بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

Sudan University Of Science And Technology
College Of Graduate Studies

An Improved Algorithm For Discrimination Of Motor Imagery Based Brain Computer Interface

خوارزمية مطورة لتصنيف الحركات التخيلية المبنية على الإتصال بين الدماغو الحاسبالآلي

A Thesis Submitted In Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY in BIOMEDICAL ENGINEERING

By:

Rania Elsadig Elmahdi Ahmed

Supervisor:

Prof. Alnazier Osman Hamza

Co-Supervisor:

Dr. Magdy Baker M. Amin

Septemper 2020

Dedication

I'd like to dedicate this work with all love and gratitude to my parents, Dr. Zubaida Abdelnabi and Dr. Elsadig Elmahdi, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.This work is also dedicated to my lovely Daughter Raghad who has inspired me; my love for you can never be quantified. Thank you. Without my family's support, help and encouragement I wouldn't have achieved any progress in my life.God bless you.

ACKNOWLEDGEMENTS

First and for most,thanks to my God the most merciful and most gracious. Second, I must give special thanks to my advisors;*Prof. Alnazier O. Hamza., Professor, Department of biomedical Engineering,* for his canalized guidance, dedication, constant inspiration and encouragement.I also extend my whole hearted thanks to my Co-supervisor *Dr. Magdy Baker M. Amin, Ph.D.,Associate professor, Faculty ofEngineering, Khartoum University,*for his kind support during the doctoral program and for having provided me with his valuable suggestions to complete the work successfully. I am also grateful to my university (Sudan University of Science and Technology), especially the Department of Biomedical Engineering. Deepest thanks to my Brothers, *Ammar Elsadig and LouiElsadig,*for their timely help, moral support and encouragement. Last but not least I must give Special thanks to my friends and colleagues who supported me all along the way.

Table of Contents

Content Page No

CHAPTER ONE
Introduction

CHAPTER TWO
Literature Survey

CHAPTER THREE
Theoretical Background

CHAPTER FOUR

Research Methodology

Theoretical Concepts

CHAPTER FIVE

Body of Research

CHAPTER SIX

Results

CHAPTER SEVEN

Analysis, Discussions and Interpretations of Results

CHAPTER EIGHT

Conclusions and Recommendations

List of Tables

List of Figures

Glossary

ALS	Amyotrophic Lateral Sclerosis
BCI	Brain Computer Interface
Coif	Coiflet Mother Wavelets

CSP	Common Spatial Pattern
Db	DaubechiesMother Wavelets
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
ECoG	Electrocorticography
EEG	Electroencephalograph
EMG	Electromyography
EOG	Electrooculography
ERD	Event Related Desynchronization
ERP	Event Related Potential
ERS	Event Related Synchronization
FBCSP	Filter Bank Common Spatial Pattern
FFT	Fast Fourier Transform
FIR	Finite ImpulseResponse
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
JAD	Joint Aproximation Diagonalization
LDA	Linear Discriminant Analysis
MI	Motor Imagery
OVO	One-versus-One
OVR	One-versus-Rest
ROI	Region of Interest
SNR	Signal-to-Noise Ratio
SOBI	Second-Order Blind Identification
SVM	Support Vector Machine

Abstract

Brain computer interface (BCI) system is one of the means of communication that allowscontrolling devices or communicating with others by active brain signals only without movement signals. The main application of this system is to help paralyzed people or disabled people to communicate with the outside world. The electroencephalography (EEG) based motor imagery is one of the methods that the BCI system uses to identify the expected behavior through brain signals. This system is considered to be accurate when it can first isolate the irrelevant brain signals from the other sources that mixed with it.These artifact signals are originating from other organs of the body or from the outside medium.Secondly, the system's ability to differentiate between different imagery movements such as to differentiates the movement of the right hand from the left.

To achieve this, two algorithms were built through which the noise signals are isolated and then discriminating between four different imagery movements (right hand, left hand, both feet, and tongue). Data set IIa from BCI competition IV was used to test both algorithms.

The first algorithm is to use anIIRband pass filter (8 - 30) Hz as preprocessing stage in order to purify the signal and extract brain waves with beta β and mu μ frequencies and then use the combined Wavelet Common Spatial (Wavelet-CSP) method to extract the most efficient features of the signal before it is presented to support vector machine (SVM) classifier.

In the second algorithm, Independent component analysis (ICA) method was used in the preprocessing stage, in order to extract the most efficient features in each electrode separately, using two software: SOBI and FastICA to compare them and determine which is faster when applied. It is proved that the method of SOBI is the fastest in the execution of the algorithm. In the second stage, the Wavelet-CSP method is used to extract the most important characteristics of the signal before presenting it to the SVM classifier.

The two algorithms proved successful with a Kappa coefficient of 0.53 for the first method and 0.55 for the second method in solving the problem of discrimination of four different imagery movements. Moreover, the second algorithm is more interestingnot because of its high performance only but also the possibility to adopt this system to be used in real time.

المستخلص

نظام إتصال المخ بالحاسب الآلى هو أحد وسائل الإتصال التى تسمح بالتحكم فى الأجهزة أو التواصل مع الآخرين، و ذلك عن طريق إشارات المخ النشطة فقط دون إشارات الحركة. التطبيق الرئيسى لهذا الإتصال هو مساعدة الأشخاص المصابين بالعجز التام أن يتواصلوا مع العالم الخارجى. إن الحركة التخيلية هي إحدي الوسائل التي يستخدمها نظام إتصال المخ بالحاسب للتعرف علي السلوك المتوقع عن طريق إشارات المخ. يعتبر ذلك النظام دقيقا عندما يستطيع أولا عزل إشارة المخ عن بقية اشارات الضجيج المختلطة بها سواء كانت اشارات من اعضاء اخري بالجسم او من الوسط الخارجي وثانيا امكانية النظام للتفريق بين الحركات التخيلية المختلفة مثل تخيل حركة اليد اليمني عن اليسرى وغيرها. لتحقيق ذلك تم بناء خوارزميتين من خلالهما يتم عزل اشارات الضجيج ومن ثم التفريق بين اربع حركات تخيلية مختلفة (يد يمنى، يد يسرى، القدمين ولسان). استخدمت البيانات رقم IIa المنشورة في المسابقة الرابعة لانظمة الاتصال بين الانسان والالة وذلك لاختبار الخوارزميتين.

الخوارزمية الاولي تم فيها استخدام مرشح تمرير نطاق الاستجابة غير المحدود (IIR band pass filter) ذو التردد (8 - 30)هيرتز كمرحلة اولية وذلك لتنقية الاشارة واستخلاص موجات المخ ذات التردد بيتا β وميوμ ومن ثم استخدام طريقة Wavelet-CSP لكي نستخلص اهم سمات الاشارة قبل تقديمها الي المصنف من النوع (Support Vector MachineSVM).

في الخوارزمية الثانية تم استخدام طريقةICA في المرحلة الاولى وذلك لاستخلاص اهم السمات في الالكترودات كل علي حدى بطريقتين هما SOBI و FastICA للمقارنة بينهما وتحديد ايهما اسرع عند التنفيذ. حيث ثبت ان هنالك فرق معنوي عالي بين زمن تنفيذ الخوارزميتين. وفي المرحل الثانية تستخدم طريقة Wavelet-CSP لكي نستخلص اهم سمات الاشارة قبل تقديما الي المصنف من النوع SVM.

أثبتت الخوارزميتين نجاحها بمعامل كابا 0.53 للاولي و 0.55 للثانية في حل مشكلة تصنيف اربع حركات تخيلية مختلفة وتعد الخوارزمية الثانية اكثر دقة حيث تعمل علي تصنيف الاشارات التخيلية وتنفيذها بسرعة.

CHAPTER ONE

Introduction

1.1 General overview

Brain computer interface (BCI) is an emerging technology for paralyzed people to communicate with external world. A BCI system can “read out” the intention of the patients and translates it into physical commands which control devices that serve the patients. However, there is a real necessity to develop high performance systems and advances in neural devices to help and support that group of people (Kavitha P., 2009).

BCI systems have been designed to help people who has severe disabilities and patients with brain diseases such as epilepsy, dementia and sleeping disorders to interact with their external environment(Lotte, 2008) BCIs are also used for controlling home appliances, lights, television, room temperature, operate the door just by thinking, controlling a robotic car, play computer games, decoding brain activity to reproduce movements in robotic arms, controlling elements in virtual reality, walking in a virtual street by thought, typing a message on computer screen by concentrating on the display, controlling a computer cursor, for spelling words etc. (VargheseJ., 2009).

A BCI system can be classified as an invasive or non-invasive BCI according to the way the brain activity is being measured within this BCI. If the sensors used for measurement are placed within the brain, the BCI is said to be invasive, if the measurement sensors are placed outside the head, the BCI is said to be non-invasive.Also according to the independence of the subjects and the numbers of classes to be classified the BCI systems can be classified into four categories: subjectdependent 2-class BCI (SD-2BCI) systems, subject-dependent multi-class BCI (SDMBCI) systems, subject-independent 2-class BCI (SI-2BCI) systems, and subjectindependent multi-class BCI (SI-MBCI) systems(Hoang, 2014).

Researches on BCI systems have been the center of interest in recent years. Increase in computing power and advances in measurement technology have led to a large variety of proof-of-concept systems is suitable for daily use by disabled subjects.

There are various BCI systems using different methods to extract the subjects' intentions from their electroencephalograph (EEG) signals (Zhang *et al.*, 2011). One of the practical BCI systems is based on motor imagery (MI). The primary phenomenon of MI based EEG is event related desynchronization (ERD) (Pfurtscheller*et al.*, 1997 & Muller *et al.*, 1999). ERD can be induced by both imagined movements in healthy people or intended movements in paralyzed patients (Kübler*et al.*, 2005).The advantage of this type of BCI systems is that no external stimulation is needed. Current development of MI-based BCI is focused on how to discriminate different MI tasks and many algorithms could be applied to get satisfied results. EEG based BCIs are capable of discriminating different types of neural activities such as the imagination of left or right hands, foot and tongue from EEG signals (Zhang*et al.*, 2007).

1.2Research Problem

BCI systems based on motor imagery are used to help people who are suffering from severe motor disabilities but are still cognitively intact, patients in the late stage of Amyotrophic Lateral Sclerosis (ALS) or locked in syndrome are not able to produce any voluntary muscle movements.

Many factors determine the performance of a BCI system. These factors include the brain signals measured, the signal processing methods that extract signal features, the algorithms that translate these features into device commands, the output devices that execute these commands, the feedback provided to the subject, and the characteristics of the subject.

Therefore, from a signal processing point of view, it is important to design a feature extraction mechanism that can learn to capture effective spatial and spectral features associated with the ERD, for each particular person. As a recent survey indicates, considerable efforts have been devoted to this topic by the signal processing, machine learning, and artificial neural networks communities.Then, in order to achieve interesting results, it is expected that the techniques and strategies could be chosen in order to address these points.

1.3 General Objectives

The main objective of this study is to open communication channel for patients who are still cognitively intact but are unable to produce any voluntary muscles movements.

Specific objectives are to

1- Develop of a statistical algorithm to be used as features to automatically select the optimal subject- specific time segments which discriminate between the spatial temporal patterns from EEG signals and corresponding neural activities.

2- Address the classification problem of signal responses by employing a novel modern pattern recognition and machine learning algorithms.

3- Through assessing classification performance, the generalization error will be estimated.

1.4 Thesis Organization

The thesis is organized into six chapters:

The first chapter entitled “Introduction” and gives a short account of the problem definition, thesis objective and its organization. A general introduction to the field of BCI research is also given.

The second chapter entitled "Literature Survey" a brief recollection of the various works exhausted both in the area of signal processing and machine learning techniques.

Thethird chapter entitled “Theoretical Background” Topic reviewed includes description of the basic BCI model, the different methods for measuring brain activity, the types of neurophysiologic signals of interest that is used in BCI system in this thesis, methods for extracting useful features from neurophysiologic signals and BCI applications.

The Fourth chapter entitled “Research Methodology – Theoretical Concepts”. This chapter presents a general exposition of the supervised machine learning problem. It also presents discussion of supervised learning algorithm that has been applied in the context of BCI. The first part describes the acquired data. In the second part the classification approach are discussed.

The fifth chapter “Body of Research” presentsthe supervised learning algorithms used in this thesis are introduced includingpreprocessing step here band pass frequency filtering or spatial filteringusing Second Order Blind Identification and Independent Component Analysis were compared.Support

Vectors Machines is used for the classification problem was discussed in details.

The sixth chapter entitled “Results”This chapter starts with a report about the classification performance that can be obtained with the machine learning algorithm used.

The seventh chapter entitled “Analysis, Discussion and Interpretation of Results” it discusses the algorithms used in the current work and how results were interpretive.

Chapter eight entitled “Conclusionsand Recommendations” it summarizes the contribution of this thesis as a conclusion to the work done. Limitations and constrains were discussed. Finally, an outlook on possible extensions of the presented work is provided as future work for this area ofresearch.

CHAPTER TWO

Literature Survey

2.1 Data Preprocessing

The main objective of pre-processing step is to clean and de-noise data acquired from the original EEG data in orderto enhance relevant information according to Bashashati *et al*., (2007). Besides the main eventthe experiments would like to acquire, there are many types of artifacts from the subjects participating in the experiment and the system. The subject artifacts are body-movement related to electrooculography(EOG), electromyography (EMG), electrocardiography (ECG), and sweating (Fatourechi *et al*., 2007).The system artifacts are electrical noise from electronic components, andcable defects. Theseartifacts make the recorded EEG signal to have a low signal-to-noise ratio (SNR). Artifacts also come from technical sources such aspower-line noises or changes in electrode impedances.

In this research, preprocessingmethods are defined as methods that try to reduce noise or artifacts. Bythis definition, pre-processing methods are artifact-removal methods only.There aretwo typical methods for removing artifacts in BCI. The first method used includes low-pass or band-pass filters which are based on the Discrete Fourier Transform (DFT),Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) demonstrated by Smith(1997).

Inmotor imagery-based BCI systems, low-pass or band-pass filters are usually used tocut off irrelevant frequencies. This method is very efficient when dealing with technical artifacts. The second method widely used for removing artifacts is IndependentComponent Analysis (ICA). This is a statistical method aimed at decomposing aset of mixed signals into its sources. Pioneer work such as that of Vigario (1997) and Vigrio *et al.*, (2000) has aimed at removing ocular artifacts from EEG signals.While ICA has been proven a robust and powerful tool for removing artifacts in theanalysis of EEG signals in Fatourechi *et al.*, (2007), reports from some authors suggestthat removing artifacts by using ICA may corrupt the power spectrum of the analyzed EEG signal Wallstrom *et al.*, (2004). Furthermore, ICA requires that artifactsbe independent of the normal activity of the analyzed EEG signal. This requirementis sometimes not easy to satisfy due to the complicated and relatively unknown operation of the brain. In this research, IIR band-pass filters, ICA, and Second-Order Blind Identification (SOBI) were comparedfor pre-processing EEG signals.

2.1.1 Frequency Filtering

To measure the signals produced by the brain, EEG are used. With the data recorded using 22 electrodes, the electrodes are placed conforming the 10-20 system,adding 3 extra electrodes to pick EOG signals (Oostenveld *et al.*, 2001).As explained in chapterthree, brain waves can be divided in four different categories based on their frequency range. Gamma waves are added as the fifth category, covering the range from (25-100)Hz as they have proven to play a role in all sensory modalities (Hughes, 2008). To remove unwanted artifacts and extract the most important information from the EEG measurements, the data is pre-processed.

Frequency filtering is an effective method to remove the noise. FIR and IIR band pass filters in range of (7 - 40) Hz is the most common method used.

Mahnaz *et al.* (2011) filtered the EEG data with band-pass filtered using elliptic filters from (8 - 35)Hz.

Habiba *et al.* (2012) & Loannis *et al*. (2018) used a zero phase forward/backward FIR band pass filter with frequency band (8 - 30)$Hz$and (7 - 15) Hz respectively.

Chuong *et al*. (2017) &Hardik *et al.* (2018) placed a Butterworth Band pass filter of order 5 to filter EEG signals. The frequencies are allowed to passbetween (8 - 30)$Hz$and (8 - 40) Hzrespectively.

Saroosh *et al*. (2017) used two Chebychev type-II filters of order 54 are applied to each segment to extract the frequency contents ofmu (μ)waves (8-13) $Hz$and thebeta (β) waves (13-30)Hz, also Thanh *et al*. (2018)implement a zero-phase 10th order Chebyshev type-II band-pass filter in the SMR bands(7-35)Hz.

In Siavash *et al*. (2018) method the EEG signals were filtered using a filter bank with nine subsequent bandpass filters, starting at 4 Hz and with a bandwidth of 4 Hz, all filters are type-II Chebyshev filters. All those methods are gathering and selecting alpha (α), mu (μ), and beta (β) waves which they are associated with a relaxed awareness and active thinking.

2.1.2 Spatial Filtering

One of the conventional and efficient approaches for detection is Independent Component Analysis. Many researchers integrated properties of ICA to upgrade method for better performance. The positive feature that popularized this method is its ability to cope with diverse artifacts such as eye blink, muscle and electrical (caused due to impedance of electrodes). ICA belongs to the blind source separation category that differentiates the EEG waveforms with maximal independence against each other. A specific pattern in the ICA components are found for eye blinks and muscle activities. In EEG signals these artifacts overlap with original source signal and thus ICA tends to distinguish and measure the overlapping projection.

ICA exploits higher-order statistical dependencies among data and discovers a generative model for the observed multidimensional data. In the ICA model, observed data variables are assumed to be linear mixtures of some unknown independent sources (independent components).A mixing system is also assumed to be unknown. Independent components are assumed to be non-Gaussian and mutually statistically independent. ICASpatial filtering proved that it is a successful way to remove noise.

Joyce *et al.* (2004) report about an extensive study on removal of ocular artifacts discussing and comparing different EOG electrode placement strategies as well as different algorithms for blind source separation. They present an automatic approach for extracting and removing ocular components after computation of ICA. Independent components are removed if:

1. Their activation is inverted in an ICA solution for EEG plus EOG channels relative to an ICA solution for EEG plus inverted EOG channels (inversion of component indicating relation to EOG channels).
2. They correlate with EOG channels above a certain threshold.
3. They show high power in low frequency bands.

Although the authors claim that their approachesare fully automatic the setting of thresholds for steps (2) and (3) make them semi-automatic at best. Results are presented graphically and in addition averages of artifact free trials are shown to be similar to averages of the same trials after removal of oculartifacts. The latter is provided as a proof that the removal of ocular artifacts.

GouyC.*et al*. (2008) presents a method to recover task-related sources from a multi-class Brain-Computer Interface (BCI) based on motor imagery. This method gathers two common approaches to tackle the multi-class problem: (i) the supervised approach of Common SpatialPattern (CSP) to discriminate between different tasks; (ii) the criterionof statistical independence of non-stationary sources used in Independent Component Analysis (ICA). Results showthat the spatial filters have to be adapted to each subject and that thecombined useof intra-trial and inter-class energy variations of brain sources yields an increase of classificationrates for four among eight subjects.

BaiX.*et al*. (2014) introduced a negative entropy-based FastICA algorithm to filter the EEG signals from the artifacts then the independent components were selected based on two things, frequency of the motor imagery-related signals are mainly in μ and β rhythms around 10Hz and 20Hz, and cross-correlation between the ICs and the raw EEG of C3, Cz, C4 in the related cortical areas.

Guillermoand Humberto (2017) evaluate different softwareimplementationsof ICA approaches using MATLAB and according to some criteria (running time, allocated memory, accuracy and scalability) targetingfour ICA algorithms: Second-Order Blind Identification (SOBI), Hyvarinen’s fixed-point algorithm (FastICA), logistic Infomax (Infomax) and Joint Approximation Diagonalization of Eigenmatrices (JADE). The outcomes have shown that SOBI’s MATLAB implementation isthe best procedure among all the analyzed techniques by drastically overcomingthespeed of the others algorithms. Moreover, its correlation grades, corresponding to Pearson and Spearman correlation coefficients respectively, indicate it as one of the more accurate algorithms.

2.2Feature extraction and classification

Common spatial pattern CSP is common technique used by most authors because of its effectiveness of discrimination between two different classes and extended to multi classes.

Ramoser *et al.* (2000) designed spatial filters for multi-channel EEG,recorded during left and right hand movement imaginary. The best classificationresults for three subjects are 90.8%, 92.7%, and 99.7%. The spatial filters are estimated from a set of data by the method of common spatial pattern and reflect the specific activation of cortical areas. The method performs a weighting of the electrodesaccording to their importance for the classificationtask. The highrecognitionrate and computational simplicity make it an optimal methodfor an EEG based brain computerinterface.

