

English proverb

Acknowledgements

Durant el procés de realitzar aquesta tesi i en el d'aprenentatge anterior hi ha hagut molta gent que m'ha donat suport. M'agradaria poder donar-los les gràcies a tots.

Primerament voldria agrair la seva ajuda al Dr. Jordi Solé, principal "culpable" de totes les aventures que he anat vivint. Tot va començar amb un viatge a Grenoble per cinc mesos que es van acabar convertint en deu, va continuar amb un treball de final de màster i ha acabat amb visites a terres tant llunyanes com el Japó i Singapur, o d'altres de més pròximes com França i Alemanya, tot treballant en el doctorat.

En segon lloc vull agrair la seva ajuda i direcció al Dr. François Vialatte, que sempre ha tingut les portes obertes del seu laboratori de París en les diferents visites que li he fet. A part, tot i la distància, sempre ha tingut temps per resoldre'm tots els dubtes que se m'han plantejat.

En un àmbit més proper a la UVic-UCC, vull agrair a tots els professors del departament el temps viscut, i també agrair les diferents oportunitats que m'han ofert per donar classe o bé per participar en l'altre gran projecte que he portat a terme, els Garrins Metàl·lics. Vull donar les gràcies, també, als companys per les bones estones viscudes. Agrair també a la UVic-UCC tot el suport rebut en forma d'ajudes, des de l'ajut predoctoral que ha permès que pugués tirar endavant aquest projecte, fins a les diferents ajudes de mobilitat que he obtingut.

Aquest projecte de tesi no seria el mateix sense les diferents estades que he realitzat, és per això que vull agrair també el suport rebut en aquestes estades i vull donar les gràcies al Dr. Justin Dauwels, Dr. Andrzej Cichocki i Dr. Elmar Lang per la bona acollida als seus centres. També agrair a tota la gent que he conegut en aquests centres que han fet que el temps fora de casa fos una experiència inoblidable.

En un àmbit més personal voldria agrair el suport rebut dels meus amics, els quals en determinades ocasions no he pogut dedicar tot el temps que m'hagués agradat, però que tot i el temps que he estat fora, a la tornada sempre m'han rebut amb els braços oberts i m'han donat suport en els moments claus.

I la família, impossible no citar-la. Gràcies als meus pares, que des del principi s'han esforçat per ensenyar-me els valors que m'han portat fins aquí, que per descomptat, no sempre ha sigut fàcil. Gràcies al meu germà Edu i la meva cunyada Eva per donar-me suport, no sols anímic, en les diferents etapes d'aquest doctorat. Agrair també a la família de la Torre Blanca i a la família Gallego tots els bons moments viscuts, en especial a la iaia Maria, puntal de la casa i la família. No em podia oblidar de la meva estimada Giorgia, que s'ha convertit en una part molt important de la meva vida.

Així doncs, moltes gràcies a tothom per ser-hi.

Abstract

Neurodegenerative diseases are a group of disorders that affect the brain. These diseases are related with changes in the brain that lead to loss of brain structure or loss of neurons, including the dead of some neurons. Alzheimer's disease (AD) is one of the most well-known neurodegenerative diseases. Nowadays there is no cure for this disease. However, there are some medicaments that may delay the symptoms if they are used during the first stages of the disease, otherwise they have no effect. Therefore early diagnose is presented as a key factor.

This PhD thesis works different aspects related with neuroscience, in order to develop new methods for the early diagnose of AD. Different aspects have been investigated, such as signal preprocessing, feature extraction, feature selection and its classification.

The present document starts with an introduction of the general framework of the thesis. This introduction is detailed before the description of the preformed work. In this introduction, first Electroencephalography (EEG) is explained. Then, AD is presented by detailing all the available information about this disease. Finally, the main changes that AD causes on EEG data are described.

Preprocessing applied to the data is presented after the introduction. First, an explanation about the main interference signals known as artifacts is done. Once artifacts are described, then a new signal processing method is defined in order to eliminate them. The proposed method is applied to simulated EEG signals which contain artifacts and different levels of SNR. Once the cleaning method is applied, cleaned data is compared with simulated clean data in order to evaluate the improvement of the data. Initially this preprocessing is only applied to simulated data.

After preprocessing, feature extraction is explained in two different chapters. First a new approach for EEG data modeling is presented. In this approach different measures are used to parametrize the changes that AD causes to EEG data. However, this approach presents an important change. Until now EEG analysis have been performed on pre-establish frequency bands. The described approach analyzes all existing frequency ranges between 1 and 30 Hz, distinguishing better AD patients and healthy subjects. During this analysis, it is discovered that AD patients present a change of synchrony that has been lightly studied in the literature. This change points out that there is an increase of synchrony in low frequencies for AD patients, whereas it has been always presented the decrease of synchrony existing in the higher frequencies. To this discovery is devoted the second chapter of feature extraction, where the two changes of synchrony are used together helping to differentiate AD patients and healthy subjects.

Chapter devoted to feature selection presents a new method to select the most suitable parameters computed on the data, and then, use them together in order to improve classifi-

cation. A new method is proposed which selects different characteristics with the objective of avoiding redundant information of the data. Detailed methodology for feature selection presents some overfitting when used. Therefore, once this methodology is described, next chapter is devoted to use some cross-validation techniques in order to adjust obtained results.

Last chapter of the document is devoted to jointly use all previous methods presented. These methods are applied to perform early diagnose of AD. This diagnose is performed by doing the preprocessing, the feature extraction and its classification.

Resum

Les malalties neurodegeneratives són un conjunt de malalties que afecten al cervell. Aquestes malalties estan relacionades amb la pèrdua progressiva de l'estructura o la funció de les neurones, incloent-hi la mort d'aquestes. La malaltia de l'Alzheimer és una de les malalties neurodegeneratives més comunes. Actualment, no es coneix cap cura per a l'Alzheimer, però es creu que hi ha un grup de medicaments que el que fan és retardar-ne els principals símptomes. Aquests s'han de prendre en les primeres fases de la malaltia ja que sinó no tenen efecte. Per tant, el diagnòstic precoç de la malaltia de l'Alzheimer és un factor clau.

En aquesta tesis doctoral s'han estudiat diferents aspectes relacionats amb la neurociència per investigar diferents eines que permetin realitzar un diagnòstic precoç de la malaltia en qüestió. Per fer-ho, s'han treballat diferents aspectes com el preprocessament de dades, l'extracció de característiques, la selecció de característiques i la seva posterior classificació.

En el present document, abans de descriure el treball realitzat, es presenta una introducció on s'explica els principals conceptes que defineixen el marc de la tesi. Primerament es fa una descripció de la tècnica d'enregistrament coneguda com a electroencefalograma (EEG). Posteriorment es presenta la malaltia de l'Alzheimer explicant la informació que es té d'aquesta fins a l'actualitat. Finalment, es presenten els principals canvis que la malaltia de l'Alzheimer produeix als enregistraments EEG dels pacients.

El preprocessament es presenta en el capítol següent a la introducció. En primer lloc es presenten els principals senyals no desitjats que afecten als enregistraments EEG. Aquests són coneguts com artefactes i són causats principalment pels músculs de la cara i del crani. Un cop s'ha presentat els artefactes es proposa un nou mètode per eliminar aquests senyals. El mètode de neteja proposat s'aplica a dades EEG simulades que contenen artefactes i diferents nivells de soroll, i un cop s'han aplicat el procediment de neteja es comparen els resultats amb dades EEG que no tenen artefactes. Amb aquest mètode es comprova quina millora presenten els senyals netejats. Inicialment es realitza només l'estudi sobre dades simulades.

Un cop s'ha treballat en el preprocessament es procedeix a treballar en l'extracció de característiques. A aquest tema se li dediquen tres capítols diferents. Per començar es presenta una nova aproximació per modelitzar les dades EEG utilitzant diferents mesures que modelen els canvis que presenten els pacients amb la malaltia de l'Alzheimer. Fins al moment els anàlisis EEG s'han limitat a estudiar l'activitat existent en diferents bandes de freqüències preestablertes. En aquesta primera aproximació, en comptes d'estudiar l'activitat EEG en aquestes bandes, s'estudia en totes les bandes existents entre 1 i 30 Hz, aconseguint millors valors en la classificació de subjectes sans i pacients amb la malaltia de l'Alzheimer. Durant aquest anàlisis es descobreix que els pacients amb aquesta malaltia

presenten un canvi en les dades que ha sigut poc estudiat en la literatura. Aquest canvi és que hi ha un augment de sincronia en les baixes freqüències quan sempre s'ha presentat que hi ha un decrement en les altes freqüències. A aquest descobriment se li dedica el segon capítol de l'extracció de característiques, en aquest, els dos canvis de sincronia són utilitzats conjuntament, la qual cosa ajuda a distingir millor entre pacients i subjectes sans.

En el capítol dedicat a la selecció de característiques es presenta un nou mètode per seleccionar els millors paràmetres que s'han extret de les dades i fer-los servir conjuntament per millorar la classificació de subjectes sans i pacients amb Alzheimer. Aquest mètode selecciona les característiques amb l'objectiu que hi hagi la mínima informació redundant en les característiques seleccionades. La metodologia presentada per la selecció de característiques presenta un cert nivell de sobreentrenament del sistema, és per això que, un cop s'ha presentat aquesta, es dedica un capítol sencer a utilitzar diferents eines de validació creuada per acabar d'ajustar els resultats obtinguts.

Finalment, un cop s'han presentat els diferents passos que es segueixen en la neurociència, s'apliquen tots conjuntament per realitzar el diagnosis d'una base de dades que conté subjectes sans i pacients amb la malaltia de l'Alzheimer. El que es realitza és, doncs, el diagnosi de si un pacient pateix aquesta malaltia o no, realitzant tot el procés, el preprocessament, l'extracció de característiques i la posterior classificació.

Contents

List of Figures

List of Tables

List of Abbreviations

Aβ	β-amyloid
AD	Alzheimer's disease
ANN	Artificial Neural Network
BCI	Brain Computer Interface
BSS	Blind Source Separation
CI	Communality Index
CR	Classification Rate
DFT	Discrete Fourier Transform
DTF	Directed Transfer Function
EEG	Electroencephalography
EMD	Empirical Mode Decomposition
EOG	Electrooculography
ffDTF	full frequency Directed Transfer Function
FFT	Fast Fourier Transform
fMRI	functional Magnetic Resonance Imaging
GC	Granger Coherence
IMFs	Intrinsic Mode Functions
LDA	Linear Discriminant Analysis
LOO	Leave-one-out
LORETA	Low resolution brain electromagnetic tomography
MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
mEMD	Multivariate Empirical Mode Decomposition
MMSE	Mini Mental State Exam

MVAR	Multivariate Autoregressive Model
NDA	Nonlinear dynamical analysis
OC	Omega Complexity
OFRFPS	Orthogonal Forward Regression with random probe and Frequency Pre-Selection
OFR	Orthogonal Forward Regression
PCA	Principal Component Analysis
PC	Partial Coherence
PDC	Partial Directed Coherence
PET	Positron Emission Tomography
PS	Phase Synchrony
RP	Relative Power
SE	Sensitivity
SNR	Signal to noise ratio
SPECT	Single-Photon Emission Computed Tomography
SP	Specificity
SSVEP	Steady-State Visually Evoked Potentials

Chapter 1

Introduction

Human brain contains about of 10^{10} neurons (Nunez & Srinivasan, 2006). Nonetheless, what makes unique this organ it is not its high number of cells but the facility to interact between them. It is well-known that the human brain controls the activity of the body. However, its study using neuroimaging techniques has represented a great advance for science.

Different approximations have been used to study the human brain. Indirect approximations of neuropsychology in s.XIX were based on the study of pathological human brains, aiming to understand the role of the zone that presented some dysfunction. These techniques were supplanted by non-invasive imaging during the s.XX. We could cite for instance electroencephalography (EEG), functional Magnetic Resonance Imaging (fMRI) or Positron Emission Tomography (PET). Even though these methods have improved the knowledge of the brain, many aspects are still to be explored. Nowadays the uses of these techniques have been proved useful to diagnose some diseases, such as neurodegenerative diseases.

Neurodegenerative disorders are a group of disorders that affect the brain. They are related with changes in the brain that leads to loss of brain structure or loss of neurons, including the death of some neurons. The most well-known diseases of this group are Parkinson's disease, AD and Huntington's disease.

AD is the most prevalent neurodegenerative disease. Nowadays there is no cure for this pathology. However, there are some treatments that may delay the symptoms if they are provided in the first stages of the disease. Therefore an early diagnosis of AD is a key issue for patients suffering from this disorder. Early diagnosis of AD is a difficult task because sometimes symptoms of the disease are confused with normal ageing effects. For this purpose, EEG has been presented as a useful technique that may facilitate the early diagnosis of AD.

EEG is one of the most used imaging methods to study the brain activity. Its economic price and its simplicity to be used in comparison with other methods make it a suitable choice for hospitals and research centers. EEG records the brain signals using electrodes attached to the scalp. EEG recordings of patients suffering from AD present some characteristic changes that can be used as biomarkers of the pathology.

This thesis will investigate the early diagnosis of AD using EEG.

1.1 Thesis Outline

The typical steps used in neuroscience for the diagnosis of brain pathologies using EEG are presented in Figure 1.1. We followed this approach for our investigations. First step is the collection of the physiological data using EEG amplifiers. Next step is preprocessing, this step includes the preparation of the data to be analyzed, either by selecting the optimal part of the recording to be analyzed or by applying some methods to improve the quality of the recorded data. The feature extraction step is applied after the preprocessing. At this point, data is modeled in order to extract relevant information. This relevant information is characterized using features, which describe the EEG signals. These features are used later for classification. For instance, in the feature selection step (also called dimensionality reduction), if a large number of features have been extracted, usually the most powerful features for the classification step are selected. Finally, the last step is classification which is usually done using supervised classifiers. In some cases the step of feature selection can be avoided, in this case the performance of extracted features is evaluated during the classification step.

The organization of this document follows the same organization as the above mentioned typical steps. The organization of the document can be summarized as follows:

- Chapter 2: a review of the principals concepts is presented in this chapter. First we describe EEG, the investigated recording method. Then the essential points of AD are described. At the final part of this chapter, the main scientific reports that perform early diagnosis of AD using EEG are detailed, showing the main effects that AD cause on EEG signals.
- Chapter 3: a new technique for the preprocessing step is explained in this chapter. This new technique is based in a multivariate decomposition. Therefore, the decomposition and its extension for multivariate data sets are introduced first. Then the preprocessing technique is defined and used in simulated signals.
- Chapter 4: this chapter presents the feature extraction that is performed on EEG signals to identify biomarkers of AD. This parametrization is based on the main effects that AD cause on EEG signals detailed in Chapter 2. Moreover, a new frequency subband decomposition approach is presented, exploring all possible frequency ranges between 1 and 30 Hz instead of using the traditional EEG frequency bands δ, θ, α and β.
- Chapter 5: in this chapter we characterize the specificities of our feature extraction approach. The new frequency approach used in Chapter 4 present some changes that have not been highlighted in the literature. This chapter studies these changes and introduces a ratio that aims to increase the differences between AD patients and healthy subjects.
- Chapter 6: this chapter is devoted to feature selection. With that aim, a new feature selection method is described to be used together with features defined in Chapter 4. Finally the evolution by using selected features is presented.
- Chapter 7: Described methodology on Chapter 6 shows that when more features are used, there is an important improvement of the obtained classification rate. However, due to the methodology used and the small size of the data bases, results have some particularities for each data base. This chapter describes the generalization of the results using cross-validation, which is a well-known machine learning approach.

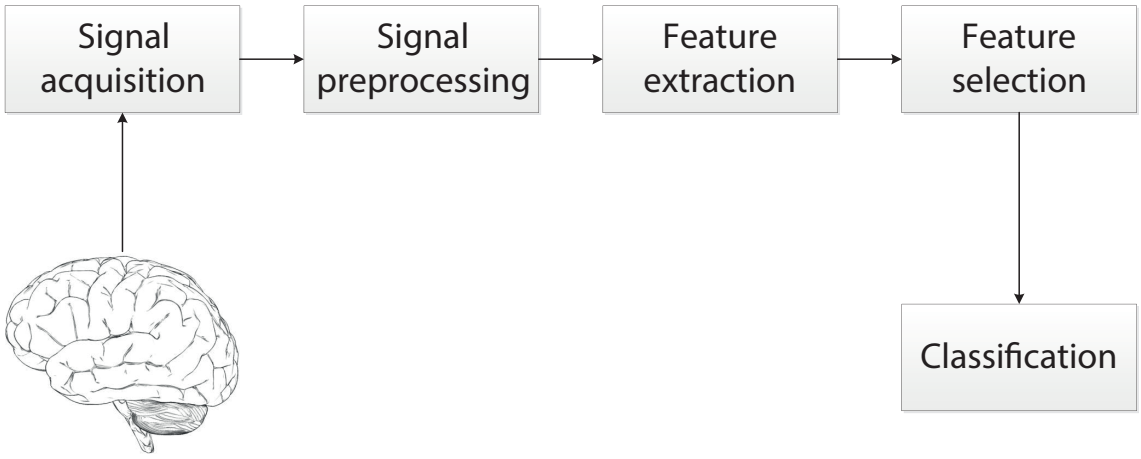


Figure 1.1: Typical steps used in neuroscience for the diagnosis of brain pathologies using EEG.

- Chapter 8: Previous chapters presented the different steps of the signal processing approach independently. Preprocessing is performed on simulated signals and the early diagnosis of AD is tested on EEG signals in which the preprocessing was done manually. This final chapter is a joint study of all the steps described in this work, by doing the preprocessing to real EEG signals and then applying the used methodology for the early diagnosis of AD.
- Chapter 9: this final chapter brings all the conclusions from the present work together, describing the new questions that have arisen from this study, and suggesting new research lines which could be addressed in future works.

1.2 Objectives of this Work

The main objective of this thesis is to present new signal processing and machine learning methods for EEG, that are used for the early diagnosis of AD. The main idea is to use existing methods from the signals processing field, and develop new ones, to analyze different EEG data sets. An important part of this work also includes the extraction of relevant features from EEG signals and its selection, which is related with the machine learning field. All main steps used in EEG data analysis are explored, from the preprocessing to classification. The main objectives of this work can be summarized as:

- Extract valuable biomarker from EEG signals to help in the early diagnosis of AD.
- Preprocessing is presented as a key issue when EEG signals are used due to the presence of undesired signals. Another objective of this work is to present a new technique for the preprocessing of EEG signals.
- Present a new frequency approach that may increase the ratio at which AD patients are recognized in comparison with healthy subjects.
- Define a feature selection technique that uses the parameters computed in the EEG signals and that facilitates the classification of patients against healthy subjects.
- Last objective is to perform a join study of all the methods presented in this work, combining all the different steps.

1.3 List of Publications

The work presented in this thesis has been partially published in (or to be publish in) different journals or conference proceedings. The list of these publications is provided below. Publications concerning the work presented in this document are presented in Section 1.3.1. Other publications presented during the time that this PhD was performed are also presented in Section 1.3.2.

1.3.1 List of publications related with this thesis

In this section the list of publications which are strictly related with the studies presented in this document.

Journals

Gallego-Jutglà, E., Solé-Casals, J., Vialatte, F. B., Elgendi, M., Cichocki, A., & Dauwels, J. 2014. A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. *Journal of Neural Engineering*, **12**(1).

Available online at: http://iopscience.iop.org/1741-2552/12/1/016018/

Gallego-Jutglà, E., Solé-Casals, J., Vialatte, F. B., Dauwels, J., & Cichocki, A. 2014. A theta-band EEG based index for early diagnosis of Alzheimer's disease. *Journal of Alzheimer's Disease*, **43**(4), 1175-1184.

Available online at: http://iospress.metapress.com/content/70525175714v9566/

Book chapters

Solé-Casals, J., Gallego-Jutglà, E., Martí-Puig, P., Travieso, C. M., & Alonso, J. B. 2013. Speech enhancement: A multivariate empirical mode decomposition approach. *Pages 192-199 of:* Drugman, T., & Dutoit, T. (eds). *Advances in Nonlinear Speech Processing*. Lecture Notes in Computer Science, vol. 7911. Springer Berlin Heidelberg.

Proceedings of peer-reviewed international conferences

Gallego-Jutglà, E., Al-Baddai, S., Al-Subari, K., Tomé, A. M., Lang, E. W., & Solé-Casals, J. 2015. Face recognition by Fast and Stable Bi-dimensional Empirical Mode decomposition. *BIOSIGNALS 2015 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing.*

Gallego-Jutglà, E., & Solé-Casals, J. 2013. Improving early diagnosis of Alzheimer's disease using synchrony measures. *Pages 167-170 of: Frontiers in Artificial Intelligence and Applications*,vol. 256.

Gallego-Jutglà, E., Lopez-de-Ipiña, K., Martí-Puig, P., & Solé-Casals, J. 2013. Empirical mode decomposition-based face recognition system. *Pages 445-450 of: BIOSIGNALS 2013 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing.*

Gallego-Jutglà, E., Rutkowski, T. M., Cichocki, A., & Solé-Casals, J. 2012. EEG signal analysis via a cleaning procedure based on multivariate empirical mode decomposition. *Pages 670-676 of: IJCCI 2012 - Proceedings of the 4th International Joint Conference on Computational Intelligence.*

Gallego-Jutglà, E., Elgendi, M., Vialatte, F. B., Solé-Casals, J., Cichocki, A., Latchoumane, C., Jeong, J., & Dauwels, J. 2012. Diagnosis of Alzheimers disease from EEG by means of synchrony measures in optimized frequency bands. *Pages 4266-4270 of: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.*

Gallego-Jutglà, E., & Solé-Casals, J. 2012. Exploring mEMD for face recognition. *Pages 498-503 of: BIOSIGNALS 2012 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing.*

Gallego-Jutglà, E., Solé-Casals, J., Rutkowski, T. M., & Cichocki, A. 2011. Application of multivariate empirical mode decomposition for cleaning eye blinks artifacts from EEG signals. *Pages 455-460 of: NCTA 2011 - Proceedings of the International Conference on Neural Computation Theory and Applications.*

1.3.2 Other publications

This section presents the list of other publications in which the author of this document has contributed during his PhD.

Book chapters

Travieso, C. M., Alonso, J. B., Orozco-Arroyave, J. R., Solé-Casals, J., & Gallego-Jutglà, E. 2013. Automatic detection of laryngeal pathologies in running speech based on the HMM transformation of the nonlinear dynamics. *Pages 136-143 of* Drugman, T., & Dutoit, T. (eds). *Advances in Nonlinear Speech Processing.* Lecture Notes in Computer Science, vol. 7911. Springer Berlin Heidelberg.

Martí-Puig, P., Solé-Casals, J., Masferrer, G., Gallego-Jutglà, E. 2013. Towards a low-complex breathing monitoring system based on acoustic signals. *Pages 128-135 of* Drugman, T., & Dutoit, T. (eds). *Advances in Nonlinear Speech Processing.* Lecture Notes in Computer Science, vol. 7911. Springer Berlin Heidelberg.

Chapter 2

Literature Review

Studying the early diagnosis of AD using EEG encompasses different fields of knowledge. Therefore, this section explores previous works in these fields dividing them into three main parts. First, the EEG recording technique is described in Section 2.1. Then in Section 2.2, AD is described, presenting the existing knowledge of this disease. Finally, the third and final section examines the most relevant studies which specifically focus on early diagnosis of AD using EEG, presenting the main changes that AD causes on EEG signals and how this can be parametrized with different measures.

2.1 Electroencephalogram

EEG was discovered in 1924 by a German scientist, Hans Berger. Since then, this recording technique has been one of the most used tools to study the brain activity and diagnose various neurological disorders. EEG has been largely used for its intrinsic simplicity and low cost compared to other recording techniques such as fMRI or PET.

EEG is the recording of the electrical fields generated in the brain. These electrical fields are generated by groups of pyramidal cells of neurons oriented perpendicularly to the surface of the head. Neurons generate ionic current flows. A group of neurons can be modeled as a microdipole when they produce synchronized electric fields (Kropotov, 2009). EEG records the electrical activity generated by different microdipoles located in the surface cortex. Approximately a group of 10^6 neurons orientated in the same direction with synchronized activity is enough to generate an electric filed which can be observable from the scalp (Nunez & Srinivasan, 2006). However, EEG is a complex combination of rhythms, recording the activity created in different parts of the brain at the same time. EEG is present from before birth (actually, non-natal brain electrical activity has also been recorded with other recording techniques) until brain death. This brain activity is related with every simple action that the human body performs, such as moving an arm or focusing the attention. The study of the brain activity has supposed an improvement of the knowledge of the brain.

Initially, studies using EEG extracted conclusions by visual interpretation and manual measurements of the EEG traces. Hence, the results were unreliable. Thanks to progresses in computerized data processing, it became possible for the EEG signal to be analyzed

digitally with parametric and nonparametric methods. Progressive developments in electrical engineering and the fascination with the human brain have attracted researchers from different scientific fields to investigate EEG recordings. Nowadays there is at least four reasons that explain why this technique is highly being used (Kropotov, 2009):

i. EEG provides a high temporal resolution which in principles cannot be achieved by other neuroimaging techniques. EEG provides a resolution of few milliseconds, whereas PET and fMRI are limited to a few seconds.

ii. The knowledge of the mechanism that generates spontaneous EEG activity has increased.

iii. EEG recording systems are cheaper than the multimillion fMRI and PET scans.

iv. New methods have emerged recently for the EEG analysis such as Blind Source Separation (BSS), or time-frequency analysis like wavelet analysis.

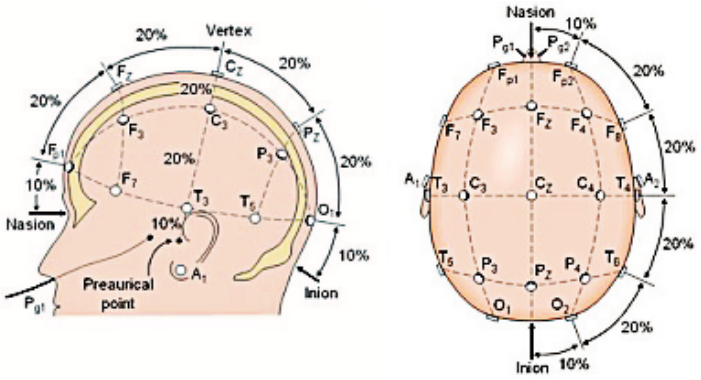
For these reasons, EEG has been presented as a helpful tool in clinical neuroscience. In particular, the low cost of an EEG recording system is the point that has extended its wide use. Consequently, several research centers and hospitals currently have their own EEG recording system. Nowadays, EEG has been used in many different applications in clinical neuroscience. Some of these applications are (Elgendi *et al.*, 2014):

- Monitoring alertness, coma and brain death (Wijdicks, 1995).
- Detecting neurodegenerative disease such as AD (Dauwels *et al.*, 2010b; Vialatte *et al.*, 2009c).
- Investigating sleep disorders (Benca *et al.*, 1999) and epilepsy (Le Van Quyen *et al.*, 2001a).
- Measuring the depth of anesthesia (Rampil, 1998).
- Testing drug effects (Blume, 2006)
- Serious games for e-learning and medical applications (Wang *et al.*, 2010).

Even though EEG has been largely used in the clinical neuroscience field, other domains have also benefit of the use of EEG. For example EEG has also been used in the Brain Computer Interface (BCI) field to communicate the human brain with a machine. Different applications can be derived from BCI systems, such as controlling video games (van Vliet *et al.*, 2012) or controlling a machine (Graimann *et al.*, 2011). Other applications related with entertainment have been given to EEG such as the shippo (Neurowear, 2014b) or the necomimi (Neurowear, 2014a) projects, in which an EEG sensor is allegedly used to evaluate the mood of the subject, and move a tail or cat's ears depending on it.

2.1.1 EEG recordings

EEG is recorded though different sensors placed on the scalp. Sensors are disk of 5 mm, generally constructed from Ag/AgCL (Silver/Silver chloride). For the EEG recordings different electrode placement systems have been proposed (10-20 system, Maudsley system,



(a) (b)

Figure 2.1: Placement of recording electrodes according to 10-20 system. (Schalk & Mellinger, 2010)

10-10 system). Nowadays the most commonly used is the 10-20 system. The International 10-20 system of electrodes placement presents an uniform coverage of the entire scalp (Fisch, 1999). This system is based on an iterative subdivision of arcs on the scalp starting from craniometric reference points: Nasion (Ns), Inion (In), Left (PAL) and Right (PAR) pre-auricular points (Schalk & Mellinger, 2010). The standard set of electrodes is detailed in Figure 2.1, presenting the location of 21 recording electrodes. The 10 and 20 indicate that the distance between adjacent electrodes is either 10% or 20% of a specified distance measured using specific anatomical landmarks, e.g. the total distance between the front and back or left and right of the head. Electrodes are numerated with a letter and a subscript. The letter specifies the anatomical area where the electrode corresponds: prefrontal or frontopolar (Fp), frontal (F), central (C), parietal (P), occipital (O), temporal (T) and auricular (A). The subscript is either the letter z, indicating zero or midline placement, or a number indicating lateral placement. Electrodes with even numbers are placed on the right side of the head, while odd numbers correspond to the left side. Number increase with increasing distance from the anterior posterior midline of the head (Fisch, 1999).

Once the electrodes are placed, different montages can be used for the recording of brain electrical potentials. EEG can be recorded using referential or bipolar montages. In referential montages, the voltage differences between all electrodes and a common electrode (reference) are recorded. In bipolar montages, however, instead of using a reference electrode, the voltage difference between two designated electrodes is recorded (i.e., each electrode pair is considered as a channel). A major disadvantage of the referential montage is that there is no single reference electrode optimal for all situations since no reference is truly inactive. Bipolar montages reduce the effects of common noise/artifacts and eliminate the influence of contaminated references (Fisch, 1999).

2.1.2 EEG rhythms

EEG is highly sensitive of subject's state and therefore EEG rhythms change depending on the subject task. However, historically five major types of continuous rhythmic EEG activities are recognized in the recordings. They are divided in different frequency bands.

The typical EEG frequency rhythms and their respective frequency bands are (Kropotov, 2009):

- δ rhythms, found in the frequency band of 1 to 4 Hz (δ band).
- θ rhythms, found in the frequency band of 4 to 8 Hz (θ band).
- α rhythms, found in the frequency band of 8 to 13 Hz (α band).
- β rhythms, found in the frequency band of 13 to 30 Hz (β band).
- γ rhythms, found in the frequency band of 30 to 100 Hz (γ band).

A description of these EEG rhythms is presented below.

δ rhythms

δ rhythms are the slowest of all the existing rhythms. However, they present the higher amplitude of all. These rhythms are visible when the subject is sleeping. The higher value is presented when the subject is in a deep sleep.

θ rhythms

θ rhythms are associated with drowsiness, childhood, adolescence and young adulthood. These are also found during problem solving, for example mathematical problems such as adding or subtracting. It is located in the prefrontal part of the cortex.

α rhythms

α rhythms are associated with relaxed states. It was the first which was discovered, because it is observable in almost all the population. It present different locations in the cortex, but is clearly visible on the occipital areas, with higher amplitude than the other rhythms. Different α rhythms are found in the human cortex: μ rhythm, α occipital rhythm and α parietal rhythm.

μ rhythms are rhythms that have been wide used in the implementation of BCI systems. They have this name due to the similarity that these rhythms present with the μ letter (with sharp negative peaks). Figure 2.2 presents an example of recording were it can be checked in electrodes C3 and C4. As it is shown in this figure, these rhythms presents some periods of activity with other of inactivity. They are found in the frequency range close to 10 Hz. The change of activity is a well-known phenomenon due to the synchronization of the groups of neurons in the motor cortex area. These groups of neurons are the ones that control the movement of arms and legs. When limbs are inactive, the μ rhythm presents activity, whereas when a subject moves his limbs the rhythm presents a decrease of amplitude, which is known as desynchronization. It has also been shown that this phenomenon is present not only when a subject physicality does an action, it also happens when a subject thinks about performing the same action. This property allows BCI systems to detect motor imagination (Schalk & Mellinger, 2010; Graimann *et al.*, 2011).

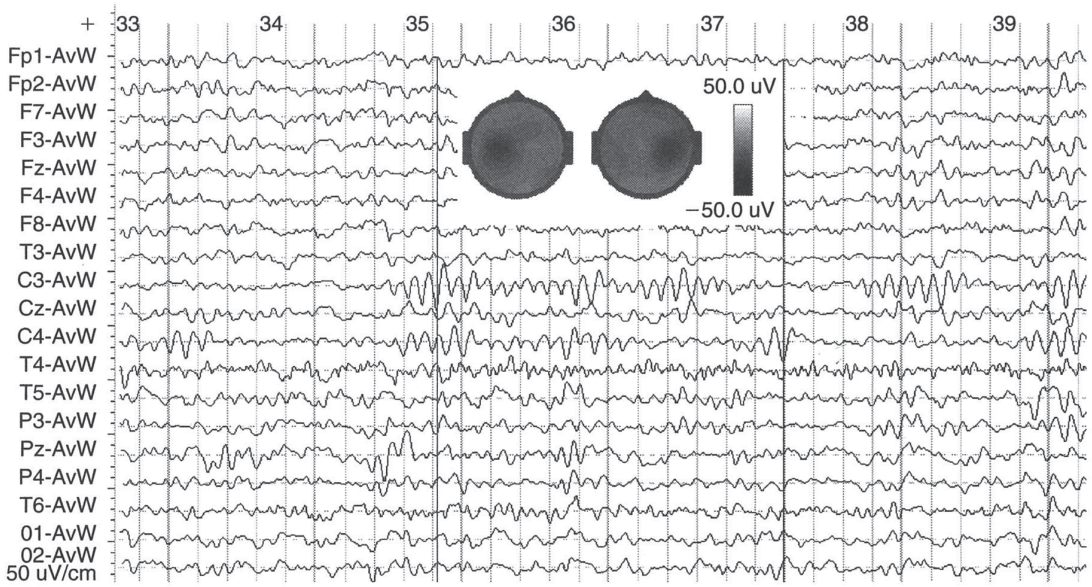


Figure 2.2: Example of EEG recording presenting the μ rhythm in electrodes C3 and C4. Two maps of potentials are taken at the moment indicated by black vertical line, illustrating that these rhythms are generated on the motor cortex area. (Kropotov, 2009)

Another example of α rhythms is the α occipital rhythms. These rhythms are visible at occipital electrodes O1 and O2. They have an amplitude higher than other rhythms as presented in Figure 2.3. These rhythms enhance their amplitude when the subject keeps his eyes closed, and they reduce the amplitude when the subject keeps his eyes opened. They decrease in response to visual stimuli.

The last of the α rhythms are found in the parietal area, with the maximum at the sensor Pz. These rhythms can be enhanced when the subject closes his eyes. However, this fact does not happen for all the population, because some subjects present a decrease of these rhythms when they close their eyes. The functional properties of the parietal rhythm are not fully understood.

β rhythms

β rhythms present low amplitude with multiple and varying frequencies, observable at different locations in the cortex. They are often associated with active, busy or anxious thinking and active concentration. There are different types of β rhythms, such as β Rolandic rhythms and β frontal rhythms.

β Rolandic rhythms are observed as spontaneous activity recorded on electrodes located close to the sensorimotor area (C3,Cz and C4). Usually present activity around the 20 Hz. β Rolandic rhythms are related with the intentionality of perform a movement, presenting desynchronization even before than the μ rhythm.

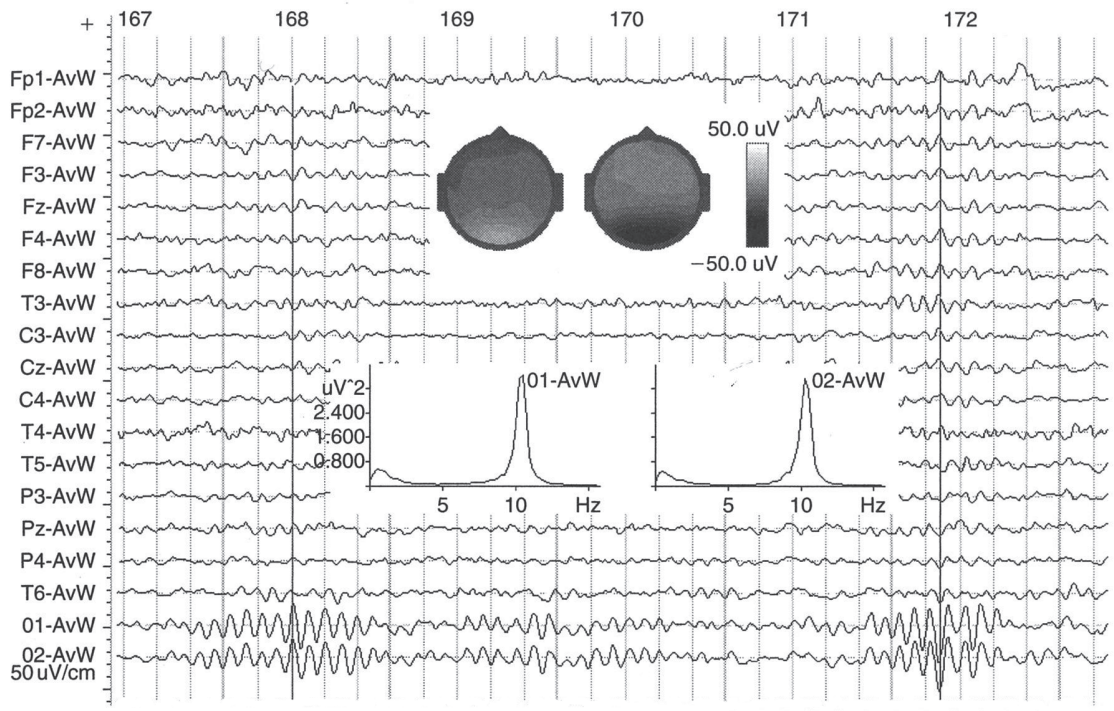


Figure 2.3: Example of EEG recording presenting an α occipital rhythm in electrodes O1 and O2. Vertical lines present the time instant at which the maps of potentials have been computed. These maps show a clear occipital location of the presented rhythms. (Kropotov, 2009)

Another type of β rhythms are β frontal rhythms. They appear during cognitive tasks related with decision making. The maximum activity is found in sensors F3, Fz and F4.

γ rhythms

Initially, γ rhythms were not studied, since old EEG recording systems could not record signals above 25 Hz. Until digital recordings systems were used, these rhythms were not known. One of the first articles describing these rhythms appeared in 1964.

γ rhythms appear to be involved in higher mental activity, including perception, problem solving, fear, and consciousness.

2.2 Alzheimer's Disease

AD is a neurodegenerative disease and the most prevalent form of age-related dementia in the modern society. With increasing life expectancy, dementia is a growing socioeconomic and medical problem. AD is positioned to become the scourge of this century, bringing with it enormous social and personal costs (Scinto & Daffner, 2000). This disease was first discovered more than 100 years ago, but research into its symptoms, causes, risk factors and treatment has gained momentum only in the last 40 years. Even though relevant

aspects of AD have been revealed, much is yet to be discovered about the changes that AD causes on patients (Alzheimer's Association, 2014).

AD causes damage of neurons in the brain. Damaged neurons no longer function normally and may die. Dead neurons cannot be replaced once lost (Scinto & Daffner, 2000). Over time, the brain shrinks dramatically, affecting nearly all its functions. AD affects patients in different ways, changing the rate of progression for each subject (Weiner *et al.*, 2013). The most initial symptom is a gradually worsening ability to remember new information. This occurs because the first neurons to malfunction and die are usually neurons in brain regions involved in forming new memories. As neurons in other parts of the brain malfunction and die, individuals experience other difficulties. The following are common symptoms of AD (Alzheimer's Association, 2014):

- Memory loss that interferes in daily life.
- Difficulties solving problems.
- Results difficult to complete familiar tasks at home, at work, or at leisure.
- Decreased or poor judgment.
- Difficulties to remember new words either speaking or writing.
- Confusion with time or place.
- Loss of the ability to retrace steps and misplacing things.
- Changes in mood and personality, including apathy and depression.
- Withdrawal from work or social activities.

AD is officially listed as the sixth leading cause of death in the United States. It is the fifth leading cause of death for people of 65 years and older (Murphy *et al.*, 2010).

Many factors have been linked to the incidence of AD, including age, gender (females are more likely to be affected), genetic factors, head injury, and Downs syndrome (Scinto & Daffner, 2000). However, experts believe that AD is rather caused by multiple factors than by a single cause. The major risk factors are (Alzheimer's Association, 2014):

- Age:

  An advanced age is the greatest risk factor for AD. Even though age is the greatest risk, is not sufficient to cause the disease.

- Family history:

  Individuals with a familiar suffering AD are more likely to later develop AD.

- APOE $\varepsilon 4$ gene:

  Research studies estimate that between 40% and 65% of people diagnosed with AD have one or two copies of the APOE $\varepsilon 4$ gene (Saunders *et al.*, 1993).

- Mild cognitive impairment (MCI):

  Patients suffering from MCI are more likely to develop AD and other dementia than people without MCI. However, not all patients suffering from MCI latter develop AD. Therefore this is a key stage for studding AD.

- Cardiovascular disease risk factor:

  It is suggested that the health of the brain is related with the heath of the heart and blood vessels. A good blood pleasure ensures that the brain receives the oxygen and nutrient necessary for its normally function.

- Social and cognitive engagement:

  Some studies suggest that remaining mentally and socially active may reduce the risk of AD and other dementia. The exact mechanism underlying this situation is unknown (Wang *et al.*, 2012).

- Education:

  People with fewer years of formal education are at higher risk for AD and other dementia, than those with more year or formal education (Kukull *et al.*, 2002).

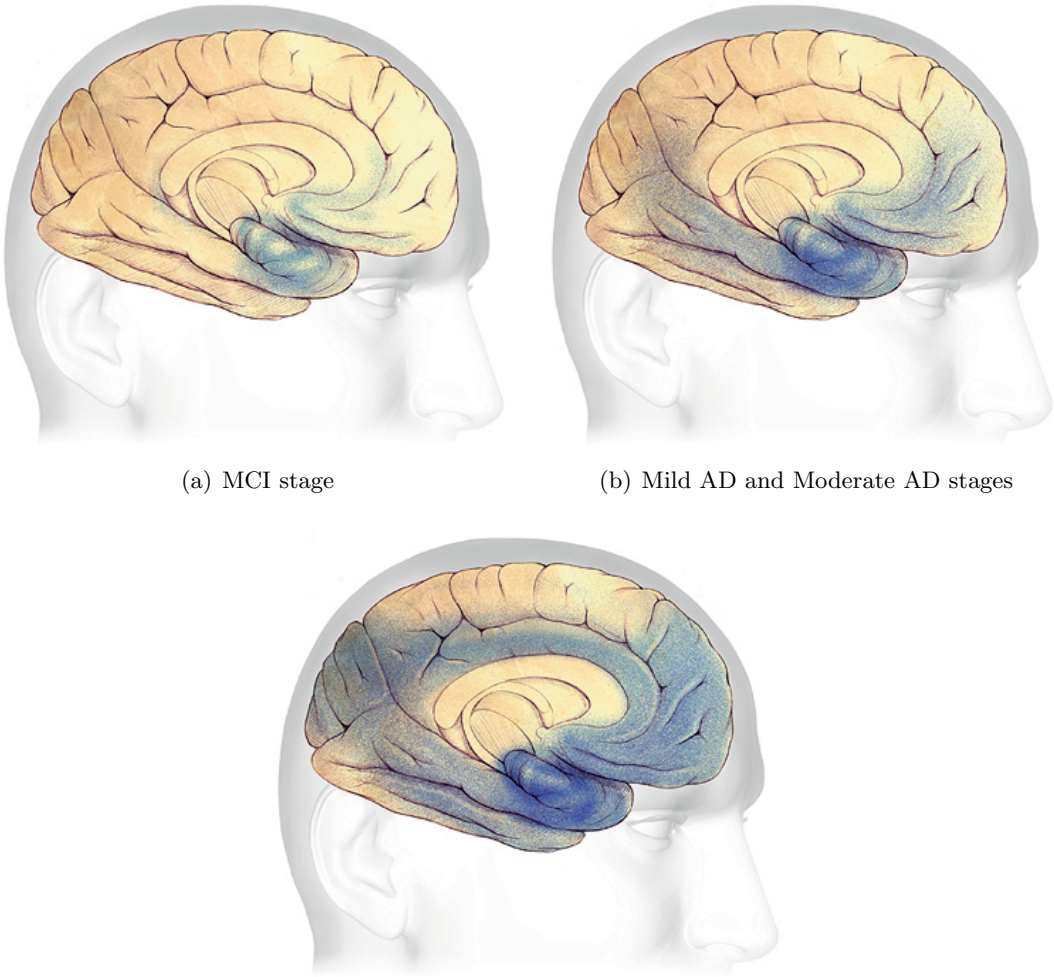
- Traumatic brain injury:

  Moderate and sever traumatic brain injuries increase the risk of developing AD.

AD does not present the same evolution in all the patients. The course of the disease depends in part on the age of the pathology declaration and on the health conditions of a person, related with the risk factors presented above.

Brain changes that are believed to contribute to the development of AD are the accumulation of β-amyloid (Aβ) plaques and neurofibrillary tangles composed of tau amyloid fibrils (Weiner *et al.*, 2013). The accumulation of the protein Aβ plaques is produced outside neurons, interfering with the communication between neurons, which happens in the synapses, and contributes to cell death. The accumulation of tau tangles is produced inside neurons. Tau tangles block the transport of nutrients to the neuron, and it is believed to also contribute to cell death (Alzheimer's Association, 2014). Currently no cure exists for Alzheimers, but administering certain medications in the early stages may delay the onset of symptoms (Dauwels *et al.*, 2010b; Alzheimer's Association, 2014).

Brain changes due to AD may begin 20 years before symptoms appear. At the beginning patients are able to function normally despite these brain changes. Afterwards the brain can no longer compensate the neuronal damage that has occurred. In this stage, patients start to show subtle decline in cognitive functions. Later, the damage due to the death of neurons is so significant that the patients start to present obvious cognitive decline. Later basic functions are impaired (Alzheimer's Association, 2014). Evolution of the disease is presented in Figure 2.4. The initial accumulation of Aβ plaques and tau tangles starts in the hippocampus, the part of the brain where memories are first formed (Figure 2.4(a)). Then more regions of the brain are affected by the accumulation of Aβ plaques and tau tangles, compromising other brain functions and presenting the different stages of the disease. At the next stage, the region of the brain where language is process is affected, compromising the ability of the patient to speak (Figure 2.4(b)). Next, the disease affects the frontal part of the brain, the regions were reasoning and planning is performed. Gradually, the person starts to lose the ability to solve problems and make plans. Then the accumulation of plaques and tangles affects the part of the brain where emotions are regulated. In this step the patient loose their control over moods and feelings. In advanced Alzheimer's disease, most of the cortex is seriously damaged (Figure 2.4(c)). Individuals lose their ability to communicate, to recognize family and loved ones and to care for themselves (Alzheimer's Association, 2014).



(a) MCI stage

(b) Mild AD and Moderate AD stages

(c) Severe AD stage

Figure 2.4: Evolution of AD in the brain. Plaques and tangles (shown in the blue-shaded areas) tend to spread through the cortex in a predictable pattern as AD progresses. (Alzheimer's Disease Association, 2014)

The presented progression of AD in the human brain is classified into four stages. The first, or preclinical, stage is MCI. MCI patients usually present some memory impairment, but retain their abilities in other cognitive domains and functional activities (Petersen *et al.*, 2009; Weiner *et al.*, 2013). Some MCI patients (between 6% and 25%) later develop AD. The next steps are characterized by growing cognitive deficits. The second and third stages are known as Mild AD and Moderate AD, while the last stage is known as Severe AD, entailing complete dependence on caregivers (Alzheimer's Association, 2014). MCI and Mild AD are key stages, an early diagnosis of AD in these stage may confer several benefits (Dauwels *et al.*, 2010b). To date, a definitive diagnosis of AD can only be made by postmortem analysis of the brain of a patient with dementia. A clinical diagnosis of AD is based on medical records, physical and neurological examination, laboratory tests, neuroimaging, and neuropsychological evaluation such as Mini-Mental State Exam (MMSE). The information of family members or close persons is used also as input. Even though the diagnosis of AD is done by combining all the existing information, it still hard to diagnose this disease because symptoms are often dismissed as normal consequences

of aging (Scinto & Daffner, 2000; Alzheimer's Association, 2014). While AD is a major public health problem, it also has a very private face that causes tremendous suffering to families. For the elderly, it is one of the most dreaded afflictions that threatens to rob them of their independence and dignity at the end of life (Scinto & Daffner, 2000; Alzheimer's Association, 2014).

2.3 Early diagnosis of Alzheimer's Disease using EEG

Previous sections presented two different topics. Section 2.1 presented the recording technique known as EEG and Section 2.2 showed the present information known about AD. This section describes the state of the art of using EEG as an early diagnosis tool to diagnose AD.

As presented in Section 2.2, AD diagnosis is performed using a combination of different tests. On the other hand, an early diagnosis may benefit the patient in different ways, such as facilitating to face the disease by giving time to inform themselves about the disease, taking economical depositions to plan for the future needs or given the patient time to take symptoms-delaying medications which are only effective in the first stages of the disease (Dauwels *et al.*, 2010b). Studies of early diagnosis of AD using EEG aim to help medical doctors to diagnose subject suffering from AD. It could be used as extra information in combination with laboratory tests and physical and neurological examination.

In recent years research groups had focused on the early diagnosis of AD using EEG (Dauwels *et al.*, 2010b; Vecchio *et al.*, 2013). There are several arguments supporting this research direction. One reason is that AD is a cortical dementia, consequently the damages induced are directly reflected in the EEG recordings (Jeong, 2004; Dauwels *et al.*, 2010b). Another reason is that the recording procedure used can be easily implemented in all clinical environments (Vecchio *et al.*, 2013). Usually when recording cortical EEG rhythms for AD diagnosis, subjects are in resting state with eyes closed. This recording procedure is used because some benefits are presented for the elder patients. As there is no required stimulation device and there is no task to perform, subjects are less fatigued and anxious to perform the task. Furthermore, resting state EEG rhythms can be recorded in highly comparable experimental conditions for all the subjects, being either MCI or Mild AD patients and healthy subjects (Babiloni *et al.*, 2011; Vecchio *et al.*, 2013).

Three different major effects had been presented that AD causes on EEG data. Slowing of EEG in AD patients, enhanced complexity of the EEG signals and perturbations in the EEG synchrony (Jeong, 2004; Dauwels *et al.*, 2010b; Babiloni *et al.*, 2011; Vecchio *et al.*, 2013). Even though these effects have been repeatedly reported in the literature, it may be not easily detectable due to the high variability presented among AD patients. Therefore, any of these effects can be used alone to facilitate a reliably early diagnosis of AD (Dauwels *et al.*, 2010b). Results obtained in the literature for each of the effects are presented below.

2.3.1 Slowing of EEG in AD patients

EEG recordings during eyes closed periods typically change across physiological aging by decreasing the power in the α band. This effect is also present when comparing AD patients (either MCI or Mild AD patients) and healthy subjects.

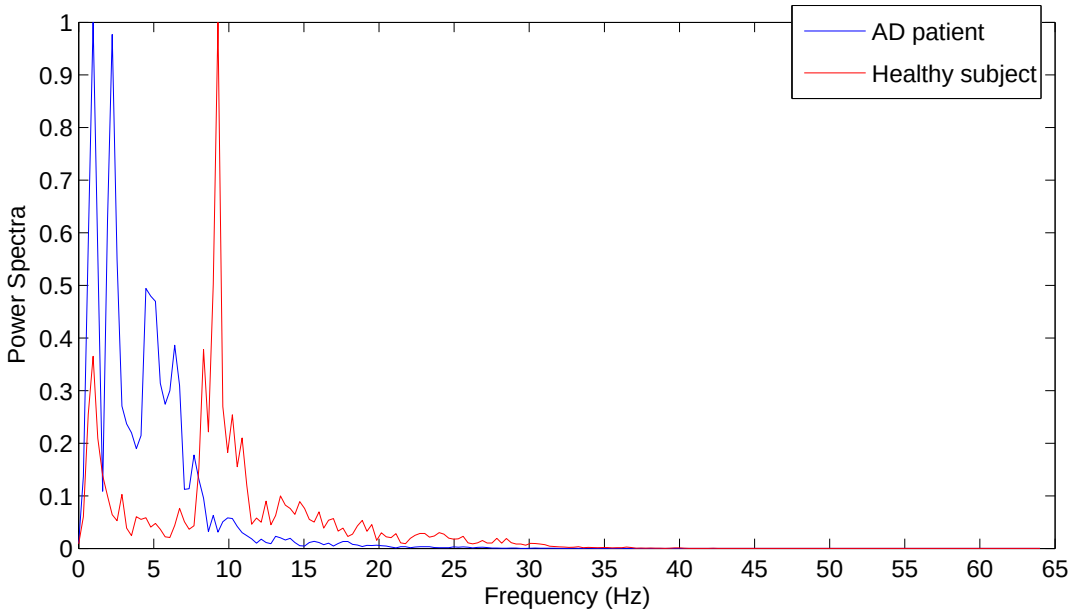


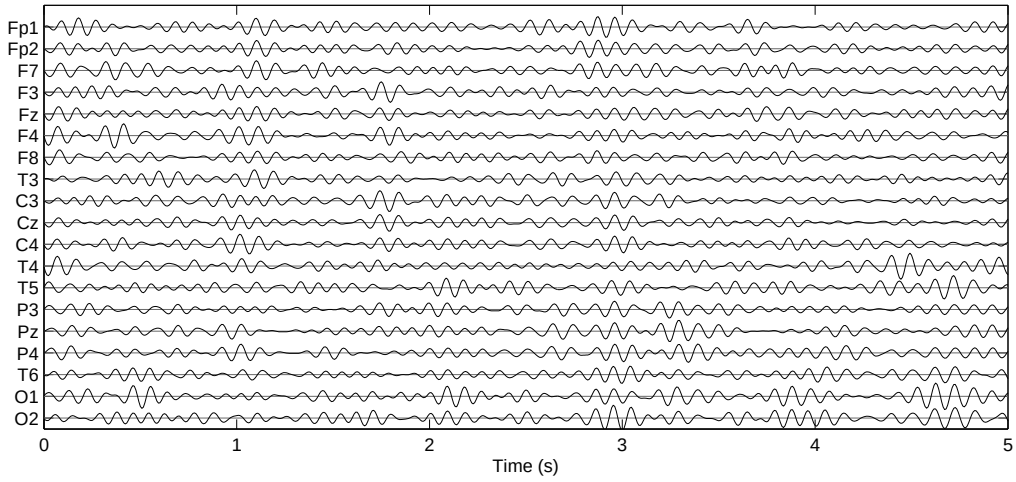
Figure 2.5: Power spectra comparison between a Mild AD patient and a healthy subject. Results presented for the electrode Fz.

Studies analyzing EEG data of AD patients have presented that there is a decrease in the power presented in the α and β bands corresponding with an increase of the power in δ and θ bands for AD patients when compared with healthy subjects (Bennys *et al.*, 2001; Jeong, 2004; Czigler *et al.*, 2008; Vialatte *et al.*, 2009c). During the process of developing AD, it is generally thought that the power decrease in β band and the power increase in θ are the earliest changes. These changes are later followed by a decrease in the α activity. Changes in the δ band are produced later, during the course of the disease (Jeong, 2004).

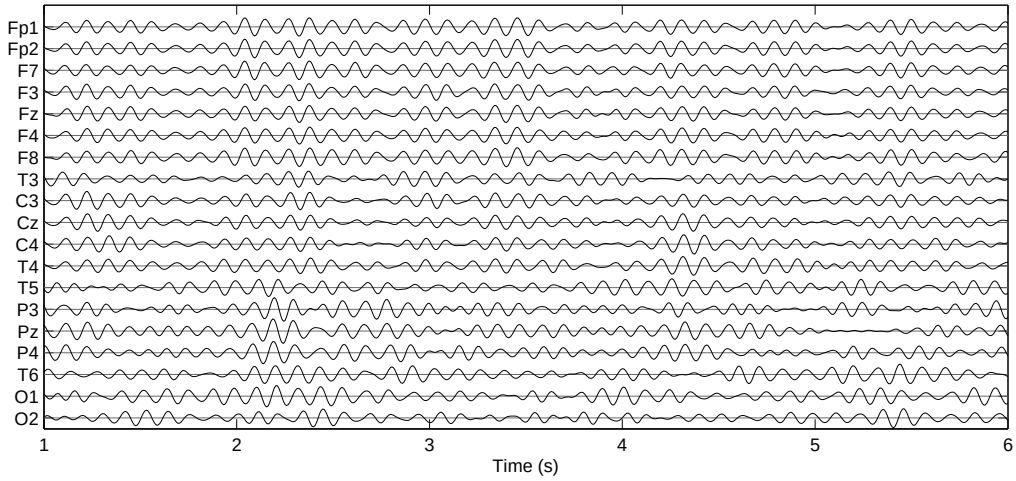
Usually to facilitate the comparison of power in the frequency domain the Fourier transform is applied. A comparison between the power spectra of an AD patient and a healthy subject is presented in Figure 2.5. This figure illustrates the clear difference existing between the power spectra of a healthy subject (red in the figure), in comparison with the power spectra of an AD patient (blue in the figure). In this figure the differences between the two power spectra are clearly presented. Healthy subject has a peak of power between 9 and 10 Hz (the α band), whereas the AD patient has different peaks of power below 8 Hz, standing for the δ an θ band, being the δ band that present the higher values of power spectra. Studies have mainly used the differences of power in the different frequency bands (δ, θ, α and β bands) to differentiate how AD patients and healthy subjects are classified (van der Hiele *et al.*, 2007; Baker *et al.*, 2008; Moretti *et al.*, 2009).

2.3.2 Perturbations in the EEG synchrony

One of the most important features of the brain is not the high number of neurons that it contains but the abundant connectivity between these. This high connectivity is reflected in synchronous activity, as neurons in anatomically connected structures tend to fire synchronously. EEG data show this synchronicity by repeated bursts at different frequencies (Kropotov, 2009).



(a) AD patient



(b) Healthy subject

Figure 2.6: Synchrony comparison in EEG recordings from an AD patient and a healthy subject. Recordings have been filtered in the α band. Higher synchrony is presented for the healthy subject.

Changes in the synchrony of AD patients have also been investigated. Results presented from the study of synchrony activity in AD patients show that synchrony in different channels of an EEG recording seems to be lower for AD patients than for healthy subjects. (Jeong, 2004; Dauwels *et al.*, 2010b; Babiloni *et al.*, 2011; Vecchio *et al.*, 2013).

In order to explain the differences of synchrony in an EEG recording Figure 2.6 shows the comparison between the recording of an AD patient (illustrated in Figure 2.6(a)) and a healthy subject (Figure 2.6(b)). For displaying purpose and aiming to facilitate the comparison between recordings, signals have been filtered with a band pass filter in the frequency range of 8 - 13 Hz (α band). Recording of the healthy subject has higher synchrony due to the higher number or bursts that it present. Furthermore, these are longer in time than the ones shown in the recording of the AD patient. Especially frontal electrodes Fp1, Fp2, F7, F3, Fz, F4 and F8 present a common activity that stars in second 2 and last until second 3.5 showing this higher synchrony. Moreover other periods of long

synchrony activity are displayed in the recording of healthy subject when compared with the recording of the AD patient.

A large variety of measures have been used in the scientific literature in order to compute the synchrony of EEG recordings. Some typical measures are Coherence, Granger Measures, state space based synchrony measures, Phase Synchrony (PS) and stochastic event synchrony among others. All these measures seek to quantify the relationships between two or more signals.

Coherence is a well-known measure that have been used to estimate the functional connectivity between areas (Locatelli *et al.*, 1998; Adler *et al.*, 2003). Coherence estimates the relationship between two time series in the frequency domain. However, Coherence does not estimate the directionality of the information and it has to be computed between pairs of signals. An extension for multivariate signals is Granger measures (Kaminski & Blinowska, 1991; Babiloni *et al.*, 2009a), which also allow to investigate the causality of interactions.

Previous measures are based on using the amplitude of the signals. However, other measures have been used for the early diagnosis of AD studying the differences in the phase of the signals. Phases between two signals may be related even if its amplitudes present no relation. An example of these measures are PS (Dauwels *et al.*, 2009a; Knyazeva *et al.*, 2010), phase lag index (Tóth *et al.*, 2014) or imaginary coherence (Hsiao *et al.*, 2013), which work for pairs of univariate signals. A measure used to compute the differences of phase between multivariate signals is global field synchronization (Koenig *et al.*, 2005; Park *et al.*, 2008).

Another type of synchrony measures that has been used can be referred to as state space based synchrony measures. This type of measures compute the synchronization in base of trajectory of the proposed signals in a state space. Signals are considered synchronous if the trajectories remain close to each other (Dauwels *et al.*, 2010b). Examples of state space based synchrony used to evaluate the differences of synchrony in AD patients are synchronization likehood (Stam *et al.*, 2005; Babiloni *et al.*, 2006; Czigler *et al.*, 2008), Omega Complexity (OC) (Bhattacharya, 2000; Czigler *et al.*, 2008) or S-estimator (Knyazeva *et al.*, 2013; Yi *et al.*, 2014).

An entirely different approach was presented recently, stochastic event synchrony (Dauwels *et al.*, 2009b,c). This new approach characterizes the interaction between spikes and transient oscillatory components. To compute stochastic event synchrony first the bumps are extracted using bump modeling (Vialatte *et al.*, 2009a,b). Each bump is considered as an event on the time-frequency plane, and then the geometry between the extracted bumps is modeled aiming to evaluate the synchrony between signals. Even though promising results have been obtained using this approach for the early diagnosis of AD (Dauwels *et al.*, 2009a), the high complexity of this measure and its number of parameters to calibrate make it an avoidable choice if the other measures are useful with the existing data.

Numerous studies have used the above presented measures and others, presenting that EEG synchrony for AD patients obtain a decreased value when compared with synchrony of healthy subjects. These EEG changes may reflect an abnormal synchronization of pyramidal neurons and a functional disconnection among cortical areas along AD process (Jeong, 2004; Babiloni *et al.*, 2009b, 2011; Vecchio *et al.*, 2013). However, there is no consensus regarding which of the measures is more effective to parametrize this functional disconnection for diagnosing AD. Studies usually only use a few measures or only one in

one specific data set, therefore it is difficult to compare between studies (Dauwels *et al.*, 2010b). Another shortcoming of these studies is that usually the data set has small number of subjects, and in some of the studies no cross-validation is done between computed values. Therefore, in some of the studies some overfitting may compromise the results.

In order to overcome the shortcoming of selecting the most appropriate measures, a study was presented recently comparing the use of different measures in a data set (Dauwels *et al.*, 2010a). In this study a total of 29 measures are used to differentiate AD patients of healthy subjects. The main result presented in this article is that similarity is found between some of the 29 measures, which lead to the hypothesis that the same information is presented for different measures. Therefore, only a small group of measures may be used to distinguish between AD patients and healthy subjects without redundant information.

2.3.3 Enhanced complexity of the EEG signals

Nonlinear dynamical analysis (NDA) of the EEG has revealed a decreased complexity of EEG patterns and reduced functional connections in AD (Jeong, 2004). NDA has offered valuable information on cortical dynamics when applied to EEG, due to the high complexity of biological signals (Jeong, 2004).

To quantify the EEG complexity different measures have been used, such as different types of entropy, or the Lempel-Ziv complexity. Entropy is a measure of the uncertainty of a random variable (Cover & Thomas, 2006). For the early diagnosis of AD using complexity measures, different types of entropy have been used, such as approximate entropy (Abásolo *et al.*, 2005), spectral entropy (Abásolo *et al.*, 2006) or tsallis entropy (Zhao *et al.*, 2007). On the other hand, the Lempel-Ziv complexity (Lempel & Ziv, 1976) has also been used for the early diagnosis of AD (Hornero *et al.*, 2009). According to these studies, EEG of AD patients display lower values of complexity (therefore is more regular), than values obtained for the healthy subjects. This difference can be used to distinguish between AD patients and healthy subjects.

Results presented in Dauwels *et al.* (2011) evaluate the use of complexity measures together with the effect of slowing of EEG (presented in Section 2.3.1). Results of this article show that for the data sets used, complexity measures are highly correlated with measures used to compute the effect of slowing of EEG. If used together for classification purpose they will only introduce redundant information to the classifier. Therefore, only by modeling one of the effects (slowing of EEG or changes in the complexity of the signal), may suffice to obtain the information contained in the recording.

Chapter 3

Using Empirical Mode Decomposition for preprocessing

This chapter focuses on improving the quality of EEG data by removing artifacts, using a new signal processing technique, Multivariate Empirical Mode Decomposition (mEMD). This technique provides a decomposition of the original EEG data into several oscillatory modes. By using this decomposition, a cleaning method is presented, and then the efficiency of the proposed method is evaluated on simulated EEG data.

The organization of this chapter can be summarized as follows. First, in Section 3.1, an introduction about the main interferences existing in the EEG data (which are known as artifacts), is presented. Section 3.2 presents the original decomposition technique Empirical Mode Decomposition (EMD), and its extension for multichannel data sets, mEMD, is then detailed in Section 3.3. Section 3.4 presents a new method to clean EEG data based on mEMD. Section 3.5 describes the methods used to simulate EEG data, and results obtained of applying the cleaning method on EEG simulated data are then presented in Section 3.6. Section 3.7 is devoted to explain other applications of the mEMD technique used for signal processing. Last section of this chapter is devoted to discussion.

3.1 Introduction

EEG signals commonly present different interference signals due to muscles artifacts, such as eye blinks or eye movements. Electric potentials due to these artifacts can be orders of magnitude larger than the EEG and can propagate across the scalp, masking and distorting brain signals (Croft & Barry, 2000).

Along years, different preprocessing techniques have been used to remove undesired signals from EEG. Some of these are filtering the data with filters, used mainly to remove the noise introduced for the power supply at 50 or 60 Hz, or by adaptive filtering, suppressing only the undesired signals. Other approaches are reducing the artifacts by asking the subject to avoid any movement during the recording, or by later rejecting the corrupted epochs that contain artifacts (Dauwels *et al.*, 2010b). On the other hand, an alternative approach known as BSS (Cichocki & Amari, 2002), has been largely used to clean artifacts from EEG signals (Joyce *et al.*, 2004; Delorme *et al.*, 2007; Vialatte *et al.*, 2009c; Solé-Casals

et al., 2010). BSS describes the EEG signals by a linear combination of a finite set of sources located within the brain. Using BSS for artifact rejection, the assumption is that artifacts are generated by a subset of extracted sources. Therefore, if sources that produce the artifacts are removed, signals can be reconstructed without affecting the EEG signals (Dauwels *et al.*, 2010b).

Even though BSS appears to be the most suitable method for artifact rejection, as any method, it presents some limitations. The key challenge is to determine what sources need to be removed. Usually this is a process that can not be done automatically. Visually inspection of the components extracted by BSS is done and then which components have to be removed is decided. This is a time consuming process, and therefore is not suitable for routine clinical EEG (Jung *et al.*, 2000). Furthermore, visual inspection is subjective (Romero *et al.*, 2009), and therefore the reliability of BSS is limited. In order to overcome this limitations, semi-automatic methods have been developed. This semi-automatic methods select the sources that have to be removed by computing statistical markers. Different approaches have been presented. However, is not well define which markers are the most appropriate for artifact detection (Dauwels *et al.*, 2010b).

New methods presented in this chapter aims to develop a system to perform the EEG signal preprocessing. This system have to be fully automatic, and therefore easy to implement for clinical EEG. The process of avoiding the visual inspection of the results is a key step that may facilitate its usage. Before this new method is defined, the main artifacts which can be found in EEG data are presented.

3.1.1 Artifacts due to eye blinks

Eye blink artifacts are generated by the movement of the eyelid along the cornea, such as during an eye blink. By friction between lid and cornea, this movement results in charge separation, with a dominantly dipolar charge distribution, and the dipole moment pointing in up-down-direction. In the EEG, this is recorded as a positive peak that lasts a few tenths of a second, is most visible in the frontopolar region (electrodes Fp1 and Fp2), but propagating to all the electrodes of the montage, becoming weaker with distance from the front. The frequency bands that are more affected by this artifacts are δ and θ, with the main portion of energy below the 5 Hz (Schalk & Mellinger, 2010).

An example of eye blinks it presented in Figure 3.1. This figure shows EEG time series that present three eye blinks. It can be observed the high impact that eye blinks produce on the frontal electrodes Fp1 and Fp2, achieving higher values than normal EEG, and how with some attenuation it affects all other electrodes of the data set. Attenuation is proportional to the distance between electrodes and the eyes.

3.1.2 Artifacts due to eye movements

This type of artifacts are produced by eye movements, due to a frictive mechanism similar to the one underlying eye blink artifacts but involving retina and cornea rather than cornea alone. Eye movement artifact affects frontal electrodes which are positioned close to the temporal area, e.g. F7 and F8. The effect created on electrodes can be symmetric or antisymmetric, depending whether the movement is vertical or horizontal, respectively.

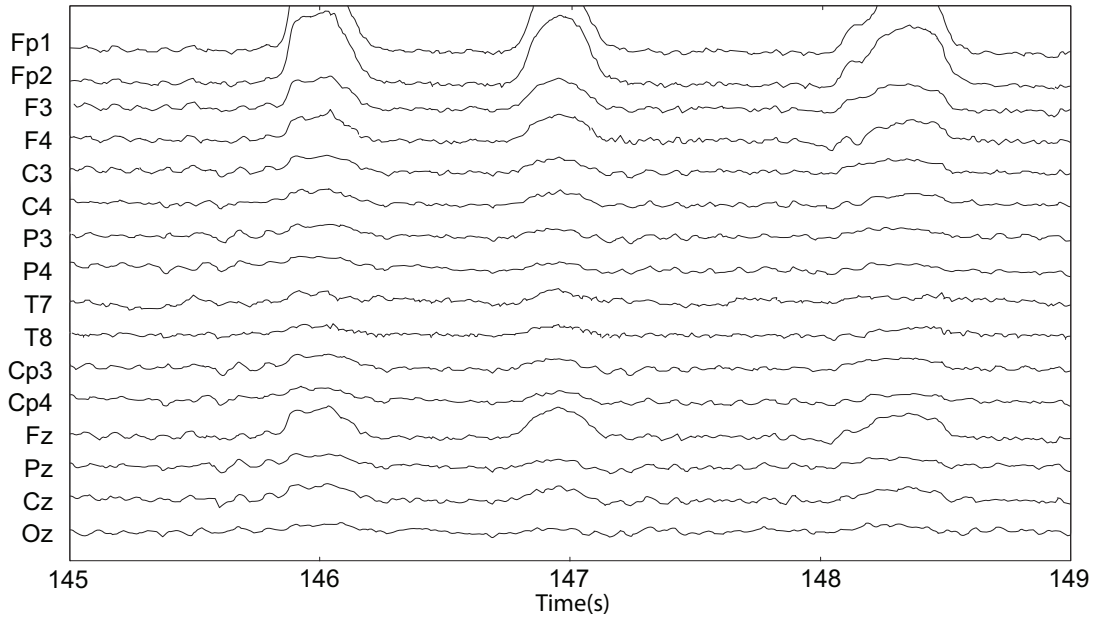


Figure 3.1: Example of EEG recording with eye blinks. Frontal electrodes Fp1 and Fp2 present clear interferences due to the activity of eye blinks. Other electrodes present the same interference, with a proportional attenuation to the distance. (Schalk & Mellinger, 2010)

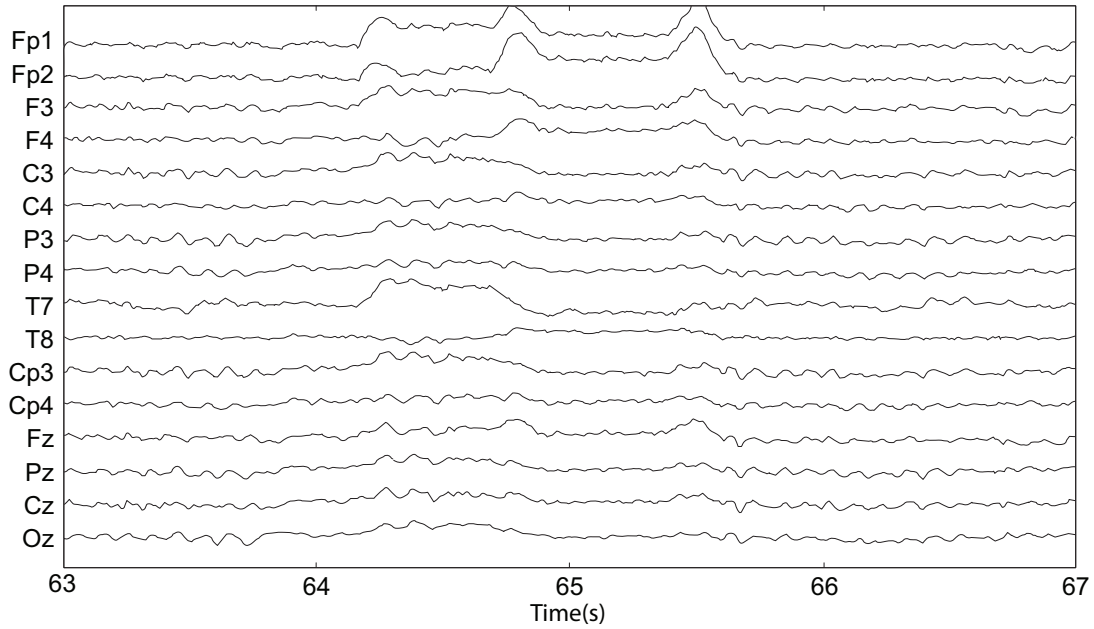


Figure 3.2: Example of EEG recording with eye movements. Distortion is clearly present in electrodes C3 and C4, which are those that present an external position. (Schalk & Mellinger, 2010)

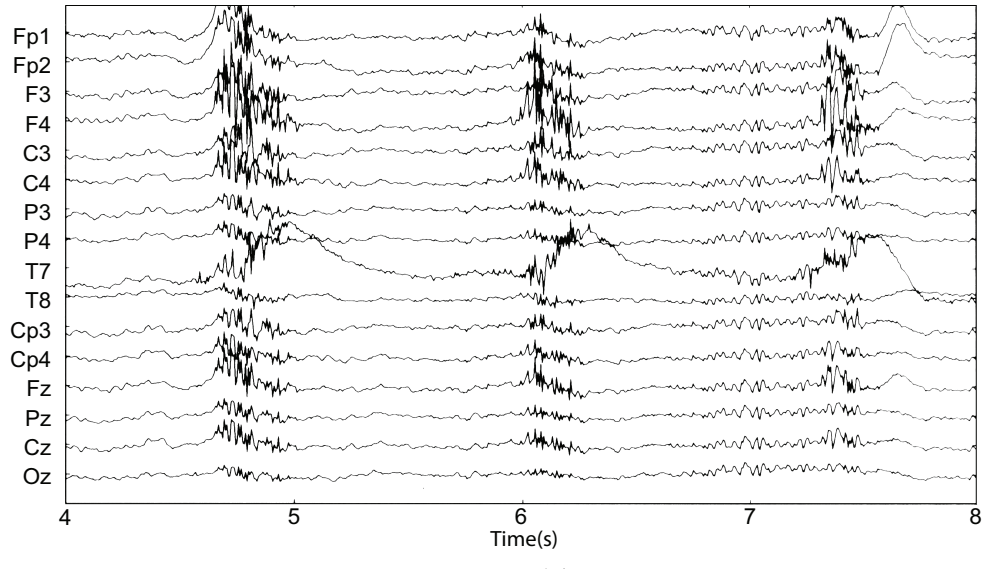


Figure 3.3: Example of EEG recording with muscle artifacts. Distortion is clearly present in all electrodes, specially those found in the central and parietal area. (Schalk & Mellinger, 2010)

The effect of eye movement artifacts is quite similar to that of blink artifacts, except that their frequency content is even lower, and amplitudes tend to be larger (Schalk & Mellinger, 2010).

Figure 3.2 present an example of EEG recording that present distortions due to lateral eye movements. In this case distortion caused by eye movements i lower in frequency and presents its maximum values in electrodes close to the eyes.

3.1.3 Artifacts due to muscle movements

Muscle artifacts are produced by the muscles that lift the eye brows, and those that close the jaw. Both group of muscles produce involuntary movements due to physiological phenomenon. The recordings of this type of artifacts can present a higher amplitude than EEG signals, making difficult to extract the EEG information. This kind of artifacts have to be prevented during the EEG recording (Schalk & Mellinger, 2010).

An example of muscle artifacts is shown in Figure 3.3. In this figure it can be observed that this kind of artifacts affect all the electrodes, and that brain activity is completely masked in the presence of these artifacts.

3.2 Empirical Mode Decomposition

This section presents the EMD technique. First, the definition is detailed and then is presented an example of application on a signal. This decomposition technique is used for signal processing analysis, nonetheless several shortcomings are found when is used in

multichannel data sets. These shortcomings are also detailed. Even though this is not the technique used for the data analysis, a clear understanding of it will facilitate the extension to multichannel data sets.

3.2.1 Theoretical definition

EMD algorithm is a method designed for multiscale decomposition and timefrequency analysis, which can analyze nonlinear and non-stationary data (Huang *et al.*, 1998). This method decomposes an original signal into a finite and often small number of oscillatory modes. These modes present an amplitude or frequency modulation. Oscillatory modes are called Intrinsic Mode Functions (IMFs). Each IMF satisfies two basic conditions (Huang *et al.*, 1998):

- The number of zero-crossings and the number of extrema must be the same or differ at most by one in the whole data set.
- At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The EMD algorithm (Huang *et al.*, 1998; Rilling *et al.*, 2003) for the signal $x(t)$ can be summarized as presented in Algorithm 3.1.

Algorithm 3.1: Empirical Mode Decomposition

i. Determine the local maxima and minima of $x(t)$.

ii. Generate an upper and a lower signal envelope by interpolationg the local maxima and minima values computed on step (i), obtaining $e_{min}(t)$ and $e_{max}(t)$.

iii. Compute the local mean $m(t)$ as:

$m(t) = \frac{e_{min}(t)+e_{max}(t)}{2}$

iv. Subtract the local mean from the data:

$s(t) = x(t) - m(t)$

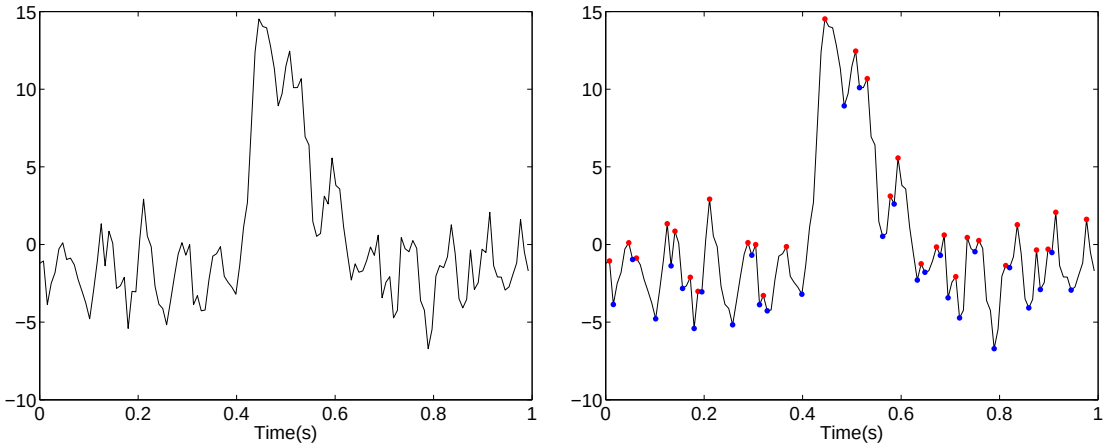
v. If $s(t)$ obeys the stopping criterion, then define $d(t) = s(t)$ as an IMF, otherwise set $x(t) = s(t)$ and repeat the process from step (i).

Then, the empirical mode decomposition of a signal $x(t)$ can be written as:

$$x(t) = \sum_{k=1}^{n} IMF_k(t) + \varepsilon(t) \tag{3.1}$$

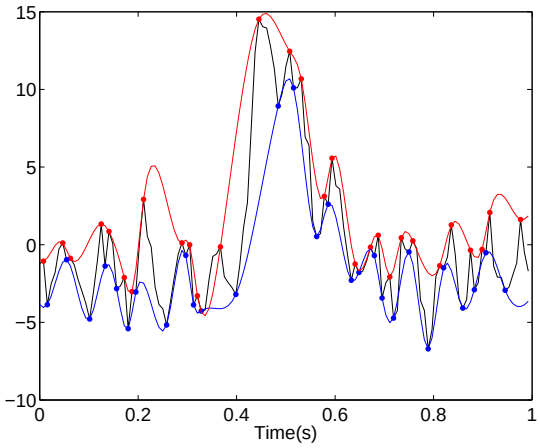
Where n is the number of extracted IMFs, and the final residue $\varepsilon(t)$ is the mean trend or a constant.

Figure 3.4 presents an example of application of the EMD on a signal $x(t)$. First, original signal $x(t)$ it is presented in Figure 3.4(a). Figure 3.4(b) shows the point (i) of the Algorithm 3.1, were the local maxima (red dots) and minima (blue dots) of $x(t)$ are computed.

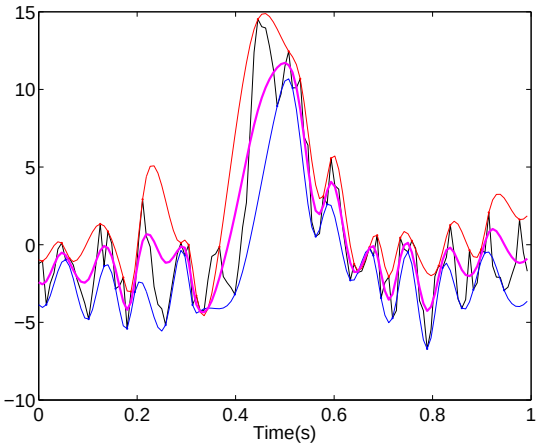


(a)

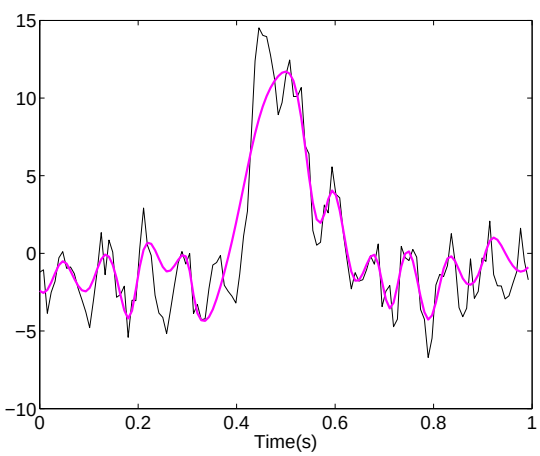
(b)



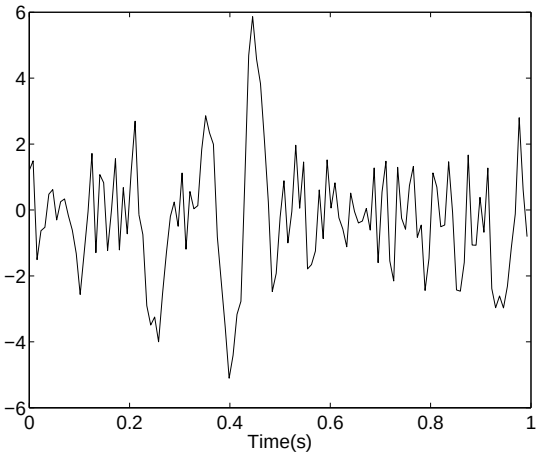
(c)



(d)

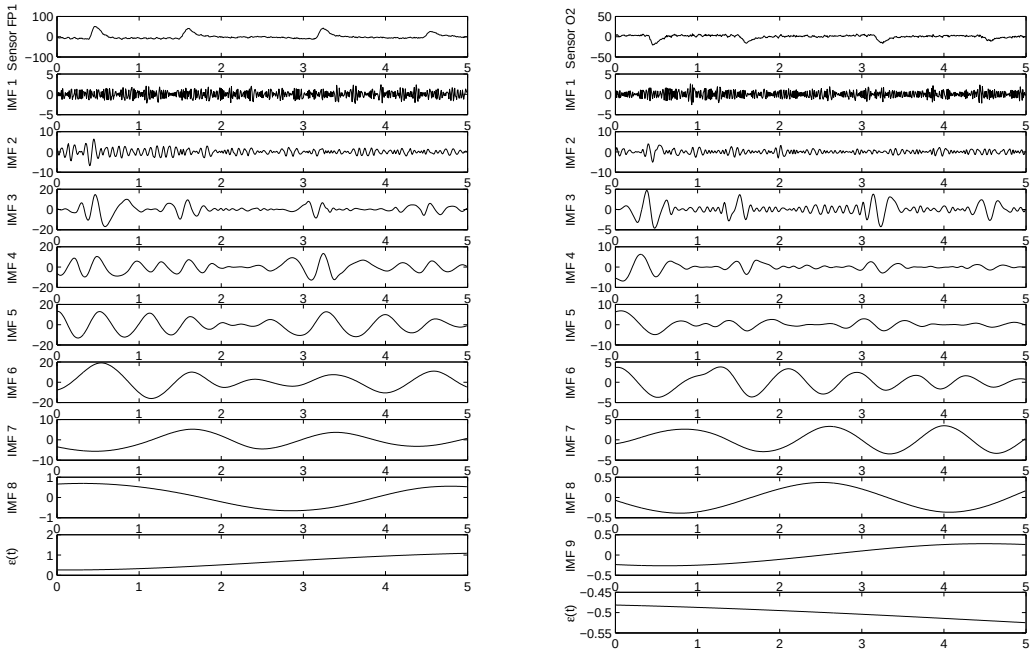


(e)



(f)

Figure 3.4: Example of one iteration of the EMD algorithm on the signal $x(t)$. (a) Initial signal $x(t)$. (b) Step (i) of the Algorithm 3.1 computing local maxima and minima values. (c) Step (ii) of the Algorithm 3.1 computing $e_{min}(t)$ and $e_{max}(t)$. (d) Step (iii) of the Algorithm 3.1 computing $m(t)$. (e) Step (iv) of the Algorithm 3.1 before substracting $m(t)$ to $x(t)$. (f) Step (v) of the algorithm presenting the signal $s(t)$.



(a) Sensor FP1 (b) Sensor O2

Figure 3.5: EMD of two EEG sensors FP1 and O2. Below each of the sensors the IMF obtained using the EMD are presented.

Then on Figure 3.4(c) the lower signal envelope ($e_{min}(t)$, in blue) and upper signal envelope ($e_{max}(t)$, in red) are computed. These curves are obtained using an interpolation method to connect the values computed previously. This procedure is defined in the step (ii) of the algorithm. Figure 3.4(d) presents the computation of the local mean ($m(t)$, in purple) as the average of $e_{max}(t)$ and $e_{min}(t)$, being the step (iii) of Algorithm 3.1. Figure 3.4(e) shows the signals $x(t)$ and $m(t)$ before being substracted. Result of computing the difference between these signals is presented in Figure 3.4(f) obtaining the signal $s(t)$. Once the signal $s(t)$ is computed, the step (v) of the Algorithm 3.1 is reached. At this point the algorithm decides if $s(t)$ obeys the stopping criterion, being an IMF or if it does not. The stopping criterion used usually is to obtain a median envelope as close as possible to zero.

3.2.2 EMD shortcomings for multichannel data sets

As presented in previous section, the EMD allows us to decompose a signal in different oscillatory modes. This decomposition technique has been successfully used previously for removing artifacts from EEG data (Rutkowski *et al.*, 2008; Diez *et al.*, 2009; Rutkowski *et al.*, 2009; Salis *et al.*, 2013). However, several shortcomings are presented when this technique is used in multichannel data sets such as EEG.

The IMFs from different time series do not necessarily correspond to the same frequency, and different time series may end up having a different number of IMFs (Mutlu & Aviyente, 2011). This second shortcoming, when using EMD for multichannel data sets, is illustrated in Figure 3.5. This figure shows five seconds of the EMD of two EEG sensors, FP1 presented in Figure 3.5(a) and O2 presented in Figure 3.5(b). These figures demonstrate that

for different time series, a different number of IMFs are obtained during the decomposition (eight for the sensor FP1 and nine for the sensor O2). On the other hand, another shortcoming is that even if the same number of IMFs is obtained during the decomposition, it may not represent the same rang of frequencies, therefore it will be difficult to compare between two modes.

These shortcomings make difficult to use this technique to work with multivariate data sets, because it is laborious to match the different obtained IMFs from different channels. To solve the presented shortcomings when working with multivariate data sets, several extensions of EMD have been proposed. These extensions are Bivariate Empirical Mode Decomposition (Molla *et al.*, 2010), to decompose two time series at the same time, and Trivariate Empirical Mode Decomposition (Rehman & Mandic, 2010a), to decompose three time series at the same time. Recently, for multivariate data sets, such as EEG, an extension of EMD to mEMD was proposed (Rehman & Mandic, 2010b).

3.3 Multivariate Empirical Mode Decomposition

The decomposition technique mEMD has been designed as an extension for multivariate data sets of EMD. This technique has been presented to deal with the shortcomings detailed in Section 3.2.2. Therefore mEMD obeys two different conditions:

- Different signals have the same number of IMFs.
- IMF i of one signal presents the same oscillation frequencies than the IMF i of another signal.

In mEMD the local mean is computed by tanking an average of upper and lower envelopes obtained from all the channels. The upper and lower envelopes, in turn are obtained by interpolating between the local maxima and minima values. However, in general for multivariate signals, the local maxima and minima may not be defined directly. To deal with these problems multiple n-dimensional envelopes are generated by taking signal projections along different direction in n-dimensional spaces (Rehman & Mandic, 2010b).

Consider a sequence of n-dimensional vectors $\{v(t)\}_{t=1}^{T} = \{v_1(t), v_2(t), \ \dots \ , v_n(t), \}$ which represents a multivariate signal with n components, and $x^{\theta_k} = \{x_1^k, x_2^k, \ \dots \ , x_n^k\}$ denoting a set of direction vectors along the directions given by angles $\theta^k = \{\theta_1^k, \theta_2^k, \ \dots \ , \theta_{(n-1)}^k\}$ on an $(n-1)$ sphere. Then, the mEMD algorithm (Rehman & Mandic, 2010b) is summarized as presented in Algorithm 3.2. Once all IMFs have been extracted, the mEMD of the input signal $v(t)$ can be written as detailed in Equation 3.1.

Figure 3.6 shows the mEMD of an EEG sensor. This decomposition has been obtained from the join decomposition of EEG simulated data, obtaining a total of 9 IMFs and $\varepsilon(t)$. This figure demonstrate that first modes present the high frequencies, and that the lasts IMFs present the low frequencies. Last value shown is $\varepsilon(t)$, which has a really low frequency oscillation term. On the other hand, Figure 3.6 shows that there is a certain relation between the presence of eye blinks in the data of the original sensor, and the presence of modulations of low frequency in IMF 4 and IMF 5. Taking into account these results, next section presents a cleaning method based on mEMD to decompose a set of original EEG data, and after computing the mEMD, IMF that present a relationship with the eye blinks are eliminated.

Algorithm 3.2: Multivariate Empirical Mode Decomposition

i. Choose a suitable pointset for sampling on an $(n-1)$ sphere.

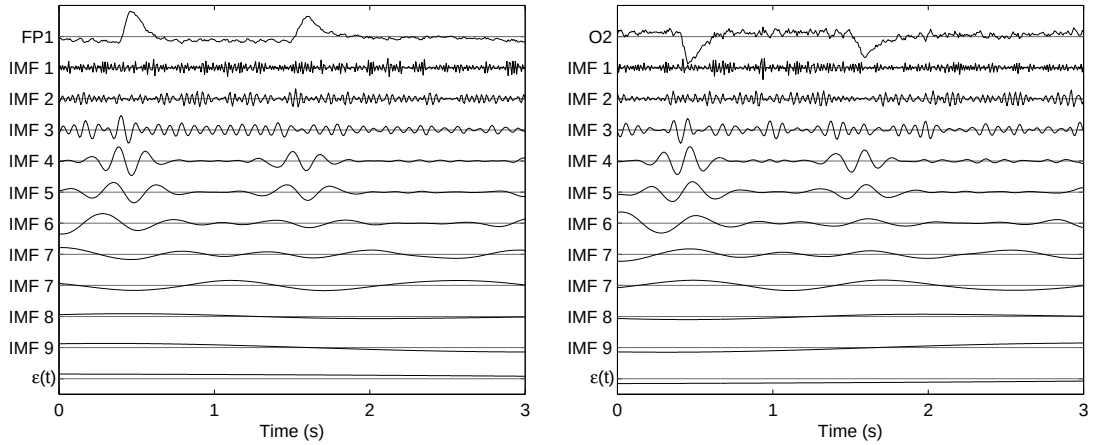
ii. Calculate a projection, denoted by $p^{\theta_k}(t)\}_{t=1}^{T}$, of the input signal $\{v(t)\}_{t=1}^{T}$ along the direction vector x^{θ_k}, for all k, giving $p^{\theta_k}(t)\}_{t=1}^{K}$ as the set of projections.

iii. Find the time instants $\{t_i^{\theta_k}\}$ corresponding to the maxima of the set of projected signals $p^{\theta_k}(t)\}_{t=1}^{K}$.

iv. Interpolate $[t_i^{\theta_k}, v(t_i^{\theta_k})]$ to obtain multivariate envelope curves $e^{\theta_k}(t)\}_{t=1}^{K}$.

v. For a set of K direction vectors, the mean of the envelope curves $m(t)$ is calculated as:
$m(t) = \frac{1}{K}\sum_{k=1}^{K} e^{\theta_k}(t)$

vi. Extract the detail $d(t)$ using $d(t) = x(t) - m(t)$. If the detail $d(t)$ fulfills the stopping criterion for a multivariate IMF, apply the above procedure to $x(t) - d(t)$, otherwise apply it to $d(t)$.



(a) Decomposition sensor Fp1

(b) Decomposition sensor O2

Figure 3.6: mEMD of two EEG signal. Values obtained for the sensors are presented in the first line. IMFs obtained are presented below. For the same sensors used in Figure 3.5 the same number of IMF have been obtained. IMF with the same number correspond to the same frequency ranges.

3.4 A cleaning method based on mEMD

As presented in Section 3.3, the mEMD of a multivariate signal $v(t)$ can be written as the sum of different IMF. Aiming to eliminate artifacts from EEG recordings, a new cleaning method is presented. The cleaning method is based on mEMD and seeks the common modes which are present in all the electrodes. The key idea is that if a mode is present in all the electrodes, it is probably due to artifacts and not to EEG signals. Therefore this mode has to be eliminated in the reconstruction process.

With the aim of improving the quality of the data, the cleaning method is detailed in Algorithm 3.3. This method present several advantages. The first is that is fully automatic, does not requires visual inspection. The second advantage is that when using this method, only the EEG signals are used during the cleaning process, and there is no need to use extra electrodes such as electrooculography (EOG) electrodes. However, a shortcoming presented for this method is that the parameter T has to be selected. In Section 3.6 different values are tested for this parameter, and a final value is presented as the optimal one. Figure 3.7 presents an example of application of the cleaning method.

Algorithm 3.3: Cleaning method

i. Apply mEMD to raw EEG data of N electrodes, in order to obtain M oscillatory modes of the multivariate data.

ii. Construct a matrix containing the same mode for all the channels. Therefore a total of M matrices are obtained.

iii. Calculate the Correlation Matrix (CM) of each one of these previous matrices, obtaining $CM \in \mathbb{R}^{N \cdot N \cdot M}$, by computing the correlation between all IMF with the same number.

iv. Compute the Communality Index (CI), containing the mean correlation of each mode for all the sensors ($CI \in \mathbb{R}^{M}$). The CI is computed using the following expression:
$CI = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |CM|$

v. Normalize CI between 0 and 1.

vi. Threshold CI in order to find which of these modes are common within all the channels. Modes with high correlation ($|r| > T$) are eliminated.

vii. Reconstruct clean signals without taking into account the eliminated modes.

3.5 Methods for EEG data simulation

This section presents the methodology used to simulate EEG data. In order to quantify the improvement that the cleaning method causes on the data, two types of data are

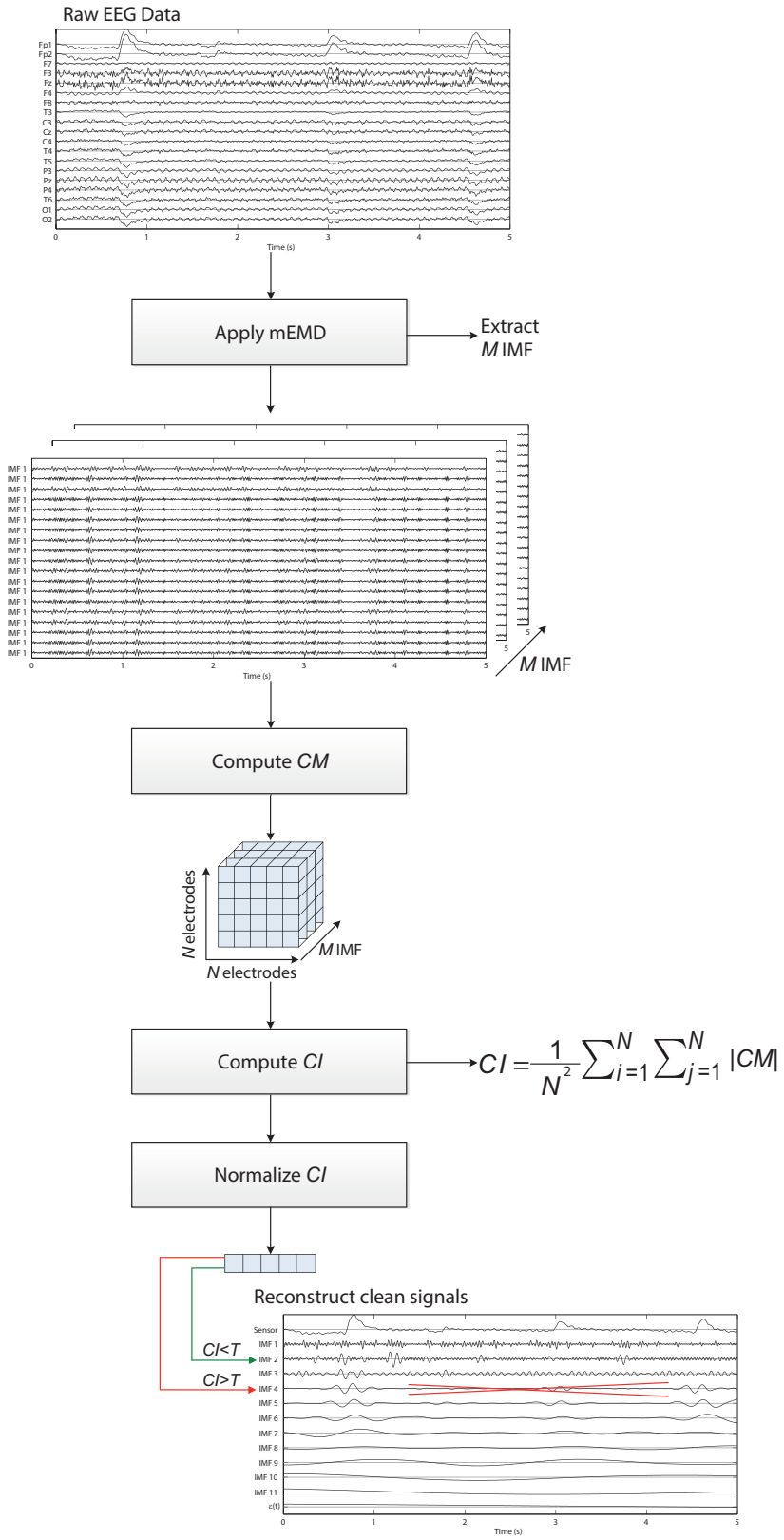


Figure 3.7: Example of reconstruction of a signal using Algorithm 3.3. After the computation of mEMD the correlation between each IMF of the same number is computed, obtaining CM. Then CI is computed and normalized. Values obtained at CI are used to select if an IMF is used in the reconstruction of the signal or not.

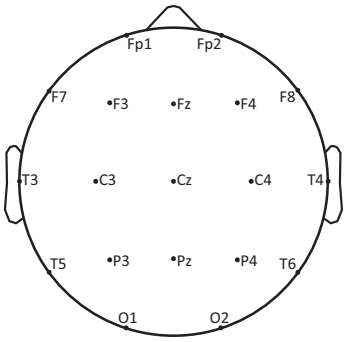


Figure 3.8: Position of the 19 electrodes used to simulate EEG data.

simulated. The first type is EEG data with eye blinks artifacts and different levels of noise. The second type is the same EEG data but without artifacts and noise.

Data simulation is done as it follows. It is known that the electric fields recorded using EEG follow this expression (Pascual-Marqui, 2002):

$$\phi = KJ + \eta(t) \tag{3.2}$$

where ϕ is a matrix containing scalp electric potentials measured at N electrodes each time interval $\frac{1}{T_s}$, J is the current density at each of the groups of neurons, also named voxels. $\eta(t)$ is white noise, uncorrelated with ϕ, generated by the electronic elements of the recording system. Finally K is a matrix containing the distances between each of the electrodes and each of the voxels. K matrix can be created with the software *low resolution brain electromagnetic tomography* (LORETA) (Pascual-Marqui, 2002).

LORETA software uses a realistic head model to create the K matrix. In this model, human cortex is divided into 2394 areas of $7mm^3$ cubic volume called voxels (Pascual-Marqui, 2002). K matrix presents the relation between each of the electrodes and each of the voxels.

For the data simulation, first a K matrix is created with the preferred number of electrodes. Then electric activity J is modeled. To simulate the electrical activity resulting from each voxel, a time series is created simulating the electrical activity of each dipole. In the present work, simulated time series are sine waves at a specific frequency. Next step consists in choosing a position for the dipole in the list of 2394 existent in the model. Then, by applying Equation 3.2 on the generated data, the simulated EEG is obtained. The last term of Equation 3.2, standing for $\eta(t)$, is simulated by adding random noise to obtain a realistic system. Adding random noise allows us to control the signal to noise ratio (SNR). In this work, different levels of SNR are tested in order to check the performance of the cleaning algorithm in different situations.

For this experiment, 10 seconds of EEG data containing 19 electrodes with a sampling frequency of 128 Hz are simulated. Different levels of SNR are tested, from -20 dB to +20 dB. Electrodes position are presented in Figure 3.8.

3.6 Results on simulated data

This section details the obtained results when applying the cleaning method on simulated EEG data. First the simulation of the data is described. Then methods used to compute

the improvement of the results are detailed. Section 3.6.3 shows the results for SNR of 0 dB. Finally in Section 3.6.4 results for all computed SNR are presented.

3.6.1 EEG data simulation

Two types of EEG data are simulated to check the improvement of the cleaning method:

- *Raw EEG data*: EEG time series containing eye blinks artifacts and different levels of white noise.
- *Clean EEG data*: EEG time series without artifacts and noise.

Equation 3.2 is used in both types of data for simulation. During data generation of both types, four dipoles are used to simulate electrical activity. For each dipole a sine wave in the traditional frequency bands is used (α, β, γ and μ). Location used for each dipole present a consistent position in the cortex according to physiological findings, as presented in Chapter 2.

In order to evaluate the improvement of the cleaning method presented in Algorithm 3.3, different levels of SNR are generated by adding noise with different amplitudes to the data. To study the statistical significance, different trials are simulated with the same SNR. Consequently, for each of the levels of SNR defined, ten trials are simulated with four simulated sine wave in different locations. For each trial the white noise is simulated again. The positions of the dipoles are also different for each of the trials. However, dipoles presenting the activity of the same frequency band are located in the same area. The temporal time series used to simulate the electrical activity of the dipoles is presented in Figure 3.9. On the other hand, the area defined for each of the dipoles is presented in Figure 3.10.

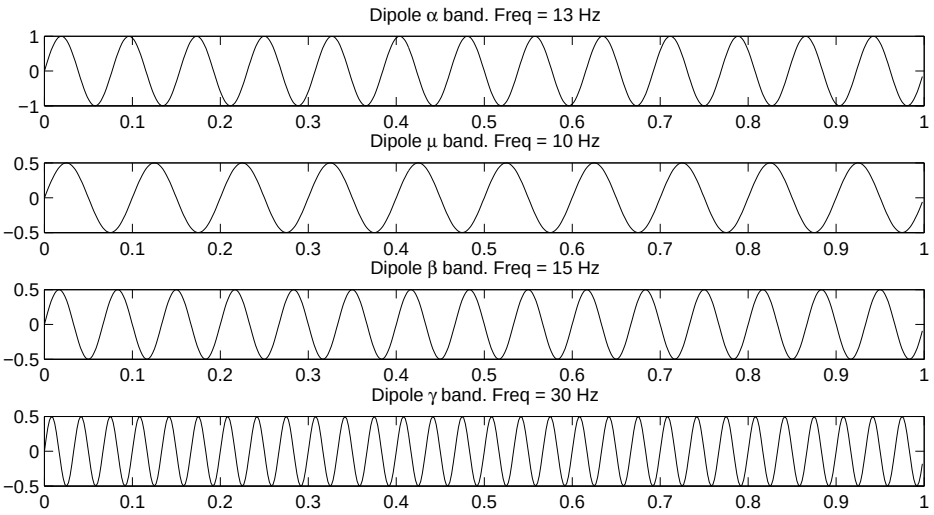
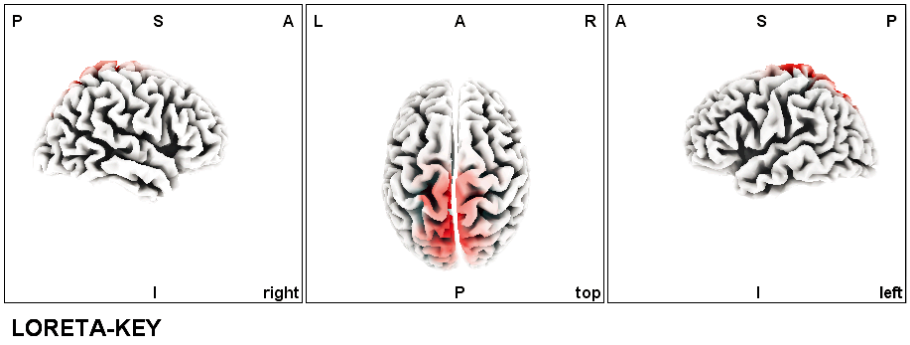
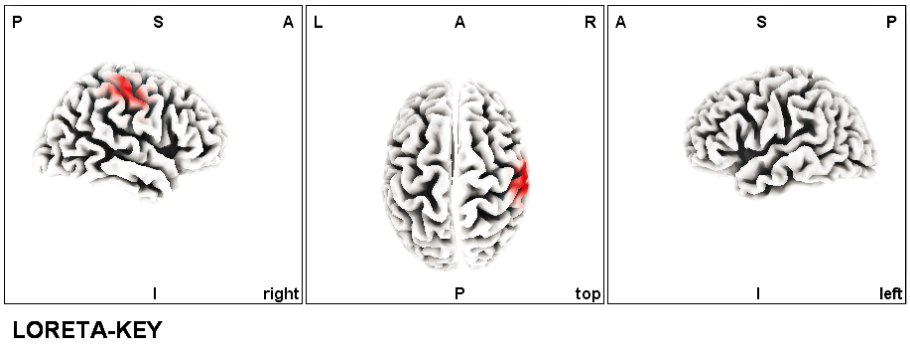


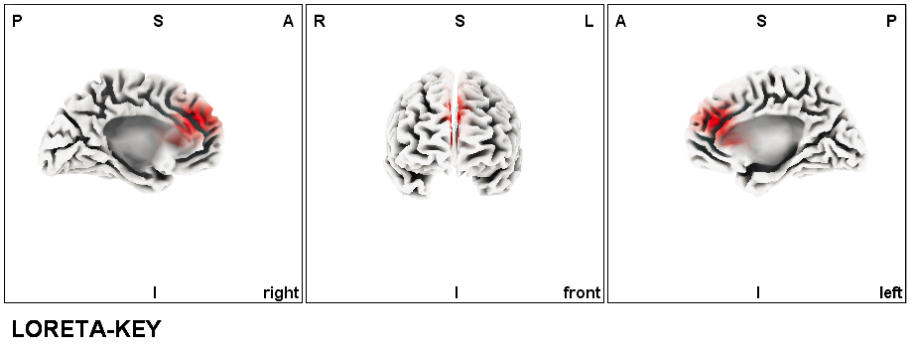
Figure 3.9: Time series used to simulate dipoles activity of EEG simulated data.