Chin*et al.* (2009) investigated the performance of three approaches proposed for the multi-class extension to the Filter Bank Common Spatial Pattern (FBCSP) algorithm. As the CSP algorithm in FBCSP was originally formulated for binary-class problems, these three proposed approaches of multi-class extension to the FBCSP algorithm: One-versus-Rest, Pair-Wise and Divide-and-Conquer. These three approaches were evaluated on the 4-class motor imagery data of BCI Competition IV dataset IIa. The experimental results showed no significant difference between the three proposed approaches. The One-versus-Rest (OVR) approach was submitted for the competition and performed the best on the evaluation data relative to the other submissions. The results also showed the multi-class FBCSP algorithm could extract features whose spatial patterns matched with neurophysiological knowledge. The variability in the performance of the multi-class FBCSP algorithm on the different classes of motor imagery action across all subjects could be due to the presence or absence of ERD/ERS effects in certain motor imagery actions.

Acombined CSP method is presented by Mahnaz *et al*. (2011), here the filtered EEG data were then used to select the optimal channels, sparse common spatial pattern(SCSP) algorithm was proposed for optimal EEG channel selection, The performance of the proposed algorithm was compared with several other channel selection methods, based on the Fisher criterion (FC), MI, SVM, CSP coefficients, and the regularized common spatial pattern (RCSP), The radial basis function was used as the kernel function of the SVM for the classification, and the hyperparameters were obtained by a grid search using cross validation on the training data, results showed that the proposed SCSP channel selection significantly reduced the number of channels, and outperformed the other existing channel selection methods in classification accuracy, also the proposed SCSP algorithm yielded an average improvement of 10% in classification accuracy compared to the use of three channels (C3, C4, and Cz).

Hsu*et al*. (2011) proposed an EEG analysis system for the single-trial recognition of MI EEG data. Associated with active segment selection and multi resolution fractal features, the Fuzzy c-means (FCM) are applied to discriminate left MI from right one. Active segment selection is an effective scheme that selects active segments in thetime-frequency domain. It makes the length of original event-related window substantiallyreduce to a 1-s segment and increases the speed of feature extraction at the same time.The multiresolution fractal features MFFVs are then obtained using proposed modifiedfractal dimension from discrete wavelet transform (DWT) data. The features are so good and discriminative that theycan enhance the classification of mental tasks. Finally, the FCM is used for the discriminant of MI EEG data. In addition, the FCM is not only a robust approach suitable forthe classification of non-stationary MI EEG signals, but is also capable of making flexiblepartitions of a finite data set. The experimental results demonstrate that the recognitionwith the FCM possesses promising potential in the application of BCI works.

In Ang *et al*. (2012) the filter bank common spatial pattern (FBCSP) algorithm is presented to classify single-trial EEG data for 2-class as well as 4-class motor imagery, where results using different feature selection algorithms and multi-class extensions to the FBCSP algorithm were compared with the CSP algorithm and other entries submitted to the BCI Com-petition IV Dataset 2a and Dataset 2b. Although other algorithms were not included in this study, prior studies on the2-class motor imagery data of the BCI Competition III Dataset IV showed that a modified SPEC-CSP algorithm using Support Vector Machines (SVM) yielded a 10×10-fold cross-validation classification accuracy of 89.5% averaged overthe 5 subjects(Wu *et al*., 2008).While the FBCSP algorithm yielded a 10×10-foldcross-validation classification accuracy of 90.3% (Ang *et al*., 2008).Although they might not be directly comparable, results fromthese prior studies suggest that the SPEC-CSP algorithm mightyield similar results as the FBCSP algorithm in Dataset 2a and 2bas well. The results on the Filter Bank Common Spatial Pattern (FBCSP) algorithm showed that it is capable of performing an autonomous selection of discriminative subject-specific frequencyrange for band-pass filtering of the EEG measurements. In the 2-class motor imagery data in Dataset 2b, even though the EEG datawas limited to 3 bipolar recordings, the FBCSP algorithm yieldedthe best performance among all the submissions by employing either the Mutual Information-based Rough Set

Reduction(MIRSR) or Mutual Information-based Best Individual Features(MIBIF) feature selection algorithm. The MIBIF feature selectionalgorithm is dependent on a Metaparameter, the number of features selected, which was set-based on the results obtained on the2-class motor imagery data from the previous BCI CompetitionIII Dataset 4a inAng et al. (2008). Hence further improvementusing the MIBIF feature selection algorithm might be possibleby optimizing the number of selected features via a nested cross-validation approach instead. In the 4-class motor imagery datain Dataset 2a, even though the FBCSP algorithm was initiallydesigned for 2-class motor imagery,the results on the 4-class motorimagery data in Dataset 2a showed that the one-versus-the-rest(OVR) and the pair-wise (PW) approaches of multi-class extension to the FBCSP algorithm could also yield relatively the bestperformance as well.

In Habibeh*et al*. (2012) after filtering, EEG signals were divided into a number of time windows, features were extracted from each time window using one versus rest (OVR) CSP algorithm, to classify these feature vectors, four fisher’s linear discriminant analysis (LDAs) were used, then the kappa score was measured and compared with the result of the best competitor of the competition IV, the result was very good in average of0.61 kappa coefficient.

Bai X.*et al*. (2014) introduced a study in which CSP filter coefficients also can be computed from wavelet coefficients. they applied Wavelet Transform to the EEG signal to obtain the wavelet coefficients of 8~30 Hz, Because the frequency bands of [7.8125, 15.625] and [15.625, 31.25] include the μ and β rhythms, the related wavelet coefficients [cD_4, cD_3] was used as the input of CSP where the 2m eigenvectors corresponding to the m largest eigenvalues were chosen to form a new matrix that are optimal for discriminating two classes. After that support vector machine classifier was adapted according to the ERD/S around [C_3, C_4, C_z] and used to classify the four classes of motor imagery. The results of the proposed method were compared with the results of the conventional CSP based on the (8-30) Hz IIR band-pass filters from BCI Competition IV in term of Kappa score, the proposed method produced a higher average k value of 0.68 than 0.52 of conventional CSP with a band-pass filter, in addition to the higher classification rate the proposed system showed more stability and adaptability.

In AlomariM. *et al*. (2014)EEG signals associated with imagined fists and feet movementswere filtered and processed using wavelet transform analysis for feature extraction. The proposed work usedNeural Networks (NNs) as a classifier that enables the classification of imagined movements into either fists orfeet. Wavelet families such as Daubechies, Symlets, and Coiflets wavelets were used to analyze the extractedevents and then different feature extraction measures were calculated for three detail levels of the waveletcoefficients. Intensive NN training and testing experiments were carried out and different network configurationswere compared. The optimum classification performance of 89.11% was achieved with a NN classifier of 20hidden layers while using the Mean Absolute Value (MAV) of the Coiflets wavelet coefficients as inputs to NN.The proposed system showed a good performance that enables controlling computer applications via imaginedfists and feet movements.

In Aydemir *et al.* (2016) researchers investigated the effect of feature extraction using different mother wavelets (MWs) on classification results of different motor imagery classes. Features were extracted from three different datasets: BCI competition II dataset Ia (Dataset 1), BCI competition II dataset III (Dataset 2), and BCI competition III dataset I (Dataset 3).Using twelve MWs, including Morlet, Shannon1-1.5, (shan1-1.5), Shannon2-3 (shan2-3), Daubechies1 (db1), Daubechies4 (db4), Symlet2 (sym2), Symlet5 (sym5), Gaussian5 (gaus5), Gaussian6 (gaus6), Meyer, Coiflet3 (coif3), and Coiflet4(coif4), and then the signals were classified using three classification algorithms, including k-nearest neighbor, support vector machine, and linear discriminant analysis. According to the obtained statistical voting results, it could be generalized that Daubechies and Shannon were the most suitable wave functions for extracting more discriminative features from the imaginary EEG/ECoG signals. The general characteristics of these wave functions should be examined to understand and reveal the cause of their superior performance. In comparison to other tested wavelets, Daubechies wavelets are compactly supported with extreme phase and the highest number of vanishing moments, which are a necessary condition for the smoothness of the wavelet function and for a given support width. On the other hand, Shannon wavelets are analytically defined, infinitely differentiable, and sharply bounded in the frequency domain. These advantages provide a very good localization of energy in the frequency domain and make those wave functions the best candidates to identify the EEG/ECoG-based BCI signals. On the contrary, it is worthwhile to mention that the Morlet wavelet presented the poorest performance due to the fact that it is a fourierbased timefrequency transformation andthus suffers from many of the shortcomings of Fourier analysis. Also it was observed that the LDA algorithm achieved much better performance than the SVM and k-NNalgorithms in terms of the obtained highest classification accuracy results. Moreover, the LDA classifier wasmore robust than the SVM and k-NN algorithms, because it had only limited flexibility (less free parametersto tune) and was less prone to overfitting.

Oana D. andAnca (2017) evaluatedthree feature extraction methods:independent component analysis, Itakura distance and phase synchronization. These features of a motor imagery paradigm based on Mu rhythm were classified with five classification methodsusing linear discriminant analysis, quadratic discriminant analysis (QDA), Mahalanobis distancebased on classifier, the k-nearest neighbors (kNN) and support vector machine (SVM).The method applied on two different databases. The algorithms are simply to apply and can be exploited by the motor imagery paradigms.In order to have a properpreparation, the subjects from their own database executed first the hand movements and then the hand movement imagination. For the subjects fromthe BCI competition 2002 databaseit is mentioned that they were well trained.Overall the highest classification rates are obtained with QDAand with kNN classifier. The results suggest that the effectiveness of the feature extraction method depends onthe classification method used.

Researchers in Saroosh *et al.* (2017) proposed the coupling of error correction output coding (ECOC) with the common spatial pattern (CSP) analysis. At first two Chebychev type-II filters of order 54 are applied to each segment to extract the frequency contents of (8-13)Hzband and the (13-30) Hz band. Then features were extracted with common spatial filters where the eigenvectors corresponding to the largest and smallest eigenvalues were selected after the normalization to imported to ECO-CSP algorithm to train seven binary classifiers, SVM and LDA classifiers were used, the performance of the classifiers was tested using 10-fold cross validation, then Kappa values were calculated (SVM=35.34, LDA=35.54).Results showed that ECO-CSP achieve similar performance in comparison to the state-of-the-art algorithms.

Joint Approximate Diagonalization (JAD) using fast Frobenius diagonalization (FFDIAG) is used by Hardik *et al.* (2018) to extend CSP algorithm for two class problems to multiclass. This method reduces the effect of noisy trials on Eigen values and Eigen vectors Frobenius norm of each covariance matrix obtained from different trials is calculated. Frobenius norm is square root of the sum of the diagonal elements of the matrix, to classify these features Self-Regulated Interval Type-II Neuro-Fuzzy Inference System (SRIT2NFIS) was implemented, it was good to handle the non-stationarity of EEG signals, then accuracies were obtained and compared with the currently state of the art algorithms for multiclass classification, results showed that the mean accuracy (54.63%) increased to 8-13%.

Also in Loannis*et al.* (2018) OVR technique is used but firstly, filtered data were down-sampled at 100 Hz and epoched for 500 msec after the visual cue with epoch duration of 3000 ms, EEG source imaging was deployed to mitigate low spatial resolution and low SNR caused by volume conduction, weighted minimum norm estimate (WMNE) method using the Brainstorm toolbox was used to solve the inverse problem, after that 24 regions of insert were defined to reduce the dimension of the source data derived from the inverse problem solution, these regions were selected according to neuroanatomical landmarks and Broadman areas, on these ROIs features extraction using common spatial filters with one versus the rest (OVR) extension was performed, these features were classified by independent ROI classification models, K-nearest neighbors (kNN), Naive Bayes, Decision Tree, and Linear Discriminant Analysis (LDA) classifier were tested and compared, and the final classification outcome was selected by an inference mechanism (majority vote). After that specific ROIs(Q=8) were selected according to parametric analysis results of the inference mechanism accuracy using LDA classifier, classification with source apace method was compared with sensor space method in terms of accuracy, kappa value, sensitivity and specificity, compared to the winner of BCI Competition IV of dataset 2. LDA had the best performance of the other used classifiers with mean accuracy 54.1%, the source method of classification achieved higher accuracy rates (43.7% to 74.5%), when compared to the sensor method (37.7% to 73.4%), the source method has mean 11.1% higher true positive rate for the left arm, 5.2% higher for right arm, and 3.3% and 1.9% better rate for foot and tongue imagery, respectively. Themean differences of sensor to source true negative rate metric are low for all classes, −1.2%, 1.9%, 2.9%, and 3.6% for left, right arm, foot, and tongue imagery, respectively. Source method improved mean accuracy by 5.6% and by 0.07 Cohen’s kappa value among all subjects, with respect to sensor method. This study supports that BCI algorithms based on source space features are superior to the sensor ones. The proposed algorithm did not reach the accuracy levels of the wining method of the BCI Competition (with 0.65 mean Cohen’s kappa value); the mean accuracy is considerably lower by 0.19 Cohen's kappa value. Where in Siavash*et al.* (2018) CSP energies were computed from the spatial filters corresponding to the 2 × NW extreme eigenvalues (NW largest and NW smallest eigenvalues) and used to be the features, after that temporal features were extracted, the signal envelope of each signal was extracted using the Hilbert transform ,after extracting the envelope three possible representations for the EEG were considered, then they concatenated all four classes forming a single matrix of signals, then convolutional neural network (CNN) was designed and optimized accordingly for the representation and classification. This system performed the best classification method in the literature on the BCI competition IV-2a 4-class MI data set by 7% increase in average subject accuracy.

A method introduced by Thanh*et al*. (2018) for multi-class MI-BCI data classification. EEG signals of data set IIa from BCI competition IV were filtered then features were extracted using CSP algorithm which is extended to multiclass using OVR approach. Then a fuzzy standard additive model with featuring Mamdani fuzzy rules(if-then), sum-product inference, and center of gravity (centroid) defuzzifier was designed for the classification, particle swarm optimization (PSO) algorithm was applied to optimize the rule structure and train the parameters of the multi-class FLS. The proposed PSO-based FLS was compared with multi-class LDA, NB, KNN, ensemble AdaBoostM2, and SVM; these algorithms were applied by using their MATLAB multi-class fitting functions. For the evaluation of the performance of the proposed approach against the other competing techniques the 10-fold cross-validation method was used, and the outcomes in terms of average and standard deviation were presented. Results shown for data set IIa FLS yet again is the best method, the average accuracy rate of the FLS across 9 subjects is 0.650, while its average kappa score is 0.533 which is much higher than 0.52 from the competition winner.

CHAPTER THREE

Theoretical Background

3. Brain Computer Interface

3.1 System Overview

For many years, people have speculated that electroencephalographic (EEG) activity or other measures of brain function might provide this new channel. Over the past decade, productive BCI research programs have begun. Facilitated and encouraged by the new understanding of brain functions and by the low-cost computer equipment, these programs have concentrated mainly in developing new communication and control technologies for people with severe neuromuscular disorders. The immediate goal is to provide communication capabilities so that any subject can control the external world without using the brain's normal output pathways of peripheral nerves and muscles.

Nowadays, such activities drive their efforts in brain (neural) signal acquisition which is a development of both invasive and non-invasive techniques for high quality signal acquisition. Algorithms and processing are playing a key point and advantage of cheap/fast computing power (i.e. Moore's Law) to enable online real-time processing where manipulate advanced machine learning and signal processing algorithms (Jorge B., 2002).

Present BCI’s use EEG activity recorded at the scalp to control cursor movement, select letters or icons, or operate a neuroprosthesis. The central element in each BCI is a translation algorithm that converts electrophysiological input from the user into output that controls external devices. BCI operation depends on effective interaction between two adaptive controllers: the user who encodes his or her commands in the electrophysiological input provided to the BCI, and the computer which recognizes the command contained in the input and expresses them in the device control. Current BCI's have maximum information transfer rates of 5-25 bits/min. Achievement of greater speed and accuracy depends on improvements in:

1. Signal acquisition: methods for increasing signal-to-noise ratio (SNR), signal-to interference ratio (S/I)) as well as optimally combining spatial and temporal information.
2. Single trial analysis: overcoming noise and interference in order to avoid averaging and maximize bit rate.
3. Co-learning: jointly optimizing combined man-machine system and taking advantage of feedback.
4. Experimental paradigms for interpretable readable signals: mapping the task to the brain state of the user (or vice versa).
5. Understanding algorithms and models within the context of the neurobiology: building predictive models having neurophysiologically meaningful parameters and incorporating physically and biologically meaningful priors (Jorge B., 2002).

The common structure of a Brain Computer Interface is the following (Fig 3-1):

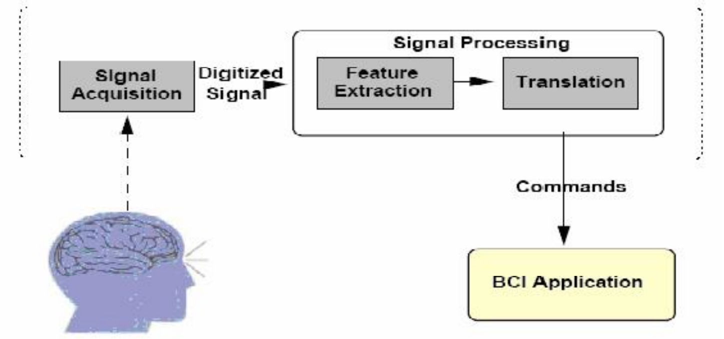


Fig.3-1: Common Structure of a BCI System (G. Schalk *et al.*, 2004).

3.2Signal Acquisition

Several non-invasive and invasive signal acquisition techniques have been used in BCI research. In non-invasive electroencephalography (EEG) and magnetoencephalography (MEG), the electromagnetic activity of the brain is measured by the electrodes placed over the skull. In invasive electrocorticograpy (ECoG), single micro-electrode (ME), micro-electrode array (MEA), and local field potentials (LFPs), the electrodes are placed surgically inside the skull to measure the cortical activity. Functional Magnetic Resonance Imaging (fMRI) several non-invasive and invasive signal acquisition techniques have been used BCI research. In non-invasive electroencephalography (EEG) and magnetoencephalography (MEG), the electromagnetic activity of the brain is measured by the electrodes placed over the skull (Fig. 3-2).

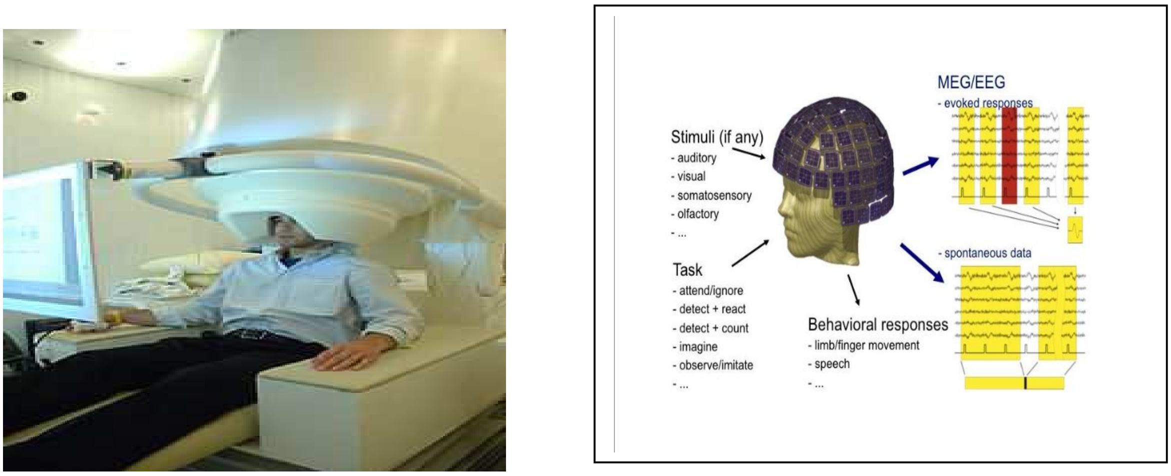


Fig. 3-2: (a) Patient undergoing a MEG (Magnetoencephalography, http://en.wikipedia.org/wiki/Magnetoencephalography).
(b) A MEG Experiment (What is Magnetoencephalography, http://ilabs.washington.edu/what-magnetoencephalography-meg)

In invasive electrocorticograpy (ECoG), single micro-electrode (ME), micro-electrode array (MEA), and local field potentials (LFPs), the electrodes are placed surgically inside the skull to measure the cortical activity(Fig 3-3).