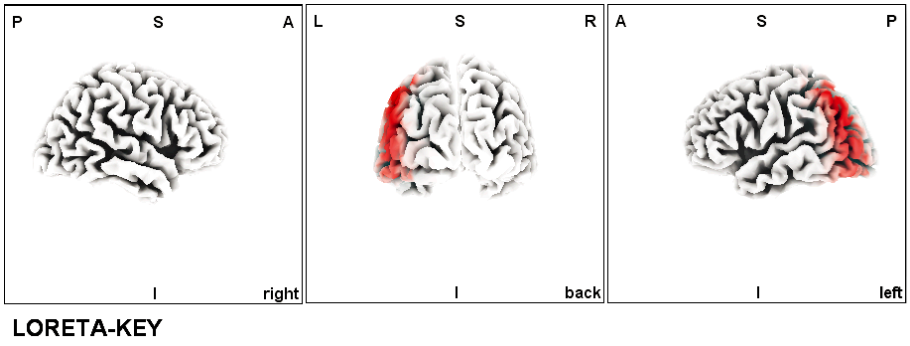
(a) Area defined for the dipoles used in the α band



(b) Area defined for the dipoles used in the μ band

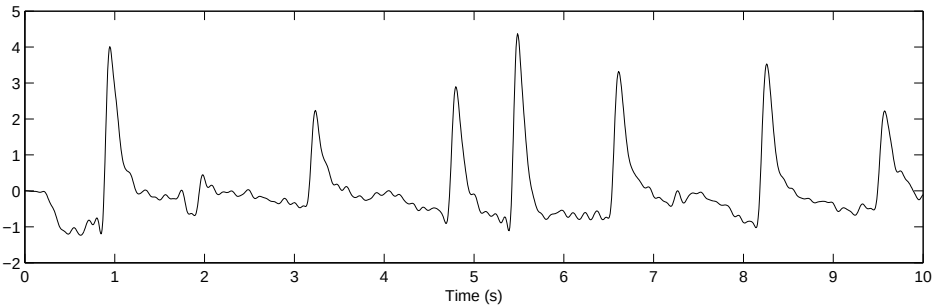


(c) Area defined for the dipoles used in the β band

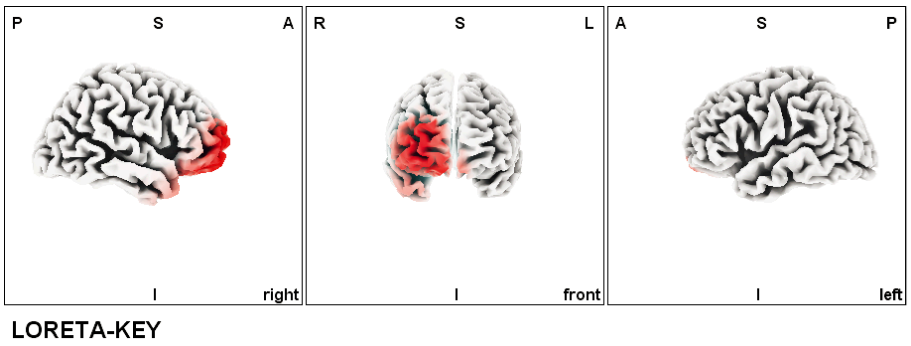


(d) Area defined for the dipoles used in the γ band

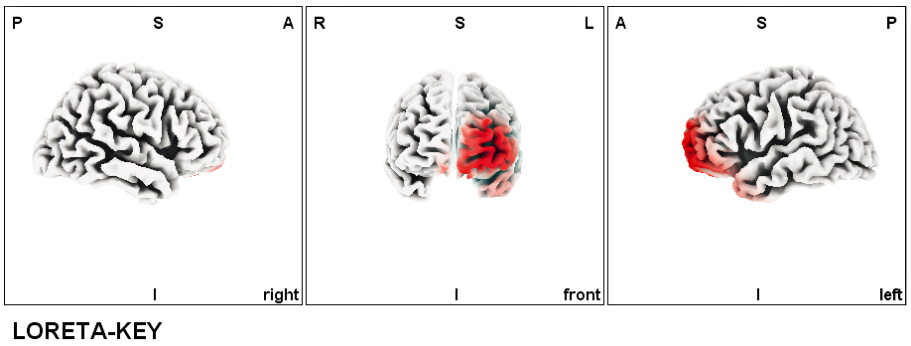
Figure 3.10: Cortex areas defined for dipoles position of EEG simulated data. In each area a dipole simulating electrical activity is simulated. L:left, R: right, A: anterior, P: posterior, S:superior, I:inferior.



(a) 10 seconds period of the time series used to create the eye blinks. Extracted from a real EEG recording using BSS.



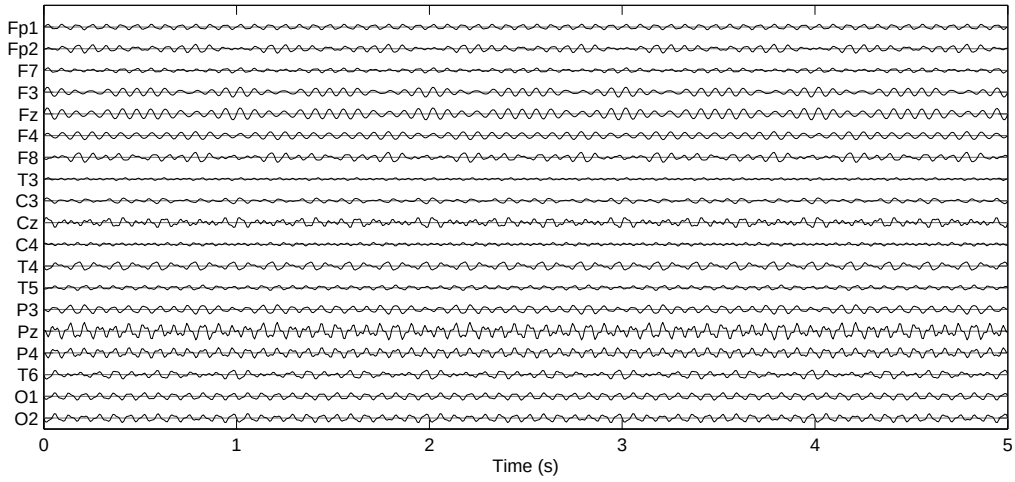
(b) Area defined for the dipoles used for the right eye



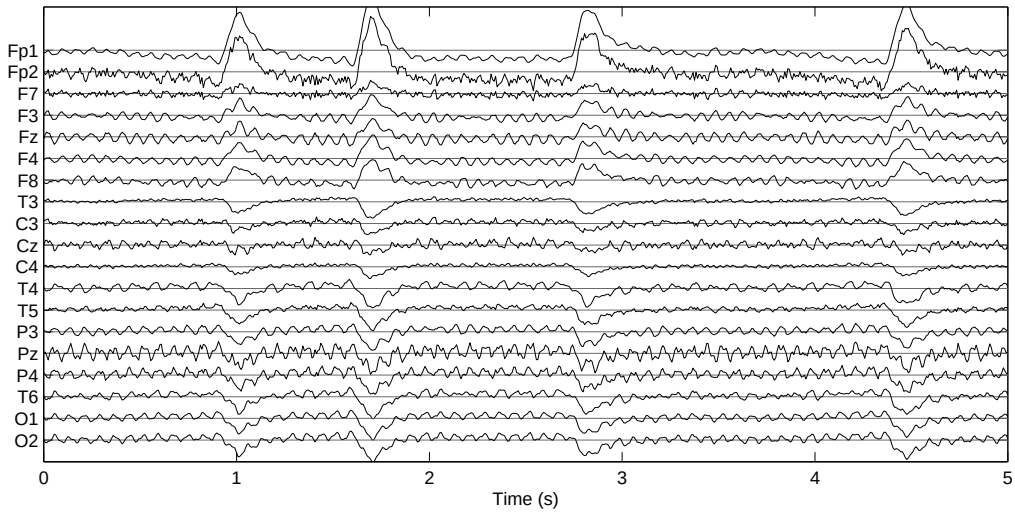
(c) Area defined for the dipoles used for the left eye

Figure 3.11: Cortex areas and source of dipoles activity defined for simulated eye blinks. L:left, R: right, A: anterior, P: posterior, S:superior, I:inferior.

Furthermore, simulated raw EEG data present eye blinks that affect the recording. The time series corresponding to eye blinks is not simulated, a real one is used. This procedure is done because a realistic aspect is aimed to be present on the EEG simulated signals. In this case, a real time series containing eye blinks was extracted from a recording. BSS was used to extract an eye blink component of a real EEG recording. An only component was extracted even though different trials are simulated for each level of SNR. However, as the length of the extracted eye blink is long, only a part of the recording is used. In order to have eye blinks in different positions of the recording, the extracted period is selected randomly for each trial. Ten seconds of the time series extracted as an eye blink are presented in Figure 3.11(a). The position of the eye blinks are also different for each trial, as is done with the other positions, an area is selected to locate the eye blinks. Selected



(a) Simulated clean EEG Data



(b) Simulated raw EEG Data

Figure 3.12: Example of simulated EEG time series used in the study. Results presented for the SNR = 0 dB

area for both eyes is shown in Figures 3.11(b) and 3.11(c), presenting the area for right and left eye respectively.

Used methodology present several limitations that must be highlighted. In real situations electrical activity due to eye blinks is not present in the cortex. This activity is present in the muscles above the skull. However, due to the model used, where only the relationship between electrodes and the cortex is defined, this electrical activity has to be approached and positioned in the cortex. Another point to be emphasized is that, even though two different spatial positions have been selected for the eye blinks, only a time series is used. This fact is due to movement of the eye blinks is the same for both eyes, and there is no delay between them.

Simulation of clean EEG data is done with the dipole activities presented in Figure 3.9, with the spatial locations defied in Figure 3.10. An example of EEG clean data is presented

in Figure 3.12(a). On the other hand, simulation of raw EEG data is done with the dipole activities presented in Figure 3.9 with the spatial locations defied in Figure 3.10. Eye blinks presented in Figure 3.11(a) are located in the areas defined in Figures 3.11(b) and 3.11(c). White noise is added to the generated signals with different levels of SNR. An example of simulated raw EEG data is shown in Figure 3.12(b). Signals presented in Figure 3.12(b) present a SNR equal to 0 dB.

3.6.2 Methods for EEG preprocessing using mEMD

Parameter T has been presented as a key choice in Algorithm 3.3, because the final quality of the data will depend on this parameter. Therefore, in order to measure the amount of improvement introduced by the cleaning method, correlation is used to compute the similarity between the time series of the raw EEG data and the clean EEG data. Correlation between the preprocessed data (cleaned EEG data) and the original clean EEG data is also computed. Then these two correlations are compared.

For two time series $x(n)$ and $y(n)$, correlation coefficient is computed as (Kropotov, 2009):

$$r = \frac{1}{N}\sum_{n=1}^{N}\frac{(x(n)-\bar{x})}{\sigma_x}\frac{(y(n)-\bar{y})}{\sigma_y} \tag{3.3}$$

where N is the length of the signals, $\bar{x}$ and $\bar{y}$ are the mean averages of the times series $x(n)$ and $y(n)$, and σ_x and σ_y correspond the standard deviation of the time series $x(n)$ and $y(n)$.

Correlation measure allows to quantify the improvement that the cleaning method produces on the raw EEG data, making it more similar to the simulated clean EEG data. In order to facilitate the comparison between correlation values, the improvement value is defined in Equation 3.4.

$$E_T = \frac{\sum_{n=1}^{N}\hat{r}_n}{\sum_{n=1}^{N} r_n} \tag{3.4}$$

where E_T is the improvement value. Parameters $\hat{r}_n$ and r_n stand for the correlation between the simulated raw EEG data and simulated clean EEG data (r_n) and correlation between cleaned EEG data and simulated clean EEG data ($\hat{r}_n$) computed at each electrode. Parameter E_T is then computed by averaging the obtained correlation at each electrode. As the term $\hat{r}_n$ is defined in the numerator, values of E_T higher than 1 present an improvement of the data during the cleaning process. On the other hand, values of E_T smaller than 1 means that some information has been lost.

3.6.3 Preprocessing results for SNR of 0 dB

This section exposes the results obtained after applying the presented cleaning method on data with a SNR = 0 dB. Once the decomposition mEMD is applied, the CI is computed. An example of CI is presented in Figure 3.13. For parameter T different values are tested (0.95, 0.90, 0.85, 0.80, 0.75 and 0.70). These values are presented in different colors in Figure 3.13. Signals after applying the algorithm are reconstructed with the IMFs that present lower value than T. Therefore the choice of this parameter defines if the algorithm is more or less strict with the correlated IMFs.

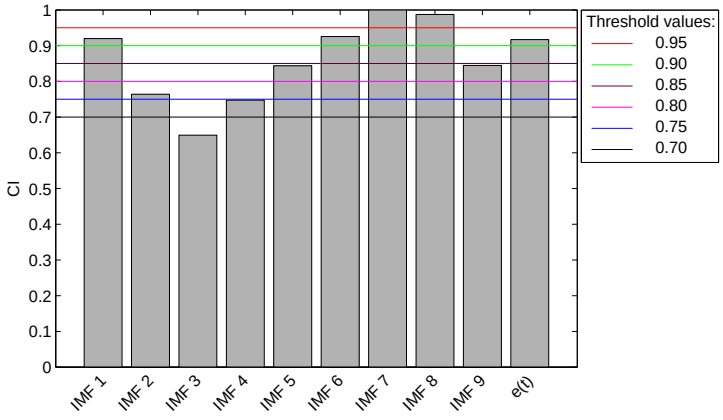


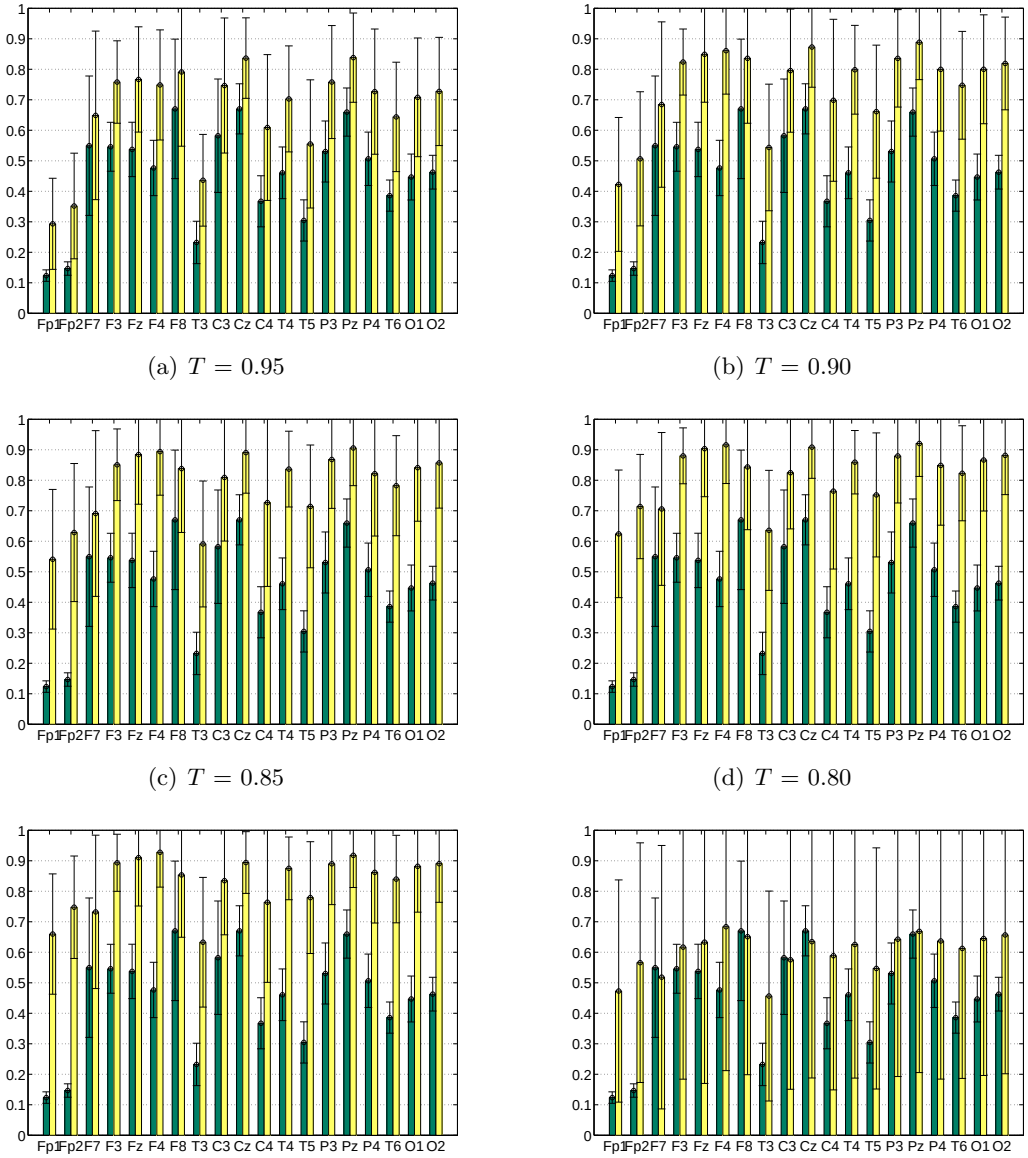
Figure 3.13: CI obtained after applying the decomposition mEMD to a 10 s of simulated raw EEG data. Different threshold values are presented.

Figure 3.13 shows that depending on the value of T, the number of IMFs used for reconstruction change. For example for a high value of T, e.g. T=0.95, only IMF7 and IMF8 are eliminated during the reconstruction process. On the other hand for a low value of T, e.g. T=0.70, all IMF are eliminated except IMF 3.

Figure 3.14 presents the correlation results obtained for each electrode with the different values of T used. In this figure, green bars stand for r_n and yellow bars present $\hat{r}_n$. In both cases, the mean value and the standard deviation computed along all the trials is computed. For all figures presented, it can be clearly seen that eye blinks and noise disturb EEG data in such a way that r_n is very low in all the electrodes. Especially frontal electrodes Fp1 and Fp2 present a low correlation ($r < 0.15$), as they are close to eyes. Using the cleaning methods to improve the quality of the data, allows us to recover an approximation of clean EEG data, and this can be observed in $\hat{r}_n$ values. Initial correlation of data with artifacts is improved with cleaned EEG data. Depending on the value of T used, the improvement changes. For $T = 0.95$ there is a small improvement. Values obtained by $\hat{r}_n$ increase as the T value decrease, until the value $T = 0.70$, where the mean obtained values for $\hat{r}_n$ clearly decrease.

In order to facilitate the comparison between the improvement compared to the values of the threshold, Figure 3.15 presents the improvement E_T obtained for each threshold value. Mean values and standard deviation obtained along all simulated trials are shown. These results show that depending on the T used, there is a higher improvement or not. Initial threshold $T = 0.95$, present a mean improvement of 1.45, then when the value of T decreases the cleaned EEG data is more similar to the original simulated clean data, as present the higher values obtained for E_T. The maximum value obtained for E_T is 1.82, obtained with $T = 0.75$. Then when $T = 0.70$ the value of E_T decreases, presenting that too much information have been eliminated, and therefore there is a decrease in the quality of the data.

The proposed cleaning method presents a clear improvement on the simulated data. Results presented in Figures 3.14 and 3.15 show this improvement by comparing the simulated raw EEG data and the cleaned EEG data, both being correlated with the original simulated clean EEG data. On the other hand, the parameter T defined in the Algorithm 3.3 is a key factor to obtain the best data cleaning. This section presents only the results when



(a) $T = 0.95$ (b) $T = 0.90$

(c) $T = 0.85$ (d) $T = 0.80$

(e) $T = 0.75$ (f) $T = 0.70$

Figure 3.14: Correlation comparison for all electrodes with different values of T. Green bars stand for the correlation between raw EEG data and clean EEG data (r_n). Yellow bars present the correlation between cleaned EEG data with the cleaning method and simulated clean EEG data ($\hat{r}_n$).

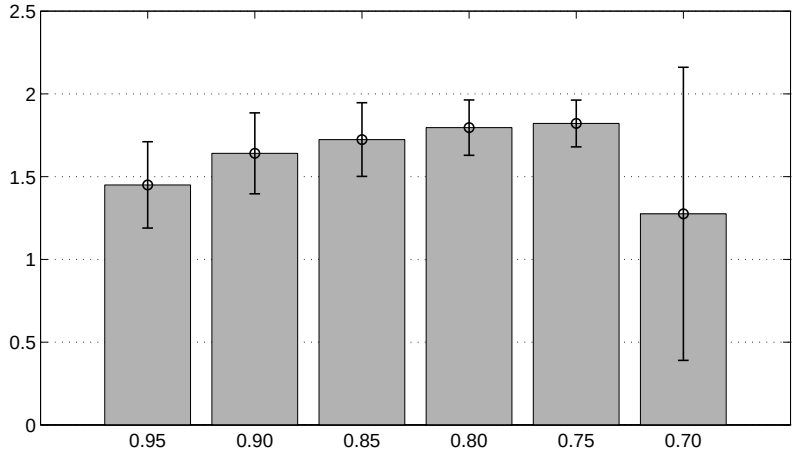


Figure 3.15: Impromement E_T presented in the simulated EEG for each of the different threshold values used.

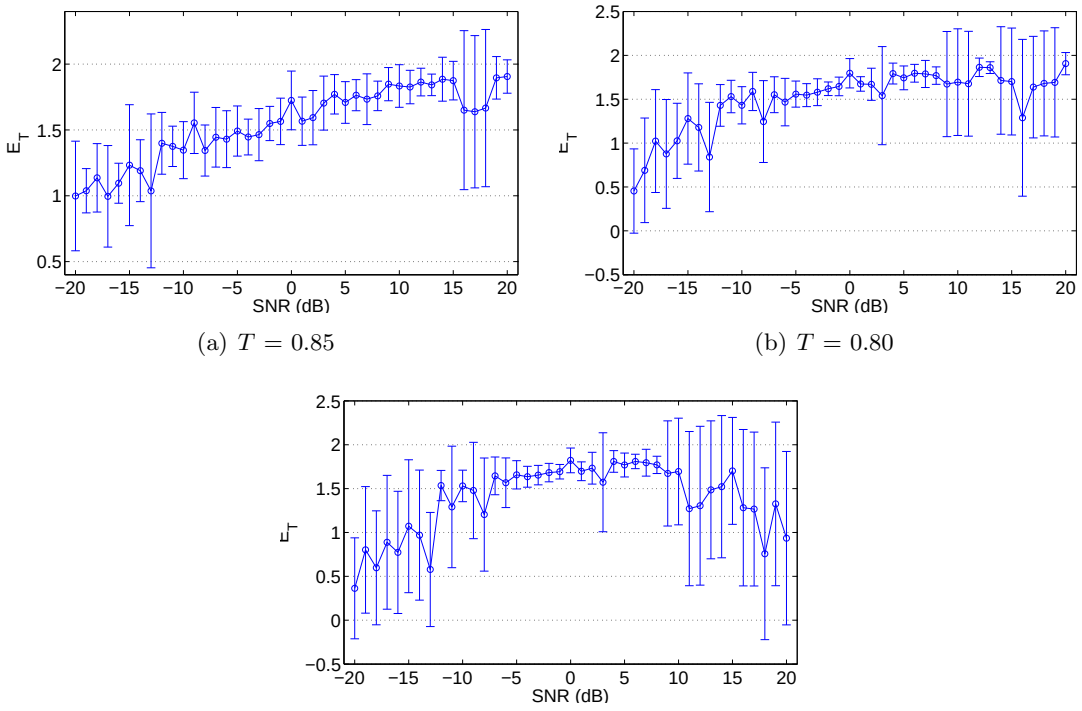
the simulated data present a SNR equal to 0 dB. Next section presents the same computations but using different values of SNR. However, only results for threshold $T = \{0.85$, 0.80 and $0.75\}$ are presented, being the values where the best improvement is found.

3.6.4 Preporcessing results for all SNR

This section presents the results for different SNR on the simulated data. The same procedure to simulate the signals with SNR equal to 0 dB is used in this case. However, values of SNR vary from -20 dB to +20 dB.

Figure 3.16 shows the evolution of the parameter E_T, presenting the mean and the standard deviation obtained. Different threshold values are presented, $T = 0.85$ (in Figure 3.16(a)), $T = 0.80$ (in Figure 3.16(b)) and $T = 0.75$ (in Figure 3.16(c)). Only three thresholds are used because these are the ones that present the best results in Figure 3.15, when the cleaning method is used with data presenting a SNR equal to 0 dB.

Results presented in Figure 3.16 show that when the SNR is low, i.e. -20 dB, there is no improvement of the quality of the data, for $T = 0.85$ the value of E_T is equal to 1, and the other values are lower than 1, presenting that the cleaning method do not improve the quality of the data but eliminates too much information. The obtained value for E_T improves as the value of SNR increases. However, for high SNR there is a decrease of E_T. Depending on the threshold used, this decrease is in different values. For $T = 0.85$ there is a small decrease of E_T when the SNR is equal to 16 dB, 17 dB and 18 dB. Values of E_T obtained with this SNR present a high variance showing that there is variability in the results. Results for $T = 0.80$ present one small decrease at SNR of 3 dB, and for higher values than 3 dB E_T stills increasing its value. However, when the SNR is equal to 9 dB, 10 dB and 11 dB, computed E_T presents a high variance, presenting that the cleaning method eliminated too much information from the simulated EEG data in some trials. For this same value of T, when the SNR is bigger than 13 dB, the obtained values of E_T are not stable. Last value of T used is 0.75. For this value the decrease of E_T is produced at SNR = 9 dB. Values higher than this SNR value show that there is a high variance in the result and the value of E_T is not stable.



(a) $T = 0.85$

(b) $T = 0.80$

(c) $T = 0.75$

Figure 3.16: Evolution of parameter E_T in function of SNR.

In order to facilitate the comparison between values of T, the mean values of E_T are presented together in Figure 3.17. Only mean values are presented without standard deviation for clarity of the figure. Figure 3.17 shows that the values obtained for the different thresholds are not too much different when the SNR is close to 0 dB, confirming the results presented in Section 3.6.3. However, for extreme values such as -20 dB and 20 dB, there is a clear decrease for threshold values $T = 0.80$ and $T = 0.75$, whereas for $T = 0.85$ the value increases. With the presented results at hand, the best possible threshold value to be selected is $T = 0.85$ for several reasons. The first is that when using $T = 0.85$ the value obtained for E_T is always equal or higher than 1, presenting that never there is a loss of information of the cleaned signal but there is always an improvement. The second reason is that, even though the obtained value of E_T is the lowest for the three thresholds when SNR is equal to 0 dB, when using extreme values of SNR the value obtained for E_T is the highest. The last reason is that when $T = 0.85$ is used in Figure 3.16, its figure is the one that presents the lower values of variance for the three thresholds, meaning that is the one with the lower variability between the simulated trials.

Aiming to test the performance of the cleaning algorithm in different scenarios, one last experiment is tested. In this experiment, two types of EEG data are simulated with different levels of SNR. One data type is the raw EEG data used in previous analysis, therefore it contains eye blinks and noise. The other type of data contains only noise with different levels of SNR. Again the improvement parameter E_T is computed comparing the correlation between the cleaned EEG data and the raw EEG data against the simulated EEG clean data. Results presenting the evolution of E_T for the data with artifacts and without artifacts with a threshold value of $T = 0.85$ are presented in Figure 3.18.

Results presented in Figure 3.18 show that whereas the improvement of E_T of data with

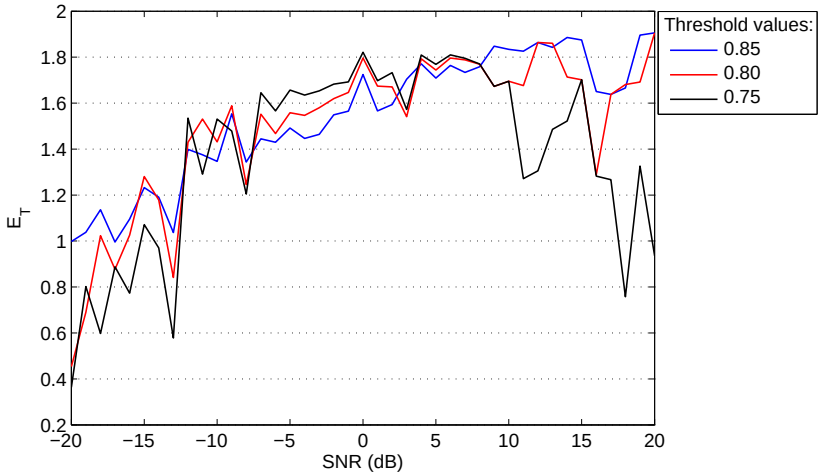


Figure 3.17: Mean improvement E_T presented in the simulated EEG for each of the different threshold values used.

artifacts (blue in Figure 3.18) present an improvement as the level of SNR increases, E_T computed for signals only containing white noise (red in the figure) present an important decrease for SNR higher than 4 dB. Values E_T computed for positives SNR values, show values close to 1 and lower. This it means that the final correlation of the reconstructed data is lower than the correlation of the raw data, presenting that an important part of the data has been eliminated during the cleaning process. On the other hand, results presented for signals with artifacts show that there is always an improvement of E_T. These results suggest that the cleaning method starts eliminating the noise when the SNR is low. But when the SNR is high, starts to eliminate the introduced artifacts presenting a similitude with the cleaned EEG data 1.6 times better than the simulated raw EEG signals.

Presented results suggest that if the data introduced to the cleaning method present a high level of interferences, being high level of noise or artifacts, the cleaning method generates an improvement of the data. However, if the data we want to clean present a low level of interference the cleaning results are not as good as before, because some important information may be lost during the cleaning, depending on the value of SNR. Therefore, if EEG data do not present a high level of noise or artifacts is better not to use the cleaning method, in order to conserve all the data.

3.7 Other applications of Multivariate Empirical Mode Decomposition

Presented methods for EEG denoising have been published in conference proceedings by the author of this documents in the publications (Gallego-Jutglà *et al.*, 2011) and (Gallego-Jutglà *et al.*, 2012b). However, the use of the mEMD has been tested in other domains.

The same procedure for denoising signals was used in (Solé-Casals *et al.*, 2013), where the cleaning method was used to improve the speech signals of two different speakers. Data simulation was done by mixing the speech recording of two subjects and random noise with different levels of SNR. Results obtained are similar to the ones obtained in this

Figure 3.18: Mean improvement E_T presented in the simulated EEG data with and without artifacts. Results presented for $T = 0.85$.

chapter, where a high improvement was obtained for small SNR, and low improvement was obtained for high SNR.

On the other hand, mEMD was also used for image processing. These results were published in conference proceedings (Gallego-Jutglà & Solé-Casals, 2012) and (Gallego-Jutglà *et al.*, 2013). In these articles mEMD is used to decompose images containing faces of different subjects. Then these decompositions are used to classify the subjects. The used data base contains 40 subjects with 10 images for each one, presenting the subject in different positions. The method used in these articles is summarized in Figure 3.19. The implemented classification system using mEMD it is summarized as it follows:

i. Five images are kept as representative for each class. The mean image of these five images is obtained for each class. These images are named as $R_i \forall 1 \leq i \leq N$, where N is the total number of classes. Remaining classes are kept for classification.

ii. For each new input image I to be classified, mEMD between I and all R_i is calculated, obtaining a total of N mEMD decompositions D_i.

iii. Then the distance between IMFs is calculated for each D_i, obtaining a vector of N values corresponding to the distances between input image I and each one of the classes.

iv. The input image I is then associated to a class as a function of some criteria on the distance.

In the presented works, different measures are used to compute the distances defined in step (iii). These measures are Correlation, the normalized Frobenius inner product and the Frobenius norm of the difference. The class association defined in step (iv) is done by two criteria as defined in Figure 3.19. Criteria defined in (Gallego-Jutglà & Solé-Casals, 2012) is to associate the new image I to the class corresponding to the minimum distance. Criteria defined in (Gallego-Jutglà *et al.*, 2013) is to associate the new image I using Artificial Neural Network (ANN). By applying the minimum distance criteria and

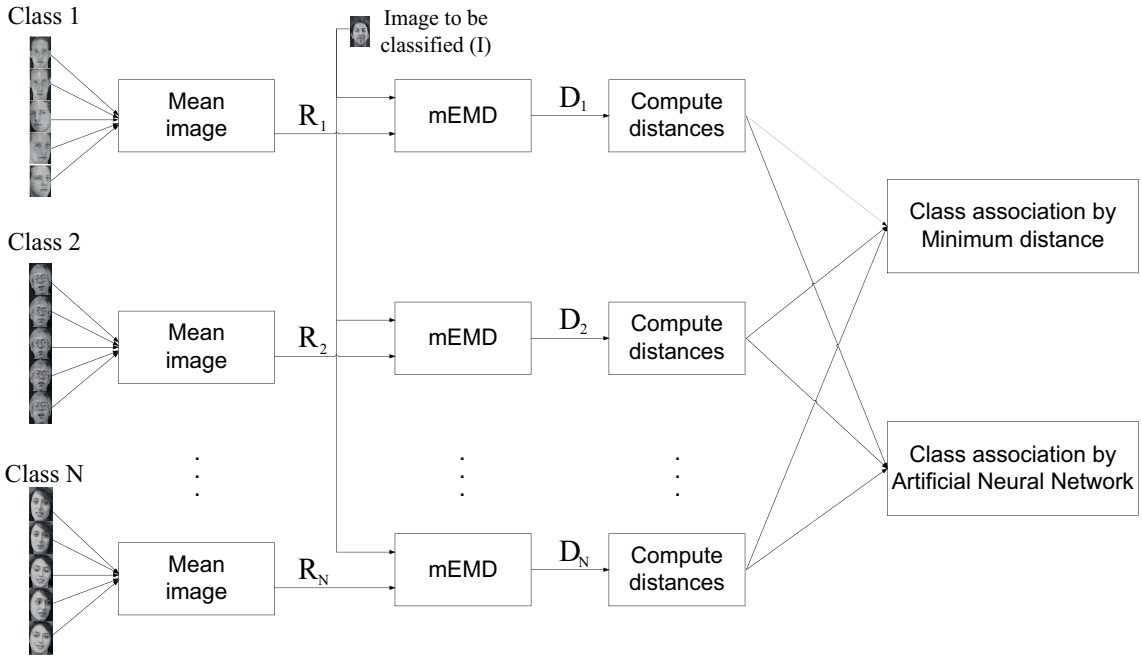


Figure 3.19: Scheme of the methods used for image processing using mEMD.

resizing the image to a smaller resolution (17 x 20 pixels in comparison with the original resolution of 92 x 112), obtained results presented a performance of classification of 82.50%. Results presented in (Gallego-Jutglà *et al.*, 2013) using a resolution of 29 x 34, present a comparison between the minimum distance criteria and the use of ANN. With the new resolution a 90% of classification performance it was obtained, when the class association based on minimum distance was used, and a 97.25% when an ANN was used. By using the three measures together as input features to an ANN, the classification results improved to 98.25%, equaling the best classification performance obtained with the same data set, but using a more simple approach.

3.8 Discussion

This chapter presents the use of the new decomposition technique mEMD to improve the quality of EEG signals. The cleaning method presents a clear improvement of the quality of the simulated data. EEG simulated data with artifacts is more similar to the original clean data once the cleaning method has been applied. The application of the cleaning method with different levels of SNR present that the selected threshold $T = 0.85$ present an improvement in all the ranges of SNR tested.

On the other hand, the usage of mEMD it has been presented as useful. These new technique for decomposition it is defined to work with multichannel data sets, therefore it explores the join information of the data. This join information helps in the identification of the information which is more present all along the data set, making easy its elimination.

Even though results on simulated data have been presented as promising, this method has not been tested on real data. As defined in Section 1.1, several steps are defined in neuroscience. Presented method in this chapter is only the first step. Chapter 8 presents

all the steps of the signal processing approach together. Therefore, in that chapter the presented cleaning method is used in real data.

A shortcoming that have been presented in this chapter is that, if the data used contains or artifacts or noise, the cleaning method clearly improve the quality of the data. However, if the data does not present artifacts or have a high SNR, some parts of the data may be erased. This is because during the cleaning, CI is normalized between 0 and 1, and then the threshold is used to select the eliminated modes. Therefore, even though before the normalization a data set present small values of CI some modes will be deleted during the reconstruction process. This opens a line for the future work where another step must be included in the cleaning method, to select if the computed values of CI are high enough to compute the normalization and to suppress some of them, or they are really low, meaning that there is no information to eliminate.

Chapter 4

Early diagnosis of AD by frequency selection

A new signal processing method for the early diagnosis of AD is introduced in this chapter. Different measures are used as feature for classification, helping in the process of distinguish between healthy subjects and AD patients. These measures are computed in different frequency ranges aiming to find a frequency range that facilitates the classification of healthy subjects and AD patients. This methodology is tested instead of using the standard frequency bands.

This chapter is organized as it follows. An introduction of the study is presented in Section 4.1. Section 4.2 introduces the three data sets used, one containing MCI patients and healthy subjects, and two containing Mild AD patients and healthy subjects. Section 4.3 explains the measures used to characterize subjects of the data bases, and then the methods used to apply these measures are described in Section 4.4. Section 4.5 shows the results of applying the presented methods on the three data sets. Obtained results are finally discussed in Section 4.6.

4.1 Introduction

An EEG recording is usually characterized by the presence of activity on specific frequency bands (δ, θ, α and β), as presented in Chapter 2. To distinguish between AD patients and healthy patients, studies traditionally have analyzed the standard frequency bands (Babiloni *et al.*, 2006; Koenig *et al.*, 2005), or extend the analysis to the entire frequency range between 4 and 30 Hz (Dauwels *et al.*, 2010a).

In this chapter a new frequency approach is presented in order to improve the early diagnosis of AD. Traditional EEG frequency bands have achieved good classification results. However, dividing the power spectra in predefined frequency ranges may not be optimal, because a more suitable frequency range may exists that facilitates how healthy subjects and AD patients are identified. The presented approach is based on working in the entire frequency range from 1 to 30 Hz, by analyzing all possible frequency ranges (e.g., 1 - 2 Hz, 1 - 3Hz, 1 - 4Hz, ... 29 - 30Hz).

On the other hand, aiming to characterize the effects that AD causes on EEG, a selection of measures has been done to parametrize this changes. In Dauwels *et al.* (2010a) a group of synchrony measures were used to distinguish between AD patients and healthy subjects. However, results presented in the same study suggest that a small group of synchrony measures may suffice to quantify the perturbations of EEG synchrony in AD patients. This fact is due to the high correlation observed between some synchrony measures. Therefore, ten of the measures defined in that work are used in the present study. Furthermore, a power measure is used to parametrize the effect of slowing of EEG. This measure is Relative Power (RP).

Even though three perturbations that AD causes on EEG data have been presented in Section 2.3, in this work only two of these effects are characterized. Results presented in Dauwels *et al.* (2011) demonstrate that RP and loss of complexity are strongly anti-correlated at low frequencies. As two of the main perturbations in EEG data are closely related, the present study investigates the early diagnosis of AD using the two other changes in EEG: slowing of EEG and changes in EEG synchrony.

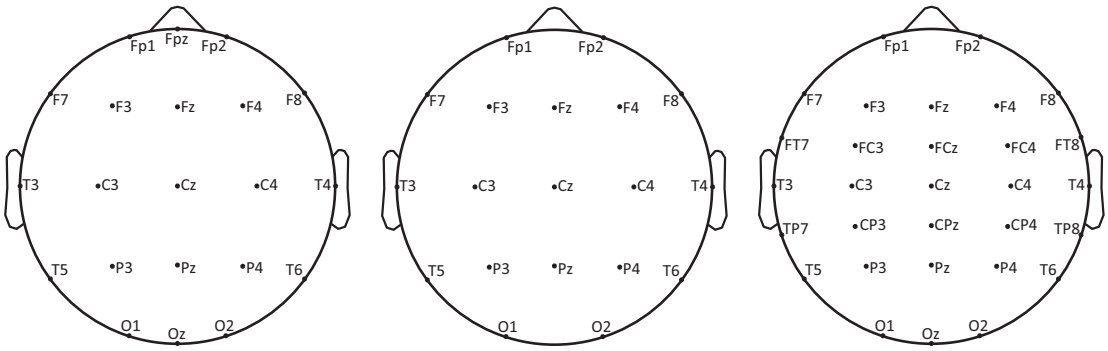
4.2 EEG data sets

In this study, we consider three different data sets. The first data set contains EEG recordings of MCI patients and healthy subjects, the second and the third contain EEG recordings of Mild AD patients. These data sets are following described.

4.2.1 The MCI Data Set: MCI patients and control subjects

The EEG data contained in this follow-up data set have been previously analyzed in a number of studies evaluating the early diagnosis of AD (Musha *et al.*, 2002; Cichocki *et al.*, 2005; Vialatte *et al.*, 2005a; Woon *et al.*, 2007; Dauwels *et al.*, 2009a, 2010a, 2011).

Patients who only complained of memory impairment were recruited. They underwent thorough neuropsychological testing that revealed a quantified and objective evidence of memory impairment with no apparent loss in either general cognitive, behavioral, or functional status. Patients did not suffer from other neurological diseases. The classification of very mild dementia impairment required a MMSE ≥ 24 and a Clinical Dementia Rating scale score of 0.5 with memory performance less than one standard deviation below the normal reference (Wechsler Logical Memory Scale and Paired Associates Learning subtests, IV and VII, ≤ 9, and/or ≤ 5 on the 30 min delayed recall of the Rey-Osterreith figure test). Fifty-three patients met these criteria. Each patient underwent Single-Photon Emission Computed Tomography (SPECT) at initial evaluation and was followed clinically for 12-18 months. Twenty-five of these fifty-three very mild AD patients developed probable or possible AD according to the criteria defined by the National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer's disease and Related Disorders Association. These subjects formed the group of patients of the MCI Data Set (age: 71.9 ± 10.2 years old), while 56 age-matched healthy subjects constituted the control group (age: 71.7 ± 8.3 years old). EEG recordings were conducted at the MCI stage. The control group consisted of healthy subjects who had no memory or other cognitive impairments. The scores of the MMSE were 28.5 ± 1.6 for the control group and 26 ± 1.8 for the MCI patients.



(a) MCI Data Set (b) Mild AD Data Set 1 (c) Mild AD Data Set 2

Figure 4.1: Electrodes positions used in each data set. (a) MCI Data Set contains 21 electrodes. (b) Mild AD Data Set 1 contains 19 electrodes and (c) Mild AD Data Set 2 contains 30 electrodes.

The EEG time series were recorded using 21 electrodes, positioned according to the 10-20 international system, with the reference electrode on the right earlobe. Positions of the 21 electrodes are presented in Figure 4.1(a). EEG was recorded with Biotop 6R12 (NEC Sanei, Tokyo, Japan) at a sampling rate of 200Hz, with analog bandpass filtering in the frequency range 0.5 - 250 Hz and online digital bandpass filtering between 1 and 30 Hz, using a third-order Butterworth filter (forward and reverse filtering).

4.2.2 The Mild AD Data Set 1: Mild AD patients and healthy subjects

The Mild AD Data Set 1 has also been analyzed previously (Goh *et al.*, 2006; Henderson *et al.*, 2006; Dauwels *et al.*, 2009a; Vialatte *et al.*, 2009c; Elgendi *et al.*, 2011; Gallego-Jutglà *et al.*, 2012a; Latchoumane *et al.*, 2012).

The Mild AD Data Set 1 consists of 24 healthy control subjects (age: 69.4 ± 11.5 years old) and 17 patients with Mild AD (age: 77.6 ± 10.0 years old). The patient group underwent a full battery of cognitive tests (MMSE, Rey Auditory Verbal Learning Test, Benton Visual Retention Test, and memory recall tests). The results from the psychometric tests were scored and interpreted by a psychologist, and all clinical and psychometric findings were discussed at a multidisciplinary team meeting. All age matched controls were healthy volunteers and had normal EEG (confirmed by a Consultant Clinical Neurophysiologist). The two groups are not perfectly age-matched, which might pose bias later on, but it was shown that no major effect was found due to this disparity (Henderson *et al.*, 2006). Patients did not suffer from other neurological diseases. The EEGs were recorded during a resting period containing various states: awake, drowsy, alert and resting with eyes closed and open. All recording sessions and experiments proceeded after informed consent was obtained of the subjects or the caregivers and were approved by local institutional ethics committees.

The EEG time series were recorded using 19 electrodes, with a reference electrode on each earlobe. Electrodes were positioned according to the Maudsley system, similar to the 10-20 international system. Positions of the 19 electrodes are presented in Figure 4.1(b). Signals were digitalized at a sampling frequency of 128 Hz. The EEGs were band-pass

	MCI Data Set		**Mild AD Data Set 1**		**Mild AD Data Set 2**	
	MCI patients	**Healthy subjects**	**Mild AD patients**	**Healthy subjects**	**Mild AD patients**	**Healthy subjects**
Subjects	22	38	17	24	28	22
Age	71.9 ± 10.2	71.7 ± 8.3	77.6 ± 10.0	69.4 ± 11.5	80.8 ± 10.5	68.9 ± 10.3
MMSE	26 ± 1.8	28.5 ± 1.6	Access not possible		19.5 ± 4.4	28.3 ± 1.6

Table 4.1: Summary of the characteristics of each data set used.

filtered with a digital third-order Butterworth filter (forward and reverse filtering) between 0.5 and 30 Hz.

4.2.3 The Mild AD Data Set 2: Mild AD patients and healthy subjects

The Mild AD Data Set 2 has not been previously analyzed. All subjects included in this data set are subjects that went to the hospital with memory complaints and suspected AD. All subjects underwent a full battery of neuropsychological tests and examinations (MMSE and others), and in addition brain imaging and blood sampling. Also they had an interview with a neurologist who observed clinical signs. The criterion used to include subjects in the Mild AD group or in the control group was based on the diagnostic made by the neurologist, who based his decision on all the available information. The control group had no diagnostic even though they suffered memory complaints. Finally the Mild AD group was formed by 32 patients (age: 80.8 ± 10.5 years old) and the control group was formed for 24 subjects (age: 68.9 ± 10.3 years old). The scores of the MMSE were 28.3 ± 1.6 for the control group and 19.5 ± 4.4 for the Mild AD patients.

The EEGs were recorded during a resting period containing various states: eyes open, eyes close and periods of stimulation with steady-state visual evoked potentials (SSVEP) at different frequencies, i.e. 1 Hz, 2 Hz, 3 Hz, 4 Hz, 5 Hz, 10 Hz, 15 Hz, 20 Hz and 30 Hz. For the present study only the eyes closed period was used. All recording sessions and experiments proceeded after informed consent was obtained of the subjects and the ethical review board of Hospital Charles Foix (Paris) validated the research project.

The EEG time series were recorded using 30 electrodes placed on the scalp according to the 10-20 international system. Positions of the 30 electrodes are presented in Figure 4.1(c). EEG was recorded with Deltamed amplifier at a sampling frequency of 256 Hz. The EEGs were band-pass filtered with a digital third-order Butterworth filter between 1 and 30 Hz.

4.2.4 Recording conditions common to all data sets

In all data sets, all recording sessions included in the analysis were conducted with the subjects in an awake but resting state. The length of the EEG recording was about 5 minutes per subject for the MCI Data Set and the Mild AD Data Set 1. For the Mild AD Data Set 2 the initial total length of the recording was 20 minutes. Approximately 5 minutes were extracted for each subject in resting state with the eyes closed. In all the recordings the EEG technicians prevented the subjects from falling asleep (vigilance

control). After recording, the EEG data was carefully inspected. As presented in Section 3.1, EEG recordings are prone to a various artifacts, for example, due to electronic noise, head movements, and muscular activity. For this study no pre-processing was done, and for each patient an EEG expert selected, by visual inspection, one 20 second segment of artifact-free EEG. Only subjects whose EEG recordings contained at least 20 seconds of artifact-free data were retained in the analysis. Specifically, for the Mild AD Data Set 2 the author of this manuscript was the responsible of all the preprocessing of the data, doing the visual inspection to select the 20 seconds of data.

As a result of this approach, the number of subjects of the MCI data set was further reduced to 22 MCI patients and 38 control subjects and the number of subjects of the Mild AD Data Set 2 was reduced to 28 Mild AD patients and 22 healthy subjects. For the Mild AD Data Set 1 no reduction was made. For the subjects eventually included in each data set, statistical differences between the ages of the patients were checked. For none of these three data sets the age difference between AD patients and control subjects is significant (t-test $p >> 0.05$). The final characteristics for each data set are summarized in Table 4.1.

4.3 EEG Measures

This study considers two types of EEG measures: a power measure and synchrony measures. As a power measure RP is selected. For synchrony, various measures are used: Correlation, Coherence, Granger Causality, OC and PS. A description of these measures is presented below.

The spectrum of EEG is helpful in describing and understanding brain activity, as presented in Section 2.3.1. The well-known effect of slowing of EEG can be easily modeled by the measure RP. RP allows us to compare the difference of power between subjects in a specific frequency range (e.g. α or θ bands), which may facilitate to distinguish between AD patients and healthy subjects. On the other hand, the use synchrony measures help to model the changes of synchrony described in Section 2.3.2 for AD patients.

4.3.1 Relative Power

RP measures the percentage of power within a specific frequency range compared to the power of the entire frequency range. It is computed using the power spectrum of the signal. The spectrum can be computed by means of the Discrete Fourier Transform (DFT) of the signal. The DFT of a signal represents his spectrum as a finite combination of complex sine wave (Bloomfield, 2000). Computation of DFT can easily be performed in a computer using the Fast Fourier Transform (FFT) algorithm (Cooley & Tukey, 1965; Frigo & Johnson, 1998). The DFT ($X(f)$) of a signal $x(n)$ with a length N is defined as:

$$X(f) = \sum_{n=0}^{N-1} x(n) e^{\frac{-jkn}{N}} \tag{4.1}$$

Fourier coefficients defined in Equation 4.1 allows us to compute the power spectrum of a signal. The power spectrum $P_x(f)$ of the signal $x(n)$ is computed as (Proakis & Manolakis, 1996):

$$P_x(f) = X(f)X^*(f) = |X(f)|^2 \tag{4.2}$$

Using Equation 4.2, the power for a specific frequency range is computed as:

$$P_x(f_1, f_2) = \sum_{f=f_1}^{f_2} P_x(f) \tag{4.3}$$

where $[f_1, f_2]$ is the band of interest and $P_x(f)$ is the power spectrum of the EEG signal $x(n)$. Finally, RP is computed as:

$$RP_x(f_1, f_2) = \frac{P_x(f_1, f_2)}{P_x(f_{min}, f_{max})} \tag{4.4}$$

where $P_x(f_1, f_2)$ is the power in the frequency range of interest $[f_1, f_2]$ for signal $x(n)$ and $P_x(f_{min}, f_{max})$ is the power in the entire frequency range $[f_{min}, f_{max}]$ for the same signal.

4.3.2 Correlation Coefficient

Correlation coefficient (r) is a measure of similarity between two signals. Correlation coefficient expression has already been defined in Equation 3.3.

The correlation coefficient r quantifies the linear correlation between two time series. r varies between -1 and +1. So, if the two time series change coherently and in the same direction in time, the correlation coefficient between these two variables is positive and close to 1. If two variables change in out of phase mode the correlation coefficient between these variables is close to -1. If it does not exist relation between the two time series $r = 0$.

4.3.3 Coherence

Coherence estimates phase synchronization between two time series in the frequency domain. Coherence is computed using the cross-spectrum of the signals. The cross-spectrum ($P_{xy}(f)$) of the signals $x(n)$ and $y(n)$ is computed as (Nunez & Srinivasan, 2006):

$$P_{xy}(f) = X(f)Y^*(f) \tag{4.5}$$

were $X(f)$ and $Y(f)$ are the DFT of the signals $x(n)$ and $y(n)$ respectively, and $Y^*(f)$ denotes the complex conjugate of $Y(f)$. When $x(n) = y(n)$ the cross-spectrum reduces to the power spectrum, as computed in Equation 4.2. The cross-spectrum can be interpreted as a measure of the covariance between two signals at one specific frequency. Using the cross-spectrum, coherence is defined as (Nunez & Srinivasan, 2006):

$$c(f) = \frac{|P_{xy}(f)|^2}{P_x(f)P_y(f)} \tag{4.6}$$

Coherence is usually interpreted as an indicator of connectivity between two brain areas (Sanei & Chambers, 2009). It takes values from the range 0 to 1 ($c(f) \in [0, 1]$), were 0 indicates no relation. Equation 4.6 follows closely the form of correlation presented in Equation 3.3. The numerator is the squared modulus of the cross-spectrum, analogous to the squared covariance. In the denominator, the power spectrum is analogous to the variance o the signal. Thus Equation 4.6 is analogous to dividing squared covariance by the variance of each channel, which is a squared correlation coefficient. Coherence measures the fraction of variance of channel x at frequency f that can be explained by a constant linear transformation of the Fourier coefficients obtained at channel y.

4.3.4 Granger Causality measures

Granger Causality is a family of measures that attempts to extract and quantify the directionality from information. This family of measures has its roots in the economics literature. The method was first proposed in 1969 by Granger and now it is known as Granger Causality (Granger, 1969). Granger measures aims to explain the relationship between two events X and Y. If events X causes event Y, then knowledge of past values of X should improve our prediction of the current value of Y. The past values of X used in the prediction are known as the model order.

Granger measures are calculated trough the multivariate autoregressive (MVAR) model of the data. The MVAR model of multivariate time series is computed as it follows: suppose that n signals $x_1(n), x_2(n), ..., x_n(n)$ are given, each stemming from a different channel. Each of these times series is normalized, with zero mean and standard deviation equal to one. We consider the following MVAR model (Kaminski & Liang, 2005):

$$x(n) = \sum_{l=1}^{p} \mathbf{A}(j)x(n-l) + e(n) \tag{4.7}$$

where $x(n) = (x_1(n), x_2(n), ..., x_n(n))^T$, p is the model order, the model coefficients $A(j)$ are $n \times n$ matrices and $e(n)$ is a zero-mean Gaussian random vector of size n. Equation 4.7 can be rewritten as:

$$e(n) = \sum_{l=0}^{p} \tilde{\mathbf{A}}(j)x(n-l) \tag{4.8}$$

where $\tilde{\mathbf{A}}(0) = I$ (identity matrix) and $\tilde{\mathbf{A}}(j) = -\mathbf{A}(j)$ for $j > 0$. Equation 4.8 can be transformed to the frequency domain (by applying the Z transform and substituting $z = e^{-j2\pi f \Delta t}$) (Proakis & Manolakis, 1996):

$$\begin{aligned} E(f) &= \tilde{\mathbf{A}}(f)\mathbf{X}(f) \\ \mathbf{X}(f) &= \tilde{\mathbf{A}}^{-1}(f)\mathbf{E}(f) = \mathbf{H}(f)\mathbf{E}(f) \\ \mathbf{H}(f) &= \left(\sum_{j=0}^{p} \tilde{\mathbf{A}}(f) e^{(-j2\pi m f \Delta t)} \right)^{-1} \end{aligned} \tag{4.9}$$

where Δt is the sampling period T_s. The matrix $\mathbf{H}$ is called the transfer matrix of the system. It contains information about relations between data channels constituting the system.

The power spectrum matrix of the signal is then given by:

$$\mathbf{S}(f) = \mathbf{X}(f)\mathbf{X}^*(f) = \mathbf{H}(f)\mathbf{E}(f)\mathbf{E}^*(f)\mathbf{H}^*(f) \tag{4.10}$$

The terms $\mathbf{E}(f)\mathbf{E}^*(f)$ denotes the variance of the noise. If we define $\mathbf{V} = \mathbf{E}(f)\mathbf{E}^*(f)$ then Equation 4.10 can be described as:

$$\mathbf{S}(f) = \mathbf{H}(f)\mathbf{V}\mathbf{H}^*(f) \tag{4.11}$$

where $\mathbf{V}$ is not dependent on frequency. Once the MVAR of the input signals is computed, the calculation of the Granger measures is done using $\tilde{\mathbf{A}}, \mathbf{H}$ and $\mathbf{S}$ matrices. Description of the Granger measures is presented below.

It has to be noted that Granger measures is a family of synchrony measures that are derived from linear stochastic models of time series. Therefore, Granger measures are multivariate, and they can be applied to multiple signals simultaneously such as EEG. However, some authors suggested that the application of Granger measures in multichannel recordings it is not computationally efficient and computation may be done from pair-wise combinations of electrodes (Sanei & Chambers, 2009). By using this strategy some causal information is lost. Therefore, in this study Granger measures are used as a multivariate measure. However, an average between some specific electrodes is done in order to avoid big data estimation. Method used are later described in Section 4.4.

Granger Coherence

Granger Coherence (GC) describes the amount of in-phase components between two signals. It is a measure derived from the MVAR model defined in Equation 4.7. It is computed through the **S** matrix defined in Equation 4.11. The computation of GC is defined as (Kaminski & Liang, 2005):

$$K_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)}\sqrt{S_{jj}(f)}} \tag{4.12}$$

where S_{ij} are elements of the spectral matrix **S**. S_{ij} is the cross-spectrum between channels i and j, S_{ii} is the power spectrum for channel i. Likewise S_{jj} for channel j. The modulus of the GC takes values from the range 0 to 1 ($|K_{ij}| \in [0, 1]$), were 0 indicates no relation.

The squared magnitude $|K_{ij}|^2 \in [0, 1]$ is an alternative to the square magnitude coherence function $c(f)$ defined in Equation 4.6. The coherence function $|K_{ij}|^2$ may be more reliable than $c(f)$ if the MVAR system is a good model for the signals at hand.

Partial Coherence

Partial coherence (PC) describes the amount of in-phase components in signals i and j at the frequency f when the influence of the other signals (the part of the signals that can be explained by linear combination of the other channels) is statistically removed. PC can be written as (Kaminski & Liang, 2005):

$$C_{ij}(f) = \frac{M_{ij}(f)}{\sqrt{M_{ii}(f)}\sqrt{M_{jj}(f)}} \tag{4.13}$$

where $M_{ij}(f)$ is the determinant of **S** with row i and column j removed. The modulus of PC takes values from the range 0 to 1 ($|C_{ij}(f)| \in [0, 1]$), similar to ordinary coherence, but it is nonzero only when the relation between channel i and j is a direct influence.

The Granger measures presented in the next sections capture causal relations. Therefore they are asymmetric or directed, being the information computed between electrode i and j different than the information computed between electrode j and i.

Directed Transfer Function

The Directed Transfer Function (DTF) is a multichannel causality measure. It is computed as (Kaminski & Blinowska, 1991; Kaminski & Liang, 2005):

$$\gamma_{ij}^2(f) = \frac{|H_{ij}(f)|^2}{\sum_{j=1}^{m} |H_{ij}(f)|^2} \tag{4.14}$$

where $H_{ij}(f)$ is the relation between input j with output i at the frequency f defined by the $\mathbf{H}$ matrix. $\mathbf{H}$ is a not symmetric matrix, which it means that transmission from channel i to channel j is different from channel j to channel i ($H_{ij} \neq H_{ji}$). DTF is a measure that represents a ratio between the inflow to channel i from channel j and all the inflows to channel i. As there is a normalization value in the denominator, DTF takes values between 0 and 1 ($\gamma_{ij}(f) \in [0, 1]$). Values of DTF close to 0 present that there is no flow from channel j to channel i at that specific frequency, and values of DTF close to 1 indicate most of the signal in channel i consists of signal from channel j.

Full Frequency Directed Transfer Function

The definition of DTF presents several difficulties for the comparison of outflows of information at different frequencies. To this end, a modified version of DTF was presented by Korzeniewska *et al.* (2003). This modified version is full frequency Directed Transfer Function (ffDTF) and it is computed as (Korzeniewska *et al.*, 2003):

$$F_{ij}^2(f) = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{j=1}^{m} |H_{ij}(f)|^2} \tag{4.15}$$

The frequency normalization appearing in the denominator assures that this value does not change with frequency. Therefore values obtained with ffDTF are only dependents on the outflow of information between channels. ffDTF as a normalize extension of DTF takes values between 0 and 1 ($F_{ij}^2(f) \in [0, 1]$).

DTF and ffDTF indicate a causal relation between channels. However, transmission of information between channels i and j may not be direct and signal may go through several other channels. If it is the case, DTF and for extension ffDTF will show that the relation exists. In some cases it is important to know what relations are direct without any other channels involved. Aiming to describe this direct relation, some direct measures were presented in the literature, those are presented below.

Partial Directed Coherence

Partial Directed Coherence (PDC) is an estimator proposed to quantify direct causal connections. It describes the interaction between two time series, when the influence due to all other $N-2$ time series is discounted. PDC it is described as (Baccalá & Sameshima, 2001):

$$P_{ij}(f) = \frac{\tilde{A}_{ij}(f)}{\sqrt{\mathbf{a}_j^\dagger(f)\mathbf{a}_j(f)}} \tag{4.16}$$

were $\tilde{A}_{ij}(f)$ is the relation between the two channels i and j, $\mathbf{a}_j(f)$ is the j^{th} column of the matrix $\tilde{A}$, † represents transposition and complex conjugate operation. PDC takes values between 0 and 1 ($P_{ij}(f) \in [0, 1]$).

Direct Directed Transfer Function

The last direct and causal measure considered is the direct Directed Transfer Function (dDTF). dDTF aims to emphasize only direct connections between channels i and j. It is defined as the product between PC and DTF, which is represented as (Korzeniewska *et al.*, 2003):

$$\chi^2_{ij}(f) = F^2_{ij}(f)C^2_{ij}(f) \tag{4.17}$$

dDTF it combines information from partial coherence function with information about direction of influence in one measure. dDTF takes values between 0 and 1 ($\chi^2_{ij}(f) \in [0, 1]$). Therefore $\chi^2_{ij}(f)$ will be nonzero when:

- There is a causal connection from channel j to channel i.
- The relation between channels i and j is direct.

4.3.5 Omega Complexity

OC is a synchrony measure for multichannel data sets. It quantifies the amount of spatial synchronization in a group of multivariate time series (Bhattacharya, 2000). Synchrony is evaluated through the covariance matrix of the data set.

Given a data set of multivariate n signals $x_1(n), x_2(n), x_3(n), ...x_n(n)$, the covariance matrix C is first computed. C is decomposed using Principal Component Analysis (PCA) as (Wackermann, 1996):

$$C = W\Lambda W^T \tag{4.18}$$

where Λ is a diagonal matrix of eigenvalues $\lambda_{(n)}$ and W are the orthogonal eigenvectors. OC is then computed using the eigenvalues. To compute the contribution of each eigenvalue, these are normalized to unit sum applying:

$$\lambda'_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{4.19}$$

The value of OC is then calculated as (Saito *et al.*, 1998):

$$\Omega = \exp\left(-\sum_{i=1}^{n} \lambda'_i \log \lambda'_i\right) \tag{4.20}$$

The argument of the exponential in Equation 4.20 is the entropy of the distribution obtained with the eigenvalues. OC presents minimum value for identical signals ($\Omega = 1$). The maximum value is obtained for independent signals ($\Omega = n$).

4.3.6 Phase Synchrony

PS computes the synchronization between two time series $x(n)$ and $y(n)$. It measures only the phase between signals, even when the amplitudes of x and y are statistically independent. PS is computed using the instantaneous phase of the signals.

The analytic signal $\psi(n)$ of a signal $x(n)$ is defined as (Rosenblum *et al.*, 1996):

$$\psi(n) = x(n) + j\tilde{x}(n) \tag{4.21}$$

where $\tilde{x}$ is the Hilbert transform of x. The instantaneous phase (ϕ_x) is then defined as:

$$\phi_x(n) = \arg\left[x(n) + j\tilde{x}(n)\right] \tag{4.22}$$

Then, the PS index for the two instantaneous phases of signals x and y (respectively ϕ_x and ϕ_y) is defined as (Lachaux *et al.*, 1999):

$$\gamma = \left|\left\langle e^{j(k\phi_x - m\phi_y)}\right\rangle\right| \tag{4.23}$$

where k and m are integers (usually $k = m = 1$) (Le Van Quyen *et al.*, 2001b), and $\langle\cdot\rangle$ is time averaging.

4.4 Methods Used to Compute Measures

This section describes the methods used to compute presented measures in Section 4.3. As defined previously, a new methodology to improve AD diagnosis is presented. This methodology investigates all possible sets of frequency ranges between 1 and 30 Hz, with the aim to find if there is a specific frequency range that improves the early diagnosis of AD. Before the computation of each measure, signals are first filtered. Then, the measures are statistically analyzed and classified.

4.4.1 Bandpass filtering and computation of the measures

All the possible frequency ranges between 1 and 30 Hz are systematically analyzed. To define the frequency ranges of study, the start frequency band (F) varies from 1 to 29 Hz, and the width (W) varies from 1 to 29 (e.g. 1 - 2 Hz, 1 - 3 Hz, 1 - 4 Hz, ... 1 - 30 Hz, ... 29 - 30 Hz). The maximum frequency of analysis ($F + W$) is limited to 30 Hz. A total of 435 frequency ranges are studied, as is next detailed:

$$W \in \mathbb{N}[1, 29] \quad \begin{bmatrix} 1-2 & & & & & \\ 1-3 & 2-3 & & & & \\ 1-4 & 2-4 & 3-4 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ 1-29 & 2-29 & 3-29 & \dots & 28-29 & \\ 1-30 & 2-30 & 3-30 & \dots & 28-30 & 29-30 \end{bmatrix}$$

$$F \in \mathbb{N}[1, 29]$$

Before measures are computed, signals are band-pass filtered with Butterworth filters. These types of filters are characterized by a magnitude response that is maximally flat in the pass-band, and they offer good transition characteristics at low coefficient orders, so they can be easily implemented (Oppenheim & Schafer, 1975).

In this study, third-order Butterworth filters are used, as it was done in (Dauwels *et al.*, 2010a). Third-order Butterworth filters offer a frequency response able to filter narrow

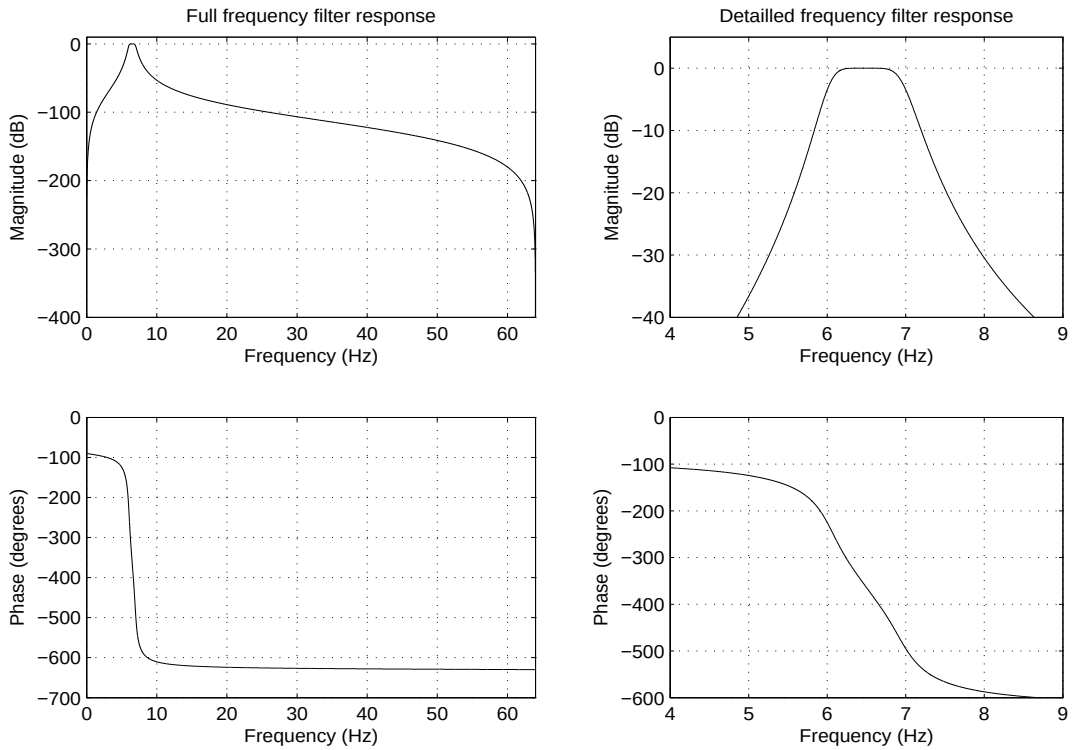


Figure 4.2: Frequency response of a third order Butterworth band-pass filter, filtering between 6 and 7 Hz with F_s=128 Hz. Left images present the frequency and phase response on the entire frequency range. Right images present in detail the frequency and phase response in the pass-band.

band frequency ranges with a bandwidth of 1 Hz, such as 1 - 2 Hz, 2 - 3 Hz, ..., 29 - 30 Hz. Even if a third-order filter offers a not-so-steep transition between the pass-band and the stop-band, for frequency ranges with $W = 1$, the frequencies F and $F + W$ offer an attenuation of 3 dB, and the adjacent frequencies, i.e. $F - 1$ and $F + W + 1$, offer an attenuation of minimum 25 dB. Figure 4.2 shows the frequency response of a filter between 6 - 7 Hz, detailing that the adjacent frequencies, 5 and 8 Hz, present a attenuation of 36 dB and 31 dB respectively.

To compute the described measures a different approach is done between the computation of synchrony measures and the power measure. However, for both types of measures only a value is computed to represent each subject. These values are then used for classification purpose. The computation is done as it follows. RP is computed for each channel independently. To obtain a global measure for each subject, the RP for all the channels are averaged as:

$$RP_s(f_1, f_2) = \frac{\sum_{i=1}^{N} RP_i(f_1, f_2)}{N} \tag{4.24}$$

were N is the number of channels contained in the data set, and s refers to the number of subject. On the other hand, computation of synchrony measures follow different approaches. As there are bivariate and multivariate measures, for each type of measure a different approach is used.

For bivariate synchrony measures, the approach defined in Dauwels *et al.* (2010a) is used. In this approach, the EEG signals are aggregated into five regions (frontal, left temporal,

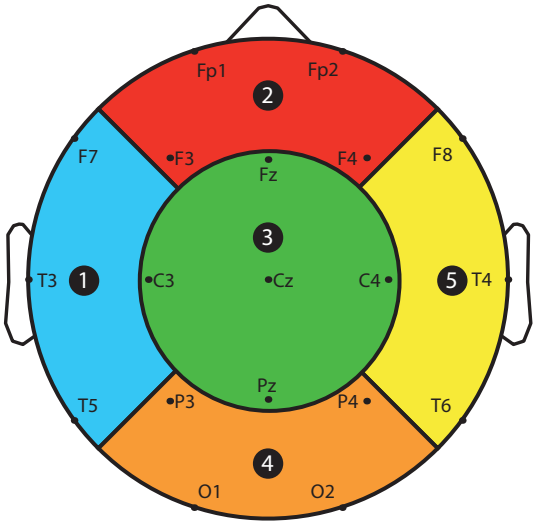


Figure 4.3: Regions defined by aggregation of electrodes: Region 1, blue, left temporal area. Region 2, red, frontal area. Region 3, green, parietal area. Region 4, orange, occipital area. Region 5, yellow, right temporal area. This example presents the aggregation for 19 electrodes.

central, right temporal and occipital). Figure 4.3 presents an example of aggregation for 19 electrodes. This strategy is useful because it facilitates the results comparison of different data sets that contain different number of electrodes. To compute the synchrony between two regions, one first computes the synchrony between each EEG signal from one region and each signal from the other, obtaining a value $Sync_{e_i,e_j}$ for each of the pairs. The next step is evaluating synchrony between regions by computing the average of these signal pairs as:

$$Sync_{R_j,R_k} = \frac{\sum_{i=1}^{N_j}\sum_{j=1}^{N_k} Sync_{e_i,e_j}}{N_j \times N_k} \tag{4.25}$$

where N_j and N_k are the number of electrodes of region j and k respectively. Using this notation, if for example we compute the synchrony between the left temporal (Region 1) and the parietal (Region 3) regions, the value $Sync_{R_1,R_3}$ is computed. To obtain this value the average of $Sync_{F_7,F_z}, Sync_{F_7,C_3}, Sync_{F_7,C_z}, Sync_{F_7,C_4}, ...Sync_{T_5,P_z}$ it is used in numerator of Equation 4.25. Then in the denominator of Equation 4.25 the value of 15 is used. This value is the number of pairs of electrodes used, in our case $N_1 = 3$ and $N_3 = 5$.

Once all $Sync_{R_j,R_k}$ are computed, the average of synchrony between these values are calculated to obtain a global synchrony value for each subject ($Sync_s$). This value is computed as:

$$Sync_s = \frac{\sum_{j=1}^{5}\sum_{j=i+1}^{5} Sync_{R_j,R_k}}{10} \tag{4.26}$$

The denominator presented in Equation 4.26 is equal to 10 due to the fact that we have aggregated the electrodes into five different regions. This approach is used for all the bivariate synchrony measures (Correlation, Coherence and PS).

A different approach is used to compute the EEG synchrony for multivariate measures. OC is applied directly to all EEG signals of the data set. However, a different strategy is followed for Granger measures. As presented in Section 4.3.4, some authors suggest that application of Granger measures in multichannel recordings it is not computationally

efficient (Sanei & Chambers, 2009). If for example an EEG data set contains 19 electrodes, when we apply the Granger measures it supposed to estimate a 19-dimensional MVAR model. To avoid this high-dimensional estimation, some authors suggest to compute Granger measures from pair-wise combinations of electrodes (Sanei & Chambers, 2009). This could be a solution, but if we apply this approach, we will be losing information about the causality, due to the fact that the influence of other channels in a specific channel it is not defined in the MVAR model. Therefore some measures such as DTF and ffDTF will not be useful. Consequently another approach is followed. In this approach we defined the same five regions that are previously defined in Figure 4.3. Then, in order to reduce the number of electrodes, the time averaging between electrodes of the same region is calculated, obtaining averaged EEG time series for each region. The Granger measures are then applied to these five averaged EEG signals. As presented previously for the bivariate measures, an only value of global synchrony measure is used to represent each subject. Hence, the values computed for the Granger measures between regions ($Sync_{R_j,R_k}$) are averaged to obtain a global synchrony measure ($Sync_s$) using Equation 4.26.

4.4.2 Statistical analysis

To evaluate the difference between populations, the Mann-Whitney test is used. This test measures the significance of the differences between two populations. In the present study, this test is computed between the values computed using the measures for the MCI patients and control subjects as well as between Mild AD patients and control subjects. Mann-Whitney test is a non-parametric test that allows investigating the statistical differences between two populations without assumptions of Gaussianity. Low p-values (close to zero, e.g., $p < 0.05$) indicate a large difference between the medians of the two populations.

EEG is highly non-stationary (Nunez & Srinivasan, 2006). It is widely accepted that the duration of a so-called quasi-stationary interval of continuous EEG recordings is expected not to exceed 2-4 s, but some authors found much longer fragments of even 12 s (Cohen & Sances, 1977) and even 25 s (Kawabata, 1976), and thus data characteristics may change over time. To deal with this problem, time segmentation is used to compute synchrony measures. Exploring different parameters like window length (in all the synchrony measures) or polynomial order (only in Granger measures) is important to select the parameters that are most effective in classifying the subjects as AD or healthy.

As a priory the best window time to use is an unknown parameter, several time window length values (L=1 s, 5 s, and 20 s) are used for each data set. Also several polynomial orders are evaluated in Granger measures (p=1, 2, ... 9). Then, the Mann-Whitney test is computed along all the possible configurations in all the possible frequency ranges. The parameter configuration that presented the lowest p-value in any of the defined frequency ranges is defined as optimal and used in further analysis for that specific data set. In section results (Section 4.5), the configuration used for each measure and for each data set is presented.

4.4.3 Separability criterion

The separability criterion ($J(F, F+W)$) is used to represent the difference between population in all the frequency ranges studied. $J(F, F+W)$ is a measure of distance between two

normal distributions inspired by the z-score (Kropotov, 2009). The distance $J(F, F+W)$ has large values when the mean difference between two populations is large, and the standard deviations of both distributions are small, therefore the two populations can then be easily distinguished. On the other hand, if there is little difference between two populations, $J(F, F+W)$ presents a value close to 0. The separability criterion is defined as:

$$J(F, F+W) = \frac{|\mu_{Ctr}(F, F+W) - \mu_{Pat}(F, F+W)|}{\sigma_{Ctr}(F, F+W) + \sigma_{Pat}(F, F+W)} \tag{4.27}$$

where $(F, F+W)$ refers to the frequency range of study. As defined in Section 4.4.1, F and $(F+W)$ refer to the start and end frequency of the study respectively; μ_{Ctr} is the mean of the control population; and μ_{Pat} is the mean of the patient population (MCI or Mild AD depending on the data set). Similarly, σ_{Ctr} and σ_{Pat} refer to the standard deviations of the control and patient (MCI or Mild AD) groups. Separability criterion is computed for each proposed measure.

4.4.4 Classification

Changes in EEG synchrony or RP allow us to distinguish between AD patients and healthy subjects. In order to quantify the number of subjects that can be classified with the presented methodology, Linear Discriminant Analysis (LDA) is used with leave-one-out (LOO) cross-validation. Classification Rate (CR) is then computed as: number of correctly classified subjects divided by the total number of subjects. Therefore, CR presents the percentage of subjects correctly classified.

LDA is a well-known scheme for feature extraction and dimension reduction. This classification method has been widely used in many applications involving high dimensional data, such as face recognition and image retrieval (Duda *et al.*, 2001). LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class distances to the within-class distance is maximized, achieving maximum discrimination. The decision of classification is done in the low dimensional space (Bishop, 2006). LDA makes the assumption that the density of the input data follows a normal distribution. Therefore, the gaussianity of the computed synchrony values was checked before the computation of the classification values by means of histograms and quantile-quantile plots.

On the other hand, presented data sets in Section 4.2 present a small number of subjects. Small number of subjects in the data base is a traditional problem that occurs in machine learning field. In order to build good models, different solutions have been presented. One of these solutions it is cross-validation. Cross-validation allows to use as much of the available data as possible for training. When using cross-validation, different approaches can be used (Bishop, 2006):

- k-fold cross-validation: In this approach data is partitioned into k groups. Then $k-1$ of the groups are used to train a model, that later is evaluated with the remaining group. This procedure is then repeated k times, using a different subset each time for validation purpose. Final CR is obtained by averaging the CR obtained in each group.
- LOO cross-validation: This approach is a particular case of the k-fold, used when the number of subjects in the data base is small, as the databases presented in this

document. In this case is considered $k = N$, where N is the total number of subjects. Therefore, all subjects except one are used for training, and the remaining subject is used for evaluation. This procedure is repeated N leaving a different subject out in each iteration. As for the k-fold cross-validation, the final CR is obtained by averaging the CR obtained for each of the subjects.

The advantage of these approaches is that all observations are used for both training and validation, and each observation is used for validation exactly once. These are the most robust evaluation methods because they try to overcome a possible overfitting. After applying cross-validation, a final value of CR is obtained because results for each of the iteration are averaged in order to produce a single estimation.

For clinical applications it is also suggested that the computation of two statistical measures, is useful. These two statistical measures are Sensitivity (SE) and Specificity (SP). SE, also call true positive rate, measures the proportion of actual positives which are correctly identified. SP, also call false negative rate, measures the proportion of negatives which are correctly identified. Therefore in the presented study for AD, SE measures the percentage of people identified as AD which suffer AD, and SP measures the percentage of people identified as healthy which are healthy.

4.5 Results

This section presents the results obtained through searching the optimal frequency range for each measure, instead of study the traditional EEG frequency bands, i.e. δ, θ, α or β. Results presented are obtained using each measure independently as a classification feature. Results for the three data sets are presented. Results for the MCI data set are shown in Section 4.5.1, results for the Mild AD Data Set 1 are described in Section 4.5.2 and results for the Mild AD Data Set 2 are presented in Section 4.5.3.

EEG non stationarity is an issue that has to be addressed when working with EEG. As presented in Section 4.4.2, for each synchrony measure the optimal configuration is used, i.e. window length and Granger order. These optimal configurations are presented in Table 4.2 and Table 4.3. Table 4.2 presents the selected configuration for the synchrony measures, with the optimal time length of the window. Table 4.3 presents the selected

Measure	**Window length**		
	MCI Data set	**Mild AD Data set 1**	**Mild AD Data set 2**
Correlation	20 sec	20 sec	1 sec
Coherence	1 sec	5 sec	1 sec
OC	5 sec	1 sec	5 sec
PS	20 sec	20sec	20 sec

Table 4.2: Optimal time window lengths used to compute the synchrony measures for the three data sets.

Granger Measures	**Window length**		
	MCI Data set	**Mild AD Data set 1**	**Mild AD Data set 2**
GC	1 sec	1 sec	5 sec
PC	20 sec	1 sec	5 sec
DTF	1 sec	20 sec	1 sec
ffDTF	1 sec	20 sec	20 sec
PDC	1 sec	1 sec	5 sec
dDTF	20 sec	20 sec	5 sec
Granger Measures	**Granger order**		
	MCI Data set	**Mild AD Data set 1**	**Mild AD Data set 2**
GC	2	9	3
PC	4	4	5
DTF	7	8	9
ffDTF	2	8	2
PDC	2	3	2
dDTF	2	8	5

Table 4.3: Optimal time window lengths and orders used to compute the Granger measures for the three data sets.

configurations for Granger measures, with the optimal time length of the window and the Granger order used to compute them.

The optimal frequency range for each measure is defined as the range in which the best CR is found; if a measure has several frequency ranges with the same CR, the one selected as the optimal frequency range is the one with the highest $J(F, F+W)$ for that measure. Results presented below are the ones obtained in the optimal frequency range.

4.5.1 Results MCI data set

Results obtained for the MCI Data Set are shown below. Table 4.4 presents the optimal frequency range and the CR computed in this frequency range. The SE and the SP computed in the optimal frequency range are also presented. The p-value computed between values of MCI patients and healthy subjects obtained in the optimal frequency range are shown in the last column. The three best results are highlighted in bold.

The best result for the MCI Data Set is obtained using RP with a value of 78.33% (SE 72.73% and SP 81.58%) in the frequency range of 2 - 9 Hz. This frequency range is close to the θ band, the one that has been largely used for the early diagnosis of AD due to the effect of slowing of EGG. The best result with a synchrony measure is 75.00% (SE 81.82% and SP 71.05%) for dDTF in the frequency range of 14 - 16 Hz. This frequency

Measures	Optimal freq. range (Hz)	CR (%)	SE (%)	SP (%)	p-value
RP	**2 - 9**	**78.33**	**72.73**	**81.58**	**0.0001**
Correlation	**1 - 8**	**71.67**	**77.27**	**68.42**	**0.0012**
Coherence	8 - 13	68.33	63.64	71.05	0.0132
GC	2 - 8	70.00	72.73	68.42	0.0021
PC	21 - 27	70.00	81.82	63.16	0.0157
DTF	6 - 27	65.00	63.64	65.79	0.2727
ffDTF	8 - 30	70.00	63.64	73.68	0.0013
PDC	1 - 2	66.67	68.18	65.79	0.0085
dDTF	**14 - 16**	**75.00**	**81.82**	**71.05**	$\mathbf{8.88 \times 10^{-5}}$
OC	8 - 10	68.33	59.09	73.68	0.0138
PS	4 - 5	70.00	90.91	57.89	0.0201

Table 4.4: Obtained results for the MCI data set. CR is presented with the optimal frequency ranges, its corresponding p-value and the computed values of SE and SP. The three best results for this data set are shown in bold.

range is inside the β band, therefore may be affected by the decrease of synchrony that AD patients suffer in the high frequencies. The third best result obtained for the MCI Data Set is 71.67% (SE 77.27% and SP 68.42%) achieved with Correlation. This result is obtained in the frequency range 1 - 8 Hz, expanding the δ and θ bands.

The best value obtained of SE is 90.91% obtained with PS. However PS presents a lower SP (57.89%), therefore the global CR is 70.00%. Best SP is obtained using RP. Interestingly most of the synchrony measures (six of the total ten) presents higher SE than SP. Furthermore, the only power measure, RP, presents higher SP than SE.

Obtained results in the optimal frequency range present always an equal or higher value of CR than the results obtained in the traditional δ, θ, α or β bands. To illustrate this, Figure 4.4 presents a comparison of the CR obtained in different frequency bands. Results presented in Figure 4.4 show that CR obtained in the optimal frequency range is always higher than the others. Only for Coherence the obtained value in the optimal frequency range and the value in the α band are equal, because the optimal frequency range is the same than the α band.

Another observation that can be extracted from Table 4.4 is that, presented results show that only four synchrony measures (Correlation, GC, PDC and PS) have the optimal frequency range in the low frequencies. All other measures found the optimal frequency range in the high frequencies, which may be related with the decrease of synchrony in the higher frequencies. To check this effect, the box plots of the values in their optimal frequency ranges are computed. Results for each measure are presented in Figure 4.5.

Presented results in Figure 4.5 are consistent with previous knowledge of changes that AD causes on EEG. Figure 4.5(a) presents the box plot of RP. It can be observed the

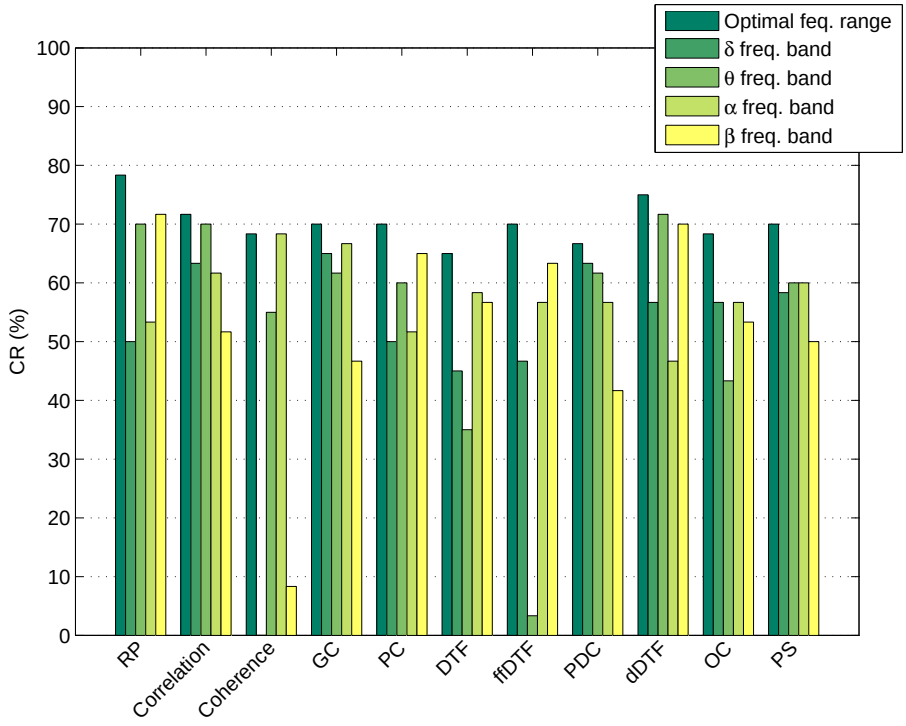


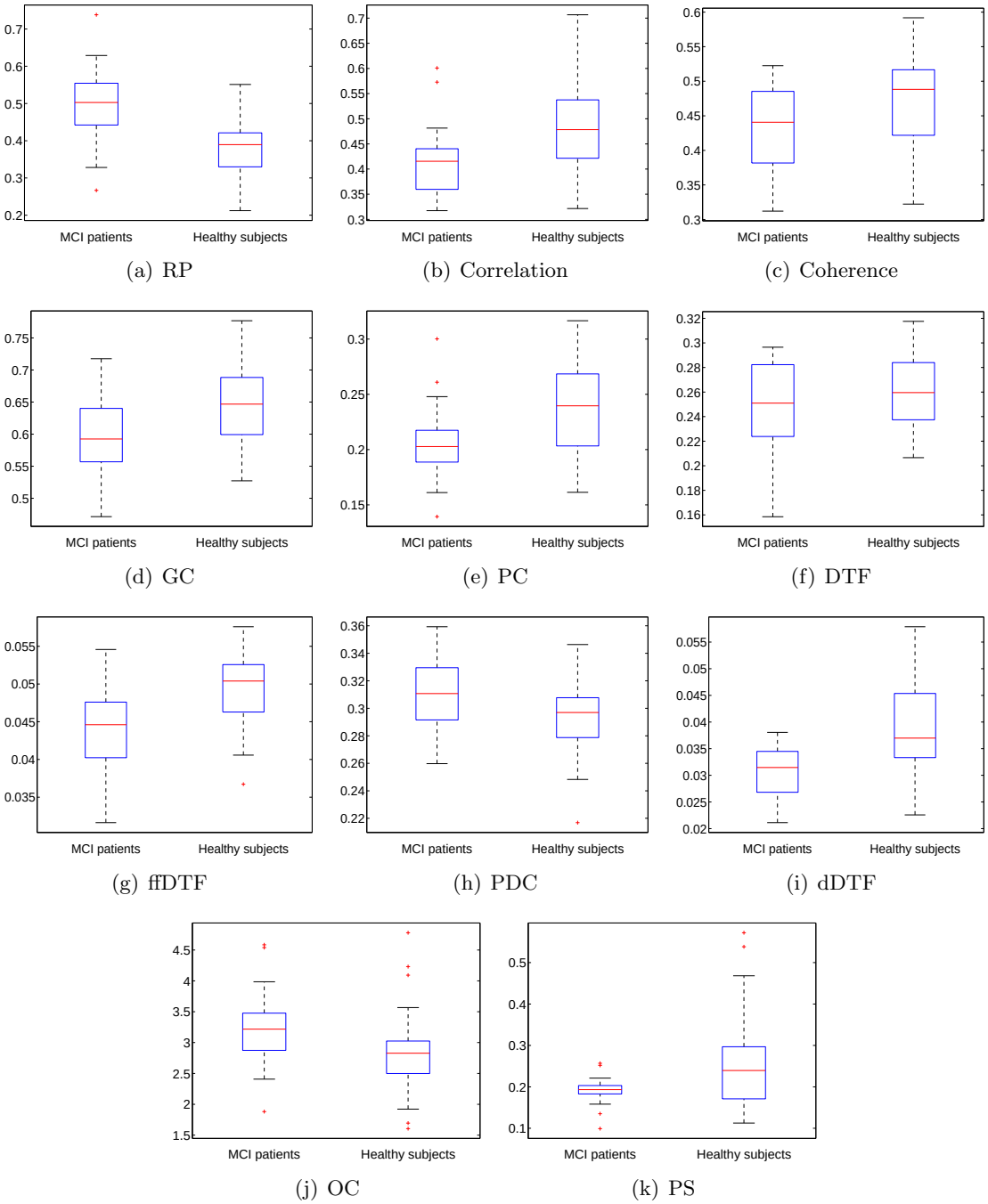
Figure 4.4: CR comparison between the optimal frequency range, δ, θ, α and β bands. Results presented for the MCI Data Set.

effect of slowing of EEG, in which the MCI patients present higher values of RP than the healthy subjects. Synchrony measures that have the optimal frequency range in the higher frequency, i.e. Coherence, PC, DTF, ffDTF, dDTF, OC, present higher values of synchrony for healthy subjects than for MCI patients, which correspond to the loss of EEG synchrony in the higher frequencies. Figure 4.5(j) present a higher value for MCI patients than for healthy subjects. However, it has to be noted that OC presents higher values when there is no synchronization of the time series. On the other hand, results obtained for PDC (Figure 4.5(h)), presents a higher synchrony value for MCI patients than for healthy subjects.

4.5.2 Results Mild AD Data Set 1

Results obtained for the Mild AD Data Set 1 are shown below. Table 4.5 presents the optimal frequency range and the CR computed in this frequency range. The SE and the SP computed in the optimal frequency range for each measure are also presented. The p-value computed between values of Mild AD patients and healthy subjects obtained in the optimal frequency range are shown in the last column. The three best results are highlighted in bold.

The best result for the Mild AD Data Set 1 is obtained using RP with a value of 97.56% (SE 94.12% and SP 100.00%) in the frequency range of 4 - 7 Hz. This frequency range is close to the θ band, which is affected for the slowing of EEG effect. The best result with a synchrony measure is 95.12% (SE 100.00% and SP 91.67%) for DTF in the frequency range of 5 - 6 Hz. The third best result obtained for the Mild AD Data Set 1 is 82.93% (SE 76.47% and SP 87.50%) obtained with GC. This result is obtained in the low frequency



(a) RP (b) Correlation (c) Coherence

(d) GC (e) PC (f) DTF

(g) ffDTF (h) PDC (i) dDTF

(j) OC (k) PS

Figure 4.5: Box plots showing the differences between MCI patients and healthy subjects in the optimal frequency range, where the best CR is achieved. Results presented for the MCI Data Set.

Measures	Optimal freq. range (Hz)	CR (%)	SE (%)	SP (%)	p-value
RP	**4 - 7**	**97.56**	**94.12**	**100.00**	$\mathbf{8.38 \times 10^{-8}}$
Correlation	23 - 24	68.29	58.82	75.00	0.0699
Coherence	8 - 13	75.61	82.35	70.83	0.0003
GC	**1 - 2**	**82.93**	**76.47**	**87.50**	$\mathbf{3.07 \times 10^{-5}}$
PC	3 - 4	68.29	70.59	66.67	0.0027
DTF	**5 - 6**	**95.12**	**100.00**	**91.67**	$\mathbf{2.64 \times 10^{-6}}$
ffDTF	1 - 2	80.49	88.24	75.00	3.00×10^{-6}
PDC	1 - 4	80.49	88.24	75.00	1.78×10^{-6}
dDTF	2 - 4	78.05	76.47	79.17	8.50×10^{-5}
OC	7 - 8	75.61	82.35	70.83	0.0005
PS	9 - 10	80.49	88.24	75.00	5.45×10^{-5}

Table 4.5: Obtained results for the Mild AD data set 1. CR is presented with the optimal frequency ranges, its corresponding p-value and the computed values of SE and SP. The three best results for this data set are shown in bold.

range 1 - 2 Hz. Obtained p-values are small for almost all the measures, which is consistent with the high classification rates obtained.

Interestingly, the best value obtained of SE is 100.00% obtained with DTF, which helps to obtain the best CR for a synchrony measure. Best SP is obtained using RP. As the results presented for the MCI Data Set, synchrony measures may help to improve the value of SE and RP may help to improve the value of SP.

Figure 4.6 presents a comparison of the CR obtained in different frequency bands. Results presented in Figure 4.6 show that CR obtained in the optimal frequency range is always higher than the CR obtained in the traditional δ, θ, α or β bands. Again, as Coherence obtained the optimal frequency range in the α band, computed values of CR in the optimal frequency range and in the α band it is the same.

Presented results in Table 4.5 show that almost all synchrony measures (except Correlation, Coherence and PS) present the optimal frequency range in the lower frequencies. To check the obtained values in their optimal frequency ranges, the box plots of each measure are computed.

Figure 4.7 presents the box plot of obtained values in their optimal frequency range for each measure. Figure 4.7(a) presents the obtained values for RP. In this figure it can be observed the effect of slowing of EEG, due to the higher values of RP that present Mild AD patients than healthy subjects. Another effect that has been largely reported as one of the causes of AD is the decrease of synchrony in the higher frequencies. However, presented results in Figure 4.7 show that for some of the synchrony measures, specifically the Granger measures (Figure 4.7(d-i)), there is an increase of synchrony for Mild AD patients in the narrow bands inside the δ and θ bands. Interestingly this increase of synchrony may be

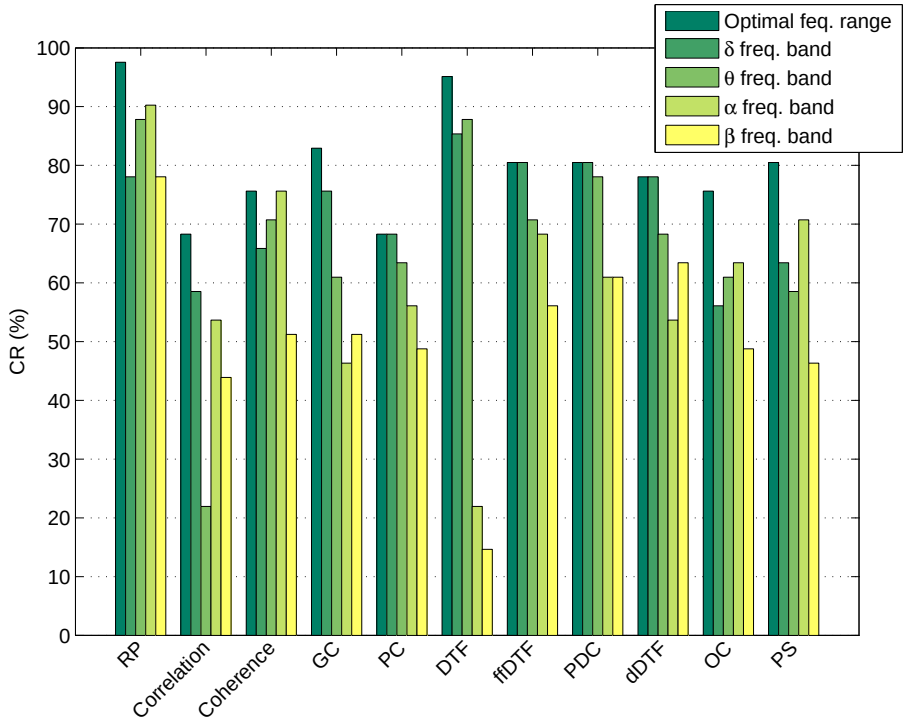


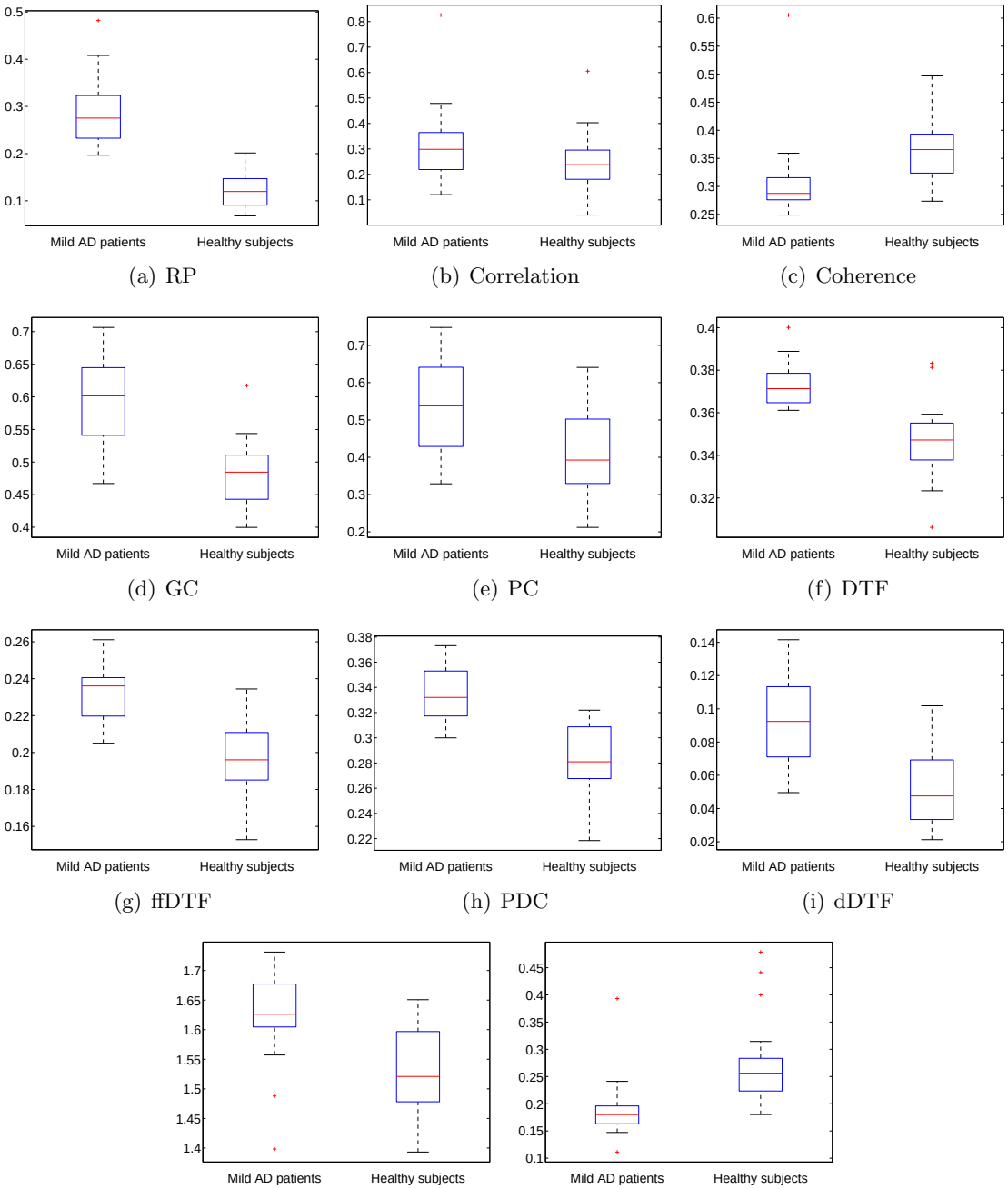
Figure 4.6: CR comparison between the optimal frequency range, δ, θ, α and β bands. Results presented for the Mild AD Data Set 1.

more discriminative than the well-known decrease of synchrony, due to the fact that all Granger measures found their optimal frequency range in the low frequencies. An example of the power of discrimination that may introduce this increase of synchrony is presented for DTF (Figure 4.7(f)), whose optimal frequency range has been obtained where there is a clear increase of synchrony for Mild AD patients. Although OC (Figure 4.7(j)) found the optimal frequency range in the lower frequencies, no such increase it is observed for it. On the other hand, Correlation and PS found the optimal frequency range in narrow bands inside the β and α bands respectively. For Correlation (Figure 4.7(b)) no decrease of synchrony it is observed for Mild AD patients, whereas for PS (Figure 4.7(k)) obtained values for Mild AD patients are lower than those obtained for healthy subjects.

4.5.3 Results Mild AD Data Set 2

Results obtained for the Mild AD Data Set 2 are shown below. Table 4.6 presents the optimal frequency range and the CR computed in this frequency range. As in the other sections, SE, SP and p-value are also presented in the table. The four best classification results for this data set are highlighted in bold.

The best CR for the Mild AD Data Set 2 is 76.00%, obtained at the same time for two synchrony measures OC, in the frequency range of 16 - 21 Hz with a SE of 71.43% and a SP of 81.82%, and for PS in the frequency range of 2 - 12 Hz with a SE of 78.57% and a SP of 72.73%. The second best result is 74.00% obtained with Correlation and DTF. Correlation (SE 71.43% and SP 77.27%) presents its optimal frequency range inside the θ band (5 - 9 Hz), and DTF (SE 75.00% and SP 72.73%) presents its optimal frequency range inside the α band (9 - 11 Hz). Classification values are the lowest obtained for any



(a) RP (b) Correlation (c) Coherence

(d) GC (e) PC (f) DTF

(g) ffDTF (h) PDC (i) dDTF

(j) OC (k) PS

Figure 4.7: Box plots showing the differences between Mild AD patients and healthy subjects in the optimal frequency range, where the best CR is achieved. Results presented for the Mild AD Data Set 1.

Measures	Optimal freq. range (Hz)	CR (%)	SE (%)	SP (%)	p-value
RP	10 - 11	72.00	82.14	59.09	0.0340
Correlation	**5 - 9**	**74.00**	**71.43**	**77.27**	**0.0018**
Coherence	9 - 10	72.00	78.57	63.64	0.0108
GC	11 - 13	70.00	75.00	63.64	0.0293
PC	7 - 8	68.00	64.29	72.73	0.0195
DTF	**9 - 11**	**74.00**	**75.00**	**72.73**	**0.0007**
ffDTF	10 - 11	72.00	64.29	81.82	0.0011
PDC	10 - 11	68.00	57.14	81.82	0.0031
dDTF	9 - 10	68.00	67.86	68.18	0.0769
OC	**16 - 21**	**76.00**	**71.43**	**81.82**	**0.0011**
PS	**2 - 12**	**76.00**	**78.57**	**72.73**	**0.0308**

Table 4.6: Obtained results for the Mild AD data set 2. CR is presented with the optimal frequency ranges, its corresponding p-value and the computed values of SE and SP. The four best results for this data set are shown in bold.

of the data sets and accordingly, the obtained p-values are the highest. RP in this data set is not as effective as when used in the other data sets.

Figure 4.8 presents again the comparison of the CR obtained in different frequency bands. Results presented in Figure 4.8 show that even if CR using the optimal frequency bands do not achieve the same high values obtained with the other data sets, obtained CR in the optimal frequency range stills being the best CR obtained with this data set, because is better than any other CR computed in the traditional EEG frequency bands.

Obtained results in Table 4.6 show that for the Mild AD Data Set 2, almost all the measures obtained their optimal frequency range in the α band. Only Correlation, PC, OC and PS obtained their values in other ranges that the α band. Specifically PS found its optimal frequency range in a range of frequencies spanning the δ, θ and α bands. To check the differences between the obtained values the box plots of the obtained values in the optimal frequency bands are presented in Figure 4.9.

Results presented in Figure 4.9 again are consistent with the previous knowledge of AD. Figure 4.9(a) presents the effect of the slowing of EEG, where in the α band the values of RP are lower for Mild AD patients than for healthy subjects. Also the effect of decrease of synchrony in the higher frequencies is present for Mild AD patients of this data set. Synchrony values computed in the α band present lower values for Mild AD patients than for healthy subjects, a clear example of it is presented for Coherence (Figure 4.9(c)) and ffDTF (Figure 4.9(g)). All other measures found in the α band (GC, DTF and PDC) also present lower values for Mild AD patients. The clear differences between Mild AD patients and healthy subjects can be observed where the optimal frequency range is found (Correlation, DTF, OC and PS). However, this differences is not enough significant to obtain a good CR. Finally, as it is noted for the other data sets, some of the synchrony

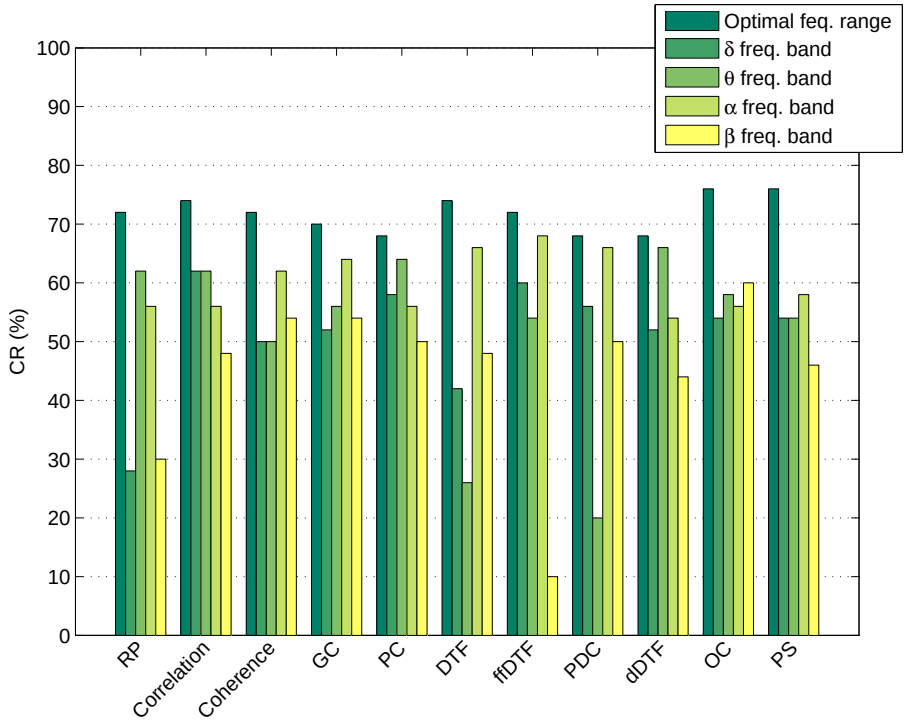


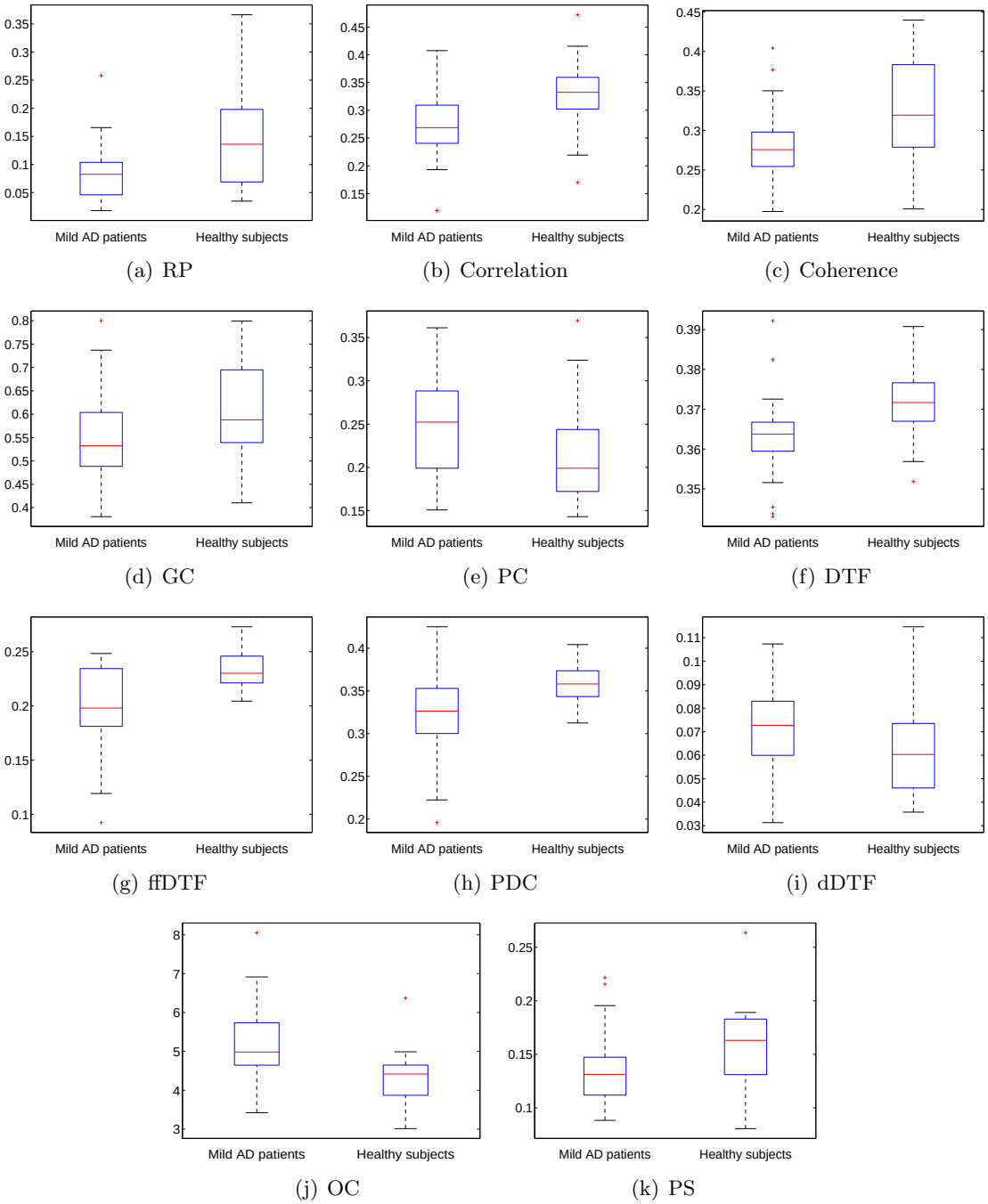
Figure 4.8: CR comparison between the optimal frequency range, δ, θ, α and β bands. Results presented for the Mild AD Data Set 2.

measures found an increase of synchrony for Mild AD patients in the θ band. For this data set, this result can be observed for PC (Figure 4.9(e)) but not for Correlation.

4.5.4 Results comparison

Presented results in Sections 4.5.1, 4.5.2 and 4.5.3, are consistent with previous results noted in the literature (Jeong, 2004; Dauwels *et al.*, 2010b; Vecchio *et al.*, 2013). For the three data sets the effect of slowing of EEG has been reported in the results. For the MCI Data Set and the Mild AD Data Set 1, RP found their optimal frequency ranges close to the θ band, therefore reported values of RP are higher for AD patients than for healthy subjects. Reported results of RP for the Mild AD Data Set 2 are found inside the α band. Therefore obtained values of RP are higher for healthy subjects when compared to values of AD patients.