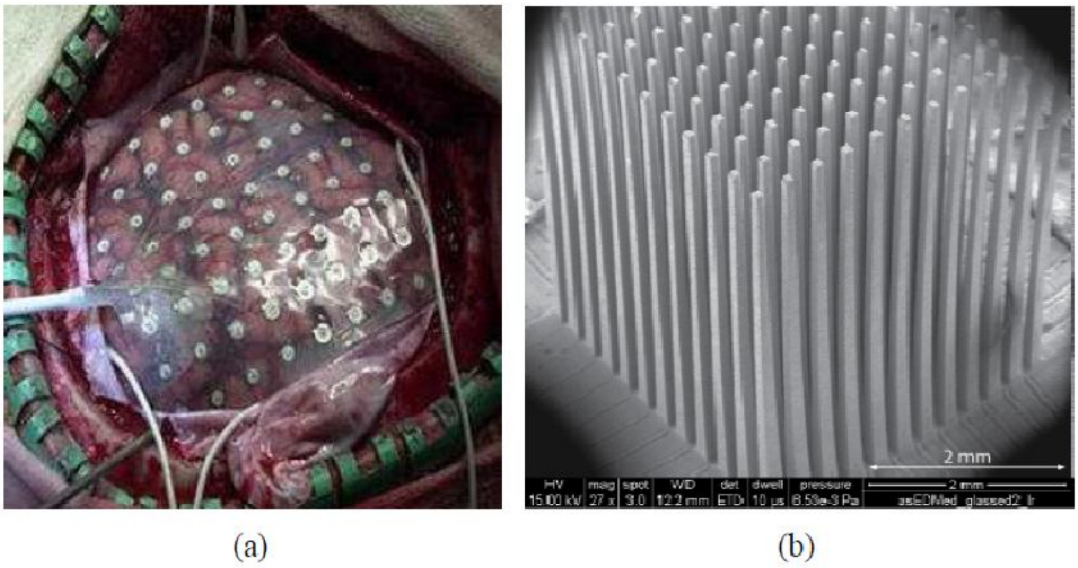


Fig.3-3: (a) ECoG electrodes over the cortex(medGadget, http://medgadget.com/archives/2007/03/the_first_comme_1.html.)
(b) Cortical microelectrode array (Precision Design Laboratory, http://www.mech.utah.edu/~bamberg/)

Functional Magnetic Resonance Imaging (fMRI) and Near Infrared Spectroscopy (NIRS), in which regional changes in cerebral blood oxygenation levels are detected non-invasively, are also used in BCI and Near Infrared Spectroscopy (NIRS), in which regional changes in cerebral blood oxygenation levels are detected non-invasively, are also used in BCI.

In this proposed EEG signal will be used to investigate the information carried out form brain to the specific organ. Next section will discuss the principles of EEG signal.

3.2.1Principles of Electroencephalography

3.2.1.1The Nature of the EEG Signals

The electrical nature of the human nervous system has been recognized for more than a century. It is well known that the variation of the surface potential distribution on the scalp reflects functional activities emerging from the underlying brain (KandelE.R. *et al.*, 1991). This surface potential variation can be recorded by affixing an array of electrodes to the scalp, and measuring the voltage between pairs of these electrodes, which are then filtered, amplified, and recorded. The resulting data is called the EEG. Fig. 3-4shows waveforms of a 10 second EEG segment containing six recording channels, while the recording sitesare illustrated in Fig. 3-5. In this experiments, the10-20 system ofelectrode placementwas used, which is based on the relationship between the location of anelectrode and the underlying area of cerebral cortex (the "10" and "20" refer to the 10%or 20% interelectrode distance) (http://faculty.washington.edu/chudler/1020.html).

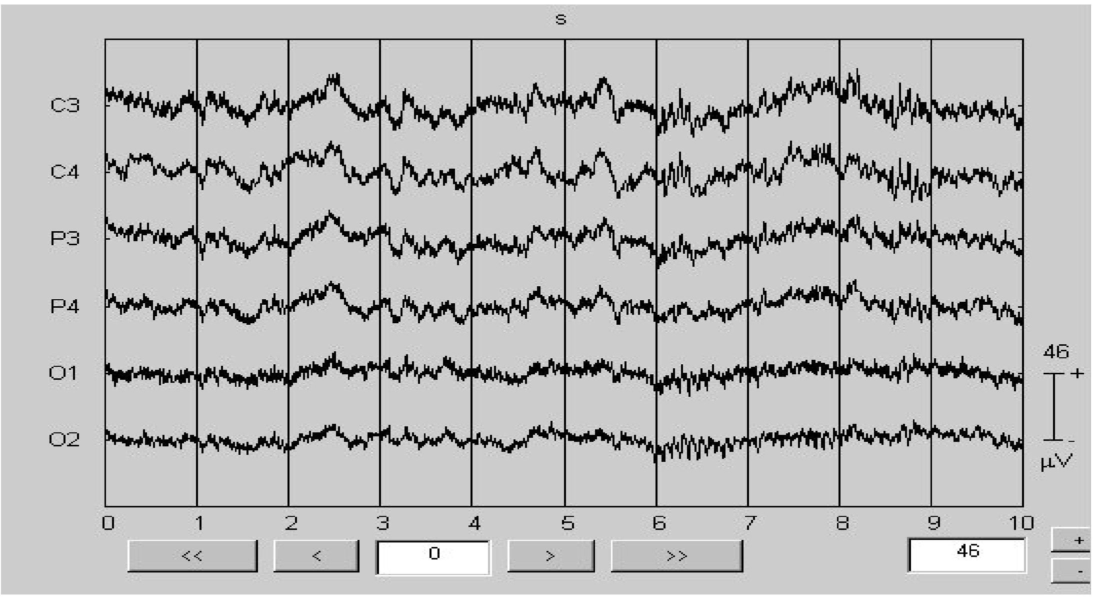


Fig. 3-4: A Segment of a multichannel EEG of an adult subject during a multiplication task (Jorge B., 2002).

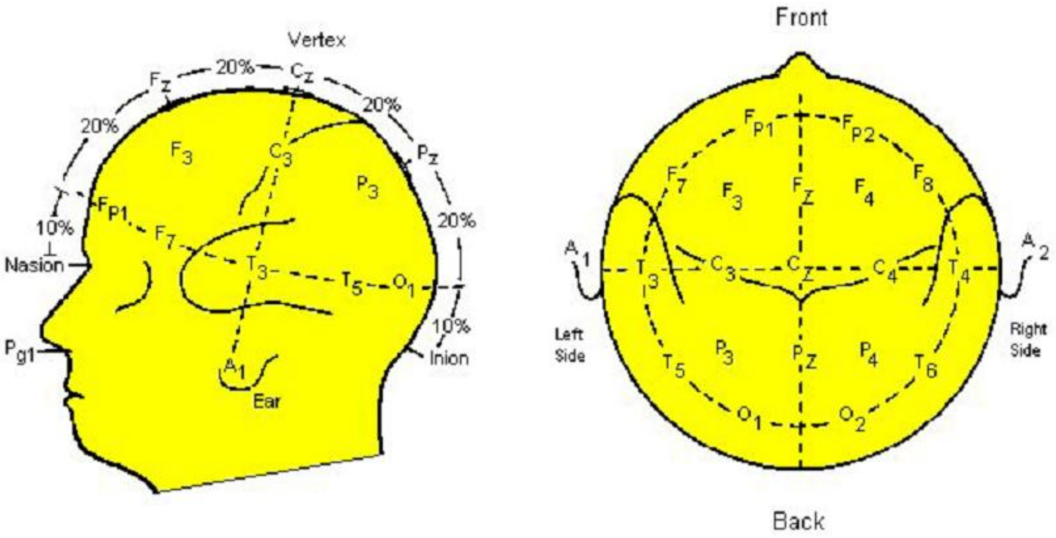


Fig.3-5: The 10-20 System of Electrode Placement (Jorge B., 2002).

Each site has a letter (to identify the lobe) and a number or another letter to identify the hemisphere location. The letters F, T, C, P, and O stand for Frontal, Temporal, Central, Parietal and Occipital. (Note that there is no "central lobe", but this is just used for identification purposes.) Even numbers (2, 4, 6, and 8) refer to the right hemisphere and odd numbers (1, 3, 5, and 7) refer to the left hemisphere. The z refers to an electrode placed on the midline. Nasion:point between the forehead and nose. Inion:Bump at back of skull

The EEG is thought to be the synchronized sub-threshold dentritic potentials produced by the synaptic activity of many neurons summed (Orrison Jr. *et al*., 1995). In its formation not all types of brain activity have identical impact. The depth, orientation and intrinsic symmetry of connections in the cortex are significant in it. As it is exposed in previous works (Orrison Jr. *et al*., 1995) (LopesF. H.*et al*., 1982), pyramidal cells are thought to cause the strongest part of the EEG signal.

Nowadays, modern techniques for EEG acquisition collect these underlying electrical patterns from the scalp, and digitalize them for computer storage. Electrodes conduct voltage potentials as microvolt level signals, and carry them into amplifiers that magnify the signals approximately ten thousand times. The use of this technology depends strongly on the electrodes positioning and the electrodes contact. For this reason, electrodes are usually constructed from conductive materials, such us gold or silver chloride, with an approximate diameter of 1 cm, and subjects must also use a conductive gel on the scalp to maintain an acceptable signal to noise ratio. This method of EEG signal recording is shown in Fig. 3-6.

Fig 3-6: EEG signal recording (Jorge B., 2002).

3.2.1.2EEG wave groups

The analysis of continuous EEG signals or brain waves is complex, due to the large amount of information received from every electrode. As a science in itself, it has to be completed with its own set of perplexing nomenclature. Different waves, like so many radio stations, are categorized by the frequency of their emanations and, in some cases, by the shape of their waveforms. Although none of these waves is ever emitted alone, the state of consciousness of the individuals may make one frequency range more pronounced than others. Five types are particularly important:

BETA: The rate of change lies between 13 and 30 Hz, and usually has a low voltage between 5-30 µV (Fig. 3-7). Betais the brain wave usually associated with active thinking, active attention, and focus on the outside world or solving concrete problems. It can reach frequencies near 50 Hz during intense mental activity.

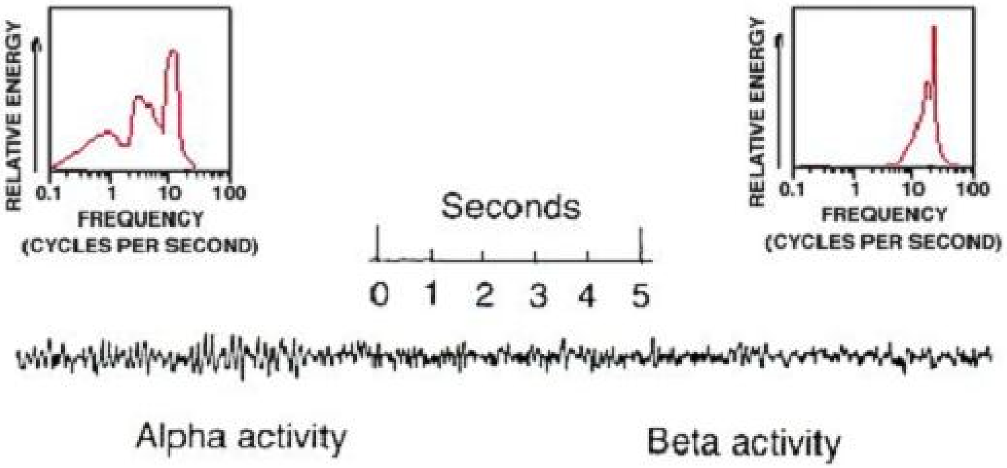


Fig 3-7: Alpha (left) and Beta (right) waves(Jorge B., 2002).

ALPHA: The rate of change lies between 8 and 13 Hz, with 30-50 µV amplitude (Fig 3-7). Alpha waves have been thought to indicate both a relaxed awareness and also inattention. They are strongest over the occipital (back of the head) cortex and also over frontal cortex. Alphais the most prominent wave in the whole realm of brain activity and possibly covers a greater range than has been previously thought of. It is frequent to see a peak in the beta range as high as 20 Hz, which has the characteristics of an alpha state rather than a beta, and the setting in which such a response appears also leads to the same conclusion. Alpha alone seems to indicate an empty mind rather than a relaxed one, a mindless state rather than a passive one, and can be reduced or eliminated by opening the eyes, by hearing unfamiliar sounds, or by anxiety or mental concentration.

THETA: Theta waves lie within the range of 4 to 7 Hz, with an amplitude usually greater than 20 µV Fig. 3-8. Theta arises from emotional stress, especially frustration or disappointment.

Theta has been also associated with access to unconscious material, creative inspiration and deep meditation. The large dominant peak of the theta waves is around 7 Hz.

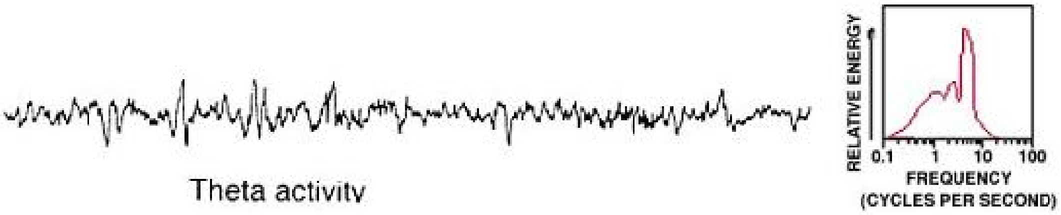


Fig 3-8:Theta wave (Jorge B., 2002).

DELTA: Delta waves lie within the range of 0.5 to 4 Hz, with variable amplitude Fig. 3-9. Delta waves are primarily associated with deep sleep, and in the waking state, were thought to indicate physical defects in the brain. It is very easy to confuse artifact signals caused by the large muscles of the neck and jaw with the genuine delta responses. This is because the muscles are near the surface of the skin and produce large signals whereas the signal which is of interest originates deep in the brain and is severely attenuated in passing through the skull. Nevertheless, with an instant analysis EEG, it is easy to see when the response is caused by excessive movement.

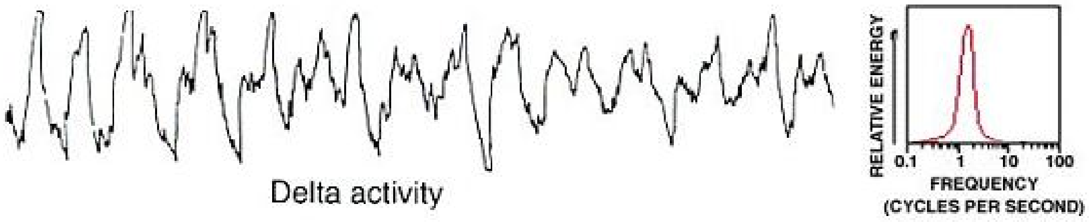


Fig 3-9: Delta wave (Jorge B., 2002).

GAMMA: Gamma waves lie within the range of 35Hz and up. It is thought that this band reflects the mechanism of consciousness the binding together of distinct modular brain functions into coherent percepts capable of behaving in a re-entrant fashion (feeding back on them over time to create a sense of stream-of-consciousness).

MU: It is in the same frequency band as in the alpha wave (Fig. 3-10), but this last one is recorded over occipital cortex.Mu wave is an 8-12 Hz spontaneous EEG wave associated with motor activities and maximally recorded over motor cortex (Fig. 3-11). They diminish with movement or the intention to move.

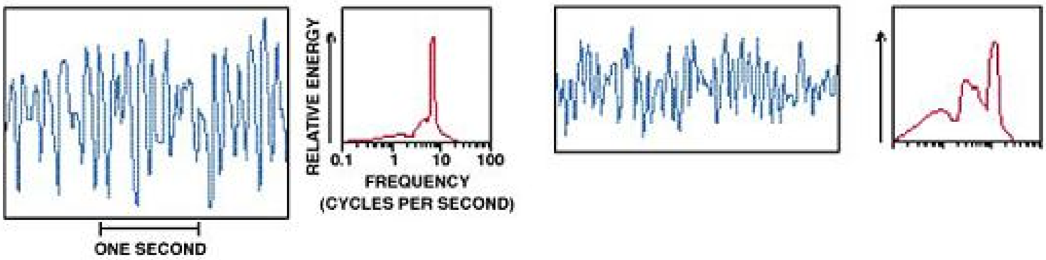


Fig 3-10: Mu (left) and alpha (right) waves (Jorge B., 2002).

Most attempts to control a computer with continuous EEG measurements work by monitoring alpha or mu waves, because people can learn to change the amplitude of these two waves by making the appropriate mental effort. A person might accomplish this result, for instance, by recalling some strongly stimulating image or by raising his or her level of attention.

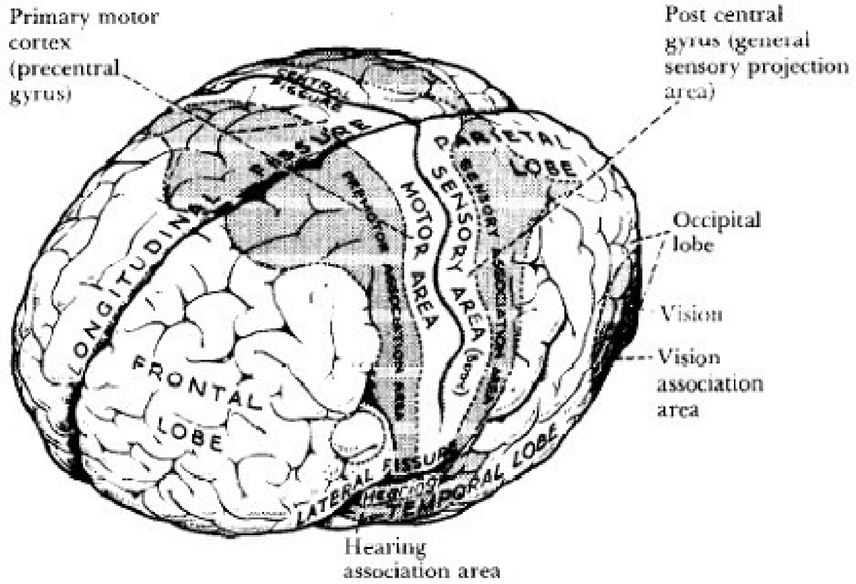


Fig 3-11: Cerebral hemispheres showing the motor areas (towards the front) and the sensory areas (towards the back) (Jorge B., 2002).

3.3 Supervised Machine Learning

To allow actual control of a BCI, the complex neurophysiologic signals have to be translated into simple commands for controlling a computer application or a device. The most straightforward approach is to map signals into command signals.The commands is probably to look at the distribution of a small number of simple features of the signals rule in order to be mapped to values that allow the discrimination of different classes of signals, i.e. the neurophysiologic signals have to be classified, and to manually specify a translation rule. This method has been used in early prototype of BCI systems.

For example, in the work described by (Wolpaw*et al.*,2000), subjects could move a cursor up and down by modifying their mu-rhythm amplitude. To translate mu-rhythm amplitude into cursor movements, different voltage ranges were fixed manually by an operator based on the characteristics of previously recorded signals. However, as noted by (Wolpaw *et al.*,2000) even if only one feature is used, it is difficult for a human to specify an optimal mapping between signals and commands. If more features are used, it quickly becomes impossible to manually design mappings. Moreover, neurophysiologic signals show a relatively large variance between subjects. This means that translation rules have to be specified for each new subject that wants to access a BCI system.

A solution to these problems that is used in almost all BCI systems is after the first step which underlay most methods for classification of neurophysiologic signals is to acquire labeled training data. Acquiring labeled training data means that the subject has to perform prescribed actions, while neurophysiologic signals are recorded. Then, machine learning algorithms are applied to infer functions that can be used to classify neurophysiologic signals. Then, a computer is used to learn the desired mapping between signals and commands. Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised machine learning problems. Algorithms that learn from set of training examples how to map inputs to desired outputs are called supervised learning algorithms.

For example in mu-rhythm BCI, the result of training can be a set of trials in which the subject has imagined left hand movement and another set of trials in which the subject has imagined right hand movement. After the training phase better results can be obtained by adopting a supervised machine learning approach to learn the desired mapping of neurophysiologic signals into commands.

Machine learning algorithms are one of the main themes of this thesis. In machine learning approach a large set of N examples or vectors $\{x1, x2, \dots\dots\dots, xN\}$ called training set is used to tune the parameters of adaptive model. The categories of the data in the training set are known in advance. The category of the training data can be expressed using target t which represents the class or the identity of each training example.

The machine learning approach contains the problem of searching for different patterns in data which makes pattern recognition algorithms represent a fundamental task in this approach. Pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories. A fundamental objective of pattern recognition is classification. If the aim is to assign each input vector to one of a finite number of discrete categories then this problem is called classification problems. If the desired output consists of one or more continuous variables, then the task is called regression. The very aim of BCI is to translate brain activity into a command for a computer. To achieve this goal, either regression (McFarlandD.*et al.*, 2005) or classification (PennyW.*et al.*, 2000) algorithms can be used.

The result of running the machine learning algorithm can be expressed as a function $f(x)$ or $y(x)$ which takes a new vector x as input and generates an output vector y, encoded in the same way as the target vectors t. The precise form of the function $y(x)$ is determined during the training phase which is also known as learning phase. Once the model is trained using a machine learning algorithm it can then determine the identity of new vector which are said to comprise a test set. The ability to categorize correctly new examples that differ from those used for training is known as ***generalization***. Therefore, generalization is a central goal in pattern recognition. For most practical applications, the original input variables are typically preprocessed to transform them into some new space of variables where, it is hopped, the pattern recognition problem will be easier to solve. Thus, machine learning is concerned with the design and development of algorithms and techniques that allow computers to be learned(BishopC. M., 2006). These machine learning techniques focuses on extracting rules and patterns out of massive data automatically. For simplicity and practical reasons, machine learning algorithms are usually divided into two modules: Feature extraction and classification.