Values obtained for the synchrony measures present the expected results due to the literature review. For the three data sets the effect of decrease of synchrony in the higher frequencies has been reported in the results. Shown results for all data sets presented higher synchrony values in the α and β bands for healthy subjects than for AD patients. However, synchrony results presented in the θ band in some cases present an increase of synchrony for the AD patients, especially for the Granger measures. This effect is really clear in the Mild AD Data Set 1, where all the Granger measures present this increase of synchrony, whereas that results obtained with the MCI Data Set and the Mild AD Data Set 2, only obtain a Granger measure with increase of synchrony, PDC for the MCI Data Set and PC for the Mild AD Data Set 1. This increase of synchrony in the θ band was first presented for the author of this document in the conference paper (Gallego-Jutglà



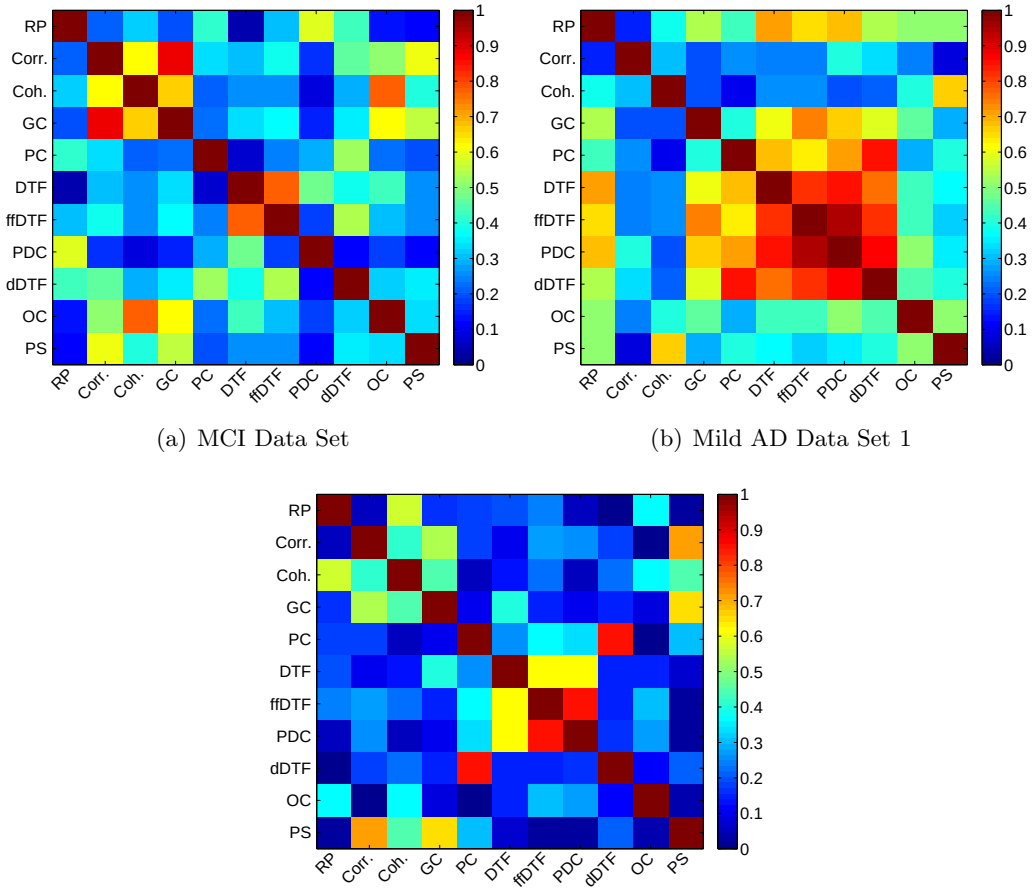
(a) RP (b) Correlation (c) Coherence

(d) GC (e) PC (f) DTF

(g) ffDTF (h) PDC (i) dDTF

(j) OC (k) PS

Figure 4.9: Box plots showing the differences between Mild AD patients and healthy subjects in the optimal frequency range, where the best CR is achieved. Results presented for the Mild AD Data set 2.



(a) MCI Data Set

(b) Mild AD Data Set 1

(c) Mild AD Data Set 2

Figure 4.10: Modulus of the correlation coefficient computed between all the measures for each data set. The optimal frequency range of each measure is used.

et al., 2012a). This discovery latter motivated the work presented in Chapter 5, where this increase of synchrony is used to facilitate the discrimination between patients suffering from AD and healthy patients.

In this chapter the parametrization of the two main effects observed in the EEG of AD patients have been investigated independently. A combination of both effects may increase the ability of the system to differentiate between AD patients and healthy subjects. However, the direct combinations of the features may lead to use irrelevant or redundant features to the target concept. To evaluate the redundant information presented by the features, correlation coefficient is computed between measures in its optimal frequency range. Figure 4.10 presents the correlation modulus of the obtained results. For each data set, high correlation ($|r| > 0.8$) is found between some measures. Figure 4.10(a) presents the obtained results for the MCI Data Set. This figure only presents a correlation value higher than $|r| > 0.8$. This value is found between Correlation and GC. Results found for Mild AD Data Set 1, presented in Figure 4.10(b), present six pairs of features with high correlation ($|r| > 0.8$). These values are found between pairs of Granger measures (DTF-PC, ffDTF-DTF, PDC-DTF, PDC-ffDTF, dDTF-ffDTF and dDTF-PDC). Finally results presented in Figure 4.10(c) for the Mild AD Data Set 2 present only two pairs of measures with $|r| > 0.8$. These values are found again between pairs of Granger measures,

PC-dDTF and PDC-ffDTF.

Interestingly the two data sets that present the lower CR (MCI Data Set and Mild AD Data Set 2) are less correlated between them than the Mild AD Data Set 1, which obtained the highest CR. This may indicate that results presented for these data sets may be easily improved by combining the power measure and the synchrony measures. However, obtained features achieve good classification rates when they are used as single feature but they introduce overfitting if we combine them directly into a multifeature classifier (Guyon & Elisseeff, 2003). Overfitting appears because the system learns redundant information of the same aspects of the data. Therefore a feature selection method is needed in order to minimize overfitting and to improve the ability of the system to distinguish between healthy subjects and AD patients.

4.6 Discussion

In this chapter, the use of synchrony measures and a frequency power measure in the whole set of frequency ranges existing between 1 and 30 Hz have been investigated. Presented results show that for some of the data sets high classification results are achieved. However, these high values of CR are not constant for all the data sets and classification results are obtained in different frequency ranges. Furthermore, the use of specific frequency ranges for each measure improves the classification performance in comparison with the results obtained in the classical frequency bands (δ, θ, α and β).

The presented results show that when using only one measure, RP is the best discriminating feature for the classification of AD patients versus healthy subjects. For the MCI Data Set, using RP we obtain a CR of 78.33%. On the other hand, the use of only a single synchrony measure achieves the best CR, 75.00% for DTF. Previous studies using this data set achieved similar results. For instance (Cichocki *et al.*, 2005), using a completely different approach BSS and RP in a different frequency range achieved a CR of 80%. In the results presented in this chapter, 78.33% is obtained without applying any decomposition technique. Results using synchrony measures present some improvement over results presented in the literature. In (Dauwels *et al.*, 2009a, 2010a), using only one measure as an input feature and LDA, the best obtained classification result was 70% using ffDTF. These studies evaluated the synchrony measures in the frequency range of 4 - 30 Hz. Presented results show that analyzing an optimal frequency range for each measure results in a better CR than using the whole frequency range. Using multiple feature classification the results improved to 78.33% in (Dauwels *et al.*, 2011). A number of other studies have also presented an improvement of CR, though only with using multiple features. In (Woon *et al.*, 2007), 88.3% of CR was obtained by dividing the time series into small windows and computing the RP in each one. The value used as a discriminative feature in that study was the maximum value of RP, the best value obtained using the values of four electrodes. In (Dauwels *et al.*, 2011), the best CR, using a combination of RP and a synchrony measure, was again 88.33%. The best CR obtained for the MCI Data Set was obtained in (Vialatte *et al.*, 2005a), achieving 93.3% using bump modeling (Vialatte *et al.*, 2009a,b), an approach completely different from the one presented here, which exploit time-frequency space information using a synchrony model, whereas the presented system only exploit the frequency information.

Using only one feature, a CR of 97.56% is obtained for the Mild AD Data Set 1. The

use of only one single synchrony measure did not improve this result, because the best classification obtained is 95.12% for DTF. In (Dauwels *et al.*, 2009a) the best CR obtained was 82.9% using only one measure and in (Dauwels *et al.*, 2011) the use of three measures as input features to a classifier achieved a CR of 95.12%. Presented results are better in both cases. In (Latchoumane *et al.*, 2012), a CR of 97.6% was obtained using multiway array decomposition in other words, the same value as is obtained in this study for RP used as individual feature classification, but using a more complex approach based on multiway array decomposition.

The presented results in frequency ranges differing from the standard bands are shown to be more discriminant. It seems that using a specific configuration and computing neural synchrony in a specific frequency range is more effective than standardizing all configurations.

Chapter 5

Increase of synchrony for the early diagnosis of AD

This chapter aims to evaluate the discriminative power of the increase of synchrony in the θ band presented in Chapter 4. To do that, a new ratio between different frequency ranges is presented.

Aiming to study the increase of synchrony in the θ band, first an introduction is done. Then Section 5.2 is devoted to present the methods used to compute the ratio, which later in Section 5.3 it is applied on the data sets to check the discriminative power of the increase of synchrony. In the last section of the chapter discussion is presented evaluating the obtained results.

5.1 Introduction

One of the changes that AD cause on EEG signals is the perturbation of synchrony. This change affects AD patients, as have been reported in Chapter 2. This change is that there is a decrease of synchrony for AD patients when compared with healthy subjects. Obtained results in Chapter 4 show consistency with the results presented in the literature, and also present an increase of synchrony for AD patients when compared with healthy subjects.

Frequency-dependent abnormalities in EEG synchrony in AD patients has been studied in detail (Jeong, 2004; Dauwels *et al.*, 2010b). Most of the studies have analyzed synchrony changes in different frequency bands (δ, θ, α and β) (Stam *et al.*, 2005), by analyzing synchrony differences between different brain regions (Tóth *et al.*, 2014), or by using multivariate measures (Yi *et al.*, 2014). These studies have mainly highlighted the decrease of synchrony in α and β bands for AD patients (Koenig *et al.*, 2005; Hsiao *et al.*, 2013).

On the other hand, other researchers presented results using EEG which are consistent with the finding of Chapter 4. Adler *et al.* (2003) found an increase of coherence between electrodes T5 and T6 and Locatelli *et al.* (1998) also found an increase of synchrony for a limited number of subjects who displayed severe cognitive problems. Both studies found this increase in the θ band. Other studies using magnetoencephalography (MEG) also found such increase of synchrony in the θ band for AD patients (Stam *et al.*, 2006).

A method enabling to use this increase of synchrony together with the well-known decrease of synchrony in high frequencies is presented. With that aim, a ratio is proposed between synchrony measures in different frequency ranges to improve the early diagnosis of AD. The present study combines the ratio with the selection of specific frequency ranges in a highly detailed analysis of all possible combinations, which clearly enhances the discrimination between AD patients and healthy subjects.

5.2 Methods

Methods used in this chapter to compute synchrony measures have already been defined in Chapter 4. In this study, however, a ratio is defined with synchrony values computed in a different frequency range. Three different data sets are used to study the increase of synchrony. These data sets are MCI Data Set, Mild AD Data Set 1 and Mild AD Data Set 2, defined previously in Section 4.2. The set of synchrony measures used are the same that have been used in Chapter 4 which have been defined in Section 4.3. Parameter configuration presented in Section 4.5 for each measure is also used in this chapter.

5.2.1 Bandpass filtering

Signals are bandpass filtered using third order Butterworth filters, as defined in Section 4.4. However, in this case two different sets of frequency ranges are defined for this study. All the possible frequency ranges inside the θ band and the α band are used, defining two sets of frequency ranges.

For the first set of frequency ranges ($\theta(f_1, f_2)$), the starting frequency f_1 varies from 4 to 7 Hz, and the width W varies from 1 to 4. The maximum frequency of analysis ($f_2 = f_1 + W$) is limited to 8 Hz. A total of 10 frequency ranges are used for the study. These frequency ranges are presented below:

$$W \in \mathbb{N}[1,4] \quad \begin{bmatrix} 4-5 & & & \\ 4-6 & 5-6 & & \\ 4-7 & 5-7 & 6-7 & \\ 4-8 & 5-8 & 6-8 & 7-8 \end{bmatrix}$$

$$f_1 \in \mathbb{N}[4,7]$$

For the second set of frequency ranges, the α band is analyzed ($\alpha(f_3, f_4)$). In this case, f_3 varies from 8 to 12 Hz, and W varies from 1 to 5. The maximum frequency of analysis ($f_4 = f_3 + W$) is limited to 13 Hz. A total of 15 frequency ranges are used for this study. These frequency ranges are presented below:

$$W \in \mathbb{N}[1,5] \quad \begin{bmatrix} 8-9 & & & & \\ 8-10 & 9-10 & & & \\ 8-11 & 9-11 & 10-11 & & \\ 8-12 & 9-12 & 10-12 & 11-12 & \\ 8-13 & 9-13 & 10-13 & 11-13 & 12-13 \end{bmatrix}$$

$$f_3 \in \mathbb{N}[8,12]$$

Synchrony measures are computed in these frequency ranges and then used to compute the ratio.

5.2.2 Synchrony Ratio

To evaluate the difference between populations representing the increase of synchrony in the θ band, the following ratio is defined by dividing a synchrony value in the θ band ($sync_\theta(f_1, f_2)$) by a synchrony value in the α band ($sync_\alpha(f_3, f_4)$):

$$\rho = \frac{sync_\theta(f_1, f_2)}{sync_\alpha(f_3, f_4)} \tag{5.1}$$

Ratio presented in Equation 5.1, aims to maximize the distance between AD patients and healthy subjects, and therefore, make easier to distinguish between populations. As presented in Section 4.5, an increase of synchrony has been discovered for AD patients in the θ band for narrow bands. Therefore the synchrony in the θ band is placed in the numerator of the ratio. On the basis of existing literature (Jeong, 2004; Dauwels *et al.*, 2010b; Vecchio *et al.*, 2013) and the results presented in Section 4.5, we also know that there is a decrease of synchrony for AD patients in the α band. Therefore the synchrony in the α band is placed in the denominator increasing the value of ρ for AD patients.

The values $sync_\theta(f_1, f_2)$ and $sync_\alpha(f_3, f_4)$ refer to the synchrony values computed in the θ band and the α band, respectively. Frequency ranges $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ are selected for each measure in order to maximize the difference between AD and healthy subjects. Given this aim, the frequency range to be used in $\theta(f_1, f_2)$ is the one that presents the highest synchrony mean value for AD patients in comparison with healthy subjects. Therefore $sync_\theta(f_1, f_2)$ is the synchrony value computed in that frequency range. The same procedure is applied to define $\alpha(f_3, f_4)$. However, as we are looking for a decrease of synchrony in AD patients in the α band, the frequency range used in $\alpha(f_3, f_4)$ is the one that presents the minimum mean value for AD patients. By selecting the frequency range $\theta(f_1, f_2)$ with higher mean value for AD patients and the frequency range $\alpha(f_3, f_4)$ with lower mean value for AD patients, we are maximizing the differences between AD and healthy subjects. Later on, this will make it easier to distinguish between these two populations. This procedure designed to select $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ is repeated for each of the measures we used.

CR is computed using LDA as performed in Chapter 4. Again LOO procedure is used. The mean success classification value in percentage, i.e. the CR, is obtained as a final result. To evaluate the difference between populations, the statistical significance of the differences between synchrony values is studied using the Mann-Whitney test.

5.3 Results using the synchrony ratio

The ratio (ρ) is used to improve the early diagnosis of patients with AD. Results presented in this section show that measures which have an increase of synchrony in the θ band, made AD patients easier to identify when compared to healthy subjects. Results for the MCI Data Set are detailed in Section 5.3.1, results for the Mild AD Data Set 1 are shown in Section 5.3.2 and results for the Mild AD Data Set 2 are presented in Section 5.3.3.

Measure	$\theta(f_1, f_2)$	**p-value** $\theta(f_1, f_2)$	$\alpha(f_3, f_4)$	**p-value** $\alpha(f_3, f_4)$	**p-value** ρ
Correlation	4 - 8	0.0003	8 - 9	0.0102	0.6844
Coherence	5 - 6	0.2661	8 - 9	0.0470	0.0111
GC	4 - 8	0.0089	8 - 11	0.0132	0.4295
PC	6 - 7	0.0324	12 - 13	0.2406	0.0006
DTF	4 - 5	0.6289	12 - 13	0.0901	0.0022
ffDTF	4 - 5	0.6289	12 - 13	0.0872	0.0021
PDC	4 - 5	0.0256	8 - 13	0.6181	0.0300
dDTF	6 - 7	0.0144	12 - 13	0.0078	8.81×10^{-7}
OC	5 - 6	0.3534	8 - 10	0.0138	0.0277
PS	4 - 8	0.1347	8 - 10	0.0093	0.1746

Table 5.1: Obtained $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ for each measure and p-values computed on each frequency rate and using ρ. Results presented for the MCI data set.

Measure	**CR (%)** $\theta(f_1, f_2)$	**CR (%)** $\alpha(f_3, f_4)$	**CR (%)** ρ	**SE (%)** ρ	**SP (%)** ρ
Correlation	70.00	65.00	53.33	36.36	63.16
Coherence	61.67	60.00	66.67	63.64	68.42
GC	61.67	63.33	53.33	59.09	50.00
PC	61.67	56.67	70.00	72.73	68.42
DTF	**53.33**	**60.00**	**75.00**	**68.18**	**78.95**
ffDTF	**53.33**	**60.00**	**75.00**	**68.18**	**78.95**
PDC	58.33	56.67	58.33	59.09	57.89
dDTF	**70.00**	**61.67**	**83.33**	**81.82**	**84.21**
OC	50.00	68.33	58.33	63.64	55.26
PS	60.00	70.00	55.00	54.55	55.26

Table 5.2: Obtained CR in $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ frequency ranges and using ρ. Results presented for the MCI data set. The three best results for this data set are shown in bold.

5.3.1 Results MCI Data Set

Results obtained for the MCI Data Set are explained below. Table 5.1 presents the frequency ranges used for each measure in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$). This table also show the p-value computed between the synchrony values of healthy subjects and MCI patients in the different scenarios, the selected frequency range in the θ band, the selected frequency range in the α band and using the ratio. Frequency ranges that do not present any increase (θ band) or decrease (α band) of synchrony for MCI patients are highlighted in blue.

Table 5.2 shows the CR obtained in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$), the CR obtained using the ρ and SE and SP obtained with the ρ. To facilitate the results comparison, Figure 5.1 also shows a comparison between the CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using ρ.

Results shown in Table 5.2 confirm the increase of synchrony. Coherence, PC, DTF, ffDTF, dDTF and OC obtain a higher CR using the ρ than the CR obtained in the $\theta(f_1, f_2)$ or $\alpha(f_3, f_4)$ band. These results can be easily compared in Figure 5.1. PDC achieves the same CR for the ρ than the CR obtained in the $\theta(f_1, f_2)$. Measures that obtain the best CR are the same that present an increase of synchrony and that obtain a lower p-value, with the exception of OC that presented a higher p-value when ρ is used. Best CR obtained for this data set is 83.33% (SE 81.82% and SP 84.21%) obtained by dDTF. The second and the third best results, obtained by DTF and ffDTF, are 75.00% (SE 68.18% and SP 78.95%). Other measures that achieve good classification results are PC with a CR of 70.00% and Coherence with a CR of 66.67%.

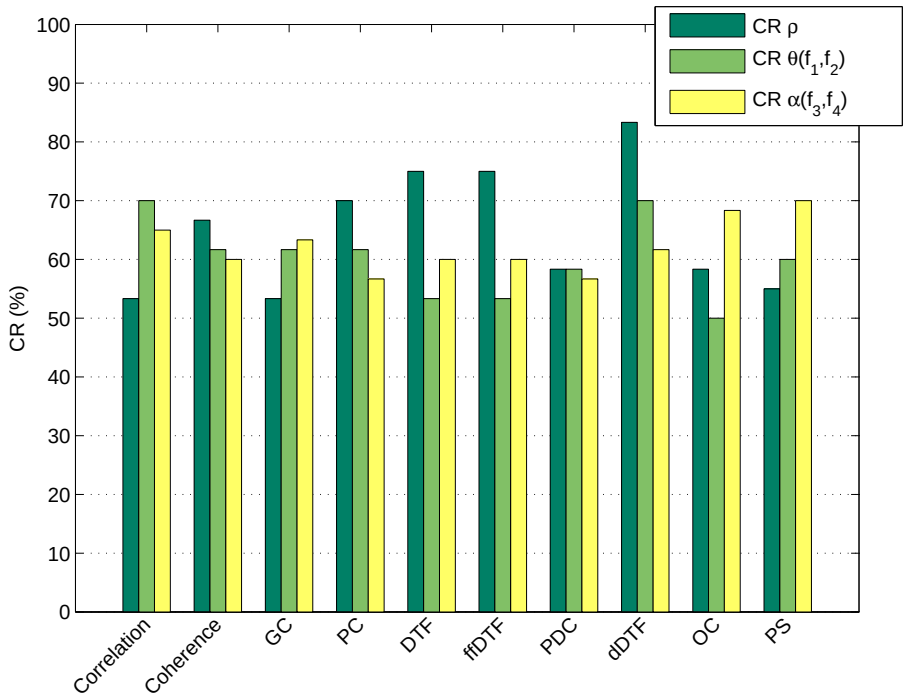
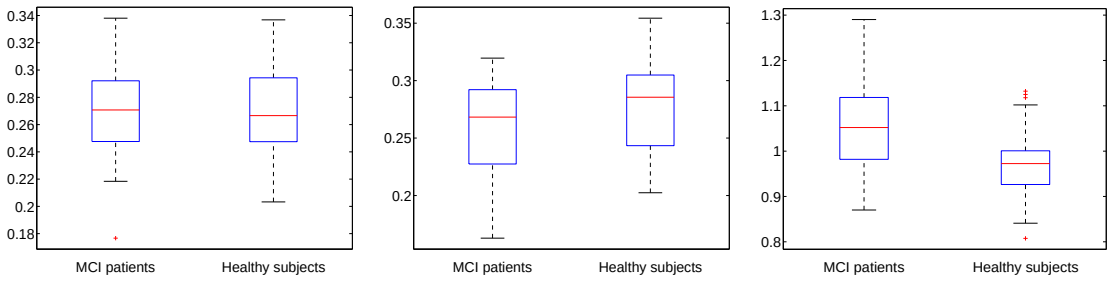
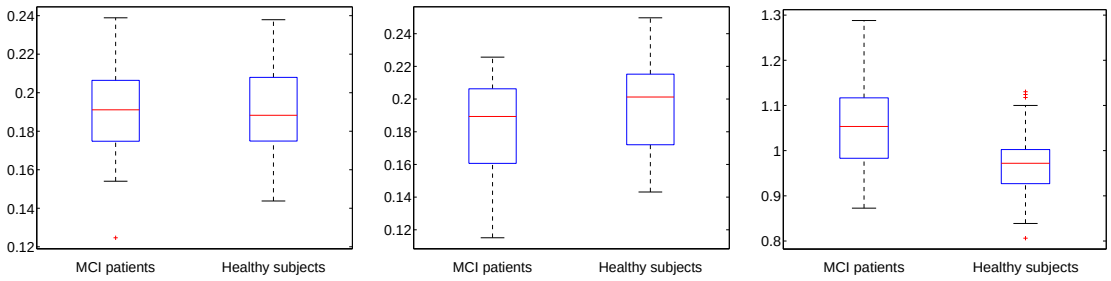


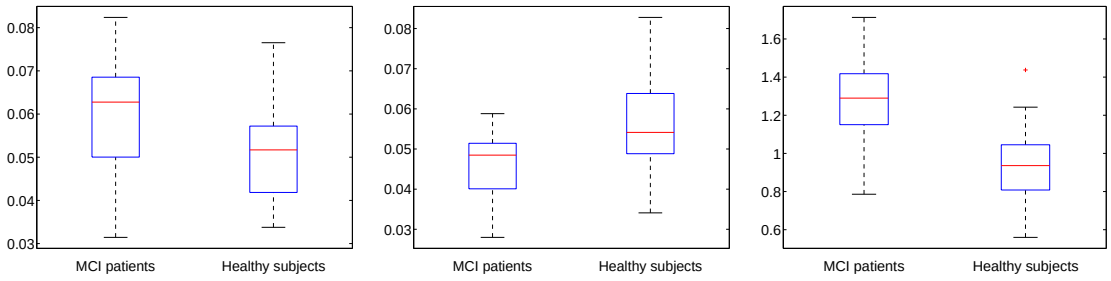
Figure 5.1: Comparison between CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using the ρ. Results presented for the MCI Data Set.



(a) DTF: $\theta(4, 5)$ (b) DTF: $\alpha(12, 13)$ (c) DTF: ρ



(d) ffDTF: $\theta(4, 5)$ (e) ffDTF: $\alpha(12, 13)$ (f) ffDTF: ρ



(g) dDTF: $\theta(6, 7)$ (h) dDTF: $\alpha(12, 13)$ (i) dDTF: ρ

Figure 5.2: Box plots presenting the three best results obtained for the MCI data set. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

Presented results in Figure 5.2 represent the phenomenon that has been detailed previously. For the three best measures, i.e. DTF, ffDTF and dDTF, the box plots of the synchrony values computed in the $\theta(f_1, f_2)$ range, the $\alpha(f_3, f_4)$ range and using the ρ are shown. For the other measures the same box plots are presented in Appendix A. Results presented in Figure 5.2 show that even though the frequency ranges in the θ band are selected because there is an increase of synchrony for MCI patients, the differences between the two populations are small, dDTF measure (Figure 5.2(g)), is the one that presented the highest difference, which is consistent with the results presented in Table 5.1 where this measure presented a small p-value. Results for α band present higher differences. In this case the effect of decrease of synchrony for MCI patients is clearly visible. Finally, results obtained when using the ρ show that the method presented in this chapter increases the distance between the two distributions. Therefore, this makes easier to classify subjects into one of the groups, as the results of the CR have shown.

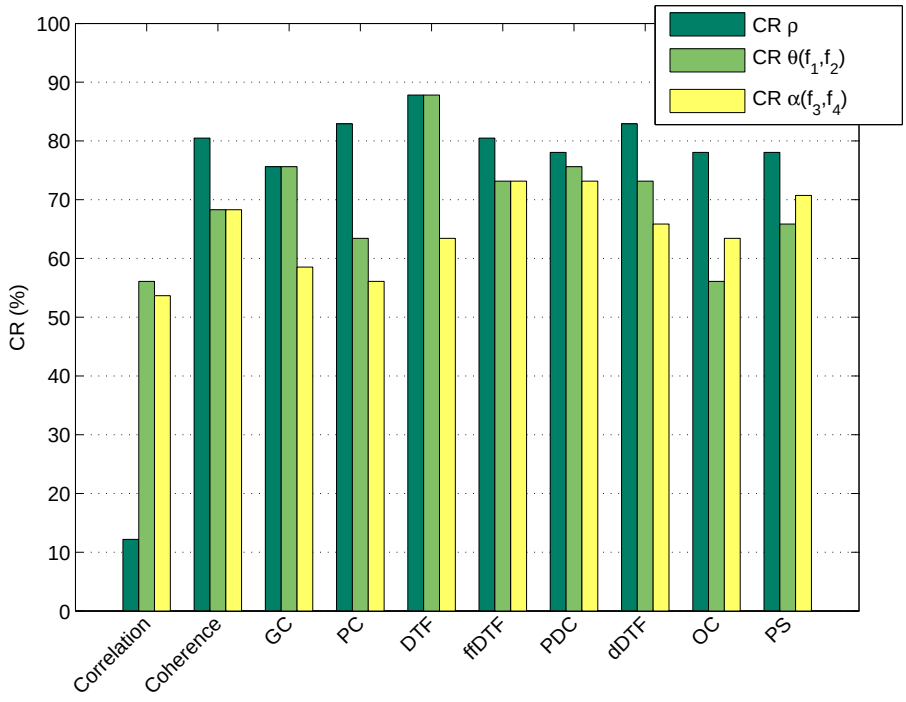


Figure 5.3: Comparison between CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using the ρ. Results presented for the Mild AD Data Set 1.

5.3.2 Results Mild AD Data Set 1

Results obtained for the Mild AD Data Set 1 are presented below. Table 5.3 presents the frequency ranges used for each measure in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$). P-values computed between the synchrony values of Mild AD patients and healthy subjects in different scenarios are shown; the selected frequency range in the θ band, the selected frequency range in the α band and using the ρ. Frequency ranges that do not present any increase (θ band) or decrease (α band) of synchrony for Mild AD patients are highlighted in blue.

Table 5.4 presents the CR obtained in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$), the CR obtained using the ρ and SE and SP obtained with the ρ. To facilitate the results comparison, Figure 5.3 also presents a comparison between the CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using ρ.

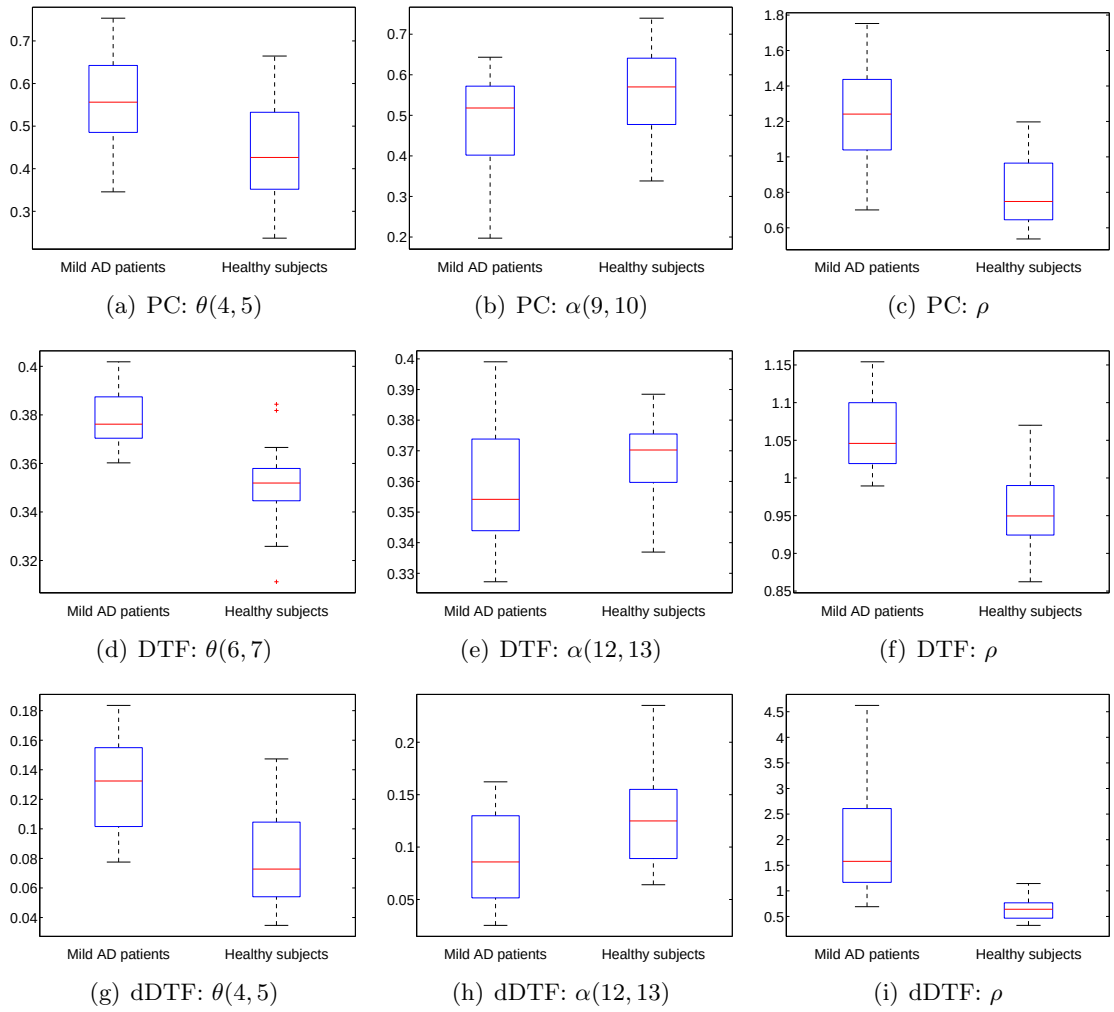
Results detailed in Table 5.3 only present Correlation as the measure in which there is no decrease of synchrony in the α band. For this measure the standard α band is used. Frequency range selected for OC in the α band is also 8 - 13 Hz. However, for this measure this is the selected frequency as the most discriminative between Mild AD patients and healthy subjects. Obtained results show that almost all measures that have an increase and a decrease of synchrony for Mild AD patients, obtain a lower p-value when the ρ is used instead of the frequency ranges $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$. DTF is the only measure that does not present a lower p-value for the ρ. This result emphasizes the idea that the ratio helps to distinguish between AD patients and healthy subjects. As results reported in Section 5.3.1, results presented for this data set show that the frequency ranges in which the increase of synchrony in the θ band was found are narrow band frequency ranges.

Measure	$\theta(f_1, f_2)$	**p-value** $\theta(f_1, f_2)$	$\alpha(f_3, f_4)$	**p-value** $\alpha(f_3, f_4)$	**p-value** ρ
Correlation	4 - 5	0.4998	8 - 13	0.6244	0.8220
Coherence	5 - 6	0.0179	10 - 11	0.0002	4.86×10^{-5}
GC	4 - 5	0.0003	10 - 11	0.2286	0.0001
PC	4 - 5	0.0091	9 - 10	0.0550	8.50×10^{-5}
DTF	6 - 7	3.41×10^{-6}	12 - 13	0.1094	4.41×10^{-6}
ffDTF	4 - 5	1.71×10^{-5}	12 - 13	0.0155	9.31×10^{-6}
PDC	4 - 5	3.88×10^{-6}	12 - 13	0.0072	1.20×10^{-6}
dDTF	4 - 5	0.0001	12 - 13	0.0155	1.78×10^{-6}
OC	4 - 6	0.2088	8 - 13	0.0206	0.0005
PS	5 - 6	0.0457	8 - 11	0.0002	2.74×10^{-5}

Table 5.3: Obtained $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ for each measure and p-values computed on each frequency rate and using ρ. Results presented for the Mild AD Data Set 1.

Measure	**CR (%)** $\theta(f_1, f_2)$	**CR (%)** $\alpha(f_3, f_4)$	**CR (%)** ρ	**SE (%)** ρ	**SP (%)** ρ
Correlation	56.10	53.66	12.20	5.88	16.67
Coherence	68.29	68.29	80.49	76.47	83.33
GC	75.61	58.54	75.61	70.59	79.17
PC	**63.41**	**56.10**	**82.93**	**76.47**	**87.50**
DTF	**87.80**	**63.41**	**87.80**	**82.35**	**91.67**
ffDTF	73.17	73.17	80.49	76.47	83.33
PDC	75.61	73.17	78.05	64.71	87.50
dDTF	**73.17**	**65.85**	**82.93**	**58.82**	**100.00**
OC	56.10	63.41	78.05	88.24	70.83
PS	65.85	70.73	78.05	70.59	83.33

Table 5.4: Obtained CR in $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ frequency ranges and using ρ. Results presented for the Mild AD Data Set 1. The three best results for this data set are shown in bold.



(a) PC: $\theta(4, 5)$ (b) PC: $\alpha(9, 10)$ (c) PC: ρ

(d) DTF: $\theta(6, 7)$ (e) DTF: $\alpha(12, 13)$ (f) DTF: ρ

(g) dDTF: $\theta(4, 5)$ (h) dDTF: $\alpha(12, 13)$ (i) dDTF: ρ

Figure 5.4: Box plots presenting the three best results obtained for the Mild AD data set 1. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

Obtained values show that for the $\theta(f_1, f_2)$, all frequency ranges presented a bandwidth of 1 Hz, being 4 - 5 Hz, 5 - 6 Hz and 6 - 7 Hz the frequency ranges selected as the ones that present higher synchrony for the Mild AD patients. Values obtained for the α band also present narrow band frequency ranges. In this case, again a bandwidth of 1 Hz is selected together with bandwidths of 3 and 5 Hz. Frequency ranges selected for the bandwidth of 1 Hz are 9 - 10 Hz, 10 - 11 Hz and 12 - 13 Hz, being this last one (12 - 13 Hz) the one selected more times.

Table 5.4 confirms the previous results. Almost all measures, except Correlation, GC and DTF present a higher CR using the ρ than the CR obtained in the $\theta(f_1, f_2)$ or $\alpha(f_3, f_4)$ range. Specifically, GC and DTF achieved the same CR using the ρ than the CR obtained in the $\theta(f_1, f_2)$. Figure 5.3 facilitates the CR comparison between measures. Measures that present better CR are the same that presented an increase of synchrony and that presented a lower p-value. Best CR obtained for this data set is 87.80% (SE 82.35% and SP 91.67%) obtained by DTF. The second and the third best results are 82.93% obtained by PC (SE 76.47% and SP 87.50%) and dDTF (SE 58.82% and SP 100%). Other measures that achieve good classification are PDC, OC and PS with a CR of 78.05%. Results presented

for DTF in Table 5.3 shown higher p-value for the synchrony values obtained using the ρ than the values obtained in the $\theta(f_1, f_2)$ range. However, the best CR is achieved with this measure. It has to be taken into account that the p-value obtained in the $\theta(f_1, f_2)$ range is significantly small (3.41×10^{-6}). P-value obtained using the ρ is also small (4.41×10^{-6}) but not as the presented in the $\theta(f_1, f_2)$ range.

Figure 5.4 present the box plots of the synchrony values computed in the $\theta(f_1, f_2)$ band, the $\alpha(f_3, f_4)$ band and using the ρ. Results for the three best measures, i.e. PC, DTF and dDTF, are presented. For the other measures the same box plots are presented in Appendix A in Figure A.3 and in Figure A.4. Results presented in Figure 5.4 show a clear increase of synchrony for the three measures in the narrow bands inside the θ band. In addition, there is a clear decrease of synchrony in the $\alpha(f_3, f_4)$ range for two of the three measures (PC (Figure 5.4(b)) and dDTF (Figure 5.4(h))). DTF (Figure 5.4(e)) presents a lower median value but a higher variance for the values of Mild AD patients than for healthy subjects. For the three measures, box plots obtained for the ρ present a higher difference between the two population than any of the other frequencies ranges.

5.3.3 Results Mild AD Data Set 2

Results using the ρ obtained for the Mild AD Data Set 2 are presented next. Table 5.5 shows the frequency ranges used for each measure in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$). P-values computed between the synchrony values of Mild AD patients and healthy subjects in different scenarios are shown; the selected frequency range in the θ band, the selected frequency range in the α band and using the ρ. Frequency ranges that do not present any increase (θ band) or decrease (α band) of synchrony for Mild AD patients are highlighted in blue.

Measure	$\boldsymbol{\theta(f_1, f_2)}$	**p-value** $\boldsymbol{\theta(f_1, f_2)}$	$\boldsymbol{\alpha(f_3, f_4)}$	**p-value** $\boldsymbol{\alpha(f_3, f_4)}$	**p-value** $\boldsymbol{\rho}$
Correlation	4 - 8	0.0050	8 - 9	0.0167	0.5001
Coherence	5 - 6	0.6887	9 - 10	0.0108	0.0081
GC	4 - 8	0.1565	12 - 13	0.0158	0.1623
PC	7 - 8	0.0195	8 - 13	0.1202	0.1250
DTF	6 - 7	0.8681	10 - 11	0.0007	0.0648
ffDTF	4 - 8	0.4518	10 - 11	0.0011	0.1299
PDC	5 - 6	0.9922	11 - 12	0.0176	0.2075
dDTF	7 - 8	0.0431	11 - 12	0.8835	0.0947
OC	4 - 8	0.1299	9 - 13	0.0769	0.2957
PS	4 - 8	0.1509	9 - 10	0.0648	0.1509

Table 5.5: Obtained $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ for each measure and p-values computed on each frequency rate and using ρ. Results presented for the Mild AD Data Set 2.

Measure	CR (%) $\theta(f_1, f_2)$	CR (%) $\alpha(f_3, f_4)$	CR (%) ρ	SE (%) ρ	SP (%) ρ
Correlation	62.00	64.00	54.00	39.29	72.73
Coherence	**46.00**	**72.00**	**68.00**	**60.71**	**77.27**
GC	**56.00**	**68.00**	**62.00**	**60.71**	**63.64**
PC	68.00	56.00	60.00	53.57	68.18
DTF	44.00	70.00	60.00	57.14	63.64
ffDTF	54.00	72.00	58.00	50.00	68.18
PDC	34.00	66.00	50.00	42.86	59.09
dDTF	60.00	12.00	58.00	46.43	72.73
OC	**58.00**	**60.00**	**64.00**	**67.86**	**59.09**
PS	54.00	66.00	20.00	17.86	22.73

Table 5.6: Obtained CR in $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ frequency ranges and using ρ. Results presented for the Mild AD Data Set 2. The three best results for this data set are shown in bold.

Table 5.6 presents the CR obtained in the θ band ($\theta(f_1, f_2)$) and α band ($\alpha(f_3, f_4)$), the CR obtained using the ρ and SE and SP obtained with the ρ. To facilitate the results comparison, Figure 5.5 also presents a comparison between the CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using ρ.

Results detailed in Table 5.5 show five measures which do not present increase of synchrony in the θ band, i.e. Correlation, GC, ffDTF, OC and PS, and one measure which does not present decrease of synchrony for the Mild AD patients in the α band, which is PC. For all these measures the standard ranges are used. Results presented for this data set show that only for Coherence a lower p-value is achieved for the ρ, for all other measures obtained p-values using the ρ are higher than the obtained in the $\theta(f_1, f_2)$ or $\alpha(f_3, f_4)$ ranges. Frequency ranges in which the increase of synchrony in the θ band for the Mild AD patients is found, are frequency ranges with a bandwidth of 1 Hz, which are 5 - 6 Hz, 6 - 7 Hz and 7 - 8 Hz. On the other hand, the decrease of synchrony for AD patients find in the α band are also find in narrow band with a band width of 1 Hz (8 - 9 Hz, 9 - 10 Hz, 10 - 11 Hz, 11 - 12 Hz and 12 - 13 Hz), but OC find the frequency range in which the differences between synchrony for AD patients and healthy subjects are higher in the 9 - 13 Hz, a frequency range close to the α band.

CR results presented in Table 5.6 show that there is only one measure that present an increase of the CR when the ρ is used, in comparison with the CR obtained in the $\theta(f_1, f_2)$ and $\alpha(f_3, f_4)$ range. This measure is OC. However, this measure did not present an increase of synchrony for the Mild AD patients. Therefore the standard θ band is used. This results comparison can be easily seen in Figure 5.5, where bars presenting the CR obtained using the ρ are always lower than the obtained for the other CR, with the exception of OC. As can be seen in Table 5.6, measures that presented the best CR are Coherence, OC and GC. Best CR using the ρ obtained for this data set is 68.00% (SE 60.71% and SP 77.27%) obtained by Coherence. The second best result it is 64.00 % obtained by OC, with 67.86% of SE and 59.09% of SP. The third best result it is obtained with GC and achieves a CR

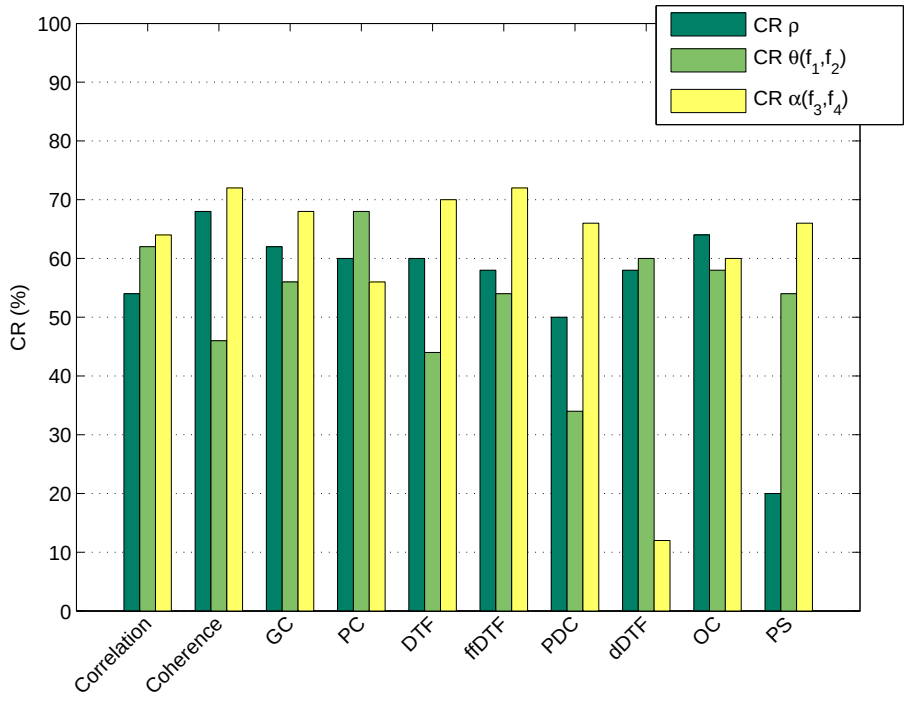


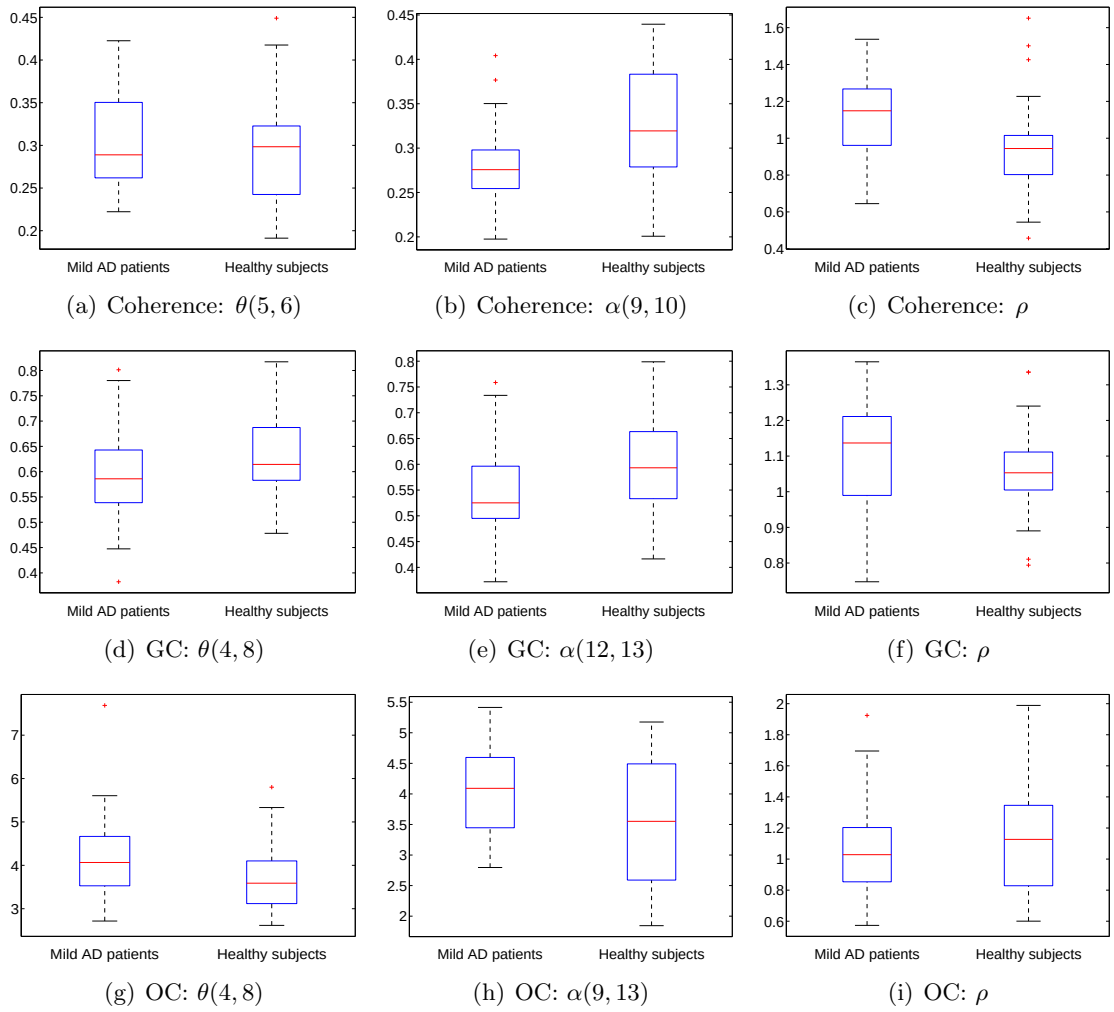
Figure 5.5: Comparison between CR obtained in $\theta(f_1, f_2)$, CR obtained in $\alpha(f_3, f_4)$ and CR obtained using the ρ. Results presented for the Mild AD Data Set 2.

of 62.00% (SE 60.71% and SP 63.64%).

As for the other data sets, the box plots of the synchrony values computed in the $\theta(f_1, f_2)$ band, the $\alpha(f_3, f_4)$ band and using the ρ are computed. These are shown in Figure 5.6. Results for the three measures that achieve the higher CR using the ρ are shown. Results presented in Appendix A in Figure A.5 and in Figure A.6, show the results for the other measures. Results detailed in Figure 5.6 show that for GC (Figure 5.6(d)) and OC (Figure 5.6(g)) no increase of synchrony is found, therefore they use the standard θ band. Results presented for the $\alpha(f_3, f_4)$ range are consistent with previous results found, due to the decrease of synchrony presented for the Mild AD patients. Results obtained using the ρ do not help to enlarge the differences between populations as happened with the other data sets. Results presented in Appendix A show that measures that in Table 5.5 presented a narrow band frequency range of 1 Hz, present higher synchrony values for Mild AD patients. However, this increase is not significant enough to help to distinguish between AD patients and healthy subjects.

5.3.4 Results comparison

Presented results show different outcome for the three data sets. Obtained results with the MCI Data Set and the Mild AD Data Set 1 present a clear improvement of the CR and the p-value when ρ is used. However Mild AD Data Set 2 does not present such improvement. Obtained results show consistence with previous results presented in the literature (Jeong, 2004; Dauwels *et al.*, 2010a; Vecchio *et al.*, 2013), and at the same time confirm previous studies (Gallego-Jutglà *et al.*, 2012a; Gallego-Jutglà & Solé-Casals, 2013) that presented a increase of synchrony for AD patients in the θ band.



(a) Coherence: $\theta(5,6)$ (b) Coherence: $\alpha(9,10)$ (c) Coherence: ρ

(d) GC: $\theta(4,8)$ (e) GC: $\alpha(12,13)$ (f) GC: ρ

(g) OC: $\theta(4,8)$ (h) OC: $\alpha(9,13)$ (i) OC: ρ

Figure 5.6: Box plots presenting the three best results obtained for the Mild AD data set 2. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

Results obtained for the three data sets show that in the narrow bands inside the θ band, synchrony measures present an increase of value for AD patients, whether MCI or Mild AD patients. For the MCI Data Set and the Mild AD Data Set 1, the frequency range 4 - 5 Hz is the one that appears most discriminative in the θ band, (three times for the MCI Data Set and six for the Mild AD Data Set 1). For these data sets, the frequency range 5 - 6 Hz is the second that appears the most, twice for each data set, followed for the frequency range 6 - 7 Hz. Results obtained for the Mild AD Data Set 2 show that only some measures present this increase of synchrony. For this data set there are two frequency ranges that appear the most, these are 5 - 6 Hz and 7 - 8 Hz, being selected each of those two times by different measures.

On the other hand, the decrease of synchrony in the α band is also presented in narrow bands frequency ranges. For MCI Data Set and Mild AD Data Set 1, the frequency range 12 - 13 Hz appears as the most discriminative, being selected four times for both data sets. Results presented for the Mild AD Data Set 2 only present the frequency range of 12 - 13 Hz being selected one time. For this data set the frequencies ranges that appear frequently are 9 - 10 Hz, 10 - 11 Hz and 11 - 12 Hz, being selected twice each one. Even

though these are different, they are also found in narrow bands frequency range of 1 Hz.

As presented in Chapter 4, Mild AD Data Set 2 is the most difficult to classify. In this section obtained results show the same tendency. Results presented for the MCI Data Set and the Mild AD Data Set 1 show that the ability to distinguish between healthy subjects and AD patients is increased when the ρ is used in comparison with the results obtained in the classical EEG frequency bands. In addition, in some cases using the ρ improves the CR obtained in the optimal frequency band which is defined in Chapter 4. Figure 5.7 presents the results comparison between the CR obtained using the ρ and CR obtained in the optimal frequency range (results presented in Chapter 4). Results for the three data sets are shown.

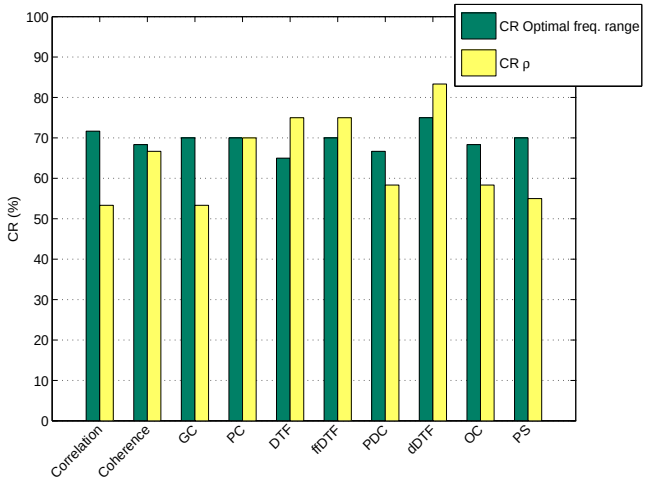
As presented in Figure 5.7(a), results for the MCI Data Set show that the three best measures obtained using the ρ (DTF, ffDTF and dDTF), the CR obtained using the ρ overcomes classification results obtained in the optimal frequency range. The best CR obtained using the ρ, 83.33% obtained with dDTF, clearly improves the CR obtained for synchrony measures in the optimal frequency range, and also improve the 81.58% obtained with RP in the frequency range of 2 - 8 Hz. Therefore the result of 83.33% obtained using the ratio is the best result obtained for the MCI data set. Results obtained for the Mild AD Data Set 1 (Figure 5.7(b)), show that four measures obtain a better CR using the ρ than the CR obtained in the optimal frequency range. These measures are Coherence, PC, dDTF and OC. Even though these measures present an improvement of the CR, the best CR obtained for a synchrony measure for the Mild AD Data Set 1, 95.12% obtained with DTF in the frequency range of 5 - 6 Hz, it is not improved. Finally, results obtained in Figure 5.7(c) for the Mild AD Data Set 2 show that obtained results using the ρ do not improve results in the optimal frequency range for each measure.