Given an input of some form, it can be analyzed to provide a meaningful categorization of its data content. A pattern recognition system can be considered as a two stage device. The first stage is feature extraction. The second is classification. Therefore, the performance of machine learning approach depends on the feature extraction and pattern recognition algorithms employed. Therefore, we are going to discuss signal processing methods used for feature extraction and pattern recognition algorithms used for classification which are essential tools in the development of improved BCI technology.

3.3.1Feature Extraction and Signal Processing

The feature extraction or preprocessing stage serves to transform raw brain signals into a representation that makes classification easy. In other words, the goal of feature extraction is to remove noise and other unnecessary information from the input signals, while at the same time retaining information that is important to discriminate different classes of signals. Features are extracted from the brain signals by signal processing methods.

The feature extraction step consists of mapping the input vector of observations onto a new feature description which is more suitable for the classification task. Therefore, as a primary stage preprocessing applied on data sets in order to extract the most significant features before introducing them to pattern recognition algorithms.

Preprocessing greatly reduces the variability within each class because the presentation of all vectors is now similar. This makes it much easier for subsequent pattern recognition algorithm to distinguish between the different classes.

Also preprocessing might be performed in order to speed up computation for example for high dimensional data where the computer must handle huge numbers and presenting these directly to a complex pattern recognition algorithm may be computationally infeasible. Instead, the aim is to find useful features that are fast to compute and yet that also preserve useful discriminatory information enabling to distinguish the target from non-target. These features are then used as the inputs to the pattern recognition algorithm. This kind of preprocessing represents a form of dimensionality reduction. Therefore, there is another related goal of feature extraction which is to reduce the dimensionality of the data that has to be classified.

Care must be taken during preprocessing because often information is discarded and if this information is important to the solution of the problem then the overall accuracy of the system can suffer. There is something to be noticed, that the new test data must be preprocessed using the same steps as the training data.

Concerning the pivotal step of feature extraction, neurophysiologic a priori knowledge can aid to decide which brain signal feature is to be expected to hold the most discriminative information for the chosen paradigm. This information helps to distinguish the parts of the signal that encode the subject's intent.

Furthermore, as the power and availability of digital signal processing techniques increases attempts are underway to design a digital signal processing system to facilitate the classification phase.

The processing of information in the brain activity is reflected in dynamical changes. Activity variations are found in time, frequency & space. Depending on the type of signals to be classified this knowledge can take many different forms. Consequently many different feature extraction methods have been described. Some basic and often used methods are described below. A more exhaustive review of feature extraction methods for BCIs can be found in A. Bashashati*et al*., 2007.

To achieve the goals of feature extraction, neurophysiologic knowledge about the characteristics of the signals in the temporal, the frequency, and spatial domain can be used. Depending on the type of signals to be classified this knowledge can take many different forms. Consequently many different feature extraction methods have been used in BCIs.

1. Time Domain Features

Time domain features are related to changes in the amplitude of neurophysiologic signals, occurring time-locked to the presentation of stimuli or time-locked to actions of the subject. One of the good examples for signals that can be characterized with the help of time domain features are the P300. A strategy that is often used to separate these signals from background activity and noise is low-pass or band-pass filtering, optionally followed by down-sampling. This strategy is reasonable because most of the energy of the P300 is concentrated at low frequencies. Low-pass filtering, together with down-sampling thus allows to remove unimportant information from high frequency bands. In addition, the dimensionality of the signals is reduced.

An alternative to filtering is to use the wavelet transform of the signals. There are systems that are based on the discrete wavelet transform (DWT) such as the work described by Bai X.,2014, as well as systems based on the continuous wavelet transform (CWT) such that used byEltaf A.*et al.*, 2014.

2. Frequency Domain Features

Frequency domain features are related to changes in oscillatory activity. Oscillatory brain activities are sinusoid like activity that occurs in many regions of the brain. This activity changes according to the state of the subjects, for example between wake and sleep or between concentrating in work and idling. Oscillatory activities in the EEG are classified in different bands or rhythms as sown in Fig.3-12 and also such as sensorimotor rhythm (i.e. the mu-rhythm oscillations) mentioned before in section (3.1.1). Such changes can be evoked by concentration of the subject on a specific mental task such as cognitive tasks (i.e. mental calculation or imagination of rotating geometric object) or motor imagery. Since the phase of oscillatory activity is usually not time-locked to the presentation of stimuli, time domain feature extraction techniques cannot be used. Instead, feature extraction techniques that are invariant to the exact temporal evolution of signals have to be used.

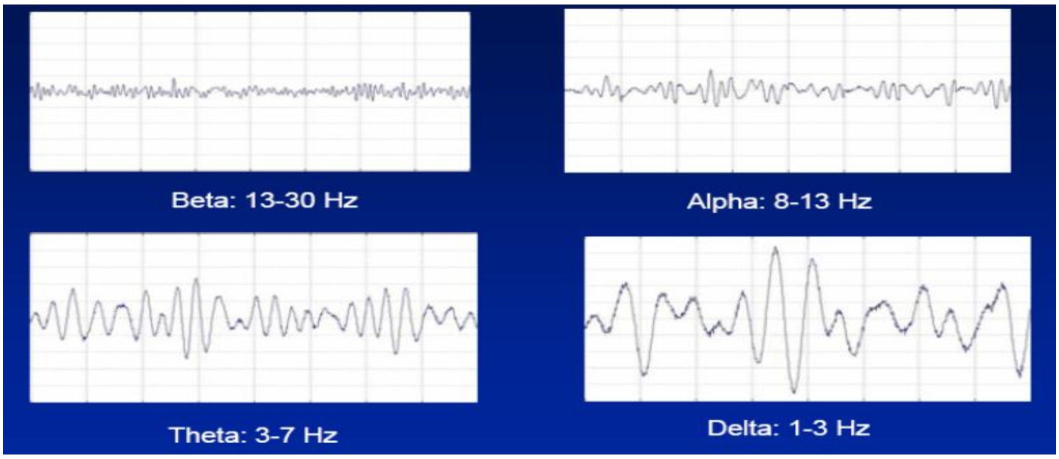


Fig.3-12: Typically observable oscillatory brain activities (Ebrahimi T., 2007)

The most commonly used frequency domain features are related to changes in the amplitude of oscillatory activity. For example in systems based on motor imagery, the band power in the mu (μ) and beta(β) frequency bands at electrodes located over the sensorimotor cortex is used as a feature such as power spectral density (PSD).

3. Spatial Domain Features

The importance for spatial filtering arises due to the poor spatial resolution of EEG measurements.As mentioned before, this poor spatial resolution is the result of a signal caused by theactivity of thousands of neurons. A simulation using a volume conductor model of the head,showed that only as little as 5% of the measured signal comes from sources directly under a1 cm diameter of the respective electrode. 50% comes from within a 3 cm diameter and 95%from within a 6 cm diameter (HoffmannU.*et al.*, 2008). This confirms that it is hard to distinguish exactly wherethe activity came from.

For left and right hand movement it is known that the main signal will be above the contralateralcorresponding primary sensorimotor cortex, but with possible effects of artifacts ornoise, the task still remains difficult (XuN.*et al*., 2004).

In spatial filtering, signals from multiple electrodes are linearly combined, which makes iteasier to locate the source origin, as the increase in signal-to-noise ratio results in being ableto extract more discriminative information from the EEG signals.

Different spatial filtering techniques are used in MI-based EEGsuch as common average referencing,Laplace filtering and common spatial patterns. The first two methods are based on channelre-referencing and are explained shortly. The last method makes use of class information andis the method used in this thesis and in many other researches.

3.4 Pattern Recognition Algorithm

For the description of the classification algorithms, one or more of the feature extraction methods mentioned before has been used to transform raw neurophysiologic signals into feature vectors.

After feature extraction, classification algorithms are used to solve two tasks. During training, the task is to infer a mapping between signals and classes. For this, the labeled feature vectors produced by the feature extraction module are used. During application of BCI, the task is to discriminate different types of neurophysiologic signals and hence to allow for control of a BCI.

Classification algorithms which are to be used in practical BCI systems ideally should fulfill the following requirements. First, algorithms should be robust with respect to outlier. This is important because neurophysiologic signals can contain many outliers and artifacts, caused for example by eye-blinks and muscle activity. Second, algorithms should be of low computational complexity during inference (i.e. learning) and prediction. Low computational complexity during inference reduces the time needed to setup a BCI system. Low computational complexity during prediction is crucial because in BCI systems data should be processed in real-time. Third, algorithms should provide confidence levels for their predictions or, equivalently, probabilistic outputs. This is important because probabilistic outputs provide a natural basis to combine information obtained from different sources and to use decision theory when taking decisions. As demonstrated later that combining information as well as taking decisions in a principled manner allow building advanced BCIs.

Support Vector Machines (SVM) classification algorithm was previously employed in different BCI systems and they yield very good results. This classificationalgorithm is usedin order to demonstrate which of these feature extraction techniques is the best based on classification performance with the multiclass based on motor imagery datasets.

In all of MI-based systems, algorithms that can infer the command a subject wants to execute from the EEG recorded during cue presentation are necessary. The input for these algorithms is the EEG recorded during presentation of cue, together with the timing of scheme. The required output is the class label of the imagined class. To compute this output in all classification algorithms that will be described later in this thesis the same general approach is employed which is:

First, for each presentation of a class cue a short EEG segment, a so-called single trial, is extracted.

Secondly, features are extracted from the single trials then classified with the help of classification algorithms. The outcome of the classification is a class label for each single trial.

Finally the classifier output is compared with the real class label to calculate the classification performance from all single trials performed by the subject.

3.5BCI Systems Subtypes

Subjects try to convey their intentions by behaving according to well-defined paradigms, e.g., motor imagery, specific mental tasks, or feedback control. In order to obtain different brain patterns to be used to drive different BCI systems.

Among the different brain activity patterns there are two main types that can be discerned which are either ***stimulus driven*** or ***user driven*** brain activities (EbrahimiT., 2007) (BlankertzB.*et al*.,http://www.audentia-gestion.fr/research.microsoft/blankertz_bbci_print.pdf). Consequently there are two subtypes of BCI systems according to the type of brain activity that can be used to derive these BCI systems.

3.5.1Stimulus Driven

In this type the brain activity is an ***evoked activity***that appears due to an external stimulus where the subject perceives a set of stimuli that could be visual or auditory or any other sensory stimuli displayed by a BCI system and can manipulate their brain activities by focusing onto one specific stimulus i.e. it is ***stimulus driven***. The changes in brain signals resulting from perception and processing of stimuli are termed event related potentials (ERPs) and will be discussed later. Event related potential is the most widely used neurophysiologic phenomenon to derive BCI systems.

The advantages of BCI systems based on ERPs are that no subject training is necessary because ERPs occur as a natural response of the brain to stimulation. This might be of a particular importance for subjects with concentration problems or for subjects not willing to go through long training phase. The drawback is that communication depends on the presentation and perception of stimuli. Therefore, for disabled subjects it is required to still have remaining cognitive abilities. Moreover, BCI systems based on ERPs have only limited application scenarios because a device to present stimuli is needed and because subjects need to pay attention to stimuli, even in the presence of other unrelated, distractingstimuli. Therefore, no feedback is currently needed and the classification accuracy depends on the natural responses of brain activity. As an example for this type the BCI system is based on evoked P300 event related. The P300 (P3) is a specific, electrical brain signal evoked event related potential (ERP) which can be recorded by electroencephalogram (EEG) as other ERPs. P300 appears as a natural response of the brain to some specific external stimuli and it is emitted within a fraction of a second approximately 300 *ms* when a subject recognizes and processes an incoming stimulus that is significant and noteworthy (Fazel-RezaiR., 2007).

BCI systems using this approach are systems based on the evoked P300 event related potential. Roughly speaking, such systems work by detecting on which particular stimulus (i.e. target stimuli) out of a random series of stimuli (i.e. including target and non-target stimuli) the subject is concentrating upon. Since different commands or actions are linked to the series of stimuli, subjects can select a command simply by concentrating on the associated stimulus.

3.5.2User Driven

In this type the brain activity is a *spontaneous activity* where the subjects control their brain activity by concentrating on a specific mental task, for example imagination of hand movement can be used to modify activity in the motor cortex which is *user driven*.

In this approach which is used in this thesis more flexible BCI systems can be imagined because no computer screen or other device is needed to present stimuli. However, to gain voluntary control over brain activity, subjects have to perform feedback training, in order to learn the production of high amplitude and easily detectable patterns of the desired neurophysiologic signals. The feedback training can take several weeks before subjects are able to reliably control a BCI. Therefore, BCI systems based on user-driven might be less suited for subjects with concentration problems or for subjects who are not willing to go through a long training phase (Erfanian A.and ErfaniA., 2004).

An example is given by so-called mu-rhythm BCIs. Mu-rhythm oscillations can be observed over the sensorimotor cortex when a subject does not perform movements. Fig.3-13 shows (Blankertz B.*et al*.,http://www.audentia-gestion.fr/research.microsoft/blankertz_bbci_print.pdf) the motor related potentials (MRPs) and event related desynchronization (ERD) which occurs during motor imagery i.e. oscillations are decreased in amplitude when movements of body parts are imagined or performed. In these systems feedback training is used to let subjects acquire voluntary control over the amplitude of the mu-rhythm, i.e. electroencephalogram (EEG) activity in the frequency range of (8-12) Hz, located over the motor cortex. Changes in mu-rhythm amplitude can be caused by for example imaginary left hand or right hand movements and are then linked to movements of a cursor or to other commands (DornhegeG.*et al*.,2004). Also the slow cortical potentials (SCP) that occurs during intentions or states of preparation and relaxation can be used as an example of spontaneous activity to derive a BCI system.

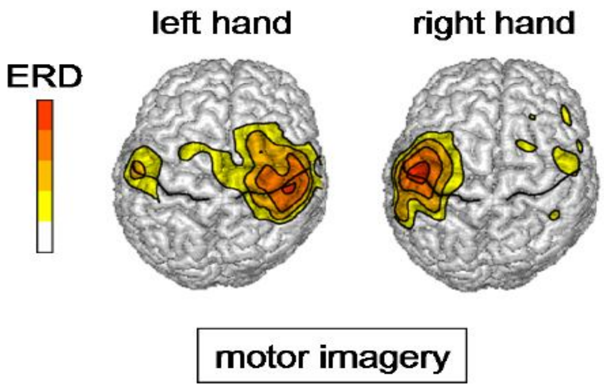


Fig.3-13: Event related Desynchronization (ERD) due to motor imagery (B. Blankertz et al,http://www.audentia-gestion.fr/research.microsoft/blankertz_ bbci_print.pdf).

Hence, there are BCI systems that are based on spontaneous activity using brain signals that do not depend on external stimuli and that can be influenced by concentrating on a specific mental task. Therefore, feedback training is necessary and classification accuracy significantly increases as subjects learn how to modulate their brain activity.

3.6Event-Related Potentials (ERPs)

Event-related potentials (ERPs) are stereotyped, spatio-temporal patterns of brain activity, occurring time-locked to an event as a response to an internal or external stimulus. More simply, it is any measured brain response that is directly the result of a thought or perception (Wikipedia Webpage.http://en.wikipedia.org/.) For example after presentation of a stimulus or detection of a novel stimulus i.e. the changes in brain signals resulting from perception and processing of stimuli.

Evoked event related potentials consist of transient voltage changes that occur in response to a sensory stimulus. These take the form of a series of negative and positive waves. Traditionally, event-related changes have been used in neuroscience for studying the different stages of perception, cognition, and action. ERPs are reliably recorded with the electroencephalogram (EEG) and can also be measured with other signal acquisition technique such as Magnetoencephalogram (MEG) or Functional Magnetic Resonance Imaging (fMRI) but in BCI usually the EEG is used for measuring such changes.

3.6.1ERD in memory and movement tasks

The alpha band rhythms demonstrate a relatively widespread desynchronization (ERD) in perceptual, judgementand memory tasks. An increase of task complexity or attention resultsin an increased magnitude of ERD. It has to be kept in mind, however,that the ERD is measured in percentage of power relative tothe reference interval and therefore it depends on theamount of rhythmic activity in this interval.

It is important to note that alpha band desynchronizationis not a unitary phenomenon. If different frequency bandswithin the range of the extended alpha band are distinguished, at least two distinct patterns of alpha desynchronization can be observed. Lower alpha desynchronization (inthe range of about 7±10 Hz) is obtained in response toalmost any type of task. It is topographically widespreadover wide areas of the scalp and probably reflects generaltask demands and attentional processes. Upper alpha (mu)desynchronization (in the range of about 10±12 Hz) is veryoften topographically restricted and develops during theprocessing of sensory-semantic information above parietooccipital areas.

The degree of desynchronization is closely linked to semantic memory processes. For example, during semantic encoding of words, good memory performers showed a significantly larger ERD in the lower alpha band ascompared to bad performers. An explanation for this finding may bethat a higher level of attention and alertness is requiredduring encoding. In contrast to the alpha band activities,the activity in the theta band may be responsible for theencoding of new information.

In an auditory memory task, no localized alpha band ERDwas found in ongoing EEG. This absence of an alpha band desynchronizationin EEG during the memory set presentation may beexplained by the anatomical localization of the auditorycortex in the supratemporal plane. EEG desynchronizationdue to direct auditory processing alone is, therefore, hard todetect in scalp-recorded EEG. On the contrary, a desynchronization localized to the auditory cortex following auditorystimuli was reported in MEG recordings.

Voluntary movement results in a circumscribed desynchronization in the upper alpha and lower beta bands, localized close to sensorimotor areas. This desynchronizationstarts about 2*s* prior to movement-onset over the contralateral Rolandic region and becomes bilaterally symmetricalimmediately before execution of movement. It is of interestthat the time course of the contralateral mu desynchronization is almost identical with brisk and slow finger movement, although both, brisk and slow movements are quitedifferent. Brisk movement is preprogrammed and it does notrequire feedback from the periphery, while slow movementdepends on the reafferent input from kinesthetic receptorsevoked by the movement itself. Analysis of alpha and beta ERD using electrocorticographic (ECoG) recordings showed that the topography of beta ERD was often more discrete and somatotopicallyspecific than that of alpha ERD.

The contralateral dominant pre-movement mu ERD is notonly independent of movement duration, but also similarwith index finger, thumb and hand movement.

The circumscribed hand area mu ERD can be foundin nearly every subject, a foot area mu ERDlocalized close to the primary foot area between both hemispheres is less frequent. Anexample of a relatively widespread foot area desynchronization in the 7±8 Hz band and a circumscribed 20±24 Hzband ERD is presented in Fig.3-14. For comparison, also ahand area mu desynchronization in the 10±11 Hz band,with a beta desynchronization in the 20±24 Hz band, isdisplayed. It is of interest that the localization of the betaERD is slightly more anterior compared to the largest muERD with hand as well as the foot movement and that the 7±8 Hz-ERD with foot movement is found not only over thefoot but also over the hand representation area. This can beinterpreted as follows: the mu rhythm is generated mainly inthe post-Rolandic somatosensory area and the central betarhythm (at least some components of it) in the pre-Rolandicmotor area.

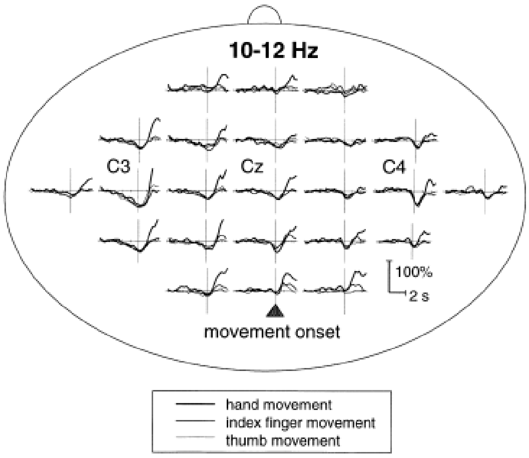


Fig. 3-14: Topographic display of grand average ERD curves from three movement experiments. Data are triggered with movement-onset (vertical line). The reference interval (base line) is marked by a horizontal line. (Pfurtscheller *et al*., 1999).

An investigation of hand area mu and betarhythms on single trial basis also revealed a slightly moreanterior focus of the beta rhythm.MEG measurements revealed a similar result. They interpreted the 10-Hz murhythm as localized in the primary somatosensory areaand the 20-Hz beta rhythm as localized in the motor area.

In an electrophysiological study with subdural electrodes literatures reported on mu rhythms not only selectively blocked with arm and leg movements, but also withface movement. It can therefore be hypothesized that besidea great variety of occipital alpha rhythms also a variety of Rolandic murhythms exist (Pfurtscheller*et al*., 1994).