5.4 Discussion

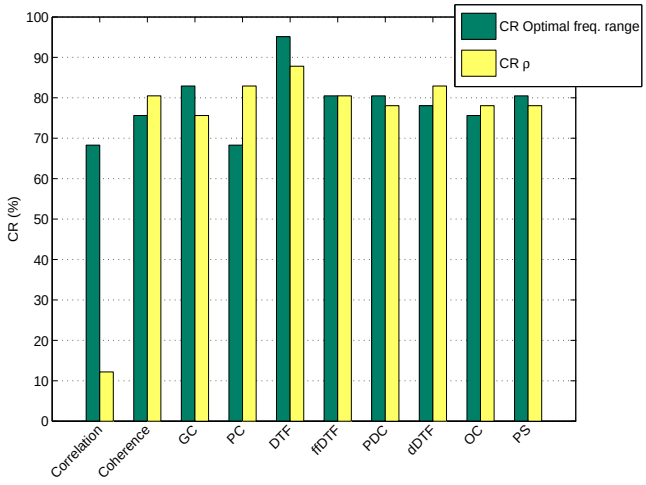
In this chapter the increase of synchrony found in AD patients has been studied. With that aim, a ratio has been presented that compares the synchrony in different frequency bands. Presented results for the MCI Data Set and the Mild AD Data Set 1 show that the ρ improves classification of the individuals as either AD patients or healthy subjects. The p-values obtained are smaller for the ρ than for the synchrony measures alone. The results obtained for the Mild AD Data Set 1 present higher CR than the results obtained for the MCI Data Set. Results obtained for the Mild AD Data Set 2 do not present this tendency, neither improvement of CR nor a decrease of the p-value.

Our results show that for the MCI Data Set a CR of 83.33% is achieved and a CR of 87.8% is obtained for Mild AD Data Set 1. Previous studies using these data sets and synchrony measures achieved comparable results. In (Dauwels *et al.*, 2009a, 2010a) the same results were achieved using the MCI Data Set, but in this case using two measures as input features to a LDA classifier. When these studies used only one measure as a single feature, the best CR decreased to 70.0 %. For the Mild AD Data Set 1 a CR of 85.0 % was achieved using two measures and LDA, but using only one measure the CR decreased to 82.9 % in (Dauwels *et al.*, 2009a). The two obtained values in (Dauwels *et al.*, 2009a) were lower than the ones obtained in this study.

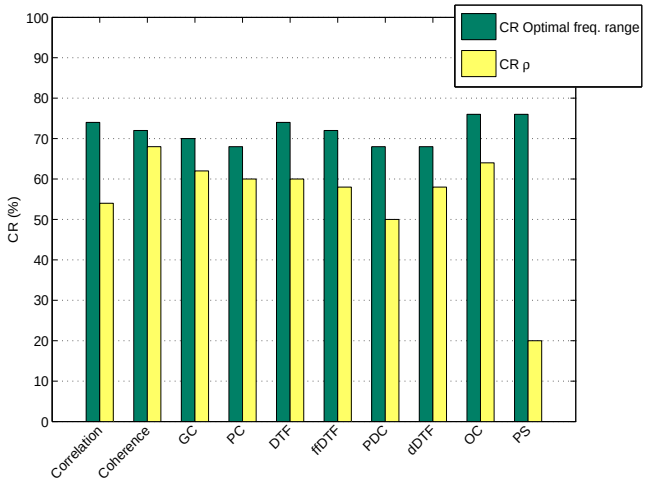
Different results were obtained in (Dauwels *et al.*, 2011) where a combination of RP and other measures were used. The best result in (Dauwels *et al.*, 2011) for MCI Data Set was



(a) MCI Data Set



(b) Mild AD Data Set 1



(c) Mild AD Data Set 2

Figure 5.7: Results comparison between the CR obtained in the optimal frequency range and using the ρ.

78.33 % using Stochastic Event Synchrony and the RP in the θ band as input features into a LDA classifier. Therefore, for the MCI Data Set, the ratio clearly improves the CR and simplifies the classification system. However, for the Mild AD Data Set 1, using RP in the θ band and ffDTF as features, the best CR obtained was 95.12 %. This result improves our best CR for this data set. Nevertheless, it is important to note that: (i) input space is two-dimensional (a vector of two features as input for the classifier), whereas we have only one-dimensional space and (ii) features used to obtain the best CR were different for both data sets, whereas we use the ratio in both data sets. Keeping in mind that the objective is to deal with the early diagnosis of AD, the proposed ratio ρ yields good classification performance on both EEG data sets, which is interesting for medical applications.

On the other hand, earlier studies were mainly focused on the decrease of synchrony for AD patients in higher frequencies (Jeong, 2004; van der Hiele *et al.*, 2007); only a few studies have presented an increase of synchrony in the θ band. In our study, such increase of synchrony in the θ band is confirmed for Mild AD patients of the Mild AD Data Set 1, and an increase of synchrony in narrow bands is also found in MCI subjects. This increase is also present for some measures of the Mild AD Data Set 2. However, this increase is not as significant as the others to help to distinguish between AD patients and healthy subjects. Other earlier studies (Adler *et al.*, 2003; Babiloni *et al.*, 2006; Knyazeva *et al.*, 2010, 2013; Kramer *et al.*, 2007) also presented an increase of synchrony in the θ band, usually in a specific region like the posterior cingulate gyrus area, or the area covered by the electrodes P3-P4, C3-C4, F3-F4 and FP1-FP2. Locatelli *et al.* (1998) found an increase of synchrony in the θ band only for a limited number of subjects who displayed severe cognitive problems. Other studies (Koenig *et al.*, 2005) reported an increase of global synchrony in the θ band using a multivariate measure. Some of the above studies highlight the decrease of synchrony in the α band, instead of the increase in θ band. Our results show an increase on the global synchrony value in the θ band.

Comparing results from different studies remains a difficult task. The most important issue is the high variability among the different methods used in the studies. Few studies consider multiple synchrony measures. Usually only one synchrony measure is considered in each paper, which is often different in each study. Besides, experimental conditions between studies may be different, e.g., different recording conditions, electrode placements, and/or patient inclusion criteria. This last condition may be a key factor to explain the variability of results found in the literature.

Chapter 6

Classification improvement through feature selection

This chapter presents a feature selection method to improve the early diagnosis of AD. As presented in the previous chapters, different features have been computed. However, the simple combination of those may lead to overfitting. In order to reduce it, a feature selection method it is presented and tested with the data sets containing MCI and Mild AD patients.

The organization of this chapter can be summarized as follows. First an introduction to feature selection is presented. Then, in Section 6.2 the new proposed method for feature selection is detailed, and in Section 6.3 this method is used aiming to improve the rate at which AD patients are distinguished from healthy subjects. Once results are shown, their statistical significance is presented in Section 6.4. Last section of this chapter is devoted to the discussion of the obtained results.

6.1 Introduction to feature selection

Feature selection is a specific term inside the main term dimension reduction. Research on this field dates back to the 60's. Since then the research in dimension reduction has been a challenging field. Nowadays machine learning and data acquisition advances demand the processing of data with thousands of features. An example is microarray processing. Using dimension reduction helps us to select the features that are more relevant to describe the data.

In real-world situations, the relevant features that may be used to explain a target are an unknown parameter. To achieve a good representation of the domain, usually many candidate features are introduced into the systems. Unfortunately, many of these features are either partially or completely irrelevant/redundant to the target concept (Dash & Liu, 1997). A relevant feature is neither irrelevant nor redundant to the target concept; an irrelevant feature does not affect the target concept in any way, and a redundant feature does not add anything new to the target concept and may affect negatively the performance of associated classification methods (John *et al.*, 1994; Rückstieß*et al.*, 2011). Therefore, applying dimension reduction to the number of irrelevant/redundant features drastically

reduces the running time of a learning algorithm and yields a more general concept. This helps in getting a better insight into the underlying concept of a real-world classification problem (Koller & Sahami, 1995; Kohavi & Sommerfield, 1995).

Nowadays, there are two major approaches for dimension reduction, feature extraction and feature selection (Levner *et al.*, 2006). Feature extraction is based in the reduction of the existing features by applying a transformation to these, obtaining a lower dimensional space which better represents the target concept (Levner *et al.*, 2006). An example of transformation to apply feature extraction it is PCA (Duda *et al.*, 2001). On the other hand, feature selection is a field of study that consists in find automatically a subset of features such that this new subset, is more relevant to the target concept (Kohavi & John, 1997).

Feature selection is also necessary in order to limit model complexity to the minimum necessary to explain the data. Attempts to improve prediction performance on training data by increasing model complexity may lead to data overfitting. If we use a high number of features to describe the data, what it could happen is that the accuracy obtained using training data is high, but this high performance does not extend to new test data; the model does not generalize to new unknown cases (Guyon & Elisseeff, 2003). On the other hand, overfitting also occurs when a model is excessively complex, such as having too many parameters relative to the number of observations, and it describes random error or noise instead of the underlying the real relationship that there is on the data (Guyon & Elisseeff, 2003; Brown *et al.*, 2012).

The presented methodology in Chapter 4 shows that a high number of features are computed to characterize the EEG data of AD patients and healthy subjects. For each patient eleven measures are computed and for each measure a total number of 435 frequency ranges are computed. When used alone, some of these measures presented good results in the differentiation between patients suffering AD and healthy subjects. With the results achieved, one could think about the possibility to increase the dimensionality of the features used in the classifier, i.e. use more features, and by testing all the possible combination of features until find an optimal. In this case we would introduce an important overfitting to the results because we could find pair of measures that perfectly explain the data set. However, when new data will be used, results would show that the selected features are not as optimal as could be.

In order to minimize the problem of overfitting and to select the features that better differentiate between AD patients and healthy subjects, feature selection is done. Next section presents a new feature selection method used to select the best frequency range for each of the measures used, and then sort the measures in the order that better describe the data set. This new method is based in a previous one call Orthogonal Forward Regression (OFR). However, a variation to select the best frequency range is added to fulfill the presented methodology in this work.

6.2 Methods for feature selection

This section presents the feature selection method used to increase the performance in the classification of AD patients and healthy subjects. First the original OFR algorithm is presented, then a methodology to control overfitting is detailed and finally the proposed variation of OFR is presented.

6.2.1 Orthogonal Forward Regression

OFR is an iterative algorithm that sorts the variables candidates based on the relevance with the output. The traditional OFR algorithm is the one presented in (Chen *et al.*, 1989). This algorithm considers the observation space, with dimension N, in which the output that has to be modeled is a vector y^p, and each of the variables is represented by a vector z_i, $i = 1 \dots p$; being p the number of parameters that will be used by the model and N the total number of subjects (i.e. number of AD patients + number of healthy subjects). This algorithm selects the input feature z_i that best correlates with the desired output, and projects the remaining features in the null space of the selected one. This procedure is repeated for all input features. The algorithm sorts the input features according to their correlation with the output. The OFR algorithm is summarized as presented in Algorithm 6.1 (Dreyfus *et al.*, 2008).

Algorithm 6.1: Orthogonal Forward Regression

i. Compute the correlation between all input features z_i and the output to be modeled y^p as:

$r^2_{y^p,z_i} = cos^2\theta_i = \frac{(y^p \cdot z_i)^2}{(y^p \cdot y^p)(z_i \cdot z_i)}$

ii. Define the selected feature (x_j) as the one that maximize the correlation between z_i and y^p:

$x_j = z_i \quad : \quad max(r^2_{y^p,z_i})$

iii. Project y^p and all other z_i onto the null space orthogonal to the selected feature x_j using Gram-Schmidt orthogonalization

$y'^p = y^p - \frac{y^p \cdot x_j}{x_j \cdot x_j} x_j$

$z'_i = z_i - \frac{z_i \cdot x_j}{x_j \cdot x_j} x_j \qquad \text{for} \qquad i = 1 \dots p$

iv. Remove x_j from the list of input feature z_i.

v. Return to (i) until all features have been selected.

Figure 6.1 presents an example of feature selection using OFR for $N = 2$ and $p = 4$. In this example an iteration of the algorithm is presented, working with four vectors acting as an input features z_i and one output vector y^p modeling the desired output. For simplicity the representation in two dimensions is presented. However, this algorithm presents no limitation in the number of dimensions to be use.

Used features are first presented in Figure 6.1(a). This figure displays four features z_i pointing to different directions in the space. Output to be modeled, i.e. y^p, it is then displayed in Figure 6.1(b). The first step defined in point (i) of the Algorithm 6.1, is to compute the correlation between the desired y^p and all existing z_i. This correlation is presented in Figure 6.1(c). In this figure, the angles between the features and the output are represented. The smallest angle, therefore the highest correlation, with the output is obtained with z_2. As defined in step (ii) of the Algorithm 6.1, this feature is kept as x_2 (presented in Figure 6.1(d)), and then all resting features are projected onto the null space

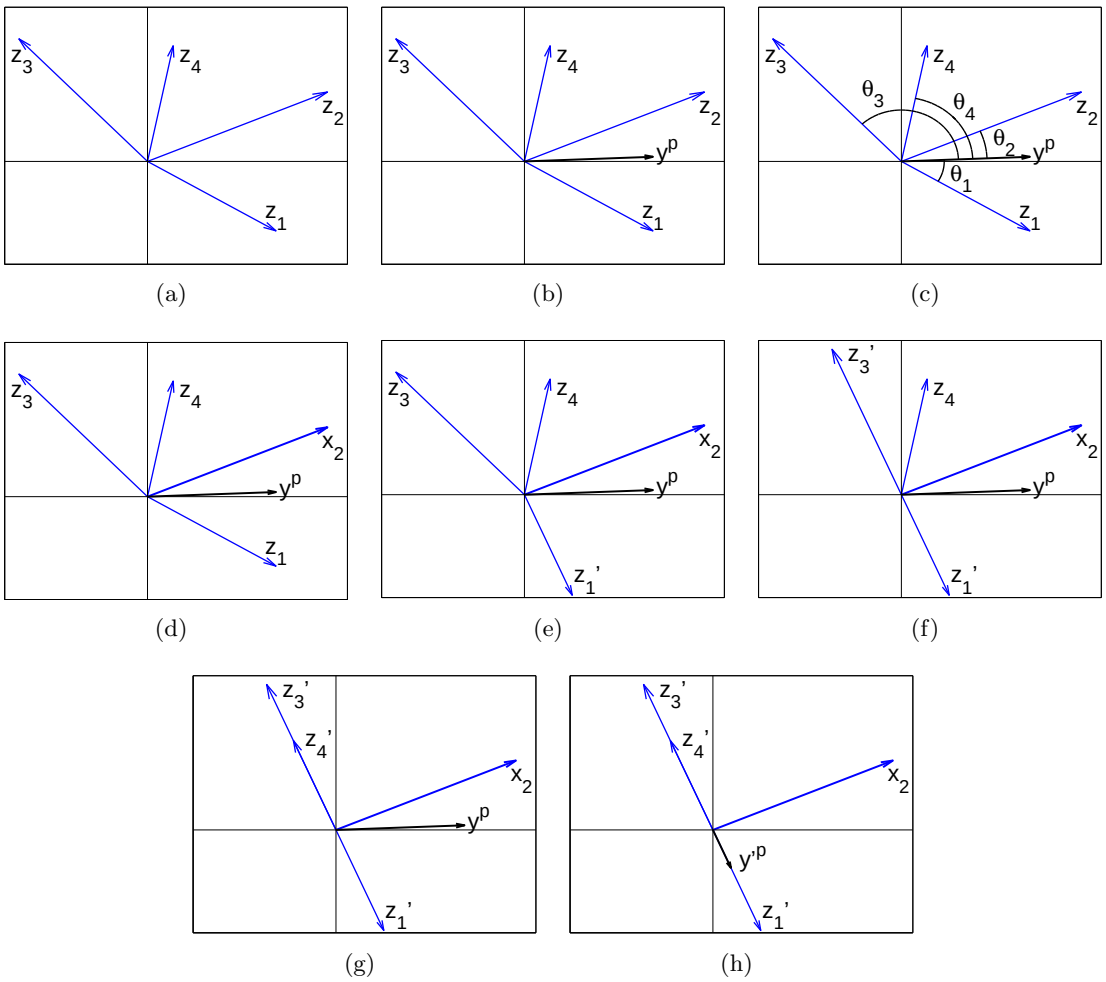


Figure 6.1: Example of one iteration of the OFR algorithm. (a) Initial input features z_i. (b) z_i and the output to be modeled y^p. (c) Step (i) of the Algorithm 6.1. (d) Step (ii) of the Algorithm 6.1. (e - h) Step (iii) of the Algorithm 6.1.

orthogonal to x_2. The Gram-Schmidt orthogonalization of the input features z_i is shown from Figure 6.1(e) to Figure 6.1(g), presenting the corresponding orthogonal feature z_i' in each figure. Finally, the last step is applying the orthogonalization to the output vector y^p as displayed in Figure 6.1(h). In this step, vector x_2 will be removed from the list of input features and the procedure will be repeated with the resting z_i'.

6.2.2 Orthogonal Forward Regression with random probe

The OFR algorithm presented in Section 6.2.1 sorts the input features based on their relevance. However, when this system is used, even though we give the selected features by the OFR algorithm to a classifier, we will introduce an overfitting to the system. Therefore, in this work, in order to control overfitting a variation of the algorithm is applied. This variation selects the optimal number of features depending on the data characteristics. This variation is known as random probe method (Stoppiglia *et al.*, 2003).

Random probe method refers to a method in which random generations of data are used

to verify that the analyzed data is more significant than random data. To compute the OFR with a random probe, one first creates a set of random probes. Then, one defines a risk level that corresponds to the risk that a feature might be kept despite being less relevant than the probe. The algorithm to compute the OFR with random probe it is summarized as presented in Algorithm 6.2 (Stoppiglia *et al.*, 2003).

Algorithm 6.2: Orthogonal Forward Regression with random probe

i. Define M sets of random probes.

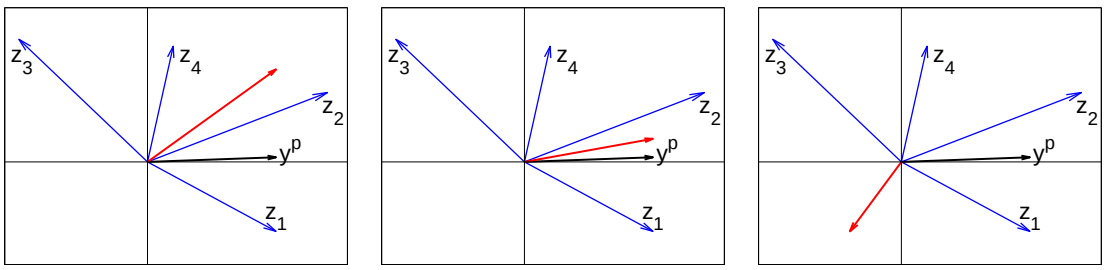
ii. Compute the OFR algorithm M times. Use a different random probe each time as input feature z_i.

iii. Compute the cumulative distribution function of the position achieved by the probe each time.

iv. Select a level of risk., e.g. 5%, 10%...

v. Select features that are ranked in a lower position than the level of risk defined.

Figure 6.2 presents an example of feature selection using OFR with random probe. For each figure a different random probe is given to the algorithm. Vectors presenting the features to be classified, z_i, are the same that have been used in Figure 6.1. Therefore, as presented in Section 6.2.1, if the OFR algorithm is applied, the first selected feature will be z_2. In examples presented in Figures 6.2(a) and 6.2(c), the random probe which is represented with the red vector, will not change this selection because the angle between the random probe and y^p it is bigger than the angle between z_2 and y^p. However, as presented in Figure 6.2(b), in some cases random probe may be more correlated with y^p than any of the z_i, in any of the iterations of the algorithm. Therefore in this case, the random data will be selected before than any of the other features, which it will mean that our system it is affected by noise. Once all the M random probes have been introduced to the OFR algorithm, the ranking of the positions of the random probes is computed to check the positions that the random probe achieved. Finally, with the position of the random probe, it is obtained which of the features are more relevant than the percentage of random data defined.



(a) (b) (c)

Figure 6.2: Example of application of the OFR with random probe algorithm. A different random probe is used in each figure, represented by the red vector.

6.2.3 Orthogonal Forward Regression with random probe and Frequency Pre-Selection

The OFR with probe algorithm has been used in previous studies to select the best features to compute classification on EEG data sets (Vialatte *et al.*, 2005a,b). However, in this study a variation of the OFR algorithm with probe is used. This variation is Orthogonal Forward Regression with random probe and Frequency Pre-Selection (OFRFPS). The main difference between OFR algorithm with probe and the variant implemented in this work lies in how the feature selection is performed. Whereas the OFR algorithm with probe presented in Section 6.2.2 sorts the given inputs and selects the optimal number of features to use, OFRFPS preselects the input features that are given to the OFR algorithm in each iteration.

As detailed in Chapter 4, a new frequency approach is presented to improve the early diagnosis of AD. However, if all features were given to a classifier overfitting would occur as presented in Section 6.1. On the other hand, optimal selected frequency range for each measure presents high correlation when compared. Therefore if the OFR algorithm is applied using those features as z_i, redundant information presented by correlated values will be lost during the orthogonalization, and no new information will be included. Consequently, a new OFR algorithm is presented to deal with these issues.

Each measure presents a set of f frequency ranges, as detailed in Section 4.4.1. The preselection is used to select the values V computed in one of these f frequency range as z_i. This process is repeated for each measure. The separability criterion $J(F, F+W)$, defined in Equation 4.27, is used to evaluate the frequency range that presents the highest differences between AD and healthy patients, and therefore is used as z_i. Then the OFR algorithm with all selected z_i is applied. On the first iteration, once the first feature x_j has been selected, the orthogonalization is not applied to the computed z_i. In this new approach, the orthogonalization is applied to values V computed in all f frequencies ranges of each measure. This process is repeated until all features have been selected. The new proposed OFRFPS is summarized as presented in Algorithm 6.3.

This new implementation of the OFR algorithm helps to reduce the redundant information, and therefore the overfitting, of the data. As in each iteration the input features are reselected, if during the orthogonalization a feature is completely decorrelated based in some others information, it will not be selected in the next iteration.

6.2.4 Feature selection for improving classification rate

Methods used to compute the CR using feature selection are the same that have already been presented in Chapter 4. The 11 measures presented in Section 4.2, RP, Correlation, Coherence, Granger Causality measures, OC and PS are used with the parameter configuration, i.e. window length and Granger order, presented in Section 4.5.

Once the OFRFPS algorithm is applied, the list of features that better characterize each data set is extracted. Then this list of features are used to compute the CR using LDA and LOO, the same methods that are presented in Section 4.4.4.

In the implementation of the OFRFPS algorithm computed in this work, random probes presented a special characteristic. Random probes are not generated using random data.

Algorithm 6.3: Orthogonal Forward Regression with random probe and frequency pre-selection

i. Define M sets of random probes with the same size of the used measures, i.e. with the same number of subjects (N) and frequencies ranges (f).

ii. Compute $J_i(F, F+W)$ for all f and for all p measures (including the used random probe).

iii. Define the input features z_i as the values V_i where the highest $J(F, F+W)$ it is achieved for each measure:

$$z_i = V_i(F, F+W) \quad : \quad max(J_i(F, F+W)) \quad \text{for} \quad i = 1 \ldots p$$

iv. Compute the correlation between all input features z_i and the output to be modeled y^p as:

$$r^2_{y^p,z_i} = cos^2\theta_i = \frac{(y^p \cdot z_i)^2}{(y^p \cdot y^p)(z_i \cdot z_i)}$$

v. Define the selected feature (x_j) as the one that maximize the correlation between z_i and y^p:

$$x_j = z_i \quad : \quad max(r^2_{y^p,z_i})$$

vi. Project y^p onto the null space orthogonal to the selected feature x_j:

$$y'^p = y^p - \frac{y^p \cdot x_j}{x_j \cdot x_j} x_j$$

vii. Project $V_i(F, F+W)$ onto the null space orthogonal to the selected feature x_j:

$$V_i'(F, F+W) = V_i(F, F+W) - \frac{V_i(F,F+W) \cdot x_j}{x_j \cdot x_j} x_j \quad \text{for} \quad F, F+W = 1 \ldots f \quad i = 1 \ldots p$$

viii. Remove x_j from the list of input feature z_i.

ix. Return to (ii) until all features have been selected.

x. Repeat from step (ii) to (ix) M times. Use a different random probe each time.

xi. Select a level of risk., e.g. 5%, 10%...

xii. Select features that are ranked in a lower position than the level of risk defined.

Instead, surrogate probes are generated with the same characteristics as the original data, with a different measure (or a synchrony measure or RP) used to generate each probe. The values of a specific measure for the two populations (AD patients and control subjects) are mixed together, and then labels for each class are assigned randomly. This process is repeated 500 times for each measure. Therefore 5.500 probes are computed and added to the feature set to quantify the degree of overfitting.

6.3 Results using feature selection

Feature selection is used to improve the early diagnosis of patients with AD. Results presented in this section show that once the algorithm of the OFRFPS it is used, the CR improves as more features are used. Results for the MCI Data Set are presented in Section 6.3.1, results for the Mild AD Data Set 1 are detailed in Section 6.3.2 and results for the Mild AD Data Set 2 are shown in Section 6.3.3.

6.3.1 Results MCI Data Set

Results obtained for the MCI Data Set are shown below. Table 6.1 presents the measures in the order that are selected by the OFRFPS algorithm. Selected OFRFPS frequency ranges and the corresponding standard frequencies are also presented in the same table.

Algorithm Order	**Features**	**OFRFPS Selected frequency ranges (Hz)**	**Standard frequency bands (Hz)**
1	RP	2 - 8	4 - 8
2	Correlation	3 - 8	4 - 8
3	Coherence	1 - 6	1 - 4
4	PDC	1 - 3	1 - 4
5	ffDTF	9 - 29	13 - 30
6	OC	24 - 25	13 - 30
7	GC	1 - 30	1 - 4, 4 - 8, 8 - 13, 8 - 13
8	DTF	4 - 5	4 - 8
9	PC	1 - 10	1 - 4, 4 - 8
10	PS	28 - 30	13 - 30
11	dDTF	1 - 2	1 - 4

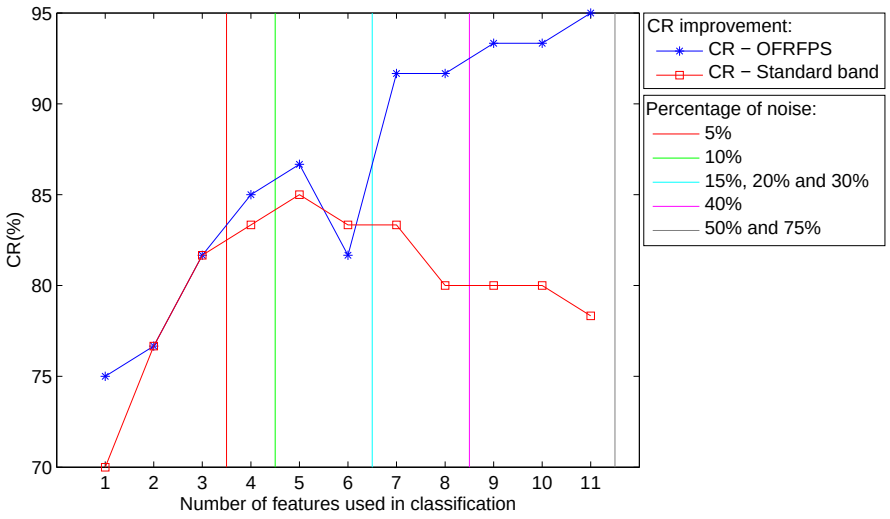
Table 6.1: Features and frequency ranges selected by the OFRFPS algorithm for MCI Data Set. The last column presents the standard frequency bands corresponding to the measures.

Results detailed in Table 6.1 show that the first measure selected by the OFRFPS algorithm is RP. Therefore, RP is the best feature for differentiating MCI patients from healthy subjects in the frequency range 2 - 8 Hz. This result is consistent with the obtained in Section 4.5.1, where the best CR is obtained with RP. However, in Section 4.5.1 the frequency range defined as optimal it is 2 - 9 Hz. This change is due to the parameter selection, that in Section 4.5.1 is only based on the best CR achieved, and using the OFRFPS it is based on the maximum value of $J(F, F + W)$.

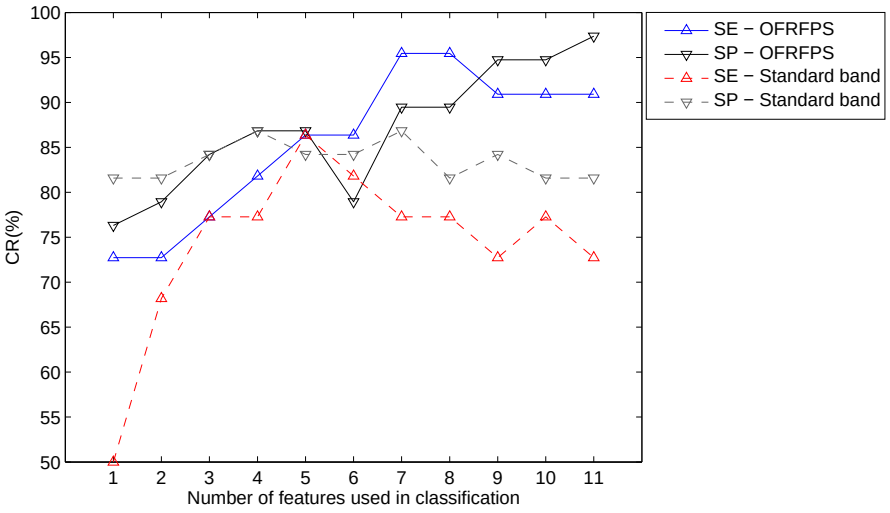
The first synchrony measure that appears in the OFRFPS results is Correlation. It appears in the range of 3 - 8 Hz, close to the 1 - 8 Hz defined as optimal frequency range in Section 4.5.1. However, using the OFRFPS it appears to be more discriminative than when it is used as a single feature, because in this results appears as the second most discriminative feature whereas when used alone it appeared as the third most discriminative feature. dDTF, which it is the synchrony measure that achieved the best CR when the optimal frequency range is used, in this case appears in the last position of the ranking. This result may indicate that some of the features selected in lower positions made dDTF lost part of its redundant information during the orthogonalization process. None of the selected frequency ranges presents the same results than any of the standard frequency ranges. Correlation and PDC are the only ones that present a close value to standard frequency ranges, 3 - 8 Hz and 1 - 3 Hz respectively. Any of the others is close to a standard band.

In order to study the improvement in the CR presented from including more of the selected features as input features to a classifier, the evolution between the obtained CR and the number of features used is examined. It is also examined the improvement of performance by computing the CR using the features and frequency range selected by the OFRFPS algorithm, and the same features in the standards frequency bands. SE and SP are also evaluated for the different numbers of input features. Figure 6.3 presents this relationship for MCI subjects. Figure 6.3(a) represents the CR evolution, comparing the CR obtained with selected OFRFPS frequency ranges and the CR computed with the standards frequency bands. Vertical lines indicate the percentage of noise introduced in the data set using the probe method. Figure 6.3(b) presents the evolution of SE and SP in comparison with the number of features used.

Results obtained using the OFRFPS selected features in Figure 6.3(a) present a clear improvement as more features are used in the classification. Using one measure obtained CR is 75.00%. When more features are used, CR achieves a 86.67% when using five features, with a level of significance lower than 15%, but then when the sixth feature is used (which is OC as presented in Table 6.1), CR drops down its value to 81.67%. However, when seven or more features are used the CR ascends again to high CR values, achieving 95.00% when the eleven features are used. However when using eleven features the level of significance is lower than 50%. Figure 6.3(a) presents that CR obtained with the OFRFPS selected features is higher than CR obtained in the standard bands. Only when there is the decrease in the sixth feature CR using OFRFPS is lower than the obtained in the standard bands. CR obtained in the standard bands starts with a CR of 70.00%, and increases its value to a maximum of 85.00% when five features are used, but then if more features are used, the obtained CR does not improve its value but decreases. Values presented for SE and SP in Figure 6.3(b) present the same evolution. There is an improvement in the values obtained as more features are used for the classification. Interestingly, SP starts being higher than SE. Then when OC is used SE has the same value and SP drops down his value and the global CR. Finally, when nine, ten or eleven features are used SP present



(a) CR evolution



(b) SN and SP evolution

Figure 6.3: Results of applying the selected OFRFPS features for the MCI Data Set.

again higher values than SE. This result indicates that the decrease of the CR is due to the difficult of OC in differentiating healthy subjects.

6.3.2 Results Mild AD Data Set 1

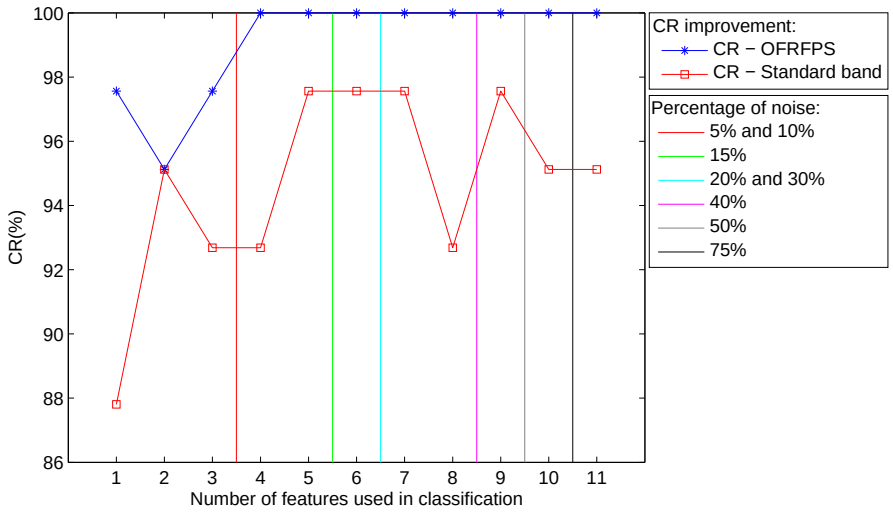
Results obtained for the Mild AD Data Set 1 are shown in this section. Table 6.2 presents the measures in the order that are selected by the OFRFPS algorithm. Selected OFRFPS frequency ranges and the corresponding standard frequencies are also presented in the same table.

Algorithm Order	Features	OFRFPS Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	4 - 7	4 - 8
2	GC	1 - 2	1 - 4
3	Correlation	9 - 10	8 - 13
4	PS	25 - 26	13 - 30
5	PC	13 - 14	13 - 30
6	dDTF	2 - 6	4 - 8
7	Coherence	5 - 6	4 - 8
8	OC	11 - 14	8 - 13
9	ffDTF	6 - 19	8 - 13
10	DTF	20 - 21	13 - 30
11	PDC	1 - 2	1 - 4

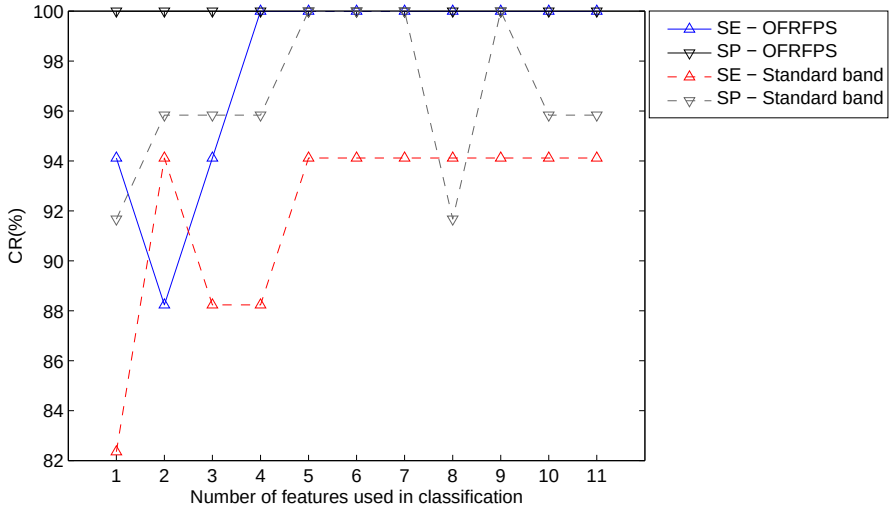
Table 6.2: Features and frequency ranges selected by the OFRFPS algorithm for Mild AD Data Set 1. The last column presents the standard frequency bands corresponding to the measures.

Results presented in Table 6.2 show that the first measure selected by the OFRFPS algorithm is RP. Interestingly, the frequency range which has been selected is 4 - 7 Hz, which is the same that is presented as optimal in Section 4.5.2. Therefore, RP and the frequency range of 4 - 7 Hz it is presented as the best feature for differentiating Mild AD patients from healthy subjects in this data set. The first synchrony measure that appears in the OFRFPS results is GC. It appears in the range of 1 - 2 Hz, again is the same that is presented as optimal in Section 4.5.2. Results presented in Section 4.5.2 showed that GC is the second synchrony measure that achieved the best CR. The first synchrony measure is DTF. In this case GC has been presented more discriminative than DTF, which achieved the tenth position. The third best feature is Correlation, which as happens in results for the MCI Data Set presented in Section 6.3.1, is presented as one of the three best features to explain this data set. Results presented in Chapter 4 show that high correlation is found in the optimal frequency range for the Granger measures (results presented in Figure 4.10, pag. 73). During the orthogonalization process this information is eliminated. This may explain the lower positions achieved for some of the Granger measures such as ffDTF, DTF and PDC which achieved high CR when used alone, but in this join analysis they are ranked in the lasts positions. Regarding the frequencies ranges selected for the OFRFPS algorithm, any of these presents the same results than any of the standards frequency bands. RP is the only that present a close value to standard α band with 4 - 7 Hz. Any of the others is close to a standard band.

The evolution between the obtained CR and the use for classification of different number of features select by the OFRFPS algorithm is examined. Figure 6.4 presents this evolution for the Mild AD Data Set 1. Figure 6.4(a) represents the CR evolution, comparing the CR obtained with selected OFRFPS frequency ranges and the CR computed with the standards frequency bands. Vertical lines indicate the percentage of noise introduced in



(a) CR evolution



(b) SN and SP evolution

Figure 6.4: Results of applying the selected OFRFPS features for the Mild AD Data Set 1.

the data set using the probe method. Figure 6.4(b) shows the evolution of SE and SP in comparison with the number of features used.

Results obtained using the OFRFPS selected features in Figure 6.4(a) present a clear improvement as more features are used in the classification. When RP alone is used, CR is 97.56%, the same value that is achieved in Section 4.5.2. Then when GC it is used together with RP the CR drops down to 95.12%, which is the lowest value obtained. Using three features the classification improves again to 97.56%, and then when four features are used, the CR reaches the values of 100% with a level of significance bigger than 10% and lower than 15%. The value of 100% persists when more features are used in classification. CR obtained using the standards frequency bands do not improve the previous CR. When two features are used is the only moment that the two CR have the same value. For all other number of features, the CR obtained using the OFRFPS parameters are always higher. CR obtained using the standard bands, present different increases and decreases of

Algorithm Order	Features	OFRFPS Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	OC	17 - 18	13 - 30
2	PS	14 - 15	13 - 30
3	DTF	10 - 11	8 - 13
4	Correlation	1 - 2	1 - 4
5	dDTF	26 - 27	13 - 30
6	GC	17 - 18	13 - 30
7	RP	10 - 11	8 - 13
8	Coherence	27 - 28	13 - 30
9	PC	23 - 29	13 - 30
10	ffDTF	15 - 16	13 - 30
11	PDC	16 - 17	13 - 30

Table 6.3: Features and frequency ranges selected by the OFRFPS algorithm for Mild AD Data Set 2. The last column presents the standard frequency bands corresponding to the measures.

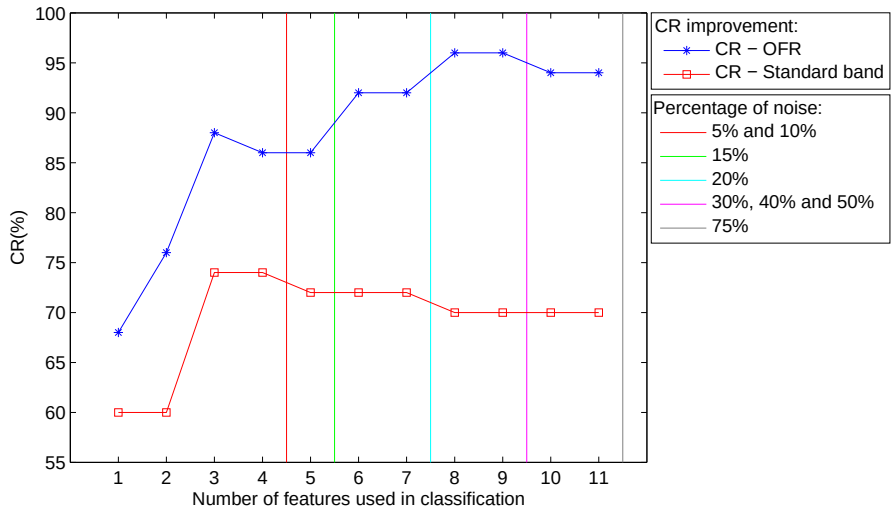
the CR value, which may demonstrate that the ensemble use is not optimal. The highest value obtained it is 97.56% when five, six, seven and nine features are used, which is the same value that using the selected features has been achieved with one or three features.

Results presented for SE and SP in Figure 6.4(b) show that the value of SP is always at 100%, and the value that changes it is the SE when the OFRFPS features are used. This may indicate that adding more synchrony measures can characterize Mild AD patients better.

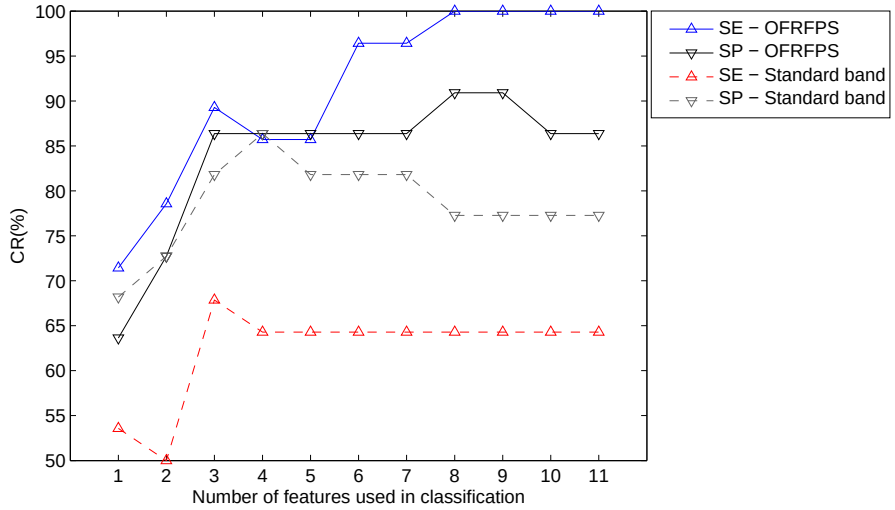
6.3.3 Results Mild AD Data Set 2

This section presents the results obtained for the Mild AD Data Set 2 using the OFRFPS algorithm. Table 6.3 shows the measures in the order that are selected by the algorithm. Selected OFRFPS frequency ranges and the corresponding standard frequencies are also presented in the same table.

Results presented in Table 6.3 show that the first measure selected by the OFRFPS algorithm is OC in the frequency range of 17 - 18 Hz. Interestingly, for this data set the first selected measure it is a synchrony measure, whereas for the MCI Data Set and the Mild AD Data Set 1 RP appears as the first feature. In Section 4.5.3 the frequency range defined as optimal for OC it is 16 - 21 Hz. In this case the selected OFRFPS frequency range is a narrow band frequency range of 1 Hz, which is inside the optimal frequency range. The second selected feature presented in Table 6.3 is PS in the frequency range of 14 - 15 Hz. This frequency range is different to the defined as optimal in Section 4.5.3. The third and fourth OFRFPS features are respectively DTF in the frequency range of



(a) CR evolution



(b) SN and SP evolution

Figure 6.5: Results of applying the selected OFRFPS features for the Mild AD Data Set 2.

10 - 11 Hz and Correlation in the frequency range of 1 - 2 Hz. Results show in Table 4.6, Section 4.5.3 pag. 70, present this four features as the four which obtained the best CR when used individually, presenting a performance in the classification of 76.00% for OC and PS, and a 74.00 % for DTF and Correlation. However, frequency ranges presented are not the ones defined as optimal. For OC and DTF the OFRFPS frequency range is a narrow band frequency range inside the optimal. For PS and Correlation the OFRFPS frequency range is found outside the optimal frequency range. Measures found in the fifth and sixth positions are a couple of Granger Measures, dDTF and GC respectively. Then RP appears in the seventh position in the frequency range of 10 - 11 Hz. For this data set, results presented for RP are not as good as the reported for the other data sets, in which RP achieved the best CR when used individually and is selected as the first measure for the OFRFPS algorithm.

The evolution between the obtained CR and the use for classification of different number

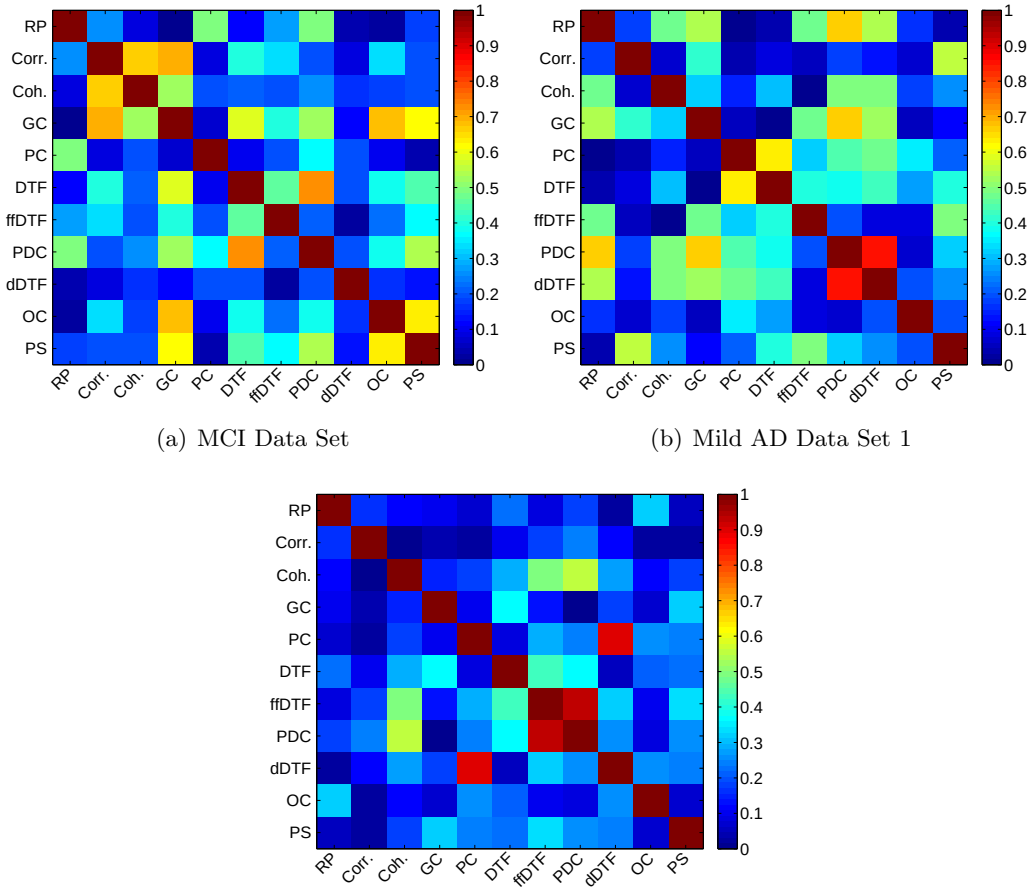
of features select by the OFRFPS algorithm is also examined. Figure 6.5 presents this evolution for the Mild AD Data Set 2. Figure 6.5(a) represents the CR evolution, comparing the CR obtained with selected OFRFPS frequency bands and the CR computed with the standards frequency bands. Vertical lines indicate the percentage of noise introduced in the data set using the probe method. Figure 6.5(b) presents the evolution of SE and SP in comparison with the number of features used.

Results obtained using the OFRFPS selected features in Figure 6.5(a) show that there is an improvement of the CR when more of the OFRFPS features are used as a input features to a classifier. When OC alone it is used CR is equal to 68.00%. Then when PS it is used together with OC the CR achieved it is 76.00%. Including DTF to the classifier improves the CR to 88.00%, but when Correlation and dDTF are used the CR drops down to 86.00%, with a percentage of significance lower than 15%. The use of GC improves the CR to 92.00%, which is a maintained when RP it is included in the used features. The best CR obtained for this data set it is 96.00% using eight and nine features, with a percentage of significance lower than 30%. The last CR obtained is 94.00% when ten or eleven features are used, and with a percentage of significance lower than 75%. This is a small decrease in the CR, a 2% that represents only one subject of the total 50. CR obtained using the standards frequency bands do not improve the previous CR. When one or two features are used a 60.00% is obtained. When three or four features are used, the CR achieves its maximum value obtained of 74.00%, but then when more features are added there is a decrease of the CR presenting a stable final value of 70.00%. Results presented in Figure 6.5(a) show that CR obtained using the OFRFPS frequency ranges are better than the obtained using the standards frequency bands. Results presented for SE and SP in Figure 6.5(b) show again that including more features helps to improve the performance. In this case, however, there is a clear improvement of the value of SE, which when eight features are used reaches the value of 100%, and the value of SP present stable values between 86.36% and 90.91%.

6.3.4 Results comparison

Results shown for the three data sets present a clear improvement of the CR when the OFRFPS parameters are used. However, the selected OFRFPS configuration for each data set it is different. Results shown for the MCI Data Set and the Mild AD Data Set 1 present RP as the most discriminative feature, ranking it in the first position for both data sets. Selected frequency ranges for the RP in each data set are similar, selecting 2 - 8 Hz in the MCI Data Set and 4 - 7 Hz in the Mild AD Data Set 1. These frequency ranges are where the effect of slowing of EEG presents the higher differences, and both are close to the θ band, presenting higher values for the AD patients. Results detailed for the Mild AD Data Set 2 present RP in the seventh position, and the selected frequency range is 10 - 11 Hz. This frequency range is also affected for the effect of slowing of EEG, however healthy subjects present higher values than the AD patients. Results obtained using RP for the Mild AD Data Set 2 in this chapter are consistent with the obtained in Chapter 4, where RP presented low CR values.

On the other hand, when synchrony measures are used obtained CR is improved for all data sets. For the three data sets Correlation appears in the top four measures, being the only measure that it is used in the four initial positions for all data sets. However, the OFRFPS selected frequency ranges do not present the same tendency, because selected



(a) MCI Data Set

(b) Mild AD Data Set 1

(c) Mild AD Data Set 2

Figure 6.6: Modulus of the correlation coefficient computed between all the measures for each data set. Frequencies selected by the OFRFPS algorithm are used.

frequency range are different for each data set. Even though all the other synchrony measures used present a high variability, some similitude can be found between Mild AD Data Set 1 and Mild AD Data Set 2. In both data sets some measures achieved different but close position. PS it is the fourth selected feature of the Mild AD Data Set 1 and the second for the Mild AD Data Set 2. Other measures, such as dDTF, Coherence and ffDTF achieved positions that varied only one number between data sets. Finally PDC appears for both data sets as the least meaningful feature, being the last selected in both cases.

To check the redundant information that the selected features provide, correlation between measures is computed as in Section 4.6. However, this time the OFRFPS selected frequency ranges are used. The modulus of the results are presented in Figure 6.6. We observe that correlation values are now lower than those presented in Figure 4.10 (pag. 73). Figure 6.6(a) shows the obtained results for the MCI Data Set. For this data set the highest correlation is obtained between PDC and DTF ($|r| = 0.72$). These values are ranked in the lasts positions of the OFRFPS features, therefore small redundancy is introduced. All the other values are smaller than 0.70. Results obtained for the Mild AD Data Set 1 presented in Figure 6.6(b) show an important decrease compared to the ones depicted in Figure 4.10. Now, only the correlation between PDC and dDTF presents a

high value ($|r| = 0.85$), in contrast with the six pairs obtained without using OFRFPS. All other pairs of correlation presents lower values that $|r| < 0.671$, demonstrating the decrease of redundant information. For the Mild AD Data Set 1, the existing high correlation between PDC and dDTF it is not meaningful, because PDC is found as the last measure in the OFRFPS algorithm, and the CR of 100% is obtained previously. Finally, results presented in Figure 6.6(c) show the correlation values obtained for the Mild AD Data Set 2. For this data set, the high correlation between pairs of Granger measures, PC-dDTF and PDC-ffDTF, found in Section 4.6 present not changes. However, a decrease of the maximum values is found for the other measures. As it happened with the Mild AD Data Set 1, this high correlation is not meaningful, because these measures are ranked in the lasts positions of the OFRFPS algorithm.

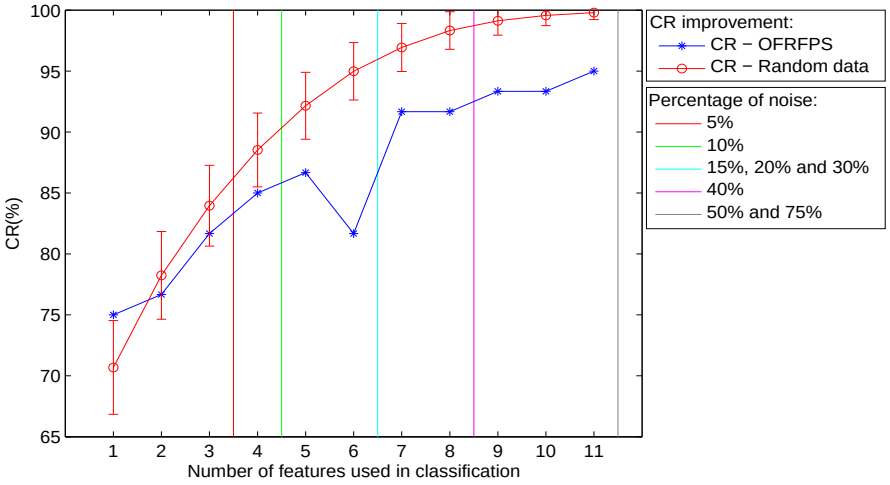
6.4 Statistical analysis

Results shown in Section 6.3 present that there is an improvement of the CR obtained when more features are used into the classifier. However, as the data sets used are small, the same subset used to compute the OFRFPS is the subset used to compute the classification. This of course, introduces an important overfitting to the system.

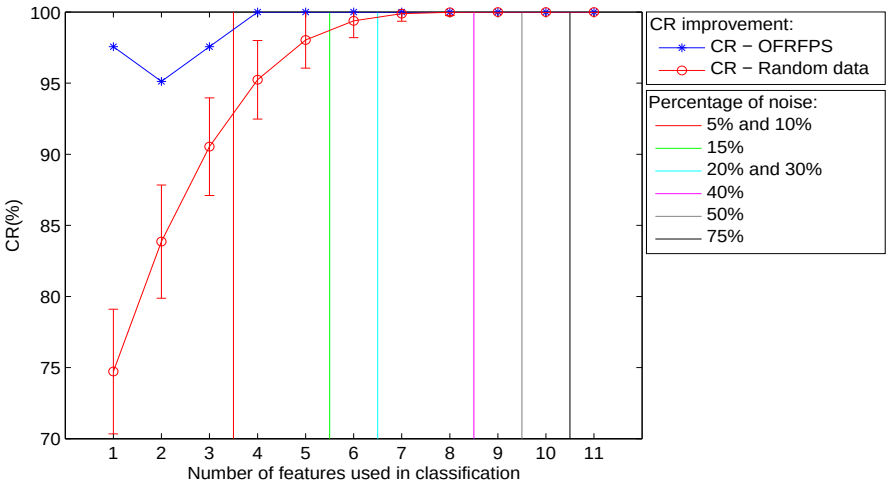
To deal with the problem of compute the overfitting introduced to the system, the random probe method is used to study the level of noise that is introduced to the system during the feature selection procedure. Moreover, to evaluate the level of overfitting introduced to the system due to the methodology used, another experiment is done using random data. In this experiment computed measures are substituted for random data following a normal distribution ($N(0, 1)$). The CR is computed with the same procedure as defined in previous sections. As the three used data sets present different number of subjects, this procedure is repeated for each data set using the same number of subjects in the data simulation. The computation of the CR using random data is repeated 1.000 times. Results presenting a comparison between the CR obtained for each data set and the mean and the standard deviation of the CR obtained using the random data and are presented in Figure 6.7. In this figure, the levels of noise defined for the random probe method are also presented.