3.7 Brain Computer-Interface Technology

As previously discussed, the brain signals have to be modified by the subjects through special “thoughts”. Further, the brain signals have to be analyzed & classified. The results of reliable classification of such a system can be used to establish a lexicon of commands that could control a variety of computer applications including for example, specialized graphical subject interface and simple word processing software. Also such a system may become a valuable tool for paralyzed patients who may ultimately use it to control a wheelchair, an artificial limb or a computer application that is used as an environment control systems.

3.7.1 Brain Computer-Interface Operation Modes

Several application scenarios exist in which a BCI could be used. However, so far no commercially available BCI application has emerged. In theory any device that can be connected to a computer or a microcontroller could be controlled with a BCI. In practice however, the set of devices and applications that can be controlled with BCI is limited. To understand this, one has to know that Brain Computer interfaces (BCI) can be divided into systems working in two modes of operations either synchronous or asynchronous modes.

1. Synchronous

In a synchronous BCI, the analysis & classification of brain potentials is limited predefined fixed to variable time windows. A characteristic features of all synchronous system is that onset of mental activity is known in advance & associated with a specific cue or trigger stimulus. In synchronous mode communication is possible only during predefined time intervals. This means the system tells the subject when it is ready to receive the next command and limits severely the possible type of applications (PfurtschellerG., http://www.eurasip.org/Proceedings/Eusipco/Eusipco2004/defevent/papers/cr1940.pdf.)

2. Asynchronous

In case of an ideal asynchronous BCI system no cue stimulus is used & the subject can intend whenever he/she wishes, a specific mental activity. The ongoing brain signals have to be analyzed & classified continuously, mental events have to be detected and discriminated from noise & non-events then transformed into a control signal as quickly & accurately as possible.

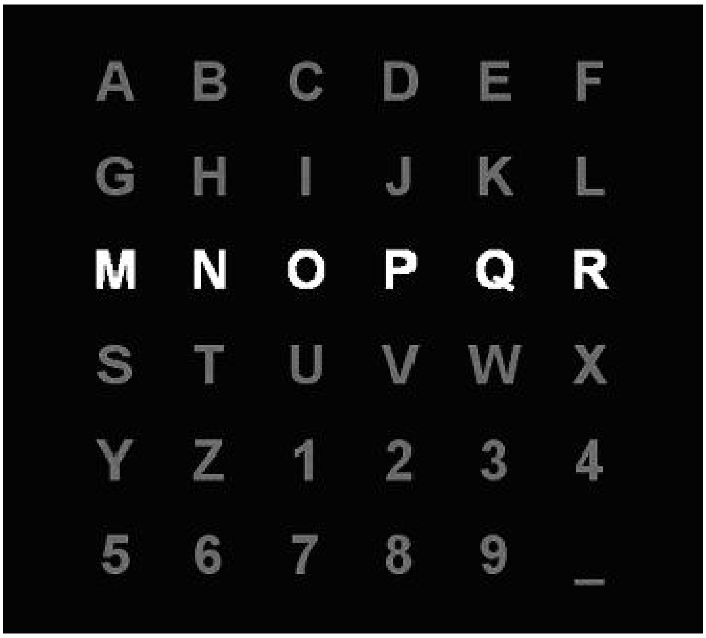
Consequently, we have to consider that the amount of information which can be transmitted with present day BCI systems is limited. As an additional obstacle most present day BCI systems function only in synchronous mode. This is possibly due to the fact that the current technology does not allow building BCI systems which can work in asynchronous mode and provide high information transfer rates. A possible approach to circumvent the problem of limited information transfer rates is to build intelligence into the application, i.e. to reduce the information needed by the application by cleverly restraining the number of commands possible in a given situation (PfurtschellerG., http://www.eurasip.org/Proceedings-/Eusipco/Eusipco2004/defevent/papers/cr1940.pdf.).

3.7.2 Brain Computer-Interface Applications

Examples for applications in which such a strategy has been implemented with the current BCIs are described below.

3.7.2.1Spelling Devices

Spelling devices allow severely disabled subjects to communicate with their environment by sequentially selecting symbols from the alphabet as shown in Fig.3-15. It is a device that helps disabled subjects to spell words by means of their brain activities. One of the first spelling devices motioned in the BCI literature is the P300 speller (Farwell L.and DonchinE., 1988).



A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	_

Fig.3-15: P300 Visual Speller (EbrahimiT., 2007).

3.7.2.2Environment Control

Environment control systems allow controlling electrical appliances with a BCI. Development of asynchronous BCI system is probably the most important research topic to advance the area of environment control systems. Bayliss J. in 2003tested if P300 could be evoked in a virtual reality environment. In the system presented by Bayliss, subjects viewed a virtual apartment alternatively on a monitor or through a head-mounted display. Control of several items in the virtual apartment as shown in Fig.3-16, for example switching on/off a lamp, was possible by concentrating on small spheres that were flashing in random order over controllable items. It was shown that virtual reality, which allows for complex, yet controllable experimental environments, is an interesting other alternative to other simpler P300 paradigms.

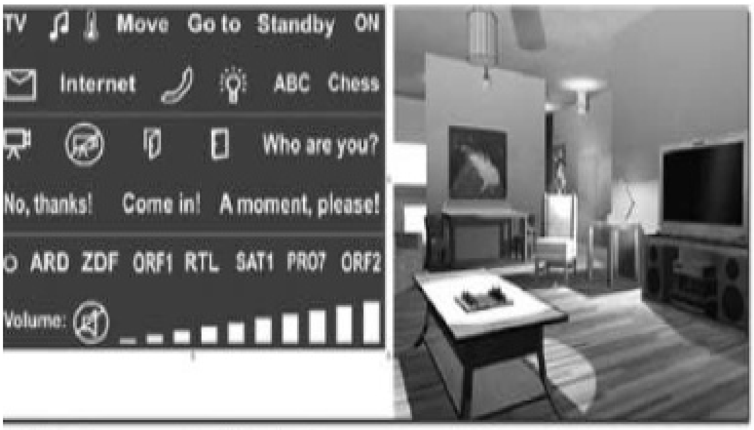


Fig.3-16: Sample scene from the virtual apartment with controllable items (BaylissJ., 2003).

3.7.2.3Cursor Control

Another P300 paradigm was presented by PolikoffJ. in 1995. The idea behind the system described byPolikoffJ.is to allow subjects to control a two-dimensional cursor with the help of the P300. To implement this idea a fixation cross with target arms in the north, east, south and west directions was displayed on a monitor as shown in Fig.3-17. At the end of each arm small crosses were displayed and temporarily replaced by asterisks. The replacement of crosses occurred in random order, and to move the cursor in a given direction subjects were instructed to count the number of asterisks appearing at the corresponding target arm. While in the study of Polikoff J. actual cursor movement was not implemented, an offline analysis showed that cursor control with the help of the P300 was in principle possible.

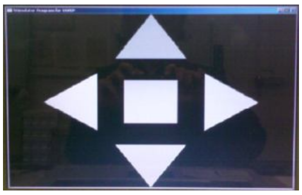


Fig.3-17: 2-Dimentional cursor control where the target detection tasks are presented in four compass positions (N, E, S, W) on a computer screen (Polikoff J., 1995).

3.7.2.4Wheel-Chair Control

Disabled subjects are almost always bound to wheelchairs. If control over some muscles remains, these can be used to steer a wheelchair. For example systems exist that allow to steer a wheelchair with only a joystick or with head movements. If no control over muscles remains, a BCI can be potentially used to steer a wheelchair. Because steering a wheelchair is a complex task and because wheelchair control has to be extremely reliable, the possible movements of the wheelchair are strongly constrained in current prototype systems. In the system presented by RebsamenB. et alin 2006 the wheelchair is constrained to move along paths predefined in software joining registered locations as shown in Fig.3-18. And a P300-based interface is used to select the desired location.

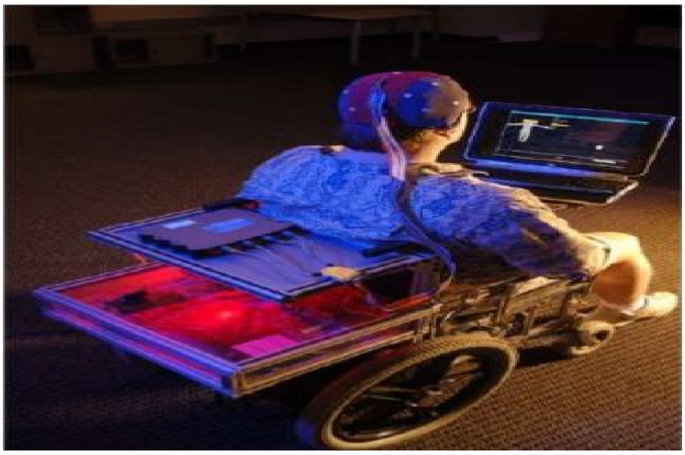


Fig.3-18: Testing environment: a lift, hall, and six guiding paths between four destinations (Rebsamen B.*et al*., 2006).

3.7.2.5Neuromotor Prostheses

The idea underlying research on neuromotor prostheses is to use a BCI for controlling movement of limbs and to restore motor function in tetraplegics and amputees. Different types of neuromotor prostheses can be envisioned depending on the information transfer rate a BCI provides. If neuronal ensemble activity is used as control signal, high information transfer rates are achieved and 3D robotics arms can be controlled as shown in Fig.3-19(Tillery S.*et al.*, 2003). If an EEG based BCI is used, only simple control tasks can be accomplished.

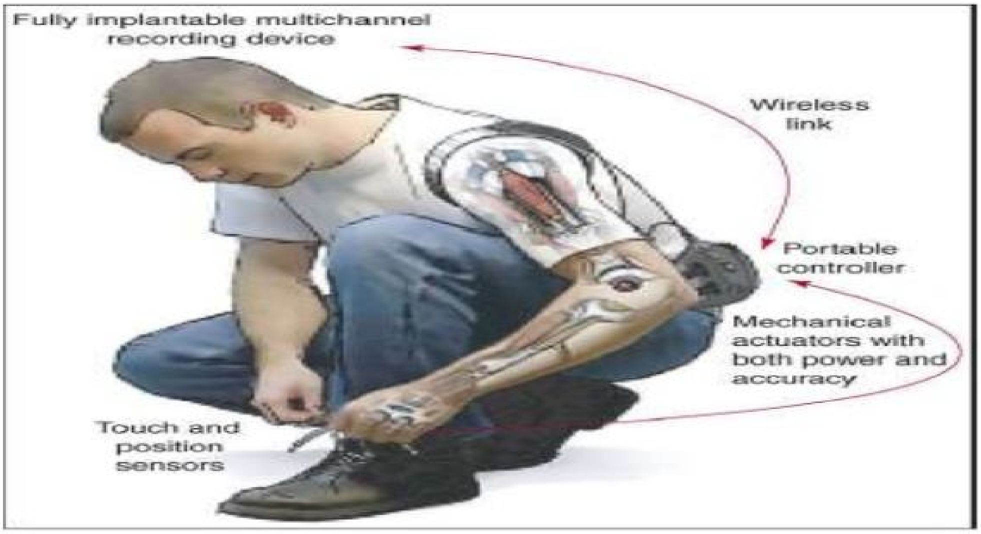

Fully implantable multichannel recording device
Wireless link
Portable controller
Mechanical actuators with both power and accuracy
Touch and position sensors

Fig.3-19: Robotics arm (Tillery S. *et al*., 2003).

3.7.2.6Gaming and Virtual Reality

Besides the applications targeted towards disabled subjects, prototypes of gaming and virtual reality applications have been described in the literature. Examples for such applications the control of an animated character in an immersive 3D gaming environment with steady-state visual evoked potentials (SSVEPs) which is shown in Fig.3-20(LalorE.*et al.*, 2005), and the control of walking in a virtual reality environment with sensorimotor rhythms (R. Leeb *et al*.,http://www.temple.edu/ispr/prev_conferences/proceedings/2005/Leeb,-%20Keinrath,- %20Friedman,%20Guger,%20Neuper,%20Garau,%20et%20al..pdf).

Fig. 3-20: On the left image the training sequence is shown and the right image shows that the character loss balance (Lalor E.*et al*., 2005).

CHAPTER FOUR

Research Methodology

Theoretical Concepts

In an EEG-Based BCI electrical signals recorded from the subject's scalp are analyzed to determine the state of the subject's brain. Thus, signal processing and classification are the main task in an EEG-Based Brain-Computer Interface and they are essential tools in the development of improved BCI technology.

More specifically, the sequence of events in a motor imagery EEG based BCI is usually as follows:

1. First, the subject chooses a command that he wants to execute with the help of BCI.
2. Then, stimuli are presented and the subject concentrates on the stimulus associated to the desired command.
3. After stimulus presentation the recorded EEG is analyzed with the help of machine learning algorithms. The goal of this analysis is to infer which stimulus was chosen as target by the subject. If the analysis is successful the command associated to the chosen stimulus is executed by the BCI systems that implement this idea.

Benchmark datasets from BCI competitions(Blankertz B., 2006) are used to show that the algorithms employed in this thesis are competitive with state-of the art electroencephalogram (EEG) machine learning algorithms. Data set 2ais provided by BCI Competition IV for the year 2008 (BlankertzB., http://www.bbci.de/competition/iv). The EEG signals are recorded from 9 subjects that are used to show the performance (i.e. classification performanceobtained) of different algorithms.In this chapter we are going to discuss the materials and methods used for recording and analysing data.

4.1. Data Description

BCI2000 is a flexible general-purpose system for brain-computer interface (BCI) research. It can be used for data acquisition, stimulus presentation, and brain monitoring applications in order to facilitate research and applications in these areas(Farwell L. *et al*, 1988).

Most demonstrations of algorithms on BCI data are evaluating classification of EEG trials, i.e., windowed EEG signals for fixed length, where each trial corresponds to a specific mental state. But in BCI applications with asynchronous feedback one is faced with the problem that the classifier has to be applied continuously to the incoming EEG without having cues of when the subject is switching her/his intention. This data set poses the challenge of applying a classifier to trial base EEG for which no cue information is given.

4.1.1. Experimental paradigm

Data recording: Twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm)were used to record the EEG; the montage is shown in Fig. 4-1 left. Allsignals were recorded monopolarly with the left mastoid serving as referenceand the right mastoid as ground. The signals were sampled with 250 Hz andbandpass-filtered between 0.5 Hz and 100 Hz. The sensitivity of the amplifierwas set to 100 μV. An additional 50 Hz notch filter was enabled to suppressline noise.

In addition to the 22 EEG channels, 3 monopolar EOG channels were recorded and also sampled with 250 Hz (see Fig. 4.1 right). They werebandpass filtered between 0.5 Hz and 100 Hz (with the 50 Hz notch filterenabled), and the sensitivity of the amplifier was set to 1 mV. The EOGchannels are provided for the subsequent application of artifact processingmethods and must not be used for classification.

A visual inspection of all data sets was carried out by an expert andtrials containing artifacts were marked.

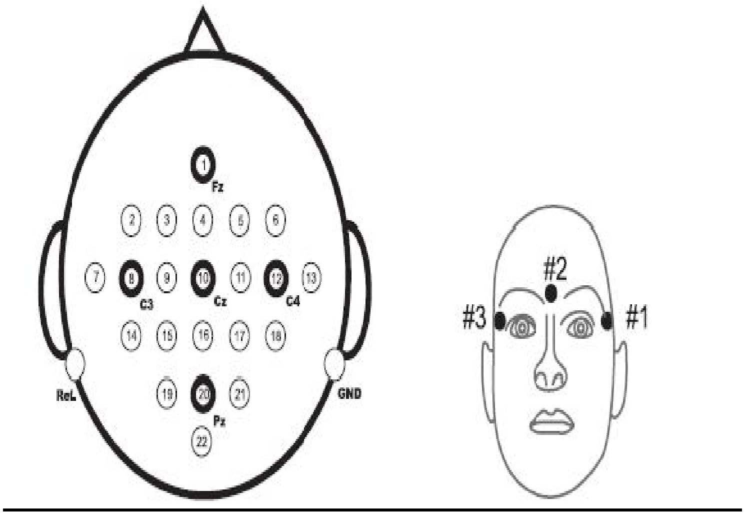


Fig. 4-1: Left Electrode montage corresponding to the international 10-20 system. Right: Electrode montage of the three monopolar EOG channels (Brunner C.*et al.*,http://www.bbci.de/competition/iv/desc_2a.pdf)

***Calibration data*:**This data set consists of EEG data from 9 subjects. The cue-based BCIparadigm consisted of four different motor imagery tasks, namely the imaginationof movement of the left hand (class 1), right hand (class 2), bothfeet (class 3), and tongue (class 4). Two sessions on different days wererecorded for each subject. Each session is comprised of 6 runs separated byshort breaks. One run consists of 48 trials (12 for each of the four possibleclasses), yielding a total of 288 trials per session.

At the beginning of each session, a recording of approximately 5 minuteswas performed to estimate the EOGinfluence. The recording was dividedinto 3 blocks: (1) two minutes with eyes open (looking at a fixation crosson the screen), (2) one minute with eyes closed, and (3) one minute witheye movements. The timing scheme of one session is illustrated in Fig.4-2.

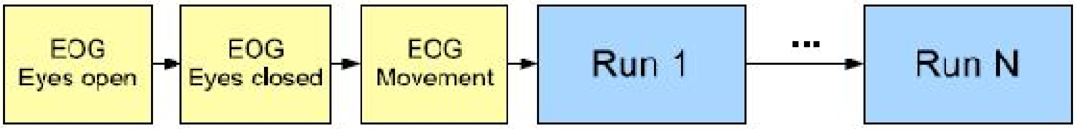


Fig. 4-2: Timing scheme of one session (Brunner C.*et al.*,http://www.bbci.de/competition/iv/desc_2a.pdf).

.

The subjects were sitting in a comfortable armchair in front of a computerscreen. At the beginning of a trial (t = 0*s*), a fixation cross appearedon the black screen. In addition, a short acoustic warning tone was presented.

After two seconds (t = 2*s*), a cue in the form of an arrow pointingeither to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared and stayed on the screen for1.25*s*. This prompted the subjects to perform the desired motor imagerytask. No feedback was provided. The subjects were ask to carry out themotor imagery task until the fixation cross disappeared from the screen at = 6*s*. A short break followed where the screen was black again. Theparadigm is illustrated in Fig.4-3.

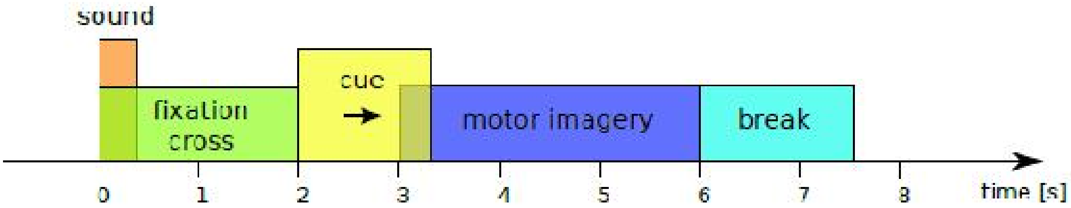


Fig. 4-3: Timing scheme of the paradigm (Brunner C.*et al.*,http://www.bbci.de/competition/iv/desc_2a.pdf).

.

Evaluation data: Among the provided two sessions the first session is used as calibration data and the second one is used as evaluation data to test the classifier and hence to evaluate the performance.

4.2.Data Analysis

This thesis focuses on BCI systems based on multiclass motor imagery. An important component of any such system, but also of other EEG based systems, are the classification or pattern recognition methods that allow discriminating EEG segments representing different types of brain activity. Hence in this thesis special emphasis is given to machine learning algorithms that learn from a set of training data how to discriminate EEG segments containing four different motor imagery classes.

As previously mentioned that machine learning algorithms are the main theme of this work, after the data acquisition phase, machine learning algorithms are applied to infer functions that can be used to classify neurophysiologic signals. For reasons of practicality and simplicity, machine learning algorithms are usually divided into two modules: feature extraction and classification.

The classification of the EEG is further complicated and difficult task due to the large EEG samples variability and also due to the presence of artifacts. Artifacts can be due to physiological or non-physiological sources. Physiological resources for artifacts include eye movements and eye blinks, muscle activity, heart activity and slow potential drifts due to transpiration. Non-Physiological sources for artifacts include power supply line noise, noise generated by the EEG amplifier and noise generated by sudden changes in the properties of the electrode-scalp interface. Artifacts often have much larger amplitude than the signals of interest. Therefore artifact removal and filtering procedures have to be applied before classification of EEG signals can be attempted.

The feature extraction module serves to transform raw neurophysiological signals into a representation that makes classification easy and removes the signals artifacts. Features (i.e. feature vectors) are extracted from the digitized EEG-signals by signal processing methods. After feature extraction, these features are translated into a control signal. During translation machine learning algorithms are used to solve two tasks. During training, the task is to infer a mapping between signals and classes. For this, the labeled feature vectors produced by the feature extraction module from labeled training data sets are used to learn a more complex decision function. During application of a BCI, the task is to discriminate different types of neurophysiological signals and hence to allow for control of a BCI.