Results presented in Figure 6.7 show that for the MCI Data Set (Figure 6.7(a)) and the Mild AD Data Set 2 (Figure 6.7(c)) CR results using random data achieve better classification results than the obtained using the measures computed for each specific data set. On the other hand, results presented in Figure 6.7(b) for the Mild AD Data Set 1 show that the CR obtained achieve better classification performance than the random data. Results presented in Figure 6.7 suggest that as the random data is classified achieving a 100% in all the figures there is a high overfitting in the presented results. Therefore, a cross-validation method has to be used in the feature selection procedure to avoid such overfitting.

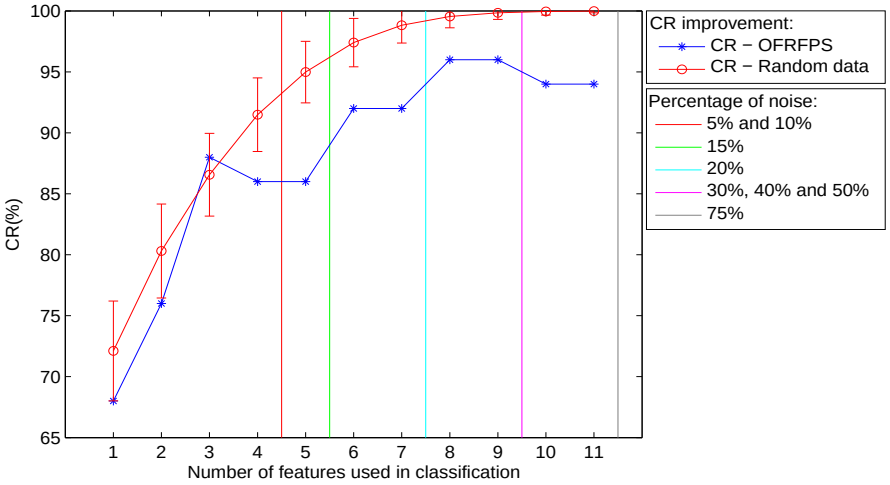
Presented results in Figures 6.7(a) and 6.7(c) also show that when the number of features used is lower than three (for the Mild AD Data Set 2) or four (for the MCI Data Set), CR obtained with the random data is not much more different than the CR obtained with the OFRFPS parameters. Therefore the level of noise is small for those configurations. This result it is consistent with the presented for the random probe levels of noise, being the CR obtained below the first level of noise (5%) the ones that do not present a big



(a) MCI Data Set



(b) Mild AD Data Set 1



(c) Mild AD Data Set 2

Figure 6.7: CR comparision between the OFRFPS features and random data.

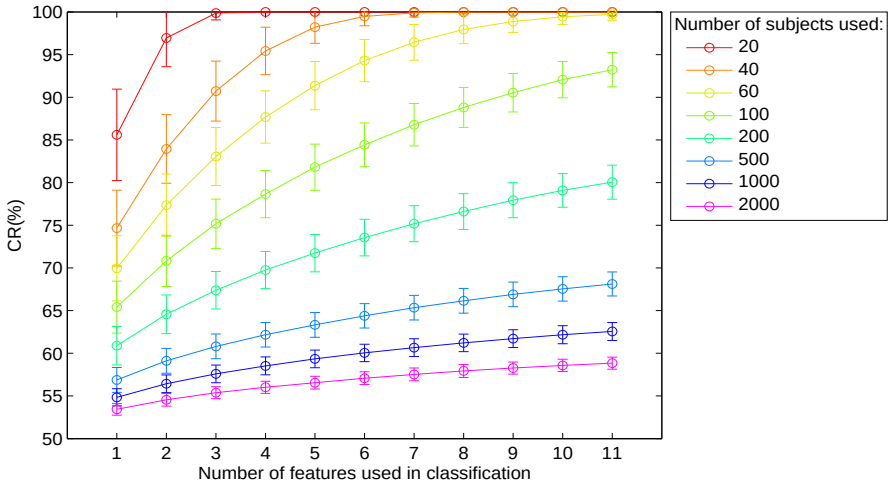


Figure 6.8: CR comparison, using random data with different number of subjects and the OFRFPS method without cross-validation.

difference. Then when the percentage of noise defined by the random probe is high, for example 20% the CR obtained with the random data achieves values close to 100% of performance, presenting that this results are highly overfitted.

Another conclusion can be extracted from results presented in Figure 6.7. CR results presented for the random data in the MCI Data Set and in the Mild AD Data Set 2, present a lower increase that those presented for the Mild AD Data Set 1. This result is due to the total number of subjects contained in the data base as presented in Figure 6.8. This figure presents the CR evolution for random data that simulates different number of subjects in the data base. Figure 6.8 shows that as more subjects are used in the random data base, results more close are to the 50% of performance. Interestingly, there is always an increase of the performance when more features are used, presenting the efficiency of the algorithm to select the optimal features, even though these are random. Figure 6.8 also present that with the methodology used, a database containing a high number of subjects would have to be used to present results more significant than the noise.

6.5 Discussion

In this chapter, it has been investigated the join use of synchrony measures and a frequency power measure in the whole set of frequency ranges existing between 1 and 30 Hz. Using this two types of measures, we are only modeling two of the three effects that AD cause on EEG data as detailed in Chapter 2. The third type of change that AD causes on EEG signals is usually modeled by complexity measures. However, this type of measures present a high correlation with RP as presented in (Dauwels *et al.*, 2011). As we use OFRFPS, each extracted feature is orthogonalized with respect to the previous extracted ones. Since RP have been selected for two of the three data sets, as the foremost discriminative feature, complexity measures would be decorrelated on this basis. The results obtained in this chapter in comparison with the ones obtained in Chapter 4 show that using a single measure, the classification is not as robust as can be with more attributes, and that a combination of RP and synchrony measures results in better classification performance.

Furthermore, the use of specific frequency ranges for each measure improves the classification performance in comparison with the results obtained in the classical frequency bands (δ, θ, α and β).

Obtained results show that combining two of the well-known changes that AD causes on EEG data can help to improve the ability to distinguish between AD patients and healthy subjects. The effect of the slowing of EEG, characterized by RP appears to be more discriminative than the changes in synchrony for the MCI Data Set and the Mild AD Data Set 1. These results agree with the results of different studies presented in (Vecchio *et al.*, 2013), where the effect of slowing of EEG can be used to predict the progression from MCI to dementia. On the other hand, combining this effect with the changes of synchrony, modeled with Coherence or Correlation for example, it has been as presented that increase the ability of differentiation between AD patients and healthy subjects. Changes of synchrony on EEG signals have been related with changes in functional connections between cortical regions (Jeong, 2004), and brain cortical and subcortical atrophy (Babiloni *et al.*, 2011). Therefore, presented results may suggest that combining the different effects that patients with AD present can make easier to differentiate between a patient with AD an another healthy.

Presented results in Section 6.3, show that there is an improvement of the CR obtained when more features are used into the classifier. However, as the data sets used are small, the same subset used to compute the OFRFPS is the subset used to compute the classification. This of course, introduces an overfitting to the system. For the MCI data set, using the eleven measures as input features for the classifier, a CR of 95.00% is achieved the best result obtained with this data set. However, the level of significance using random probes is 50% as presented in Figure 6.3(a), and has been confirmed for Figure 6.7(a). This indicates that these results are overfitted. For the Mild AD Data Set 1, a CR of 100% was achieved using four features. The level of significance at which this value was obtained is less than 15% (presented in Figure 6.4(a)), which indicates that those measures are able to clearly identify AD patients in an advanced stage of the disease. Results presented of the Mild AD Data Set 2 obtained a CR of 96.00% and the level of significance using random probes is between 20% and 30 % as presented in Figure 6.5(a). Presented results in Figure 6.7(c) also show an important overfitting in this data set.

Finally, obtained results may indicate that MCI is a stage difficult to identify in comparison with the Mild AD stage. In the case of MCI, patients start to present some memory impairments but preserve other cognitive domains, whereas in the Mild AD stage subjects begin to display some cognitive deficits. On the other hand, differences observed between Mild AD 1 and Mild AD 2 data sets show that the Mild AD Data Set 2 is more difficult to classify. These two data sets both contain Mild AD patients and healthy subjects. However, as presented in the data set description in Section 4.2, the healthy subjects used in both data sets are different. Subjects used as healthy in the Mild AD Data Set 1 are healthy volunteers and had normal EEGs, whereas that healthy subjects used in the Mild AD Data Set 2 are subjects with memory complaints and suspected AD which went to the hospital and a neurologist determined if they had AD or not. This issue may explain the differences obtained using the two data sets, and show that the results obtained with the Mild AD Data Set 2 clearly show that used methods are adequate for the diagnosis of AD.

Chapter 7

Feature selection using cross-validation

This chapter is devoted to improve results presented in Chapter 6. Results obtained in the previous chapter achieved high classification results. However, it has been showed that some of these are overfitted. To deal with this problem this chapter details the use of cross-validation during the feature selection to reduce overfitting.

This chapter presents an introduction on Section 7.1 and then different methodologies are described. Section 7.2 applies the LOO cross-validation to the features selection, and the k-fold validation is then applied in Section 7.3. Last section of this chapter is devoted to discussion.

7.1 Introduction

Results presented using feature selection clearly improves the ratio in which AD patients and healthy subjects are classified. However, as presented in Section 6.4, using the OFRFPS method improves the CR when using the computed features and also when using random noise.

The presented overfitting is due to the small number of subjects in the data bases. To deal with this shortcoming the usage of cross-validation has been defined in Section 4.4.4. Usually cross-validation is applied in the classification step, a part of the subjects are used for training the system and the other part are used to validate the system.

Cross-validation presented in this chapter is applied in the same way. During the OFRFPS algorithm, the cross-validation is done by leaving a group of subjects out of the study. This present some variability on the input data and show if the presented approach is stable along the variations. Once the cross-validation has been used and the features had been selected, the classification is done as in Chapters 4, 5 and 6, by doing LOO cross-validation using all subjects except one to train a classifier, and the remaining subject is used to evaluate the performance.

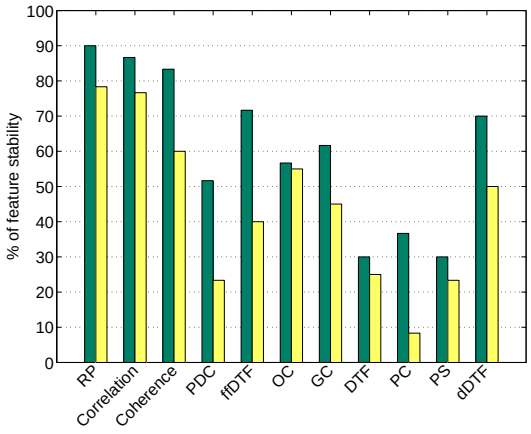
7.2 Feature cross-validation using Leave-One-Out

Presented results in Chapter 6 show the need of use feature cross-validation in the feature selection process. In order to study stability across subjects who present the selected OFRFPS features, an experiment is carried out using feature selection through LOO cross-validation. This cross-validation is performed by leaving a different subject out of the study in each iteration, with the aim of checking whether the selected features are stable all along the data set. Therefore, the OFRFPS algorithm is computed in each data set the same number of times that subjects the data base has, leaving one different subject out each time. Figure 7.1 presents the results obtained for the three data sets. In this figures, for each measure, blue column stands for the % of times that each measure is selected in the same order as by the OFRFPS algorithm and green column stands for the % of times that a feature is selected in the same order and with the same frequency range as by the OFR algorithm. Features are listed in the same order as that obtained using the OFRFPS algorithm.

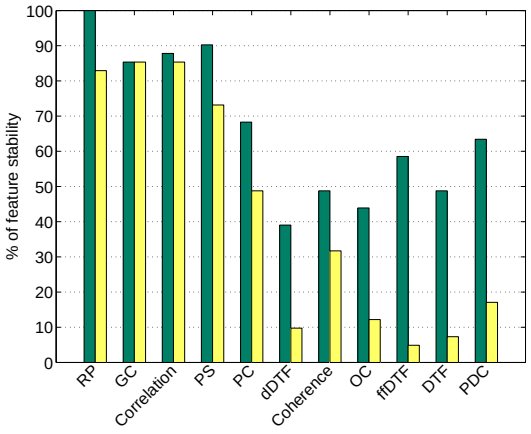
Results presented in Figure 7.1(a) show the obtained results for the MCI Data Set. In this figure is shown that RP is the most stable feature selected by the 90.00% of the iterations as the first feature. Then Correlation and Coherence are selected an 86.67% and an 83.33% respectively. For all the other features there is variability in the percentage of feature stability, varying from the 30.00% obtained with DTF and PS to the 71.67% obtained with ffDTF. Values obtained for the stability of frequency selection follows the same pattern that the selection of feature. The three first measures are the higher and then it changes depending on the feature. Results for the Mild AD Data Set 1 are presented in Figure 7.1(b), for this data set RP is selected as first feature for all the iterations, achieving a 100.00% of feature stability. Then GC, Correlation and PS are stable across the variation of subjects, achieving 85.37%, 87.80% and 90.24% respectively, and then other features present different percentages values. As for the MCI Data Set, results presented for the Mild AD Data Set 1, present high stability in the frequency selection for the first three features, but then there is a decrease of the value for other measures. Results for the Mild AD Data Set 2 are presented in Figure 7.1(c). This figure present a lower stability in the percentage of times that a feature is selected being PDC, the last of the features selected by the OFRFPS, the one that is presented as most stable with a value of 86.00%. The second that presents more stability it is OC, ranking a value of 84.00%. Measures that are ranked between second and sixth positions, i.e. PS, DTF, Correlation, dDTF and GC, present a percentage of stability lower or equal to 50.00%. This result suggests that there is a high variability between subjects, and that these are divided into two groups. Interestingly, this data set present high similitude between the percentage of ranking the feature and the frequency range which in the first five measures, present the same value or lower only a 2.00%.

Results at hand show that there is stability in the feature selection, when then the OFRFPS algorithm is repeated leaving one subject out each time. This result suggest that if cross-validation is applied, it may affect more at the random data than at the computed measures, and therefore CR of the random data will be drooped down when cross-validation will be used.

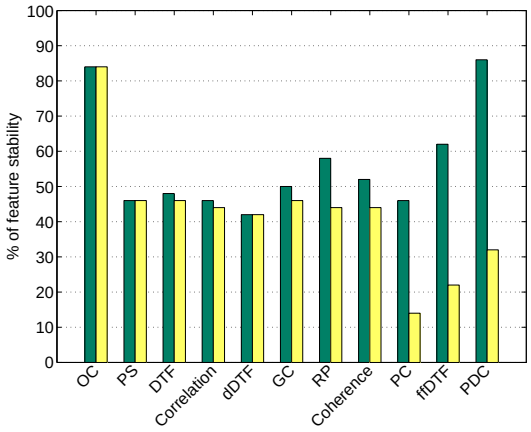
Once the cross-validation is used, for the MCI Data Set the selected features are the same that have been defined in Table 6.1 (pag. 100). Therefore the use of LOO cross-validation do not affects the performance of the classification. Results presented for the Mild AD Data



(a) MCI Data Set



(b) Mild AD Data Set 1

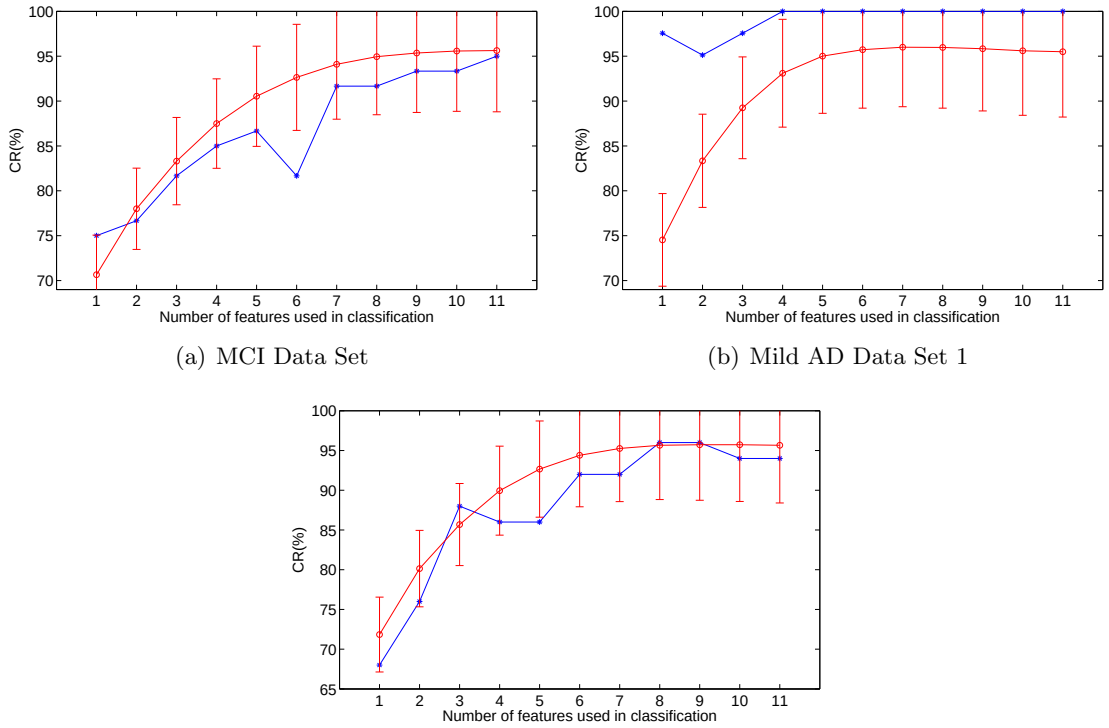


(c) Mild AD Data Set 2

Figure 7.1: Results of computing the LOO feature cross-validation in the OFRFPS algorithm. Percentage of stability is presented for each data set. Features are presented in the same order than the OFRFPS select each.

Set 1 are similar to those presented in Table 6.2 (pag. 103). When LOO cross-validation is applied for this data set, the order of the selected features is the same than the one presented in Table 6.2. However, selected frequency range present a small difference after applying cross-validation. Measures ranked in positions eight (OC), nine (ffDTF) and ten (DTF), changed the selected frequency range. OC enlarged the frequency range from 11 - 14 Hz to 11 - 17 Hz. ffDTF reduced the frequency range from 6 - 19 Hz to 18 - 19 Hz and DTF changed the frequency range from 20 - 21 Hz to 25 - 26 Hz. Selected features for the Mild AD Data Set 2 are the same that have been presented in Table 6.3 (pag. 105), being selected without using cross-validation. The only results that present some variation when using LOO cross-validation are for Mild AD Data Set 1. These results are consistent with results shown in Figure 7.1(b), when lasts positions of the OFRFPS algorithm showed low percentage of selection, whereas high percentage of selection are obtained for the other data sets.

Selected features using LOO cross-validation are used to discriminate between AD patients and healthy subjects. Figure 7.2 presents the CR obtained after applying the features selected by LOO cross-validation (in blue), and also applying LOO cross-validation to



(a) MCI Data Set

(b) Mild AD Data Set 1

(c) Mild AD Data Set 2

Figure 7.2: CR comparison between the OFRFPS features and random data. Features used to compute the classification are selected using LOO cross-validation.

simulated data (in red). For this simulation also 1.000 realizations of random data are performed. Obtained results for each data set are presented. Presented results show that the curves describing the CR obtained with the AD data do not descent in comparison with the ones obtained in Chapter 6. Results presented for the MCI Data Set (Figure 7.2(a)) and for the Mild AD Data Set 2 (Figure 7.2(c)) still present a smaller value than the mean value obtained with random data. However, using LOO cross-validation the distance between CR obtained using random data and the CR obtained using the features has decreased. Results presented in Figure 7.2(b) for the Mild AD Data Set 1 show that these results are more significant than the noise. CR obtained for the Mild AD Data Set 1, even though has presented some changes in the parameter selection, still achieving a 100% of classification.

Figure 7.2 presents the effect that cross-validation has created to the random data. If information presented in this figure is compared with the one presented in Figure 6.7, is easy to check that the maximum value obtained for the classification of random data in that figure is 100% with a really small variance, whereas results presented in Figure 7.2 achieves a mean value of 95%.

7.3 Feature cross-validation using *k*-fold

Results shown in the previous section using LOO cross-validation presented a decrease of the values achieved by the random data. However, results still presenting some overfitting.

MCI Data Set				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**20-fold Features**	**20-fold freq. ranges (Hz)**
5	ffDTF	9 - 29	ffDTF	9 - 30
6	OC	24 - 25	GC	18 - 25
7	GC	1 - 30	OC	24 - 25
8	DTF	4 - 5	dDTF	5 - 6
9	PC	1 - 10	PC	7 - 8
10	PS	28 - 30	PS	11 - 12
11	dDTF	1 - 2	DTF	10 - 11
Mild AD Data Set 1				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**20-fold Features**	**20-fold freq. ranges (Hz)**
6	dDTF	2 - 6	OC	24 - 26
8	OC	11 - 14	dDTF	2 - 6
9	ffDTF	6 - 19	ffDTF	8 - 12
Mild AD Data Set 2				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**20-fold Features**	**20-fold freq. ranges (Hz)**
11	PDC	16 - 17	PDC	1 - 2

Table 7.1: Differences between results of the OFRFPS algorithm without cross-validation and using 20-fold cross-validation. Differences between configurations are presented.

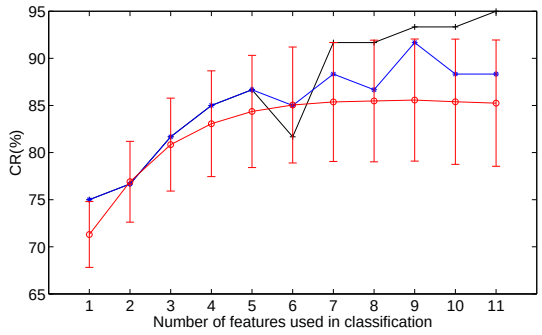
In order to reduce the overfitting the k-fold cross-validation is used in this section together with the OFRFPS algorithm.

Aiming to use the maximum number of subjects contained in the data set, the 10-fold and 20-fold are used. Another approach such as 5-fold could be used. However, as the total number of subjects for each data base is small, using the 5-fold cross-validation would suppose lose the 20% of the subjects. This high loss of subjects would make not significant the results.

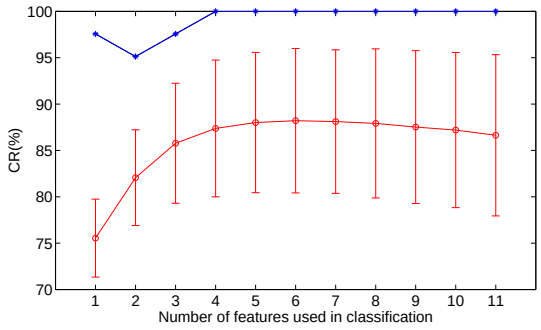
7.3.1 Feature cross-validation using 20-fold

The 20-fold cross-validation is used to select the features that are more stable along the different iterations of the OFRFPS algorithm. Selected features using 20-fold cross-validation are similar to the ones obtained without using cross-validation presented in Table 6.1, Table 6.2 and Table 6.3 for MCI Data Set, Mild AD Data Set 1 and Mild AD Data Set 2 respectively. Table 7.1 present only the differences in the configuration obtained when 20-fold cross-validation is used.

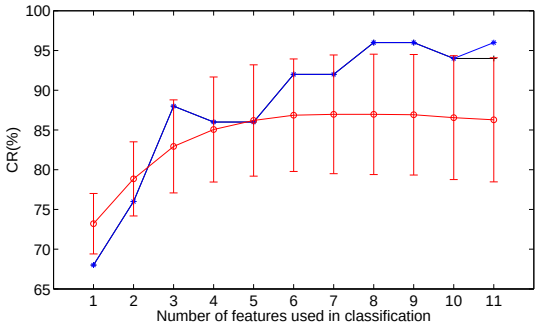
Results presented in Table 7.1 show that the data set that presents higher variability



(a) MCI Data Set



(b) Mild AD Data Set 1



(c) Mild AD Data Set 2

Figure 7.3: CR comparison between the CR obtained using 20-fold cross-validation features (in blue), features selected without cross-validation (in black) and random data (in red).

in the results when using 20-fold cross-validation is the MCI Data Set. For this data set, some measures such as ffDTF, PC and PS appear in the same position that when cross-validation is not used but in different frequency ranges. OC and GC features have interchanged position 6 and position 7, changing the selecting frequency range for GC. DTF and dDTF, had also interchanged positions, presenting that these two measures may obtain similar values for MCI patients, because when one of these is selected in the eighth position, the other one is selected as the last. Results presented for the data sets containing Mild AD patients present small changes in the configurations. For the Mild AD Data Set 1 only two features have interchanged positions, dDTF and OC, whereas ffDTF has changed the selected frequency range. Last results presented correspond to the Mild AD Data Set 2, which present all the features in the same order than the obtained without cross-validation, nevertheless the frequency range of the last selected feature PDC has changed.

CR is also tested using the 20-fold cross-validation selected features. Results for the three data sets are presented in Figure 7.3. For each data set in blue is presented the CR obtained using the 20-fold cross-validation features, in black is presented the CR obtained for the features without cross-validation and in red it is presented the mean CR obtained after applying the 20-fold cross-validation to random data. For this simulation 1.000 sets of random data are simulated and classified. Comparison between the CR obtained using the 20-fold cross-validation features and the features without cross-validation is not clear for Figure 7.3(b) and Figure 7.3(c), because the two CR lines are the same. For Figure 7.3(c), results are the same for all the positions except for the last. This differences are consistent

with the results presented in Table 7.1, where only small changes are presented.

Results presented in Figure 7.3 for the three data sets present a clear decrease of the CR obtained using random data in comparison with the obtained without cross-validation (presented in Figure 6.7) that reached the 100%. When using cross-validation, the best CR obtained using random data is close to 88% for the Mild AD Data Set 1. CR obtained for the other data sets is higher than 85%. Comparing all these CR it can be seen that whereas the CR of the features has only decrease in the MCI Data Set when using 20-fold cross-validation, results for the random data have clearly descend. Therefore, that using cross-validation the statistical significance of the results has increased. On the other hand, almost all obtained CR using 20-fold cross-validation are higher than the mean value of CR using random data. Therefore, results for MCI Data Set and Mild AD Data Set 2 still achieving the same results that random data, but present a lower overfitting.

Feature cross-validation using 10-fold

The 10-fold cross-validation is used to select the features that are more stable along the different iterations of the OFRFPS algorithm. Differences between the configuration obtained for each data set are presented in Table 7.2. In this case more changes are presented than when the 20-fold is used.

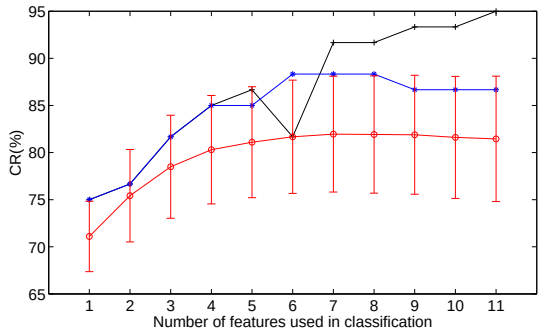
Results presented in Table 7.2 for the MCI Data Set present that almost all the features presented (only DTF and PC do not), the position that is selected by the algorithm when the 10-fold cross-validation is used is the same that when cross-validation it is not used, nevertheless selected frequency ranges are different. For the MCI Data Set only the first and the second features are stable to cross-validation. Coherence and PDC which achieved the third and the fourth position, present a frequency range that is close to the selected without cross-validation. Results presented for the Mild AD Data Set 1 present that the first four features are the more stables, being equal to the configuration obtained in Chapter 6. Features obtained in the position five and above, present not stable values neither in the position achieved or the frequency ranges selected. Results obtained for the Mild AD Data Set 2 present also different configurations when the 10-fold cross-validation is used. For this data set only OC is stable at the cross-validation. Interestingly, Correlation is selected in the second position, PC is selected in the third positions and PS is selected in the fourth.

The CR is computed using the 10-fold cross-validation features. Results in blue are presented in Figure 7.4. The same figure present the CR obtained without cross-validation (black in the figure) and the mean and the standard deviation of the CR obtained for random data also applying the 10-fold cross-validation (red in the figure). For this simulation 1.000 sets of random data are simulated and classified.

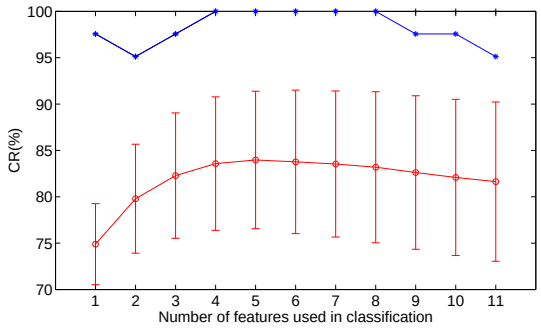
Presented results in Figure 7.4 show a clear decrease of the mean CR obtained with the random data. Therefore, the level of overfitting has clearly been reduced with this procedure. However, at the same time, the CR obtained with the feature has decrease in comparison with the results obtained without cross-validation. Results presented in Figure 7.4(a) show that the CR obtained using the 10-fold cross-validation features is the same that the obtained without, until the fourth feature is used. When the fifth feature is used, the CR obtained using cross-validation maintains its value. Best value obtained with this data set is a CR of 88.33% using six features. In this point the computed results

MCI Data Set				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**10-fold Features**	**10-fold freq. ranges (Hz)**
3	Coherence	1 - 6	Coherence	1 - 9
4	PDC	1 - 3	PDC	2 - 3
5	ffDTF	9 - 29	ffDTF	8 - 19
6	OC	24 - 25	OC	17 - 18
7	GC	1 - 30	GC	27 - 28
8	DTF	4 - 5	PS	8 - 9
9	PC	1 - 10	PC	7 - 8
10	PS	28 - 30	DTF	5 - 20
11	dDTF	1 - 2	dDTF	10 - 30
Mild AD Data Set 1				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**10-fold Features**	**10-fold freq. ranges (Hz)**
5	PC	13 - 14	ffDTF	1 - 30
6	dDTF	2 - 6	Coherence	6 - 8
7	Coherence	5 - 6	DTF	9 - 10
8	OC	11 - 14	dDTF	1 - 17
9	ffDTF	6 - 19	PC	4 - 7
10	DTF	20 - 21	OC	5 - 7
Mild AD Data Set 2				
Algorithm Order	**Features**	**Frequency ranges (Hz)**	**10-fold Features**	**10-fold freq. ranges (Hz)**
2	PS	14 - 15	Correlation	6 - 7
3	DTF	10 - 11	PC	26 - 28
4	Correlation	1 - 2	PS	14 - 15
5	dDTF	26 - 27	GC	18 - 19
6	GC	17 - 18	Coherence	18 - 20
7	RP	10 - 11	RP	14 - 18
8	Coherence	27 - 28	dDTF	25 - 30
9	PC	23 - 29	PDC	8 - 10
10	ffDTF	15 - 16	ffDTF	13 - 14
11	PDC	16 - 17	DTF	8 - 11

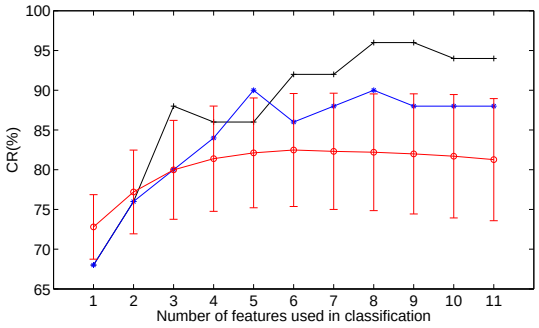
Table 7.2: Differences between results of the OFRFPS algorithm without cross-validation and using 10-fold cross-validation. Differences between configurations are presented.



(a) MCI Data Set



(b) Mild AD Data Set 1



(c) Mild AD Data Set 2

Figure 7.4: CR comparison between the CR obtained using 10-fold cross-validation features (in blue), features selected without cross-validation (in black) and random data (in red).

present a higher value than the standard deviation of the random noise, meaning that the level of overfitting is low. Results presented for the Mild AD Data Set 1 still being similar to the obtained without cross-validation. Only when the number of features used is equal to nine the CR obtained decreases to 97.56%. Results presented for the Mild AD Data Set 2 using cross-validation, differs from the obtained without it as presented in Figure 7.4(c). CR using one or two features is the same, but when three or four features are used the CR is lower. Then when five features are selected the CR reaches its maximum value of 90.00%, being higher than the obtained without cross-validation with this data set. In this same point, the CR obtained is higher than the standard deviation of the CR obtained with random data, showing that the level of overfitting is low.

7.4 Discussion

Results presented in this section show that when different cross-validations approach have been used, the level of overfitting in the data has been reduced. However, as the data sets used present a small number of subjects, the level of overfitting still being high.

Results obtained using LOO present a small decrease in the CR obtained with random data and almost the same configuration than when no cross-validation is used. This is due to the fact that only one subject is extracted each time, therefore, the configuration is quite similar. On the other hand, results presented for k-fold cross-validation (with k=10

and k=20) present variance in the obtained results. Nevertheless these results have to be taken carefully due to the small number of subjects contained in the data base after the segmentation.

A common result that have can be extracted with the results presented in previous sections, is that for all the configurations the first features are the more stables. Based on this criterion, for the MCI Data Set using four features an 85.00% of success in the classification can be achieved with a low level of overfitting, because the three tested configuration reached this value. For the Mild AD Data Set 1 results are similar, using the four first feature, the obtained CR is 100.00% and it has been achieved with all the configurations. Results presented for the Mild AD Data Set 2 are difficult to interpret due to the high variability presented in the 10-fold cross-validation configuration. Therefore, the obtained configuration using LOO and 20-fold cross-validation can be used with a small number of features, such as three. With this configuration the CR obtained is 88.00%. This number of features obtained is consistent with results presented in Chapter 7 when the random probe method is used.

Chapter 8

Preprocessing to improve classification

This section presents some of the analysis performed in this work together. As defined in the Introduction (Section 1.1), all the steps used in neuroscience are used in this thesis. Therefore a good way to test the cleaning method presented in Chapter 3 in real data is by repeating the studies but with cleaned data.

In this chapter the cleaning method first presented in Chapter 3 is applied in order to obtain cleaned data. Then, measures presented in Chapter 4 are used to parametrize the data. This data is then classified individually and using the OFR method presented in Chapter 6. Presented results show that using the cleaning method some improvement in the final CR is obtained in comparison with the CR obtained with raw data.

8.1 Methods

Methods used for the signal processing applied are presented in this section. In this work, EEG signals of the Mild AD Data Set 2 are used for cleaning and for evaluation. Figure 8.1 summarizes the used methods.

Data of the Mild AD Data Set 2 is used because is the only data set in which the raw data with any type of preprocessing is available. For this study the first 20 seconds are selected for each subject, independently if they presented a high or low SNR. In order to obtain IMF that does not expand along the entire spectrum, signals are band-pass filtered with a digital third-order Butterworth filter between 1 and 30 Hz. This is the same preprocessing applied to data defined in Section 4.2. However, data used in the previous study was selected by visual inspection, therefore the cleanest period it was selected for each subject, and now data is selected directly for all subjects.

Once signals are cleaned, measures defined in Chapter 4 are computed in the cleaned data and in the raw data as presented in Figure 8.1. Window time and Granger order configurations are the same that the ones defined in Table 4.2 and Table 4.3. Results between different types of data are then compared. The OFRFPS algorithm is also computed on both types of data. Results are also compared. To compute the data cleaning, Algorithm 3.3 is used with the threshold value of $T = 0.85$, as is the best value obtained when used with simulated data.

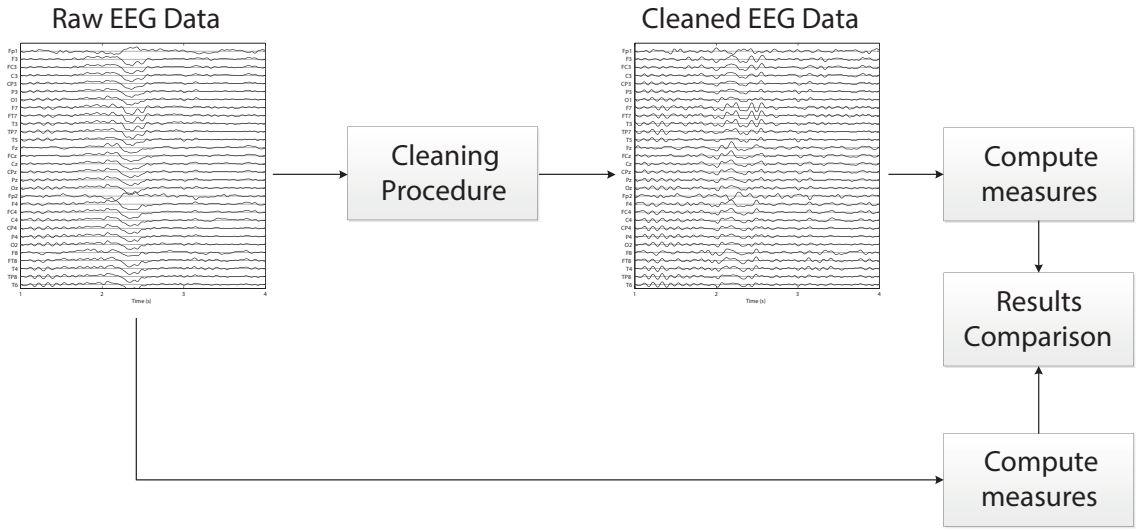


Figure 8.1: Methods used to compare raw data of a recording with data cleaned using the cleaning method.

8.2 Preprocessing results with AD data

Results obtained after applying the methods described in this thesis are presented in this section. First measures are computed on all existing frequency ranges. CR computed for each data type and the results obtained in Chapter 4 are compared in Figure 8.2. The corresponding p-values for these data sets are presented in Table 8.1.

Results presented in Figure 8.2 show that, for some measures, when the cleaning method is applied they obtain a higher CR that when the measures are computed directly on raw data. Presented results show that for RP, Coherence, GC and DTF there is an increase in the CR of 2%. On the other hand PC, PDC and dDTF present a decreased value for the cleaned data, obtaining higher values for the raw data. These results show that in some cases there is an small improvement of the data and in some others there is a deterioration. Decrease in the value of the CR is always found when some Granger measures are used, whereas in other cases GC and DTF, also Granger measures, present an improvement of the CR. Presented results suggest that some of the IMFs eliminated during the reconstruction process are parametrized by some of the Granger measures, whereas other measures are not affected. Even though some measures present an improvement of the CR when the cleaning method is used, CR are not comparable with the results obtained in Chapter 4. Only for GC obtained CR using the cleaned data is higher than the one obtained with visual selected data, in all other cases obtained CR is smaller. Data used in Chapter 4 was selected by visual inspection and therefore a period of 20 seconds without any kind of interference is used.

Results presented in Table 8.1 present the frequency ranges at which the CR where computed and its corresponding p-value. Interestingly for some of the measures such as RP, Coherence, PC, DTF, ffDTF and dDTF there is a decrease of the computed p-value for measures computed with the cleaned data. However, this decrease of the p-value is not corresponded by an increase of the CR for the measures PC, PDC and dDTF. Obtained frequency ranges are the same between raw and cleaned data for some measures (Cor-

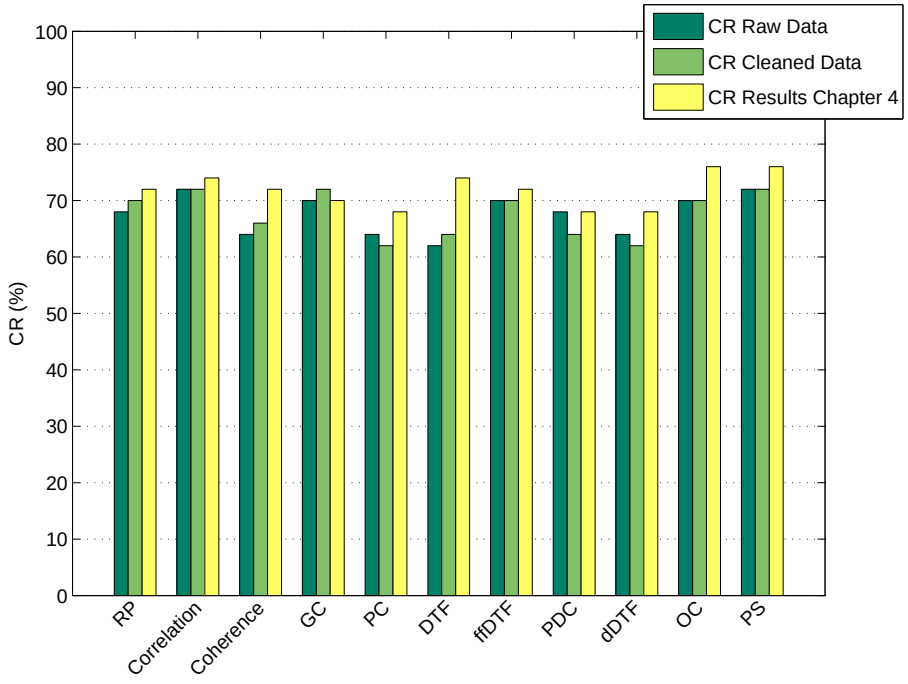
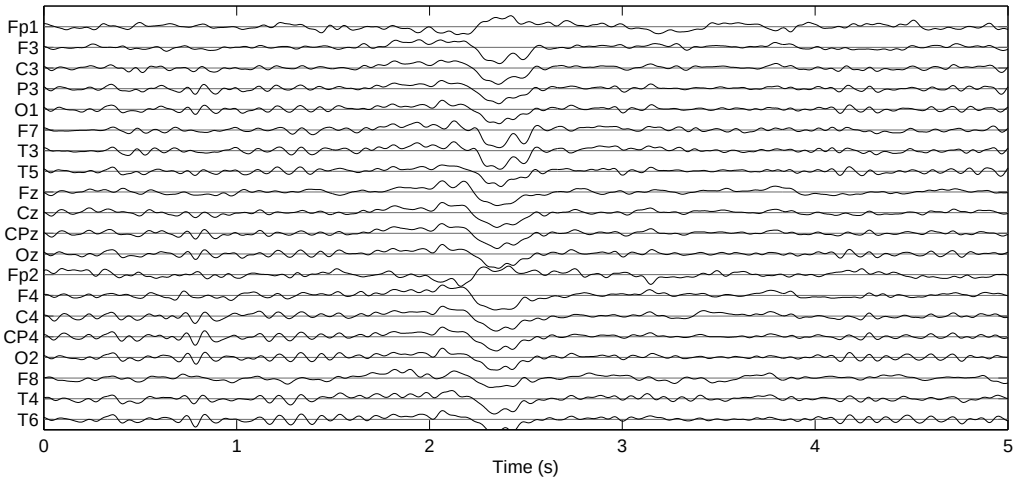


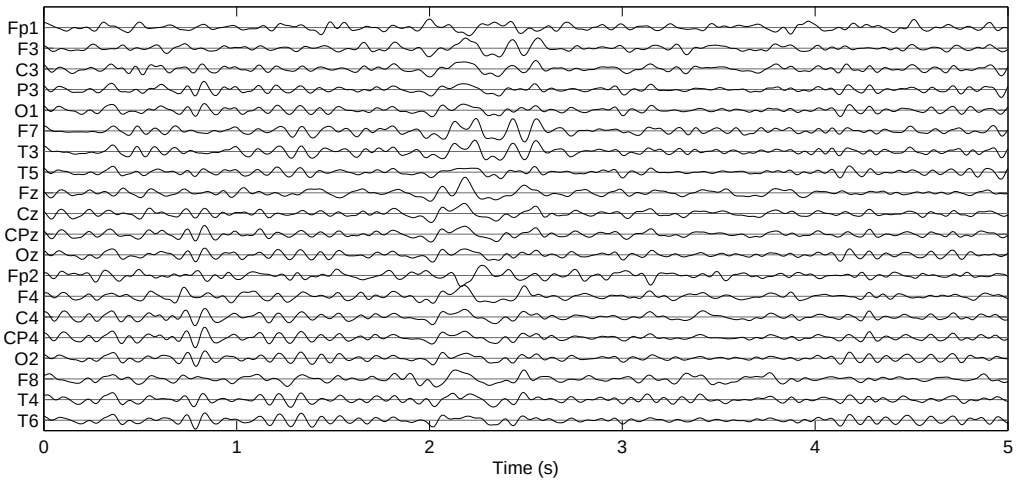
Figure 8.2: CR comparison between the raw data, the cleaned data and results of Chapter 4. Results presented for the Mild AD Data Set 2 when measures are used individually to compute the CR.

Measures	**Raw data freq. range (Hz)**	**p-value raw data**	**Clean data freq. range (Hz)**	**p-value clean data**
RP	9 - 10	0.0542	9 - 14	0.0340
Correlation	26 - 30	0.0411	26 - 30	0.0411
Coherence	8 - 9	0.0909	8 - 28	0.2957
GC	8 - 9	0.0054	8 - 9	0.0054
PC	4 - 18	0.5909	7 - 10	0.2370
DTF	1 - 6	0.5511	9 - 11	0.0909
ffDTF	9 - 13	0.0451	7 - 10	0.0411
PDC	17 - 21	0.0374	3 - 24	0.1299
dDTF	13 - 30	0.4756	9 - 10	0.0185
OC	19 - 28	0.0081	19 - 28	0.0091
PS	9 - 12	0.0060	9 - 12	0.0060

Table 8.1: Optimal frequency ranges and computed p-values using the raw data and the cleaned data.



(a) Raw data



(b) Clean data

Figure 8.3: Five seconds period of EEG time series comparison between (a) raw data and (b) cleaned data. To facilitate the understanding of the figure, only 20 of the original 30 sensors are presented.

Algorithm Order	Raw data		Cleaned data	
	Features	**OFRFPS Selected frequency ranges (Hz)**	**Features**	**OFRFPS Selected frequency ranges (Hz)**
1	PS	9 - 10	PS	9 - 10
2	dDTF	12 - 15	DTF	8 - 9
3	PDC	6 - 8	GC	10 - 12
4	OC	13 - 14	Correlation	11 - 12
5	Coherence	10 - 16	OC	8 - 9
6	DTF	9 - 13	Coherence	3 - 4
7	PC	13 - 15	dDTF	13 - 15
8	Correlation	21 - 22	RP	3 - 4
9	RP	10 - 11	ffDTF	13 - 14
10	GC	8 - 9	PC	13 - 14
11	ffDTF	8 - 9	PDC	6 - 7

Table 8.2: Features and frequency ranges selected by the OFRFPS algorithm using raw and cleaned data.

relation, GC, OC and PS). For some other measures such as RP and Coherence, the frequency range of the cleaned data is expanded in comparison with the frequency range obtained with raw data. Finally, frequency ranges obtained for Granger measures present the biggest differences presenting different results between cleaned and raw data.

Relating obtained results with results presented in Chapter 3, it have to be pointed out that selected data is just the first 20 seconds of each recording. Therefore for each of the subjects a different level of artifacts is presented, and of course a different level of SNR it would be computed. Usually for these recordings there is not a high presence of artifacts in the recordings, they present some small changes as the example presented in Figure 8.3. In this figure, raw time series (Figure 8.3(a)) show a clear artifact at second 2.5. This artifact is corrected for the cleaned data. For the other recording the situation is quite similar, and therefore the SNR computed for these recordings would be high. At the view of these results, small improvement presented in the results is normal due to the small room for improvement which data presents.

In order to check the evolution of the CR when measures are used together, the OFRFPS algorithm is computed as in Chapter 6. The most discriminative features are selected and then CR is computed. Selected features for the raw and the cleaned data are presented in Table 8.2. CRs obtained using these features are presented in Figure 8.4. This figure also presents a comparison with results obtained using visual inspected data.

Results presented in Table 8.2 show that for the two types of data PS is the most discriminative feature, whereas results obtained in Chapter 6 (Table 6.3, pag. 105) show that the most discriminative feature is OC. However, the second features selected is PS. Obtained

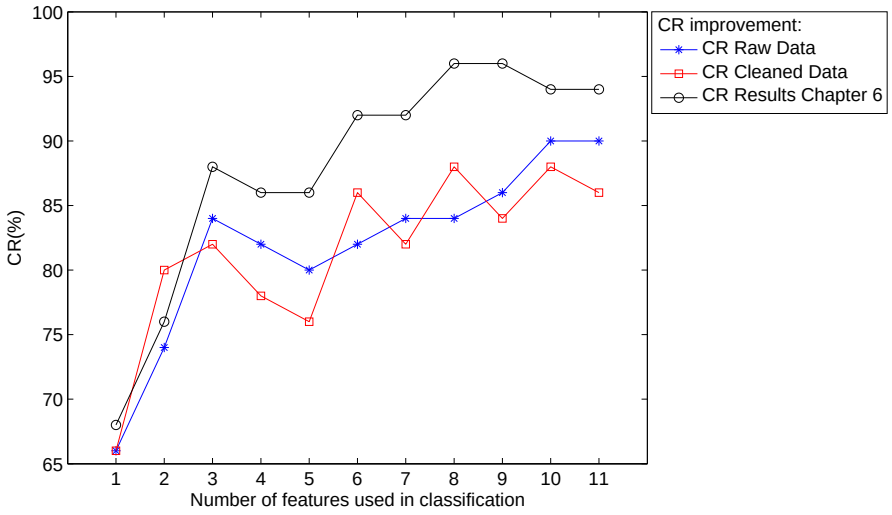


Figure 8.4: CR comparison between the raw data and the cleaned EEG data. Results presented for the Mild AD Data Set 2 when the CR is computed using the OFRFPS selected features.

results using the cleaned data are more similar to results obtained in Table 6.3 than results obtained using the raw data. Interestingly the five first selected features when the cleaned data is used are the same that results presented in Chapter 6, even though they present different order and Coherence have been included instead of dDTF. This result is significant because, as presented in Chapter 7, the three or four first features are usually the more stables when cross-validation is applied. On the other hand results using raw data do not present any relation with results obtained in Chapter 6, and for example a feature that has been prove that does not improve the CR when used, PDC, is ranked in the third position. Two conclusions can be extracted of these results. First conclusion is that when the cleaning method is used, the selected configuration is more similar to the one obtained visually inspecting the data, therefore some improvement is presented in the data when the cleaning method is used. Second conclusion is that bivariate measures such as PS, Correlation, Coherence and RP presents changes in the position of only one or two positions, whereas multivariate measures such as OC or the Granger measures present important changes in the position achieved. Therefore the cleaning method present an important perturbation to multivariate measures, meanwhile bivariate measures are slightly perturbed.

The evolution of the CR presented in Figure 8.4 show that results between raw data and cleaned data are similar. However, results presented for the cleaned data present high oscillation in the obtained CR. These results emphasizes the conclusion presented before that bivariate measures are less affected than multivariate measures. CR obtained using cleaned data starts improving from a 68% until a 82% when three features are used, in this case PS and two Granger measures, DTF and GC. However, when four and five features are used CR drops down to 76%. Interestingly the fourth measure is Correlation that even though being a bivariate measure do not improve the CR. The fifth measure used is OC, which is multivariate. From measures six to measure nine, when bivariate measures are used there is an improvement of the results and for multivariate measures there is a decrease. Results obtained for raw data present an important decrease when PDC, OC and Coherence are used, but then there is an improvement of the obtained CR until a

final value of 90%. Results obtained with visual inspected data in Chapter 6 are higher than the other almost in all the cases, only when two features are used CR obtained with cleaned data is higher.

In this study no cross-validation is performed as it is done with data inspected visually in Chapter 7. Presented results therefore present an overfitting, but the main purpose of this study is to show that some improvement is performed with the cleaning method when this is used on real data. Presented results show that some improvement can be found in the CR and in the p-value when the cleaning method is used, even if this is small.

8.3 Discussion

This section has presented the use of the cleaning method proposed in Chapter 3 in real data. At the view of obtained results it can be pointed out that the cleaning methods improve the quality of the data, but it is not as good as using the visual inspection for the preprocessing, as results on CR has presented. However, the procedure of using visual inspection for preprocessing is really time consuming and it can not be replicated exactly. Therefore, when the preprocessing has to be done, it is a compromise between the time to do the preprocessing and the aimed quality of the data. If the data is for some medical application, such as data used in previous chapters to work on the early diagnosis of AD, the best option is to check by visual inspection the data, because this data needs to have a higher quality. However, if some application needs some simple technique to apply denoising it can be used. For example for some BCI implementation it could be used.

One of the main problems that the cleaning method presents is that it uses mEMD to decompose the data. This decomposition is a really good technique for decomposition, but its main problem is that it is really slow when used with long periods of data, i.e. 15 s or 20 s, with lot of EEG channels to decompose. However, when used with short windows time and small number of channels, it is faster. Therefore the idea of a BCI implementation using this technique could be possible if some considerations are taken into account, such as use the minimum number of channels possible, like the ones which are close to the eyes, and small time windows.

On the other hand, the shortcoming presented in Chapter 3 that some part of the data may be lost if the cleaning procedure is applied with data that present a low level of noise or artifacts, may introduce some bias to the results computed in this chapter. Therefore if the cleaning method aims to be used it is necessary to add some step that decide if the level of noise is higher enough to apply the cleaning method or not.

Presented results on real data evaluate the performance of the cleaning method based on the final CR obtained after applying it. It have to be taken in account that CR is based on synchrony measures and RP. Synchrony measures present some clear perturbation when some IMFs are eliminated. Therefore, this cleaning method has to be tested in other data sets where synchrony is not a key issue.

Chapter 9

Conclusion and Future Work

This thesis has presented new signal processing and machine learning methods for the early diagnose of AD. The main steps used in EEG data analysis have been explored, from the preprocessing until the classification. The main conclusions extracted from this study are presented below.

Preprocessing is an important step that facilitates the posterior analysis of the data. A new cleaning method based on mEMD has been presented. This cleaning method has been evaluated on simulated data and on real data. Results presented using simulated data with different levels of SNR, show that for signals presenting high level of noise and artifacts, the cleaning method improves the quality of the data when compared to data without artifacts. However, for signals with a low level of noise or artifacts, the cleaning method is not as effective. Results obtained after using the cleaning method on real data presents the same characteristics, obtaining a small classification improvement when compared with data without cleaning. Results shown in this work open a new line of work in which this cleaning method can be tested in other data sets used for other purpose, such as for BCI applications (Graimann *et al.*, 2011) or for some other applications such as EEG sonification (Elgendi *et al.*, 2014). However, as pointed out in Chapter 8, in order to apply this method in real data an important step has to be added. This pre-cleaning step would be used to decide if the data is clean enough, avoiding unnecessary cleaning. Another line of work could be to develop an optimized implementation of mEMD to improve the computation time of the algorithm.

Furthermore, different biomarkers indicative of changes that AD causes on EEG data have been identified. Afterwards, these measures have been used individually to distinguish between healthy subjects and AD patients in different stages (MCI and Mild AD). Single features have been used to compute CR in order to obtain the optimal frequency range that best discriminates between AD patients and healthy subjects. A multiple feature classification approach, which is call OFRFPS has also been presented, aiming to optimize the CR. An increase of synchrony in the narrow bands inside the θ band has also been introduced. This change of EEG signals was previously presented for some authors in the literature, but it were not studied enough as the well-known decrease of synchrony in the α and β bands. Presented results show that this increase of synchrony can also be used to correctly classify AD patients and healthy subjects.