The programming language used in this work is MATLAB 18.0 Mathwork, Inc., because it has robust tool boxes that can help in this work, for example the Statistical Pattern Recognition Toolbox (STPRTool).

4.2.1.Preprocessing Phase

The key task is to detect the ERD in electroencephalography (EEG) related to the imagined class accurately and instantly. The idea behind selecting these features is a trial to minimize the classification error while getting the most efficient discrimination. To improve the classification process, many techniques were employed to obtain the most discriminate features to distinguish the class that the subject imagining.

At first, the data were preprocessed with several methods in order to do comparison between different machine learning algorithms using MI datasets.

The preprocessing and enhancement techniques are used to enhance the signals and remove the noise also filtering the signals to obtain the desired frequencies and also to reduce the dimensions of the signals. EEG signal is what mathematicians call a nonstationary time-series. They can be characterized as time-domain or frequency-domain or both. Therefore, activity variations can be found in time, frequency & space. Consequently many different feature extraction methods have been employed. The preprocessing operations were applied in the order in the following order using MATLAB:

1. Single trial extraction for both calibration and evaluation data.
2. Discrimination between different classes of calibration data.
3. Time windowing for both calibration and evaluation data.
4. Frequency filtering or spatial filtering for both data and make a comparison between them.

Next sections will discuss these points in details..1/

4.3 Spatial Filtering

4.3.1 Independent Component Analysis

In the case of EEG signals, the idea underlying the application of independent component analysis (ICA) is that the signals measured on the scalp are a linear and instantaneous mixture of signals from independent sources in the cortex, deeper brain structures, and noise. Unfortunately, the raw EEG recorded from the subject's scalp not only contains the desired evoked ERP but also ongoing activity of the brain and muscular and/or ocular artifacts such as eye blinks. As a result, the signal to noise ratio is very low and the classification task is not easy (BaiX. *et al.*, 2014). These artifacts were removed using blind source separation (BSS) such as independent component analysis (ICA) to get enhanced signal to noise ratio and artifact-free EEG signal (Bai X. *et al*., 2014). ICA has been mainly used in MI-based BCIs as a spatial feature extraction method. In such systems ICA is used to separate multichannel EEG into several components corresponding to sources in the brain or noise. By retaining only components that have anERD like spatial distribution or show ERD like waveforms, the signal to noise ratio can be improved. Consequently, classification can be performed with improved accuracy.

ICA can be seen as an extension to principal component analysis (Hyvarinen A. *et al*., 2001). ICA is much more powerful technique, however, capable of finding the underlying factors of components or sources from multivariate (multidimensional) statistical data. What distinguishes ICA from other methods is that it looks for components that are both statistically independent and non Gaussian (Hyvarinen A. *et al*., 2001).

In reality, however, the data often does not follow a Gaussian classification distribution, and the situation is not as simple as to assume that the data is Gaussian (Hyvarinen A. *et al*., 2001). If the data is Gaussian, it is simple to find components that are independent, because for Gaussian data, uncorrelated components are always independent. For example, many real-world data sets have super Gaussian distributions. This means that the random variables take relatively more often values that are very close to zero or very large. In other words, the probability density of the data is peaked at zero and as heavy tails (large values far from zero), when compared to Gaussian density of the same variance.

The first thing to note is that independence is a much stronger property than uncorrelatedness. Uncorrelatedness in itself is not enough to separate the components. Principal component analysis (PCA) gives components that are uncorrelated but little more. The starting point of ICAis to find statistically independent components in the general case where the data is non Gaussian (Hyvarinen A. *et al.*, 2001).

The first step in all algorithms using ICA is to compute independent components from the training data. However, the major drawback of such methods is that they are not specifically designed to separate brain waves and they are supervised. Indeed, after the decomposition in independent components (IC) it is necessary to select the components that present well the ERD. This can be done either manually i.e. by inspecting the data or by defining criteria that allow to automatically select ERD like components such as the spatiotemporal selection algorithm applied by BaiX.*et al.* in 2014.

In spatial feature extraction methods apply a spatial filtering after processing takes place. Spatial filtering corresponds to building linear combinations of the signals measured at several electrodes. Denoting by $S(t)$the signal from N_e electrodes at time t, spatial filtering can be expressed as follows in Eq. (4.6):

$$\hat{S}(t) = CS(t) \quad (4.6)$$

Here the $N_s \times N_e$ matrix C contains the coefficients for$N_S$spatial filters and the vector $\hat{S}(t)$contains the spatially filtered signals at time t.

Independent component analysis (ICA) algorithm can be used to compute the coefficients of spatial filters from training data. In ICA algorithms it is assumed that a set of multi-channel signals $S(t)$is generated by linearly mixing a set of source signals $x(t)$:

$$S(t) = Wx(t) \quad (4.7)$$

The goal is to compute a matrix Wthat allows one to reconstruct the source signals x by multiplying S with W^{-1}. To achieve this without having information about W, one assumes that the source signals are statistically independent. The ICA algorithm thus computes Wsuch that the signals $S(t)$ multiplied with $W^{-1}$are maximally independent.

When the algorithm is applied to new data, the data is projected on the retained independent components and then classified.

FastICA MATLAB package is used to calculate ICA components of the data sets.

4.3.1.1 Second Order Blind Identification

In this Work, ICA based Second Order Blind Identification with Robust Orthogonalization (SOBI-RO) is also used to remove ocular artifacts. The method uses time structure when the independent components (ICs) are time signals; this is in contrast to basic of ICA model which is mixed random variables. ICs may contain more structure than simple random variables such as the autocovariances (covariances over different time delays) of the ICs (Hyvarinen A.*et al.*, 2001), the standard mixing model:

$$x{=}Hs\ (k)(4.8)$$

Where (k) is mixed signals and H is mixing matrix.

Before setting time delayed covariance matrices of mixed signals, formulating the robust orthogonalizationmust be done first as:

$$\bar{x}(k) = Q\ x\ (k)(4.9)$$

By using several time lags, up to 100 number of time lags, the time delayed covariance matrices of mixed signal for preselected time delays$(p_1, p_2,\ldots\ldots,)$are defined as:

$$R_{\bar{x}}(p_i) = \frac{1}{N}\sum_{k=1}^{N} \bar{x}(k)\bar{x}^T\,(k - p_i) = QR_{\bar{x}}(p_i)Q^T (4.10)$$

And then, the orthogonalized mixing matrix $A=QH$, perform Joint Aproximation Diagonalization (JAD):

$$R_{\bar{x}}(p_i) = QR_x(p_i)Q^T = AR_S(p_i)A^T = UD_iU^T (4.11)$$

For$i=(1,2,3,\ldots L,\ )$JAD reduces the probability of un-identifiability of a mixing matrix caused by an unfortunate choice of time delay$p$. Then orthogonal mixing matrix can be estimated as$\hat{A}=Q\hat{H}=U$and diagonal matrix(D_i). Finally, the estimated of source signals as (CichockiA.*et al*., 2002):

$$\hat{S}(k) = (k)\ (4.12)$$

And the mixing matrix as:

$$\hat{H}=Q+U (4.13)$$

SOBI MATLAB package is used to calculate ICA components of the data sets. The data entered to SOBI tool is a matrix X which contains the training or calibration examples of each subject concatenated for each electrode of the 25 electrodes.

SOBI-RO also was tried without data reduction (i.e. using the whole 25 electrodes). And was tried in order to reduce the number of electrodes from 25 to 18 electrodes (i.e. minimum number of electrodes used without removing useful information).

4.4. Feature Vector Construction

The feature extraction module serves to transform raw EEG signals into a representation that makes classification easy and removes the signals artifacts. Features (i.e. feature vectors) are extracted from the digitized EEG-signals by signal processing methods. After feature extraction, these features are translated into a control signal. During translation machine learning algorithms are used to solve two tasks. During training, the task is to infer a mapping between signals and classes. For this, the labeled feature vectors produced by the feature extraction module from labeled training data sets are used to learn a more complex decision function. During application of a BCI, the task is to discriminate different types of neurophysiological signals and hence to allow for control of a BCI.

In this thesis combinedWavelet Analysis and conventional CSP, namedWavelet-CSP, to solve a multi-class issue by simplifying itinto several binary problems. According to ERD/ERS, thedifference between two kinds of motor imagery is higher inthe frequency domain than that in the time domain.Consequently, the Wavelet-CSP directly used the waveletcoefficients of the feature band instead of the samplingvalues of the signals as the input of the CSP to extract thefeatures(Bai X. *et al.*, 2014). Different wavelet mothers will examine to address the efficiency of the most discriminating featuresfor separating different classes. In addition, since CSP is sensitive to noise, theclassification success rate will debase along with thedeterioration of the signal quality. Thus, this thesis designedan ICA-filter and IIR filter to remove noise before the Wavelet-CSP as mentioned in the previous section.

4.5.Wavelet - CSPMethod:

4.5.1 Discrete Wavelet Transform (DWT)

A wavelet is a mathematical function used to decompose a given function or continuous-time signal into different frequency components and study each component with a resolution that matches its scale. Wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as daughter wavelets) of a finite length or fast decaying oscillating waveform (known as mother wavelet). Wavelet transforms are classified into continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

For many signals, the low-frequency content is the most important part. It iswhat gives the signal its identity. The high-frequency content, on the otherhand, imparts flavor or nuance. Consider the human voice. If you remove thehigh-frequency components, the voice sounds different, but you can still tellwhat’s being said. However, if you remove enough of the low-frequencycomponents, you hear gibberish.

It is for this reason that, in wavelet analysis, it is often speak of approximationsand details.

The approximations are the high-scale, low-frequency components of thesignal. The details are the low-scale, high-frequency components. The filteringprocesses, at its most basic level, see Fig. 4-4:

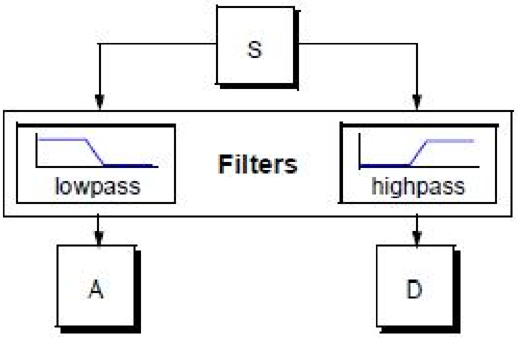


Fig. 4-4: The original signal, S, passes through two complementary filters and emerges as two signals(Misiti Y. *et al.*, 1995).

The decomposition process can be iterated, with successive approximationsbeing decomposed in turn, so that one signal is broken down into manylower-resolution components. This is called the wavelet decomposition tree in Fig 4-5.

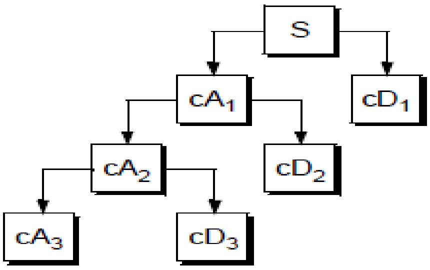


Fig. 4-5: Wavelet Decomposition Tree (Misiti Y. *et al.*, 1995).

The signal (S) is decomposed into multi-resolution subsets of coefficients a detailed coefficient subset(cD_i) and approximation subset (cA_i) at the level (i) as:

$$S = cA_i + \sum_{j=1}^{i} cD_j = cA_i + cD_i + cD_{i-1} + \ldots.. + cD_1 \quad (4.15)$$

Because the differences between the time-domain signalsare very small during imagination, it is hard to find a globaloptimal projection matrix through the sampling values.However, experiments show that the differences of thefrequency values in μ and β rhythms are more apparentaccording to ERD/ERS and the optimal projection matrixbased on the frequency values can be easier to train. Thus, thenovel Wavelet-CSP algorithm that uses the waveletcoefficients as the input of the CSP instead of the samplingvalues was employed in this study.

4.5.2.Common Spatial Pattern (CSP)

Common Spatial Patterns (CSP) algorithm isone of the frequently used algorithms to extractthose components of the EEG/MEG data that provide mostinformation on the intention of the BCI-user. CSP was first proposed in the context of EEG/MEG analysis in (Koles, 1991), and introduced to the BCI community in (Ramoser*et al.*, 2000). Given EEG/MEG data of two different classes, e.g., motor imagery of the left and right hand, the CSP algorithm computes spatial filters that maximize the ratio of the variance of the data conditioned on one class and the variance of the data conditioned on the other class. In this way, spatial filters can be designed that extract those components of the EEG/MEG data that differ maximally (in terms of the variance) between conditions. Such spatial filters are especially suited for BCIs utilizing motor imagery paradigms, in which the intention of the user is typically inferred from frequency specific changes in variance of EEG/MEG components. Excellent classification results have been reported using CSP for pre-processing in non-invasive BCIs based on motor imagery (e.g., in one of the winning entries of the BCI competition IV (Pfurtscheller G., http://www.bbci.de/competition/iv/results/#winners), and improvement of the CSP algorithm, especially its extension to the spectral domain, is an active area of research (Moritz*et al.*, 2008).

4.5.2.1.Common Spatial Patterns (CSP) analysis in 2-class BCI systems

A CSP-filter is calculated based on the potential differences V_j, with a dimensionality of $N_{ch}\times T_j$, with N_{ch} the number of EEG channels and T_j the number of samples for that trial. Each trial V_j is labeled, implicating that the use of CSP-filters is a supervised technique.

To CSP-filter the signal, a projection matrix is calculated to transform the original EEG data as shown in Fig.4-6.

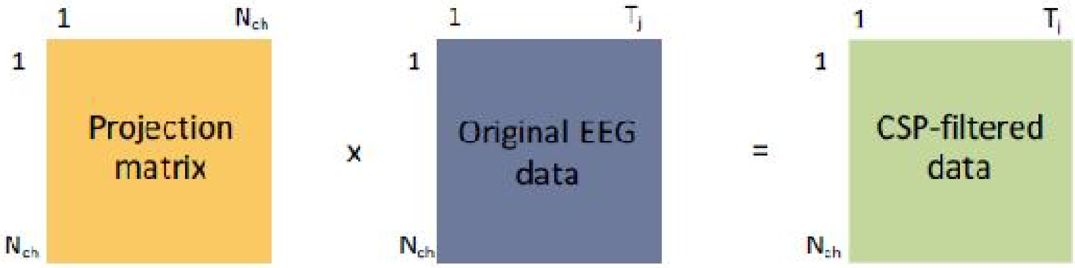


Figure 4-6: The transformation of the original EEG data with CSP-filters is a linear transformation.N_{ch} is the number of EEG channels and T_j the number of samples for that trial.

The projection matrix W will have as much filters as there are channels and the columns of the matrix will carry the weights to make linear combinations of the original EEG channels, thereby deciding which EEG-channels carry the most information.

The first half of the projection matrix will maximize the variance for class one and minimize it for class two, while the second half of the projection matrix will maximize the variance for class two and minimize it for class one.

Under the assumption that the signal is band-pass filtered, the projection matrix is constructed as follows:

Starting with the potential differences V_j from trial j, the covariance matrices are calculated for both classes with C_1 holding the left hand trials and C_2 holding the right hand trials:

$$\Sigma_1 = \sum_{j\in C^1} \frac{V_j V_j^T}{trace\ (V_j V_j^T)} (4.16)$$

$$\Sigma_2 = \sum_{j\in C^2} \frac{V_j V_j^T}{trace\ (V_j V_j^T)} (4.17)$$

The overall covariance matrix is composed as$\Sigma = \Sigma_1 + \Sigma_2$.

This covariance matrix is diagonalized and the eigenvalues and eigenvectors can be found inE and M respectively.

$$M^T \Sigma M = E (4.18)$$

To make sure that $U\Sigma U^T = I$, the whitening transformation is performed as follows (MullerJ.*et al*., 1999):

$$U = P^{-1/2} M^T (4.19)$$

Next, R_1 is calculated and diagonalized, with D and Z containing the eigenvalues and eigenvectors respectively.

$$R_1 = U\Sigma_1 U^T (4.20)$$

$$Z^T R_1 Z = D \quad (4.21)$$

It is important that the eigenvalues on the diagonal of D are sorted in ascending order, $\geq$ 0and $\leq$ 1.

With Z sorted according to D, the filters W can be calculated as:

$$W = Z^T U \quad (4.22)$$

The original EEG-data can be transformed to V^{CSP} = W $V^{Original}$, whereas each row ofV^{CSP}can be seen as a new CSP-channel. Due to the sorting of D, the first filter-pair, asshown in Fig.4-7, contains the most discriminative information. The second filter-pairwill contain less discriminative information. When looking at the variances of each channel, the first row of a CSP-filtered signal from class one will have the highest variance and thelast row the lowest, whereas if the same CSP-filter would be applied on a class two signal, the first row of this CSP-filtered signal would show a low variance and the last row the highestvariance. This results in the biggest difference in variance being between the first and thelast channel of the CSP-filtered signal. By extracting more filter pairs or CSP-channels, moreinformation is available, but with increasing amount of channels, this information becomesless discriminative as illustrated in Fig.4-7.

If the projection matrix W, as mentioned before, was able to properly maximize the variancefor one class, while minimizing it for the other, it will be much easier to correctly classify atrial.

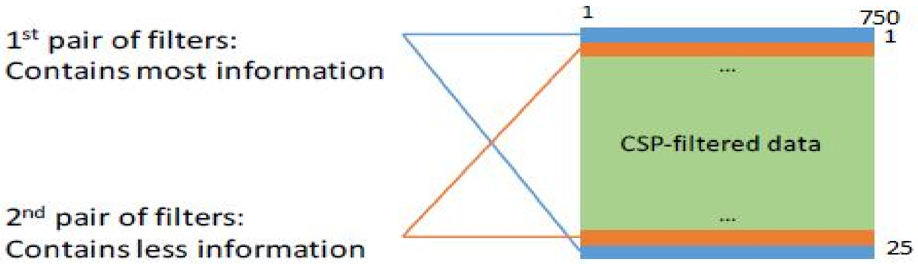


Fig.4-7: The first filter-pair extracts the most discriminative information; the second filter-pairextracts less discriminative information. 750 is the number of samples in a trial, 25 isthe amount of EEG-channels used for measurements.

Different from the conventional 2-class BCIs, the present system has to account for multiclass MI. the next subheading is presenting the CSP multiclass MI.

Based on the amount of features needed, an amount of CSP-channels, also called filter pairs,is used. When selecting a pair of filters, the outermost channels are chosen, as those filterscorrespond to the highest and lowest eigenvalues by construction, and thus contain mostinformation. In this study the systemselectsthemaximal2andtheminimal2eigenvalues and thecorrespondingeigenvectorsfor W.

4.5.2.2.CSP-based extensions for multi-class BCI systems

The main idea of CSP-based methods in dealing with a Multiclass BCI problem is to convertthe multi-class problem into multiple 2-class problems, apply the 2-class CSP analysisto solve these problems, and then combine the 2-class problem solutions together toform the solution of the original multiclass BCI problem. Based on that idea, there aretwo well-known strategies. The first is called one-versus-the-rest (CSP 1vsN). The second is called pairwise that considers all possible pairs of N classes (CSP pairs).

4.5.2.3.One-versus-the-Rest CSP

In this strategy, one class is selected and the $N-1$ remaining classes are assumedto have very similar covariance matrices to form the other class. The 2-class CSPanalysis is then applied to covariance matrix of the selected class and the estimatedcovariance matrix of the remaining $N-1$ classes. There are *N* possible options to select a class; therefore, there are *N* 2-class problems that need to be solved. The final spatial filters are formed by a combination of selected spatial filters of these *N*2-class problems. As a result, the length of the feature vectors of *CSP1vsN* methodin *N*-class BCI is $Q \times K \times N$ where Q is the number of the selected spatial filters. Letz^n be the feature vector solving the n-th problem in which class n is separated fromother classes as shown in Fig. (4-7). the feature vector of CSP 1vsN is concatenatedfrom *N* feature vectors as shown in Eq. next.

$$z = (z^1, z^2, \dots, z^N) \quad (4.23)$$

4.5.2.4.Pair-wise CSP:

In this strategy, all the possible pairs of N classes are to be solved by applying conventional 2-class CSP method, and there are total $\frac{N \times (N-1)}{2}$2-class problems thatneed to be solved. Similarly to the CSP 1vsN strategy, the final spatial filters area union of all selected spatial filters of the $\frac{N \times (N-1)}{2}$2-class problems. Let $z^{i,j}$ be thefeature vector solving the problem of the pair of class i and class j as shown in Fig.(4-7). the feature vector of CSP-pairs is concatenated from $\frac{N \times (N-1)}{2}$feature vectorsas shown as in Eq. below.

$$z = (z^{1,2}, \dots, z^{1,N}, z^{2,3}, \dots, z^{2,N}, \dots, z^{N-1,N}) \quad (4.24)$$

4.6. Classification Phase

Present day BCI systems still have shortcomings that prevent their widespread application. In particular, these shortcomings are caused by limitations in the functionality of the pattern recognition algorithms used for discriminating brain signals in BCIs. Pattern recognition methods that allow discriminating EEG segments representing different types of brain activity are considered as important component of BCI EEG based systems (Tangermann*et al*., 2012). A BCI system can be considered as a pattern recognition system with special focus on the classification algorithms used to design them. Hence, in this thesis special emphasis is given to pattern recognition algorithm.

The supervised machine learning algorithms in turn subdivided into parametric and non-parametric classification methods. The parametric methods such as Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM), the goal of LDA is to create a hyperplane that separates both classes with the class of thefeature vectors depending on the side of the vector regarding to the hyperplane. SVM'suse the same principle to discriminate between classes, but the hyperplane is selected basedon a maximisation of margins. Regardless of the fact that SVM's have good generalization properties and are insensitive to overtraining and the curse-of-dimensionality. The other type is non-parametric methods of classification such as the Euclidian distance classifier, and the K-NN classifier.

In the following section the algorithm for learning classifiers from training data are discussed and for performing classification of new data not used during training. The pattern recognition algorithm employedissupport vector machine which is presented in the next section.

4.6.1. Support Vector Machine (SVM)

Tobegin with, consider a two class classification problem with 2-dimentionalfeatures. Let the circles and triangles in Fig.4-8represent observationsbelonging to two different classes. Using these observations, many separatinghyperplanes can be selected as classifier for the problem as it is seen in the Fig.4-8.

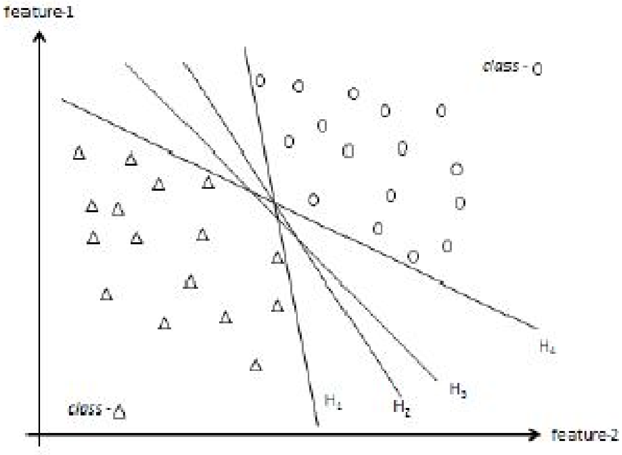


Fig.4-8: Separating hyperplanes possible to be selected as classfier for the problem (Erman, 2011)

Among all these hyperplanes, SVMs try to find the optimum one which is calledOptimum Separating Hyperplane (H_{osh}). H_{osh} is optimum in terms of its generality androbustness. It discriminates the classes such that the margin between the classboundaries is maximized. The class boundaries are determined by theobservations closest to H_{osh}, which are called support vectors. Support vectors, $H_{osh}$and the margin width are shown in Fig.4-9.

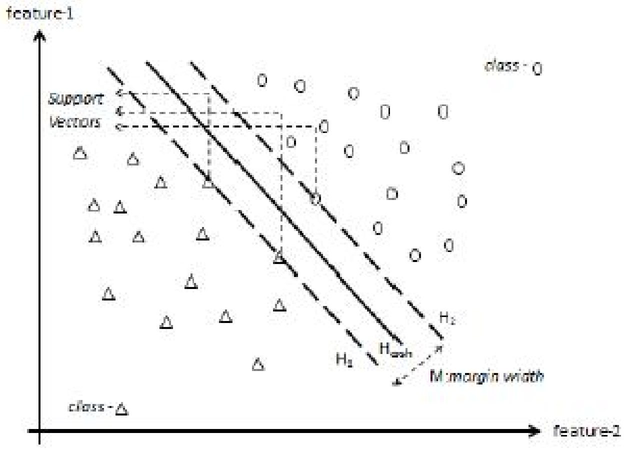


Fig.4-9: Optimum separating hyperplane maximizes the margin width determined by the support vectors (Erman, 2011).

The objective function of the SVM algorithm to be minimized can be expressedas follows:

$$\frac{1}{M} + C\sum_{k=1}^{R}\xi_k (4.26)$$

Where:

M: The margin width,

ξ : The distance of the misclassified observation to its class boundary (Fig.4-10),

C: Tradeoff parameter between the addends.

The first term in Eq. (4.26) is due to maximize the margin width and the second term is to minimize the distance of the misclassified observations to their class boundary. *C*, the tradeoff parameter between the terms, is selected by handaccording to the problem.

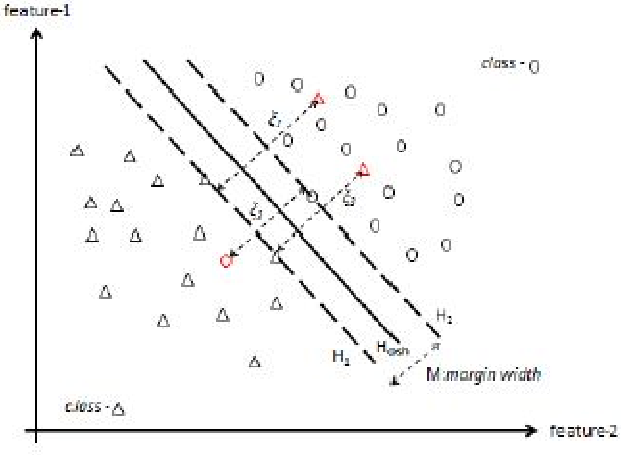


Fig.4-10: The objective of the SVM algorithm is to maximize the margin width andminimize the distance of the misclassified observations to their class boundary (Erman, 2011).

When the observations are separable, they are separated in their original space bythe H_{osh}. Otherwise, they can be mapped to a higher dimensional space in whichthey are separable. This situation is illustrated for 1-dimensional feature case below. The observations in Fig. 4-11 are separable by a 1D H_{osh}. However it is not the case in 1D for the observations in Fig. 4-12. They become separable only when they are mapped to a higher dimensional space as it is seen in Fig. 4-13.

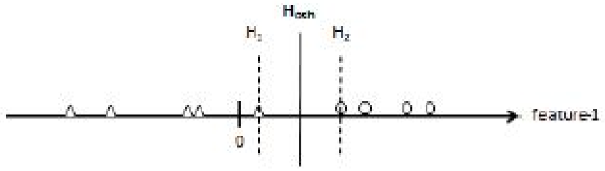


Fig.4-11: Observations separable in 1D

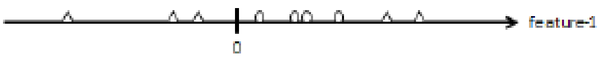


Fig.4-12: Observations not separable in 1D

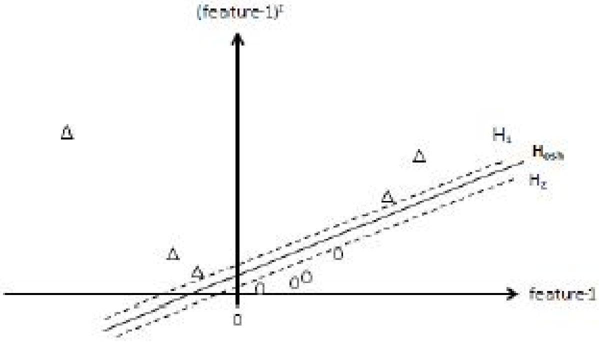


Fig.4-13: Observations those are not separable in 1D being separable in 2D.

The functions that map the observations to a higher dimensional space are calledKernel functions. Some examples of the Kernel functions used in SVM are givenbelow.

1. **Linear Kernel:**

$$K\left(x_i, x_j\right) = x_i^T x_j (4.27)$$

2. **Polynomial Kernel:**

$$K\left(x_i, x_j\right) = \left(\gamma x_i^T x_j + r\right)^d \quad , \qquad \gamma > 0 (4.28)$$

3. **Radial Basis Function Kernel:**

$$K\left(x_i, x_j\right) = exp\left(-\gamma \left|x_i - x_j\right|^2\right), \qquad \gamma > 0 (4.29)$$

4. **Sigmoid Function:**

$$K\left(x_i, x_j\right) = tanh\left(\gamma x_i^T x_j + r\right) (4.30)$$

Here, γ, r, and d are kernel parameters to be adjusted for the specific classification problem(Hsu*et al*., 2010). More information is discussed in Appendix A.

4.7.Evaluation

In order to analyze the performance of BCI systems, several evaluation techniques can be used. Classification accuracy(*Acc*), Cohen's Kappa Coefficient (κ), and Nykopp's information transfer are usually used to analyze the performance in the experiments. They are commonly used in the BCI competitions to compare the results of different research groups (B. Blankertz, http://www.bbci.de/competition/iv). One of the main limitations of the classification accuracy is it does not consider the off-diagonal elements in the confusion matrix. Also the weight of a class in the calculation depend on the number of samples from that class

The following section is introducing the Cohen's Kappa Coefficient which is used for evaluation the competitors results in BCI competition IV data set 2.

4.7.1. Cohen's Kappa Coefficient

When the limitations of the classification accuracy is considered, Cohen's kappacoefficient, κ, serves a more reliable and sensitive evaluation criteria. In thecalculation of κ, the classification accuracy, *Acc* (overall agreement), and thechance agreement,p_e, is used together. The definition of $p_e$is given below:

$$p_e = \frac{\sum_{i=1}^{M} n_{:i} n_{i:}}{N^2} \quad (4.31)$$

Where $n_{:i}$ and $n_{i:}$ are the sum of the i^th column and the i^th row of the confusionmatrix, respectively. Then, the kappa coefficient, κ, is calculated as it is givenbelow:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4.32)$$

The maximum value that the kappa coefficient can take is 1 (perfectclassification). The value changes depending on the correlation between thepredicted classes and the actual classes (A. Schl, http://www.citeulike.org/user/jmetzen/article/3211329).

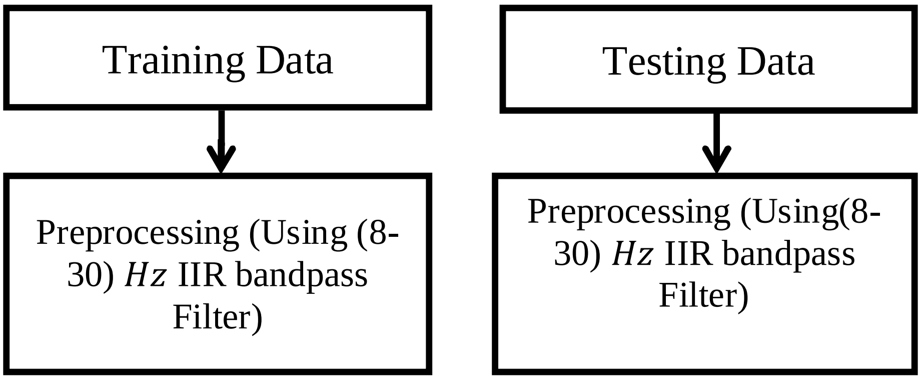
CHAPTER FIVE

Body of Research

The work done is represented as the preprocessing phase; features extraction method and the classification algorithm employed in this study are shown in the block diagrams shown below. Figs.5-1& 5-2show the processing phase which takes raw data as input and yields feature vectors as output. The processing steps are:

1. Single trial extraction.
2. Time windowing
3. Freq. (IIR Band-pass filtering) or Spatial(Applying ICA or SOBI).
4. Feature vector construction (different Wavelet mother's coefficients + CSP).
5. SVM Classifier for Six features pairs.
6. Voting.

Based on the above stages the following block diagrams Figs.(5-1& 4-15) show the structure of the proposed systems:


Training Data
Testing Data
Preprocessing (Using (8-30) Hz IIR bandpass Filter)
Preprocessing (Using(8-30) Hz IIR bandpass Filter)

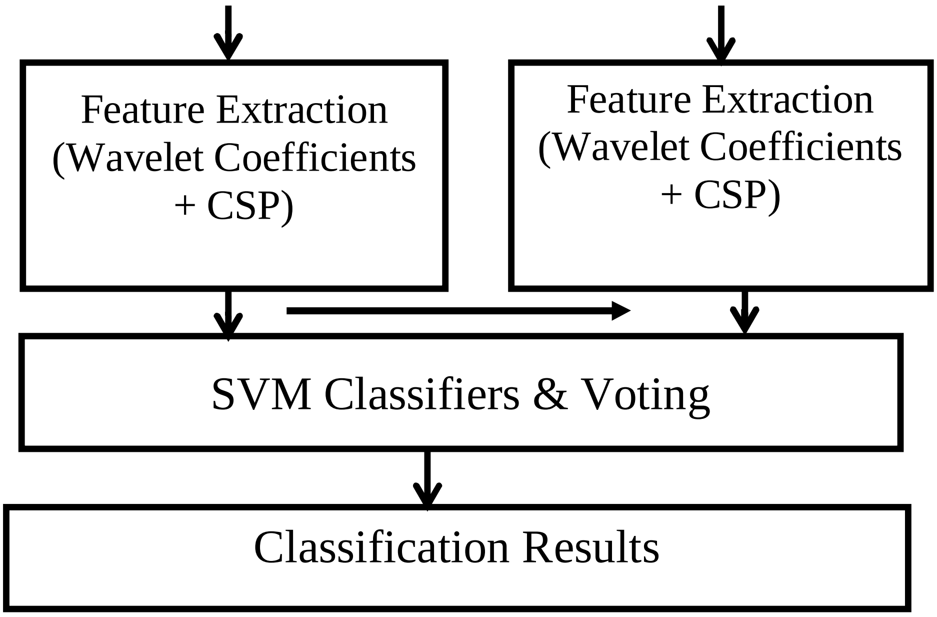
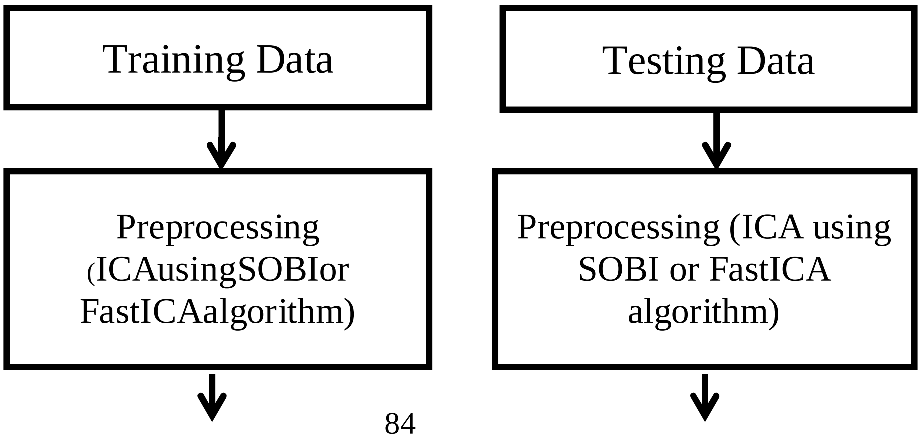

Feature Extraction (Wavelet Coefficients + CSP)
Feature Extraction (Wavelet Coefficients + CSP)
SVM Classifiers & Voting
Classification Results

Fig.5-1: Block Diagram for Proposed Training and Testing Data Algorithm Depending on (8-30) HzIIR Bandpass Filter


Training Data
Testing Data
Preprocessing (ICAusingSOBIor FastICAalgorithm)
Preprocessing (ICA using SOBI or FastICA algorithm)

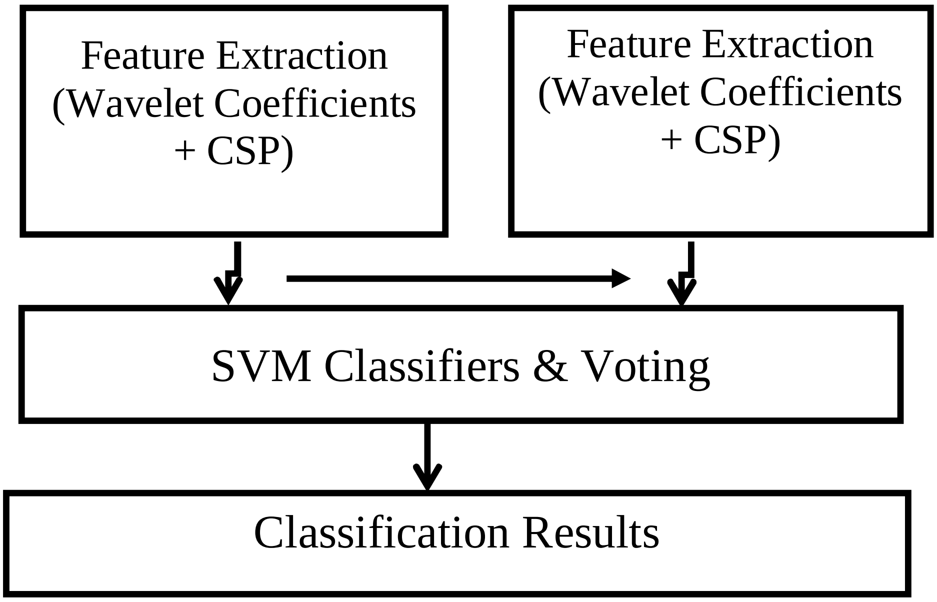

Feature Extraction (Wavelet Coefficients + CSP)
Feature Extraction (Wavelet Coefficients + CSP)
SVM Classifiers & Voting
Classification Results

Fig.5-2: Block Diagram for Proposed Training and Testing Data Algorithm Depending on ICA

Matlab codes were constructed to implement the two algorithms andMatlab version 2018 was used.

5.1 Data Sets

Data sets stored in the General Data Format for biomedical signals(GDF), one file per subject and session. All files are listedin Table 5-1. Hence the algorithm will be implemented in MATLAB, the GDF files was loaded using the open-source toolboxBioSig, available at http://biosig.source-forge.net/.

Table 5-1: List of all files contained in the data set

Subject	Training Data	Testing Data
1	A01T.gdf	A01E.gdf
2	A02T.gdf	A02E.gdf
3	A03T.gdf	A03E.gdf

4	A04T.gdf	A04E.gdf
5	A05T.gdf	A05E.gdf
6	A06T.gdf	A06E.gdf
7	A07T.gdf	A07E.gdf
8	A08T.gdf	A08E.gdf
9	A09T.gdf	A09E.gdf

The followingcommand in Octave/FreeMat/MATLAB is used to loaded a GDF file with the BioSig toolbox:

```
[s, h] = sload('A01T.gdf');
```

The workspace will then contain two variables, namely the signals‘ s ‘ anda header structure’ h’. The signal variable contains 25 channels (the first 22are EEG and the last 3 are EOG signals). The header structure containsevent information that describes the structure of the data over time. Thefollowing fields provide important information for the evaluation of this dataset:

h.EVENT.TYP

h.EVENT.POS

h.EVENT.DUR

h.ArtifactSelection

h.SampleRate

h.Label

The position of an event in samples is contained in h.EVENT.POS. The correspondingtype can be found in h.EVENT.TYP, and the duration of thatparticular event is stored in h.EVENT.DUR. The types used in this data setare described in Table 5-2 (hexadecimal values, decimal notation in

parentheses).Additional information provided such as sampling rate in h.SampleRate and
cell of channel labels in h.Label.

Note that the class labels (i. e., 1, 2, 3, 4 corresponding to event types769, 770, 771, 772) are only provided for the training data and not for thetesting data.The trials containing artifacts as scored by experts are marked as eventswith the type 1023. In addition, h.ArtifactSelection contains a list of alltrials, with 0 corresponding to a clean trial and 1 corresponding to a trialcontaining an artifact.

Table 5-2: List of event types (the first column contains decimal values and the second hexadecimal values).

Event	Type	Description
276	0×0114	Idling EEG (eyes open)
277	0×0115	Idling EEG (eyes closed)
768	0×0300	Start of trial
769	0×0301	Cue onset left (class 1)
770	0×0302	Cue onset right (class 2)
771	0×0303	Cue onset foot (class 3)
772	0×0304	Cue onset tongue (class 4)
783	0×030F	Cue unknown
1023	0×03FF	Rejected trial
1072	0×0430	Eye movements
32766	0×7FFE	Start of a new run

The recorded EEG signals ‘s’ is organized in one matrix so-called *Signal* of size total number of samples (t) x 25 channel. For training and testing files the size of *Signal matrix* is shown in Table 5-3

Table 5-3: Matrix Size of Training and Testing Signals

Subject	Training Data Size	Testing Data Size
1	672528×25	687000×25
2	677169×25	662666×25
3	660530×25	648775×25
4	600915×25	660047×25
5	686120×25	679863×25
6	678980×25	666373 ×25
7	681071 ×25	673135×25
8	675270 ×25	687792 ×25
9	673328 ×25	675098 ×25

The data foldercontains 18 GDF files, 9 of them are forcalibration and the others are for evaluation. The total size of the folder is 574 MB.