Methodology presented for selecting the optimal frequency range by computing all possible frequency ranges between 1 and 30 Hz it has been presented as more efficient that

computing the CR in the classical frequency bands (δ, θ, α and β). However, the results obtained using this methodology are not constant for the different data sets tested. The data sets analyzed in this study were obtained in different experimental conditions in different hospitals, with different EEG systems and slightly different protocols. Therefore, significant variations in the experimental conditions can be expected. Consequently, comparing these data sets is a challenging task, and it has been performed as an independent study of each database separately. The calibration and normalization of features between data sets is worth future investigation.

The joint use of the effects of slowing of EEG and changes in the synchrony has been presented as useful in two of the three data sets. Satisfactory classification results have been obtained, even if as later presented in Chapter 7 these results are partly due to overfitting. This joint use of information present the same problem that when the features are used individually, the different configurations obtained for the data sets. On the other hand, when using OFRFPS algorithm only two types of measures have been used. Complexity measures due to the correlation presented with RP have not been used. However, results presented for the Mild AD Data Set 2 show that RP is not as effective as have been presented in the other data sets. Therefore, future work will analyze the joint use of the three effects that AD cause on EEG signals, aiming to improve the rate at which AD patients and healthy subjects of the Mild AD Data Set 2 are classified. The methodology presented could be also explored for the early diagnose of other neurodegenerative diseases.

The use of the ρ defined to study the increase of synchrony for AD patients in the narrow bands inside the θ band has been shown useful. However, in future works other aspects could be further explored. For instance, controlling if a better CR could be achieved by combining the proposed ratio with different synchrony measures. Furthermore, the differences of synchrony between AD patients and healthy subjects might be enlarged by using some other function rather than a ratio.

Another line of investigation that can be followed is to investigate changes in the inter-region synchrony instead of global synchrony; this would allow us to identify which regions exhibit the strongest perturbations in synchrony, and therefore to obtain more insight on how to differentiate AD patients from healthy subjects. It may also be studied how the spatial averaging approach affects the synchrony, due to the fact that it may affect the values of inter-region and global synchrony. By studying the inter-region synchrony, it opens the way to work with other signal processing techniques which also work with feature selection. For example the use of tensor decomposition (Phan & Cichocki, 2010; Cichocki, 2011) has been presented as a prerequisite for classification of data with large dimension such as brain recordings or multiview images. Some approach for the early diagnose of AD using tensor decomposition have already been presented for Latchoumane *et al.* (2012). However, in this case a new approach is suggested by using the synchrony values computed between all the regions as a matrix for each subject. Then using tensor decomposition the common information between healthy subjects and AD patients could be extracted, using only the different information between each class as features to a classifier. Taking into account the small number of subjects contained in the data sets, the number of features given to the classifier should be restricted to a minimum in order to avoid overfitting.

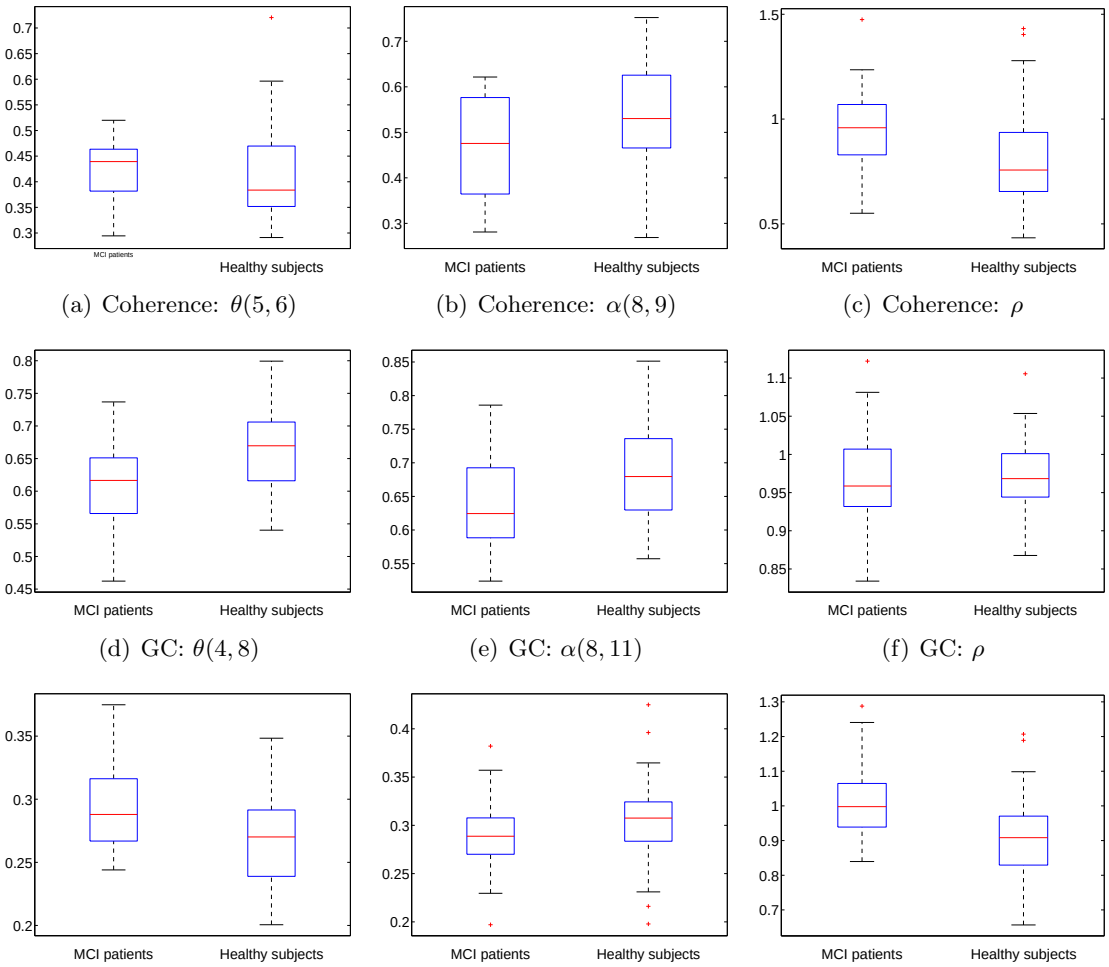
Another line of research would be to analyze the use of SSVEP on subjects suffering from AD. Indeed, some studies have already shown some changes related with aging and neurodegenerative disorders, some of those articles are presented in (Vialatte *et al.*, 2010).

However, as the recordings corresponding to Mild AD Data Set 2 present different stages of the subjects (eyes open, eyes close and periods of SSVEP at different frequencies), the parameters extracted using SSVEP could be used together with the measures used and the ρ. Nevertheless, as so many parameters would be extracted of the data, some feature selection method would have to be used, such as the OFRFPS or some other variation, which would define the number of measures that had to be used in order to minimize overfitting.

Finally, it must be noted that the two data sets used are rather small. A larger database is needed in order to generalize our results. Data sets containing different types of dementia, and optimally the evolution of MCI subjects to Mild AD, could significantly facilitate the early diagnosis of AD.

Appendix A

Complementary results using the ratio

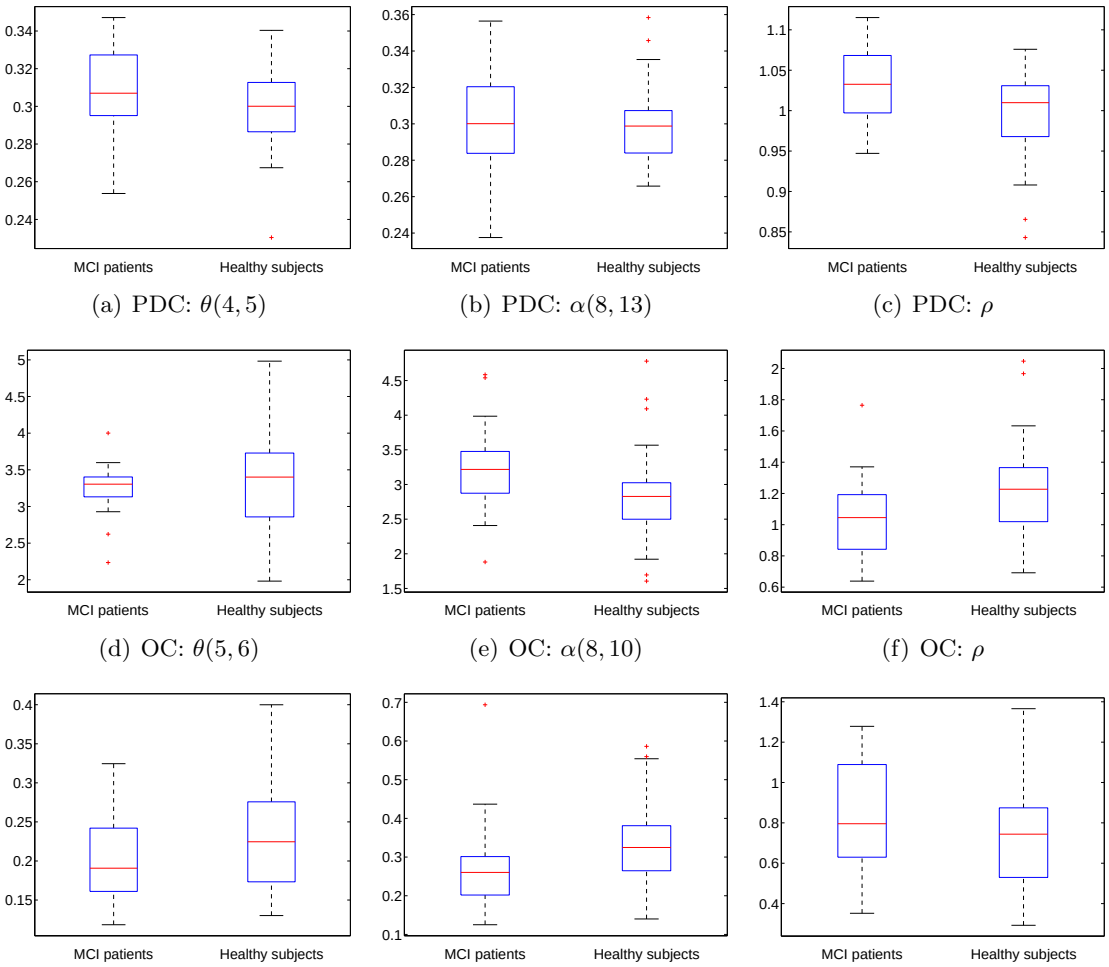


(a) Coherence: $\theta(5, 6)$ (b) Coherence: $\alpha(8, 9)$ (c) Coherence: ρ

(d) GC: $\theta(4, 8)$ (e) GC: $\alpha(8, 11)$ (f) GC: ρ

(g) PC: $\theta(6, 7)$ (h) PC: $\alpha(12, 13)$ (i) PC: ρ

Figure A.1: Box plots presenting the results obtained for the MCI Data Set for Coherence, GC and PC measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

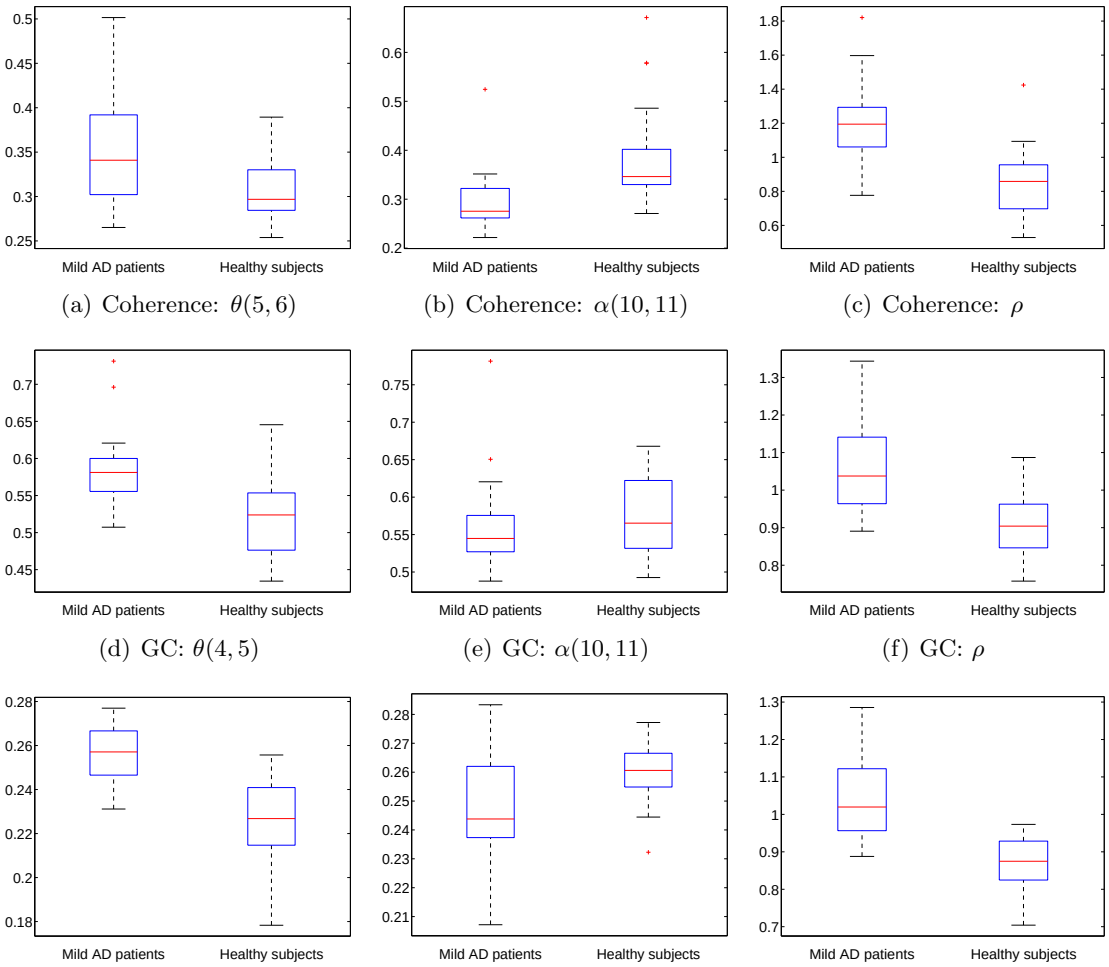


(a) PDC: $\theta(4, 5)$ (b) PDC: $\alpha(8, 13)$ (c) PDC: ρ

(d) OC: $\theta(5, 6)$ (e) OC: $\alpha(8, 10)$ (f) OC: ρ

(g) PS: $\theta(4, 8)$ (h) PS: $\alpha(8, 10)$ (i) PS: ρ

Figure A.2: Box plots presenting the results obtained for the MCI Data Set for PDC, OC and PS measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

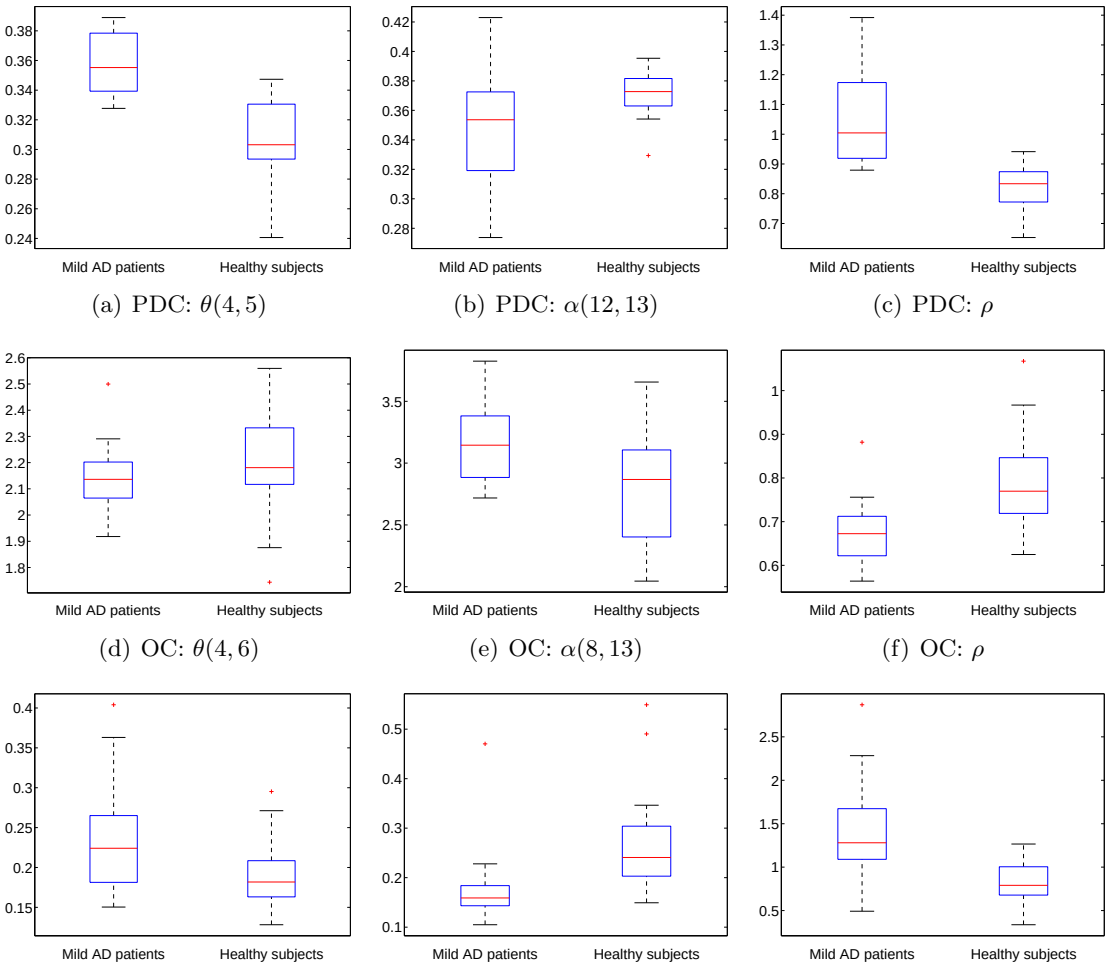


(a) Coherence: $\theta(5, 6)$ (b) Coherence: $\alpha(10, 11)$ (c) Coherence: ρ

(d) GC: $\theta(4, 5)$ (e) GC: $\alpha(10, 11)$ (f) GC: ρ

(g) ffDTF: $\theta(4, 5)$ (h) ffDTF: $\alpha(12, 13)$ (i) ffDTF: ρ

Figure A.3: Box plots presenting the results obtained for the Mild AD Data Set 1 for Coherence, GC and ffDTF measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

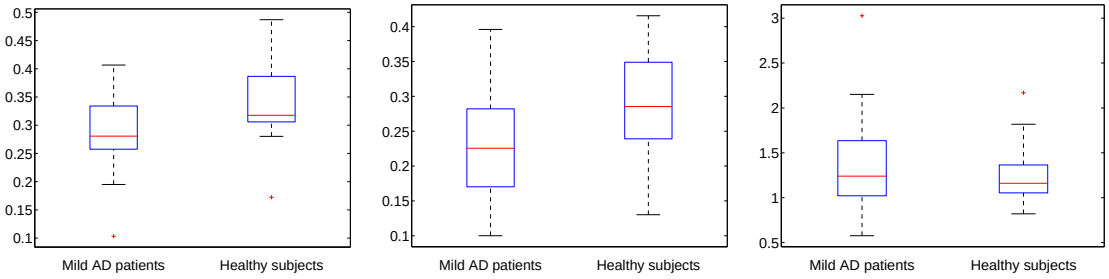


(a) PDC: $\theta(4,5)$ (b) PDC: $\alpha(12,13)$ (c) PDC: ρ

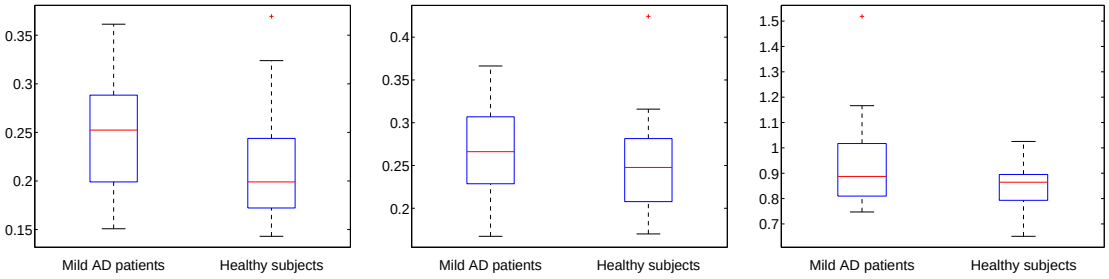
(d) OC: $\theta(4,6)$ (e) OC: $\alpha(8,13)$ (f) OC: ρ

(g) PS: $\theta(5,6)$ (h) PS: $\alpha(8,11)$ (i) PS: ρ

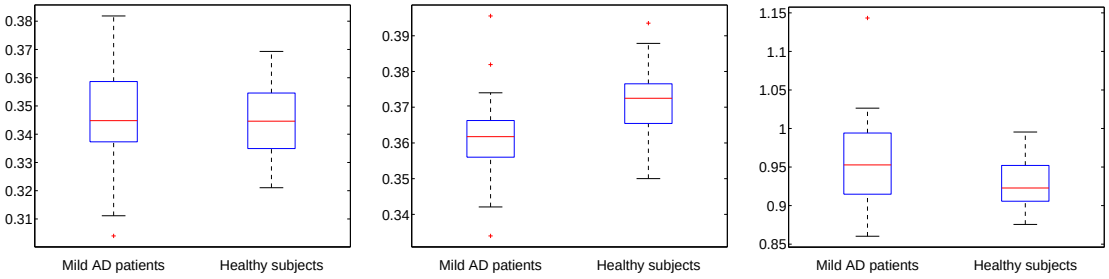
Figure A.4: Box plots presenting the results obtained for the Mild AD Data Set 1 for PDC, OC and PS measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.



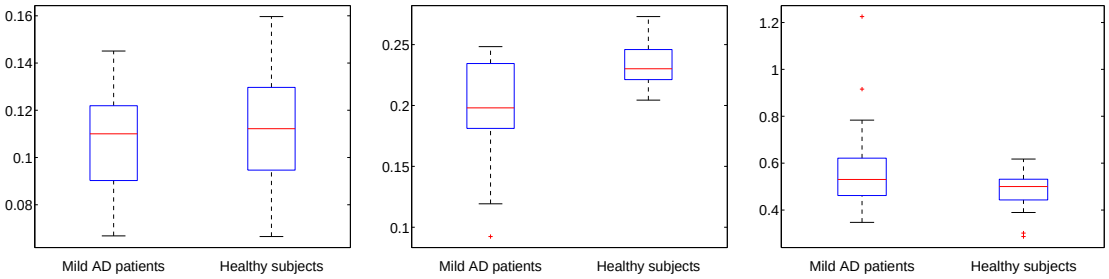
(a) Correlation: $\theta(4, 8)$ (b) Correlation: $\alpha(8, 9)$ (c) Correlation: ρ



(d) PC: $\theta(7, 8)$ (e) PC: $\alpha(8, 13)$ (f) PC: ρ

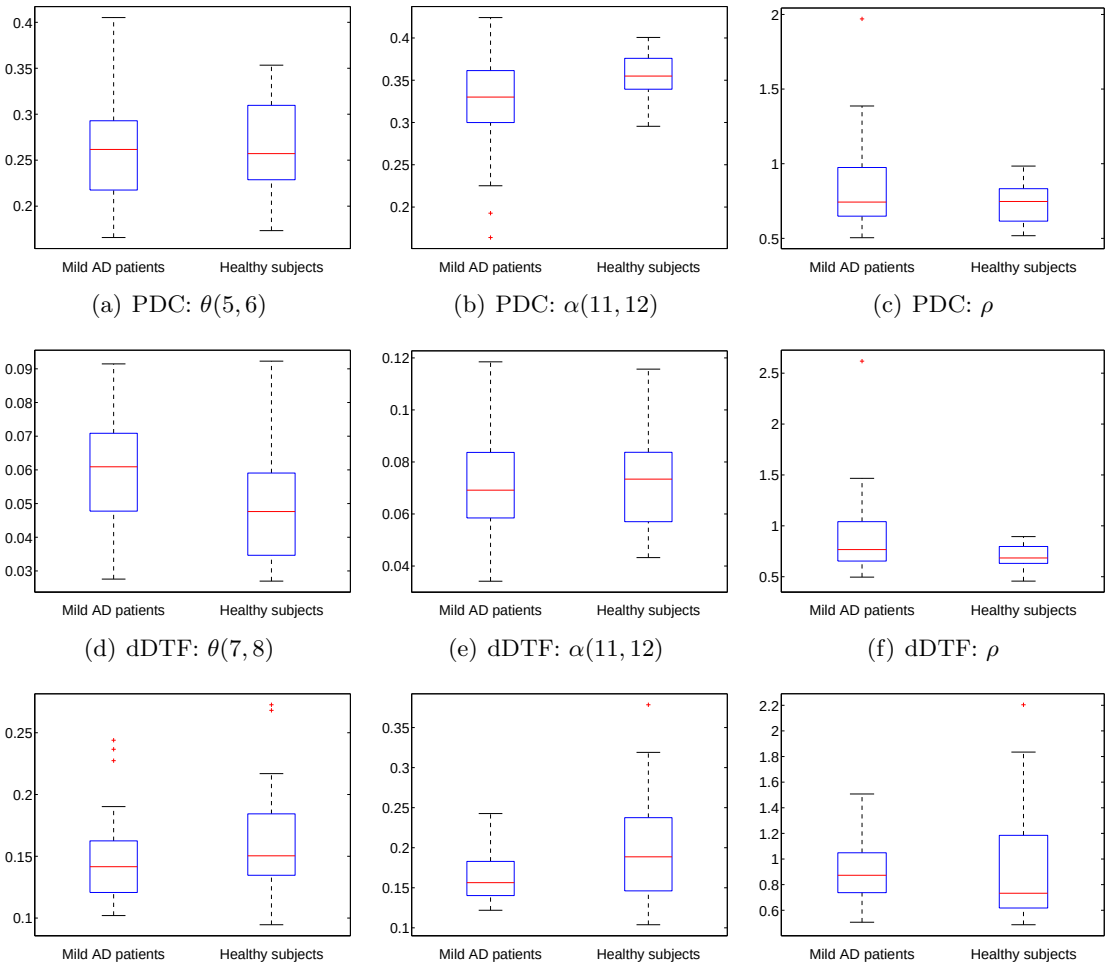


(g) DTF: $\theta(6, 7)$ (h) DTF: $\alpha(10, 11)$ (i) DTF: ρ



(j) ffDTF: $\theta(4, 8)$ (k) ffDTF: $\alpha(10, 11)$ (l) ffDTF: ρ

Figure A.5: Box plots presenting the results obtained for the Mild AD Data Set 2 for Correlation, PC, DTF and ffDTF measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.



(a) PDC: $\theta(5,6)$ (b) PDC: $\alpha(11,12)$ (c) PDC: ρ

(d) dDTF: $\theta(7,8)$ (e) dDTF: $\alpha(11,12)$ (f) dDTF: ρ

(g) PS: $\theta(4,8)$ (h) PS: $\alpha(9,10)$ (i) PS: ρ

Figure A.6: Box plots presenting the results obtained for the Mild AD Data Set 2 for PDC, dDTF and PS measures. Results in the $\theta(f_1, f_2)$ range, $\alpha(f_3, f_4)$ range and using the ρ are presented.

Bibliography

Abásolo, D., Hornero, R., Espino, P., Poza, J., Sánchez, C. I., & de la Rosa, R. 2005. Analysis of regularity in the EEG background activity of Alzheimer's disease patients with Approximate Entropy. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, **116**(8), 1826–1834.

Abásolo, D., Hornero, R., Espino, P., Alvarez, D., & Poza, J. 2006. Entropy analysis of the EEG background activity in Alzheimer's disease patients. *Physiological measurement*, **27**(3), 241–53.

Adler, G., Brassen, S., & Jajcevic, A. 2003. EEG coherence in Alzheimer's dementia. *Journal of neural transmission (Vienna, Austria : 1996)*, **110**(9), 1051–1058.

Alzheimer's Association. 2014. 2014 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, **10**(2), e47–e92.

Alzheimer's Disease Association. 2014 (September). *Evolution of AD in the brain*. http://www.alz.org/braintour/progression.asp.

Babiloni, C., Ferri, R., Binetti, G., Cassarino, A., Dal Forno, G., Ercolani, M., Ferreri, F., Frisoni, G. B., Lanuzza, Ba., & Miniussi, C. 2006. Fronto-parietal coupling of brain rhythms in mild cognitive impairment: a multicentric EEG study. *Brain research bulletin*, **69**(1), 63–73.

Babiloni, C., Ferri, R., Binetti, G., Vecchio, F., Frisoni, G. B., Lanuzza, B., Miniussi, C., Nobili, F., Rodriguez, G., Rundo, F., Cassarino, A., Infarinato, F., Cassetta, E., Salinari, S., Eusebi, F., & Rossini, P. M. 2009a. Directionality of EEG synchronization in Alzheimer's disease subjects. *Neurobiology of aging*, **30**(1), 93–102.

Babiloni, C., Pievani, M., Vecchio, F., Geroldi, C., Eusebi, F., Fracassi, C., Fletcher, E., De Carli, C., Boccardi, M., Rossini, P. M., & Frisoni, G. B. 2009b. White-matter lesions along the cholinergic tracts are related to cortical sources of EEG rhythms in amnesic mild cognitive impairment. *Human brain mapping*, **30**(5), 1431–43.

Babiloni, C., Vecchio, F., Lizio, R., Ferri, R., Rodriguez, G., Marzano, N., Frisoni, G. B., & Rossini, P. M. 2011. Resting state cortical rhythms in mild cognitive impairment and Alzheimer's disease: Electroencephalographic evidence. *Journal of Alzheimers Disease*, **26**, 201–214.

Baccalá, L. A., & Sameshima, K. 2001. Partial directed coherence: A new concept in neural structure determination. *Biological Cybernetics*, **84**(6), 463–474.

Baker, M., Akrofi, K., Schiffer, R., & O'Boyle, M. W. 2008. EEG Patterns in Mild Cognitive Impairment (MCI) Patients. *The open neuroimaging journal*, **2**, 52–55.

Benca, R. M., Obermeyer, W. H., Larson, C. L., Yun, B., Dolski, I., Kleist, K. D., Weber, S. M., & Davidson, R. J. 1999. EEG alpha power and alpha power asymmetry in sleep and wakefulness. *Psychophysiology*, **36**(04), 430–436.

Bennys, K., Rondouin, G., Vergnes, C., & Touchon, J. 2001. Diagnostic value of quantitative EEG in Alzheimers disease. *Neurophysiologie Clinique/Clinical Neurophysiology*, **31**(3), 153–160.

Bhattacharya, J. 2000. Complexity analysis of spontaneous EEG. *Acta Neurobiologiae Experimentalis*, **60**(4), 495–501.

Bishop, C. M. 2006. *Pattern recognition and machine learning*. New York :: Springer.

Bloomfield, P. 2000. *Fourier analysis of time series: an introduction*. 2nd edn. Wiley - Interscience.

Blume, W. T. 2006. Drug effects on EEG. *Journal of Clinical Neurophysiology*, **23**(4), 306–311.

Brown, G., Pocock, A., Zhao, M. J., & Luján, M. 2012. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The Journal of Machine Learning Research*, **13**(1), 27–66.

Chen, S., Billings, S. A., & Luo, W. 1989. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, **50**(5), 1873–1896.

Cichocki, A. 2011. Tensors decompositions: New concepts for brain data analysis? *Journal of Control, Measurement, and System Integration (SICE)*, **47**(7), 507–517.

Cichocki, A., & Amari, S. 2002. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons.

Cichocki, A., Shishkin, S. L., Musha, T., Leonowicz, Z., Asada, T., & Kurachi, T. 2005. EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clinical Neurophysiology*, **116**(3), 729–737.

Cohen, B. A., & Sances, A. 1977. Stationarity of the human electroencephalogram. *Medical & Biological Engineering & Computing*, **15**(5), 513–518.

Cooley, J. W., & Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, **19**(90), 297–301.

Cover, Thomas M., & Thomas, Joy A. 2006. *Elements of information theory*. 2nd edn. New York :: John Wiley & sons,.

Croft, R. J., & Barry, R. J. 2000. Removal of ocular artifact from the EEG: a review. *Clinical neurophysiology*, **30**(1), 5–19.

Czigler, B., Csikós, D., Hidasi, Z., Anna Gaál, Z., Csibri, E., Kiss, Éva, Salacz, P., & Molnár, M. 2008. Quantitative EEG in early Alzheimer's disease patientspower spectrum and complexity features. *International Journal of Psychophysiology*, **68**(1), 75–80.

Dash, M., & Liu, H. 1997. Feature Selection for Classification. *Intelligent Data Analysis*, **1**, 131–156.

Dauwels, J., Vialatte, F. B., Latchoumane, C., Jeong, F., & Cichocki, A. 2009a. EEG synchrony analysis for early diagnosis of Alzheimer's disease: A study with several synchrony measures and EEG data sets. *Pages 2224–2227 of: Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.*

Dauwels, J., Vialatte, F. B., Weber, T., & Cichocki, A. 2009b. Quantifying statistical interdependence by message passing on graphs-part I: one-dimensional point processes. *Neural computation*, **21**(8), 2152–2202.

Dauwels, J., Vialatte, F. B., Weber, T., Musha, T., & Cichocki, A. 2009c. Quantifying statistical interdependence by message passing on graphs-part II: multidimensional point processes. *Neural computation*, **21**(8), 2203–2268.

Dauwels, J., Vialatte, F. B., Musha, T., & Cichocki, A. 2010a. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *NeuroImage*, **49**(1), 668–693.

Dauwels, J., Vialatte, F. B., & Cichocki, A. 2010b. Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing? *Current Alzheimer Research*, **7**(6), 487–505.

Dauwels, J., Srinivasan, K., Ramasubba Reddy, M., Musha, T., Vialatte, F. B., Latchoumane, C., Jeong, J., & Cichocki, A. 2011. Slowing and Loss of Complexity in Alzheimer's EEG: Two Sides of the Same Coin? *International journal of Alzheimer's disease*, **2011**(Jan.), 539621.

Delorme, A., Sejnowski, T., & Makeig, S. 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, **34**(4), 1443–1449.

Diez, P. F., Mut, V., Laciar, E., Torres, A., & Avila, E. 2009. Application of the empirical mode decomposition to the extraction of features from EEG signals for mental task classification. *Pages 2579–2582 of: Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2009. IEEE.

Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M. B., Badran, F., & Thiria, S. 2008. *Apprentissage statistique: Réseaux de neurones - Cartes topologiques - Machines à vecteurs supports.* Paris: Éditions Eyrolles.

Duda, R. O., Hart, P. E., & Stork, D. G. 2001. *Pattern Classification.* John Wiley & Sons,.

Elgendi, M., Vialatte, F. B., Cichocki, A., Latchoumane, C., Jeong, J., & Dauwels, J. 2011. Optimization of EEG frequency bands for improved diagnosis of Alzheimer disease. *Pages 6087–6091 of: Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE.*

Elgendi, M., Dauwels, J., Rebsamen, B., Shukla, R., Putra, Y., Gamez, J., ZePing, N., Ho, B., Prasad, N., Aggarwal, D., Nair, A., Mishuhina, V., Vialatte, F. B., Constable, M., Cichocki, A., Latchoumane, C., Jeong, J., Thalmann, D., & Magnenat-Thalmann, N. 2014. From Auditory and Visual to Immersive Neurofeedback: Application to Diagnosis of Alzheimers Disease. *Pages 63–97 of: Neural Computation, Neural Devices, and Neural Prosthesis.* New York, NY: Springer New York.

Fisch, B. J. 1999. *Fisch and Spehlmann's EEG Primer: Basic Principles of Digital and Analog EEG*. Third revised and enlarged edition edn. Elsevier.

Frigo, M., & Johnson, S. G. 1998. FFTW: An adaptive software architecture for the FFT. *Pages 1381–1384 of: Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 3. IEEE.

Gallego-Jutglà, E., & Solé-Casals, J. 2012. Exploring mEMD for face recognition. *Pages 498–503 of: BIOSIGNALS 2012 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*.

Gallego-Jutglà, E., & Solé-Casals, J. 2013. Improving early diagnosis of Alzheimer's disease using synchrony measures. *Pages 167–170 of: Frontiers in Artificial Intelligence and Applications*, vol. 256.

Gallego-Jutglà, E., Solé-Casals, J., Rutkowski, T. M., & Cichocki, A. 2011. Application of multivariate empirical mode decomposition for cleaning eye blinks artifacts from EEG signals. *Pages 455–460 of: NCTA 2011 - Proceedings of the International Conference on Neural Computation Theory and Applications*.

Gallego-Jutglà, E., Elgendi, M., Vialatte, F. B., Solé-Casals, J., Cichocki, A., Latchoumane, C., Jeong, J., & Dauwels, J. 2012a. Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands. *Pages 4266–4270 of: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*.

Gallego-Jutglà, E., Rutkowski, T. M., Cichocki, A., & Solé-Casals, J. 2012b. EEG signal analysis via a cleaning procedure based on multivariate empirical mode decomposition. *Pages 670–676 of: IJCCI 2012 - Proceedings of the 4th International Joint Conference on Computational Intelligence*.

Gallego-Jutglà, E., Lopez-de Ipiña, K., Martí-Puig, P., & Solé-Casals, J. 2013. Empirical mode decomposition-based face recognition system. *Pages 445–450 of: BIOSIGNALS 2013 - Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*.

Goh, C., Ifeachor, E., Henderson, G., Latchoumane, C. F. V., Jeong, J., Bigan, C., Besleaga, M., Hudson, N., Capotosto, P., & Wimalaratna, S. 2006. Characterisation of EEG at different stages of Alzheimers disease (AD). *Clinical Neurophysiology*, **117**, 138–139.

Graimann, B., Allison, B. Z., & Pfurtscheller, G. 2011. *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Springer.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**(3), 424–438.

Guyon, I., & Elisseeff, A. 2003. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, **3**, 1157–1182.

Henderson, G., Ifeachor, E., Hudson, N., Goh, C., Outram, N., Wimalaratna, S., Del Percio, C., & Vecchio, F. 2006. Development and assessment of methods for detecting dementia using the human electroencephalogram. *IEEE transactions on bio-medical engineering*, **53**(8), 1557–1568.

Hornero, R., Abásolo, D., Escudero, J., & Gómez, C. 2009. Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **367**(1887), 317–336.

Hsiao, F. J., Wang, Y. J., Yan, S. H., Chen, W. T., & Lin, Y. Y. 2013. Altered oscillation and synchronization of default-mode network activity in mild Alzheimer's disease compared to mild cognitive impairment: an electrophysiological study. *PloS one*, **8**(7), e68792.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Snin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **454**(1971), 903–995.

Jeong, J. 2004. EEG dynamics in patients with Alzheimer's disease. *Clinical Neurophysiology*, **115**(7), 1490–1505.

John, G. H., Kohavi, R., & Pfleger, K. 1994. Irrelevant Features and the Subset Selection Problem. *Pages 121–129 of: Machine Learning: Proceedings of the Eleventh International Conference*.

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. 2004. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, **41**(2), 313–325.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. 2000. Removing electroencephalographic artifacts by blind source separation. - Cerca amb Google. *Psychophysiology*, **37**(2), 163–178.

Kaminski, M. J., & Blinowska, K. J. 1991. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, **65**(3), 203–210.

Kaminski, M. J., & Liang, H. 2005. Causal Influence: Advances in Neurosignal Analysis. *Critical Reviews in Biomedical Engineering*, **33**(4), 347–430.

Kawabata, N. 1976. Test of statistical stability of the electroencephalogram. *Biological Cybernetics*, **22**(4), 235–238.

Knyazeva, M. G., Jalili, M., Brioschi, A., Bourquin, I., Fornari, E., Hasler, M., Meuli, R., Maeder, P., & Ghika, J. 2010. Topography of EEG multivariate phase synchronization in early Alzheimer's disease. *Neurobiology of aging*, **31**(7), 1132–1144.

Knyazeva, M. G., Carmeli, C., Khadivi, A., Ghika, J., Meuli, R., & Frackowiak, R. S. 2013. Evolution of source EEG synchronization in early Alzheimer's disease. *Neurobiology of aging*, **34**(3), 694–705.

Koenig, T., Prichep, L., Dierks, T., Hubl, D., Wahlund, L. O., John, E. R., & Jelic, V. 2005. Decreased EEG synchronization in Alzheimers disease and mild cognitive impairment. *Neurobiology of aging*, **26**(2), 165–171.

Kohavi, R., & John, G. H. 1997. Wrappers for feature subset selection. *Artificial Intelligence*, **97**(1-2), 273–324.

Kohavi, R., & Sommerfield, D. 1995. Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology. *Pages 192–197 of: Proceedings of First International Conference on Knowledge Discovery and Data Mining.*

Koller, D., & Sahami, M. 1995. Toward optimal feature selection. *Pages 284–292 of: 13th International Conference on Machine Learning.*

Korzeniewska, A., Maczak, M., Kamiski, M., Blinowska, K. J., & Kasicki, S. 2003. Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. *Journal of Neuroscience Methods*, **125**(1-2), 195–207.

Kramer, M. A., Chang, F. L., Cohen, M. E., Hudson, D., & Andrew, J. S. 2007. Synchronization measures of the scalp electroencephalogram can discriminate healthy from Alzheimer's subjects. *International journal of neural systems*, **17**(02), 61–69.

Kropotov, J. D. 2009. *Quantitative EEG, Event-Related Potentials and Neurotherapy.* Elsevier Inc.

Kukull, W. A., Higdon, R., Bowen, J. D., McCormick, W. C., Teri, L., Schellenberg, G. D., van Belle, G., Jolley, L., & Larson, E. B. 2002. Dementia and Alzheimer Disease Incidence. *Archives of Neurology*, **59**(11), 1737.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. 1999. Measuring phase synchrony in brain signals. *Human Brain Mapping*, **8**(4), 194–208.

Latchoumane, C. F. V., Vialatte, F. B., Solé-Casals, J., Maurice, M., Wimalaratna, S. R., Hudson, N., Jeong, J., & Cichocki, A. 2012. Multiway array decomposition analysis of EEGs in Alzheimer's disease. *Journal of Neuroscience Methods*, **207**(1), 41–50.

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon, P., DHavé, M., Adam, C., Bernard, R., Varela, F.o, & Baulac, M. 2001a. Anticipation of epileptic seizures from standard EEG recordings. *The Lancet*, **357**(9251), 183–188.

Le Van Quyen, M., Foucher, J., Lachaux, J. P., Rodriguez, E., Lutz, A., Martinerie, J., & Varela, F. J. 2001b. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of Neuroscience Methods*, **111**(2), 83–98.

Lempel, A., & Ziv, J. 1976. On the complexity of finite sequences. *In: Information Theory, IEEE Transactions on.*

Levner, I., Bulitko, C., & Lin, G. 2006. Feature Extraction for Classification of Proteomic Mass Spectra: A Comparative Study. *Pages 607–624 of: Feature Extraction.* Studies in Fuzziness and Soft Computing, vol. 207. Springer Berlin Heidelberg.

Locatelli, T., Cursi, M., Liberati, D., Franceschi, M., & Comi, G. 1998. EEG coherence in Alzheimer's disease. *Electroencephalography and clinical neurophysiology*, **106**(3), 229–237.

Molla, Md. K. I., Tanaka, T., Rutkowski, T. M., & Cichocki, A. 2010. Separation of EOG artifacts from EEG signals using bivariate EMD. *Pages 562–565 of: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing.* IEEE.

Moretti, D. V., Fracassi, C., Pievani, M., Geroldi, C., Binetti, G., Zanetti, O., Sosta, K., Rossini, P. M., & Frisoni, G. B. 2009. Increase of theta/gamma ratio is associated with memory impairment. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, **120**(2), 295–303.

Murphy, S. L., Xu, J. Q., & Kochanek, K. D. 2010. Deaths : final data for 2010. *National Vital Statistics Reports*, **61**.

Musha, T., Asada, T., Yamashita, F., Kinoshita, T., Chen, Z., Matsuda, H., Uno, M., & Shankle, W. R. 2002. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. *Clinical Neurophysiology*, **113**(7), 1052–1058.

Mutlu, A. Y., & Aviyente, S. 2011. Multivariate Empirical Mode Decomposition for Quantifying Multivariate Phase Synchronization. *EURASIP Journal on Advances in Signal Processing*, **2011**(1), 1–13.

Neurowear. 2014a (June). *Neocomimim, a brainwave sensor to control cat's ear.* http://neurowear.com/projects_detail/necomimi.html.

Neurowear. 2014b (June). *Shippo, a brainwave sensor to control a tail.* http://neurowear.com/projects_detail/shippo.html.

Nunez, P. L., & Srinivasan, R. 2006. *Electric fields of the brain: the neurophysics of EEG.* 2nd edn. Oxford university press.

Oppenheim, A. V, & Schafer, R. W. 1975. *Digital signal processing.* Prentice-Hall International.

Park, Young-Min, Che, Hee-Jae, Im, Chang-Hwan, Jung, Hyung-Tae, Bae, Sung-Man, & Lee, Seung-Hwan. 2008. Decreased EEG synchronization and its correlation with symptom severity in Alzheimer's disease. *Neuroscience research*, **62**(2), 112–117.

Pascual-Marqui, R. D. 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, **24 Suppl D**, 5–12.

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., Smith, G. E., & Jack Jr, C. R. 2009. Mild cognitive impairment: ten years later. *Archives of Neurology*, **66**(12), 1447.

Phan, A.-H., & Cichocki, A. 2010. Tensor Decompositions for Feature Extraction and Classification of High Dimensional Datasets. *Nonlinear Theory and Its Applications, IEICE*, **1**, 37–68.

Proakis, J. G, & Manolakis, D. G. 1996. *Digital Signal Processing: Principles, Algorithms and Applications.* Upper Saddle River, New Jersey: Prentice-Hall.

Rampil, I. J. 1998. A primer for EEG signal processing in anesthesia. *Anesthesiology*, **89**(4), 980–1002.

Rehman, N., & Mandic, D. P. 2010a. Empirical Mode Decomposition for Trivariate Signals. *IEEE Transactions on Signal Processing*, **58**(3), 1059–1068.

Rehman, N., & Mandic, D. P. 2010b. Multivariate empirical mode decomposition. *Proceedings of the Royal Society A*, **466**(2117), 1291–1302.

Rilling, G., Flandrin, P., & Gonçalvés, P. 2003. On empirical mode decomposition and its algorithms. *Pages 8–11 of: IEEE-EURASIP workshop on nonlinear signal and image processing*, vol. 3.

Romero, S., Mañanas, M. A., & Barbanoj, M. J. 2009. Ocular reduction in EEG signals based on adaptive filtering, regression and blind source separation. *Annals of biomedical engineering*, **37**(1), 176–91.

Rosenblum, M. G., Pikovsky, A. S., & Kurths, J. 1996. Phase synchronization of chaotic oscillators. *Physical Review Letters*, **76**(11), 1804–1807.

Rückstieß, T., Osendorfer, C., & van der Smagt, P. 2011. Sequential feature selection for classification. *Pages 132—-141 of: AI 2011: Advances in Artificial Intelligence*. Springer Berlin Heidelberg.

Rutkowski, T. M., Cichocki, A., Toshihisa, T., Ralescu, A. L., & Mandic, D. P. 2008. Clustering of spectral patterns based on EMD components of EEG channels with applications to neurophysiological signals separation. *Pages 453–460 of: Advances in Neuro-Information Processing*. Springer-Verlag.

Rutkowski, T. M., Cichocki, A., Tanaka, T., Mandic, D. P., Cao, J., & Ralescu, A. L. 2009. Multichannel spectral pattern separation - An EEG processing application-. *Pages 373–376 of: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE.

Saito, N., Kuginuki, T., Yagyu, T., Kinoshita, T.o, Koenig, T., Pascual-Marqui, R. D., Kochi, K., Wackermann, J., & Lehmann, D. 1998. Global, Regional, and Local Measures of Complexity of Multichannel Electroencephalography in Acute, Neuroleptic-Naive, First-Break Schizophrenics. *Biological psychiatry*, **43**(11), 794–802.

Salis, C. I., Malissovas, A. E., Bizopoulos, P. A., Tzallas, A. T., Angelidis, P. A., & Tsalikakis, D. G. 2013. Denoising simulated EEG signals: A comparative study of EMD, wavelet transform and Kalman filter. *Pages 1–4 of: 13th IEEE International Conference on BioInformatics and BioEngineering*. IEEE.

Sanei, S., & Chambers, J. A. 2009. *EEG Signal Processing*. John Wiley & Sons.

Saunders, A. M., Strittmatter, W. J., Schmechel, D., St. George-Hyslop, P. H., Pericak-Vance, M. A., Joo, S. H., Rosi, B. L., Gusella, J. F., Crapper-MacLachlan, D. R., Alberts, M. J., Hulette, C., Crain, B., Goldgaber, D., & Roses, A. D. 1993. Association of apolipoprotein E allele 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, **43**(8), 1467–1472.

Schalk, G., & Mellinger, J. 2010. *A Practical Guide to BrainComputer Interfacing with BCI2000*. Springer Science & Business Media.

Scinto, L. F. M., & Daffner, K. R. 2000. *Early diagnosis of Alzheimer's disease*. Totowa, NJ: Humana Press,.

Solé-Casals, J., Vialatte, F. B., Pantel, J., Prvulovic, D., Haenschel, C., & Cichocki, A. 2010. ICA cleaning procedure for EEG signals analysis: Application to Alzheimer's disease detection. *Pages 485–490 of: BIOSIGNALS 2010 - Proceedings of the 3rd International Conference on Bio-inpsired Systems and Signal Processing, Proceedings*.

Solé-Casals, J., Gallego-Jutglà, E., Martí-Puig, P., Travieso, C. M., & Alonso, J. B. 2013. Speech enhancement: A multivariate empirical mode decomposition approach. *Pages 192–199 of: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7911 LNAI. Springer Berlin Heidelberg.

Stam, C. J., Montez, T., Jones, B. F., Rombouts, S. A. R. B., Van der Made, Y., Pijnenburg, Y. A. L., & Scheltens, Ph. 2005. Disturbed fluctuations of resting state EEG synchronization in Alzheimer's disease. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, **116**(3), 708–715.

Stam, C. J., Jones, B. F., Manshanden, I., van Cappellen van Walsum, A. M., Montez, T., Verbunt, J. P. A., de Munck, J. C., van Dijk, B. W., Berendse, H. W., & Scheltens, P. 2006. Magnetoencephalographic evaluation of resting-state functional connectivity in Alzheimer's disease. *NeuroImage*, **32**(3), 1335–1344.

Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. 2003. Ranking a random feature for variable and feature selection. *The Journal of Machine Learning Research*, **3**, 1399–1414.

Tóth, B., File, B., Boha, R., Kardos, Z., Hidasi, Z., Gaál, Z. A., Csibri, E., Salacz, P., Stam, C. J., & Molnár, M. 2014. EEG network connectivity changes in mild cognitive impairment - Preliminary results. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, **92**(1), 1–7.

van der Hiele, K., Vein, A. A., Reijntjes, R., Westendorp, R. G. J., Bollen, E., van Buchem, M. A., van Dijk, J. G., & Middelkoop, H. A. M. 2007. EEG correlates in the spectrum of cognitive decline. *Clinical neurophysiology*, **118**(9), 1931–1939.

van Vliet, M., Robben, A., Chumerin, N., Manyakov, N. V., Combaz, A., & Van Hulle, M. M. 2012. Designing a brain-computer interface controlled video-game using consumer grade EEG hardware. *Pages 1–6 of: 2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*. IEEE.

Vecchio, F., Babiloni, C., Lizio, R., De Vico Fallani, F., Blinowska, K., Verrienti, G., Frisoni, G., & Rossini, P. M. 2013. Resting state cortical EEG rhythms in Alzheimer's disease: Toward EEG markers for clinical applications: A review. *Supplements to Clinical Neurophysiology*, **62**, 223–236.

Vialatte, F. B., Cichocki, A., Dreyfus, G., Musha, T., Rutkowski, T. M., & Gervais, R. 2005a. Blind source separation and sparse bump modelling of time frequency representation of EEG signals: New tools for early detection of Alzheimer's disease. *Pages 27–32 of: 2005 IEEE Workshop on Machine Learning for Signal Processing.*

Vialatte, F. B., Cichocki, A., Dreyfus, G., Musha, T., Shishkin, S. L., & Gervais, R. 2005b. Early Detection of Alzheimers Disease by Blind Source Separation, Time Frequency Representation, and Bump Modeling of EEG Signals. *Pages 683–692 of: Artificial Neural Networks: Biological Inspirations ICANN 2005*, vol. 3696.

Vialatte, F. B., Solé-Casals, J., Dauwels, J., Maurice, M., & Cichocki, A. 2009a. Bump time-frequency toolbox: A toolbox for time-frequency oscillatory bursts extraction in electrophysiological signals. *BMC Neuroscience*, **10**.

Vialatte, F. B., Dauwels, J., Solé-Casals, J., Maurice, M., & Cichocki, A. 2009b. Improved sparse bump modeling for electrophysiological data. *Pages 224–231 of: Lecture Notes in Computer Science*, vol. 5506 LNCS. Springer Berlin Heidelberg.

Vialatte, F. B., Solé-Casals, J., Maurice, M., Latchoumane, C., Hudson, N., Wimalaratna, S., Jeong, J., & Cichocki, A. 2009c. Improving the Quality of EEG Data in Patients with Alzheimers Disease Using ICA. *Advances in Neuro-Information Processing*, 979–986.

Vialatte, François-Benoît, Maurice, Monique, Dauwels, Justin, & Cichocki, Andrzej. 2010. Steady-state visually evoked potentials: focus on essential paradigms and future perspectives. *Progress in neurobiology*, **90**(4), 418–38.

Wackermann, J. 1996. Beyond mapping: Estimating complexity of multichannel EEG recordings. *Acta Neurobiologiae Experimentalis*, **56**(1), 197–208.

Wang, H. X., Xu, W., & Pei, J. J. 2012. Leisure activities, cognition and dementia. *Biochimica et biophysica acta*, **1822**(3), 482–491.

Wang, Q., Sourina, O., & Nguyen, M. K. 2010. EEG-based "serious" games design for medical applications. *Pages 270–276 of: Cyberworlds (CW), 2010 International Conference on.*

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Shen, L., Siuciak, J. A., Soares, H., Toga, A. W., & Trojanowski, J. Q. 2013. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, **9**(5), e111–194.

Wijdicks, E. F. M. 1995. Determining brain death in adults. *Neurology*, **45**, 1003–1011.

Woon, W. L., Cichocki, A., Vialatte, F. B., & Musha, T. 2007. Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings. *Physiological Measurement*, **28**(4), 335.

Yi, R., Zhan, X. R., Tang, J., Zhang, L. M., Liu, X. M., & Dong, Q. 2014. EEG Synchronization Evaluation: A New Diagnostic Tool for Predicting the Progression of Alzheimer's disease. *Clinical EEG and neuroscience*, Mar., 1550059413502779.

Zhao, P., Van-Eetvelt, P., Goh, C., Hudson, N., Wimalaratna, S., & Ifeachor, E. 2007. Characterization of EEGs in Alzheimer's disease using information theoretic methods. *Pages 5127–5131 of: Conference of the IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2007.