A more detailed description of the dataset can be found in the BCI competition IV online web site document (Brunner C.*et al*., http://www.bbci.de/competition/iv/desc_2a.pdf).

5.2.Preprocessing Phase

5.2.1. Single Trial Extraction

To build high-level features that can be fed to a classifier deduced from a part of the EEG signal which occurs after each row/column start sound which is considered as single trial and have to be processed in Eq. 5.1.

$$\text{Single trial length} = \text{round}\left(t_d * \text{sampling frequency} / 1000\right) \qquad (5.1)$$

Where t_d corresponds to the duration of the signal of interest after beep sound in milliseconds.

As previously mentioned in the data description section (4.1) the collected signals are digitized at 250 Hz. Therefore, the sampling frequency is 250 Hz and the selected t_d = 7500 ms. by applying the formula:

Single trial length = round ((7500 ms / 1000)×250 Hz)

= 1875 samples or measurements (i.e. points hereafter mentioned as samples). (5.2)

Thus each single trialof calibration or evaluation consists of 1875 samples per channel. For all 25 channels each single trial is consists of (1875×25) samples.

5.2.2. Discrimination between Different Classes of Calibration Data

As mentioned in Table (5.1) for the calibration data recording each subject performed 72 trials distributed randomly between the four different classes (tongue, foot, right and left hand), so in this step similar class cued data are separated from each other using MATLAB selection and iteration statements. Accordingly four arrays were developed.

The number of samples for eachtrial is 1875 and the number of trials per class is 72 so the array for each class calculated as following:

Number of samples per class = number of samples per trial (1875) × number of trials (72) = 135000 samples. (5.3)

That means the array for all channels is (135000 samples ×25 number of channels) per class. (5.4)

Note that the rejected trials will be removed from the calibration data to eliminate the undesired features from data.

5.3. Time Windowing

Time domain features are related to changes in the amplitude of neurophysiologic signals, occurring time-locked to the actions of the subject. One of the good examples for signals that can be characterized with the help of time domain features are the ERD. A strategy used here is to extract the EEG signal segments which are related with ERD during each trial imagining time see Fig 4-3, thesesegments are created by a shifting short-time window of a 3 seconds length running through the data stream starting at t= 3 and end at t=6 for each trial that mean the final number of samples per trial segments are:

Number of samples for each trial segment = Time window (3) s × Number of samples per second (250) = 750 Samples. (5.5)

Thus each single trial segment of calibration or evaluation consists of 750 samples per channel. For all 25 channels each single trial segment consists of (750×25) samples.

5.4. Frequency Filtering

Filtering is a crucial step in noise reduction since certain types of artifact occur at known frequencies. The provided signals are originally band-pass filtered from (0.5-100)Hz as previously mentioned in data description section. After extracting single trials extra filtering was applied over the resulting single trials segments in order to remove residual noise.

As mentioned in section (3.6) the primary phenomenon of MI EEG is event-related desynchronization (ERD) or event-related synchronization which is the attenuation, or increase of the rhythmic activity over the sensorimotor cortex generally in the μ(8–14) Hz and β(14–30) Hz rhythms. The ERD/ERS can be induced by both imagined movementsin healthy people or intended movements in paralyzed patients. It is noteworthy that another neurological phenomenon called Bereitschafts potential is also associated with MI EEG but non-oscillatory. In this thesis ERD/ERS features are considered only.

Feature extraction of ERD/ERS is, however, a challenging task due to its poor, low signal to noise ratio. Therefore, spatial filtering in conjunction with frequency selection (via processing in either temporal domain or spectral domain) in multi-channel EEG has been highly successful for increasing the signal to noise ratio.

Signal has been filtered from unnecessary information including electrooculography (EOG) artifacts, to extract only mu (μ) and beta (β) waves this was accomplished by applying 5th order Chebyshev type II, IIR band pass filter (8 - 30Hz) (Soroosh *et al.*, 2017) (Siavash *et al.*, 2018).

In the first algorithm,the above filter applied to each subject data; As a result a set of calibration and evaluation data with the same size of signals for the nine subjects will be constructed. The next spatial filtering step has been performed to compare frequency with spatial filteringin removing artifacts and classification performance.

5.5.Independent Component Analysis

In the second algorithm, Independent component analysis (ICA) method was used in the preprocessing stage, in order to extract the most efficient features in each electrode separately, using two software: SOBI and FastICA to compare them and determine which is faster when applied.As a result a set of calibration and evaluation data with the same size of signals for the nine subjects will be constructed.

The data entered to the software tool is a matrix X which contains the training or calibration examples of each subject concatenated for each electrode of the 25 electrodes.

In this work ICA was tried without data reduction (i.e. using the whole 25 electrodes). And also ICA was tried in order to reduce the number of electrodes from 25 to 18 electrodes (i.e. minimum number of electrodes used without removing useful information).

5.6. Wavelet - CSP Method:

The sample frequency f_S in our data (S) is 250Hz, the sub-frequency bands of all components will be calculated as [0, $f_s/2^{i+1}$], [$f_s/2^{i+1}$, $f_s/2^{i}$], [$f_s/2^{i}$, $f_s/2^{i-1}$], …., [$f_s/2^{2}$, $f_s/2$], successively, Table5-4 presented the frequency ranges of wavelet transformation at four levels which results in four details and one approximation,we decided to extract the vectors of features from cD_3& cD_4 only which provide proper presentation for the μ and the β.

Table 5-4: Frequency ranges at four level wavelet transformations:

Signal Component	Frequency Range
CD_1	62.50 - 125
CD_2	31.25 - 62.5
CD_3	15.63 - 31.25
CD_4	7.81 - 15.625
CA_4	0 - 7.81

The coefficients of different Wavelet Daubechies (2 -20) and Coiflets (1 – 5) were examinedin calculating common spatial filtersto discriminate between four imagined classes, where four detail levels of the wavelet coefficients were calculated.

In this study CSP pairs are implemented and when applying the above equations, CSP filter with (25 ×25) features is resulting for the four different classes. Consequences, there will be six different CSP pairs:(class 1&2, class 1&3, class 1&4, class 2&3, class 2&4, and class 3&4).

If all EEG-data points would be used, the dimensionality of the data would be too high to give to a classifier, so the most relevant features are extracted. As CSP is designed to discriminate between conditions by optimizing the variances, the log variances of the CSP-filtered signal can be used as features f_i, i being the respective CSP-channel.

$$f_i = log(VAR(V_i^{CSP})) \quad (5.25)$$

That means the feature vector size for each CSP pair is ((number of selected maximum and minimum eigenvectors) 4 × (number of trials per 2-class) 144). Six pairs of feature vectors (4×144) will be carried out to the classifier.

5.7. Classification

In this study, MATLAB implemented SVM toolbox is utilized for SVM classification (“LIBSVM - A Library for Support Vector Machines,” http://www.csie.ntu.edu.tw/~cjlin/libsvm/). Six SVM models were built to discriminate between the six features pairs (class 1&2, class 1&3, class 1&4, class 2&3, class 2&4, and class 3&4).

5.8. Voting

Based on the above systems the output of the classifiers must be voted to determine the most frequent value of each classifier result. Mod Matlab function is used to majority voting for classification.

CHAPTER SIX

Results

Many of the characteristics of BCI systems depend critically on the employed machine learning algorithm. Important characteristics that are influenced by the machine learning algorithm are classification accuracy as long as the amount of time and user intervention necessary for setting up a classifier from training data.

This chapter encompasses resultsof the different methodologies implemented, as well as exemplifications of the results and the evaluation and the observations that were made throughout. Moreover, we are going to compare execution time between different methods implemented in this work to suggest the suitable method and to reveal the thesis contribution.

In this section the classification performance yielded on the test data of competition IV will be demonstrated. Before learning classifiers and before performing machine learning algorithms, the data were preprocessed with two different methods (frequency and spatial filtering). This was done because the goal of this study is to do fair comparison between the two different preprocessing methods for multiclass motor imagery dataset.

6.1. Results for Preprocessing Stage

6.1.1. Frequency Filtering Results

Test sets for the nine subjects have been processed similarly to the training set using 5th order Chebyshev type-II, IIR bandpass filter (8 - 30)Hzas described in section (5.4). Figs. (5-1, 5-2 and 5-3) show the performance of the IIR-filter and a contrast in the frequency domain between the (8-30) Hzband-pass filtered EEG and the raw EEG datain C_3, C_Z and C_4 respectively.

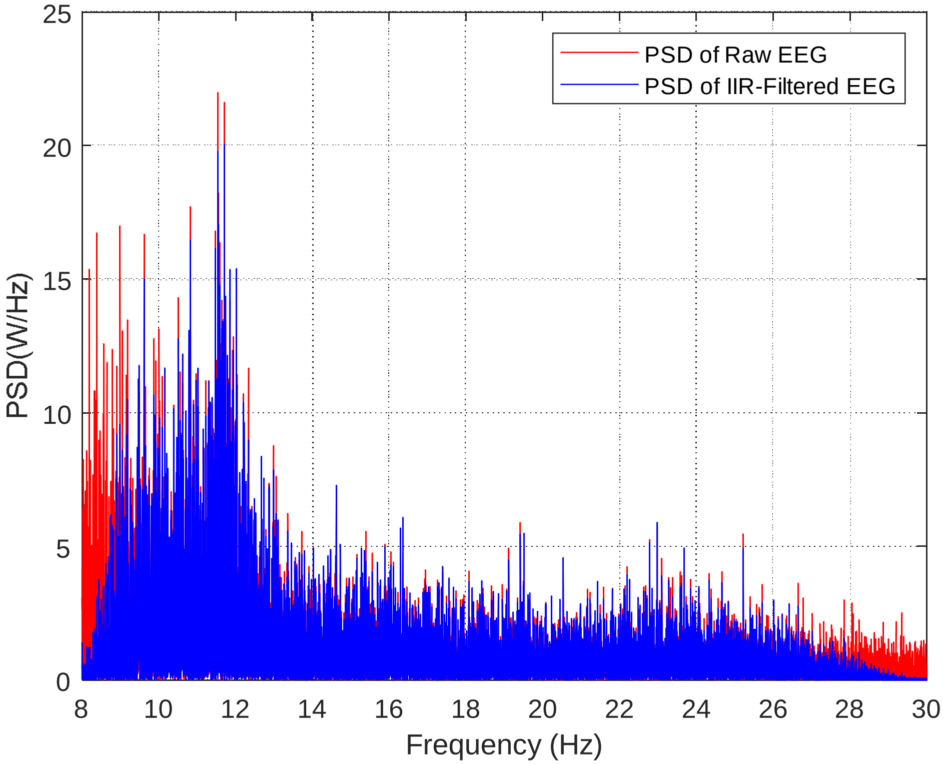


Fig. 6-1:Power spectral density(PSD) of C_3 band-pass filtered compared with EEG and the Raw EEG in (8-30) Hz

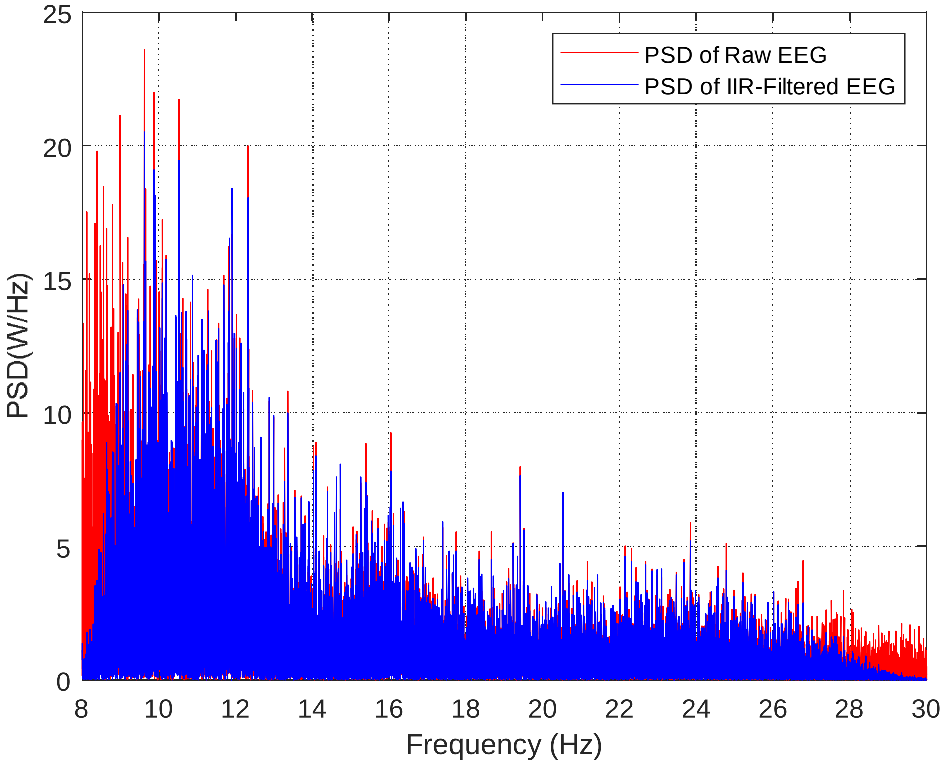


Fig. 6-2:Power spectral density(PSD) of C_Z band-pass filtered compared with EEG and the Raw EEG in (8-30) Hz

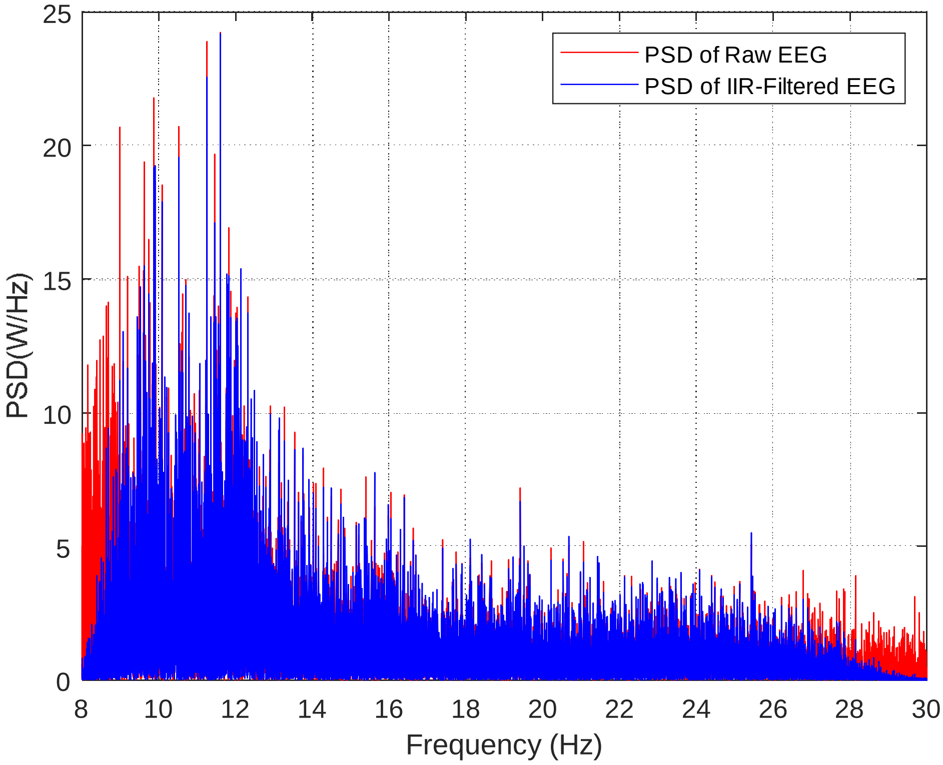


Fig. 6-3:Power spectral density(PSD) of C_4 band-pass filtered compared with EEG and the Raw EEG in (8-30) Hz

6.1.2. Results of ICA algorithms runningtime

As mentioned in section (5.5) two methods of blind source separation were applied to the EEG data to select the faster one. ICA and SOBI were applied to the raw data using FastICA & SOBI toolboxes respectively to get the results.

Running time factor to achieve the ICA using FastICA software and SOBI toolbox is addressed in this work. Running Time (latency) is defined as the amount of time passed from the initiation of the algorithm sequence to the retrieval of the mixing matrix; this benchmark aims to measure the computational load of the algorithm. As the time duration can be variable because of the computer capacity (processor frequency, RAM, etc.), results have been transformed into a percent form showing how “slow” an algorithm is in relation to the faster one.For running time measurements a MATLAB function was developed in order to structurethe running time of each subject. This functionreturned the time between the beginning and the end of eachalgorithm in seconds. This MATLAB function uses*(tic-toc)* function.

As mentioned in the data description, calibration dataconsists of 288 trials (72 for each of the four possible classes) for each of the nine subjects.This makes 9 tests to be analyzed for each subject.

Figs (6-4) and Table (6-1) present the differences in running time when using FastICA or SOBI toolboxes in each subject.

Table 6-1: Time period to execute FastICA and SOBI toolboxes in each subject for calibration data.

Subjects	Time period to execute FastICA Software (s)	Time period to execute SOBI Software (s)
Sub1	45.3	18.7
Sub2	41.0	18.4
Sub3	43.0	17.9
Sub4	33.3	20.3
Sub5	33.0	18.9
Sub6	39.1	18.6
Sub7	35.6	18.1
Sub8	35.1	18.3
Sub9	35.6	18.5

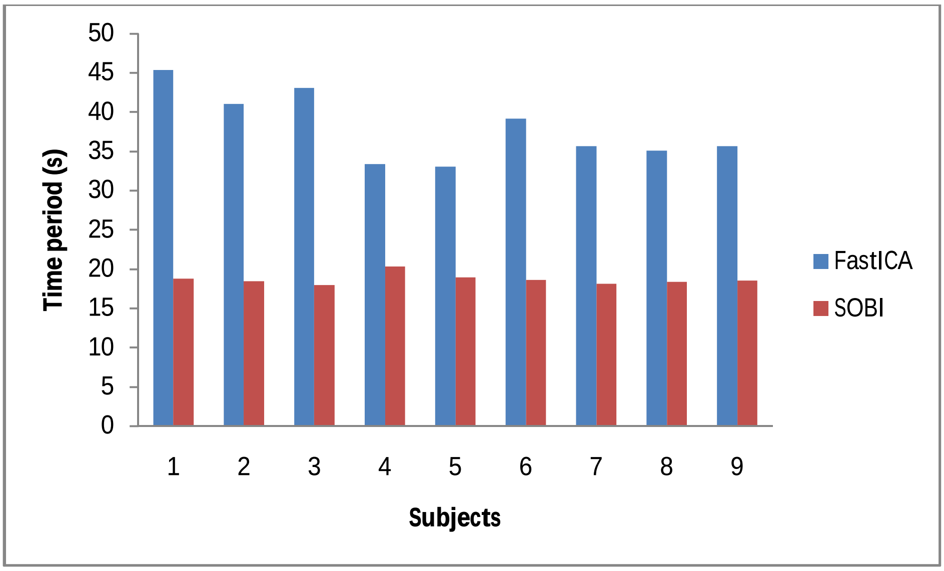


Fig.6-4: Running time for FastICA software compared to SOBI toolbox in each subject for calibration data.

To assess the significant of differences and correlation between running times of algorithms a paired t-testwas computed as shown in the Tables((6-2) – (6-4)). Results showedthe running timemean ± Std.is38 ±1.45 *s* in FastICA while it is 18.63 ± 0.23*s* in SOBI.

Table 6 - 2: Statistical evaluation parameters

Algorithm	Mean (s)	Number of Samples	Std. Deviation	Std. Error Mean
FastICA	38.000	9	4.3549	1.4516
SOBI	18.633	9	0.6946	0.2315

Table 6 - 3: Paired correlation parameters

Pair 1	Number of Samples	Correlation	Sig.
ICA & SOBI	9	-0.440	0.236

Table 6 - 4: Paired differences parameters

Pair 1	Paired Differences					t	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
				Lower	Upper		
ICA & SOBI	19.3667	4.7021	1.5674	15.75	22.98	12.356	0.000

6.2. Classification Results:

InTables6-5 and 6-6, the kappa coefficient values are given for each subject using all channelsunder IIRfiltering for preprocessing. For feature extraction different db and coif wavelet functions are used.

Table 6 - 5: results of using IIR filter combined with 10 different wavelet db mother functions

Sub	**Sub1**	**Sub2**	**Sub3**	**Sub4**	**Sub5**	**Sub6**	**Sub7**	**Sub8**	**Sub9**	**Avg.**
Db										
2	0.72	0.30	0.69	0.42	0.26	0.26	0.64	0.78	0.69	**0.53**
4	0.75	0.25	0.69	0.45	0.30	0.26	0.67	0.70	0.51	0.51
6	0.73	0.24	0.65	0.44	0.22	0.26	0.60	0.68	0.44	0.47
8	0.73	0.26	0.67	0.44	0.18	0.19	0.58	0.68	0.47	0.47
10	0.68	0.29	0.63	0.48	0.20	0.26	0.65	0.74	0.47	0.49
12	0.74	0.30	0.69	0.44	0.19	0.25	0.65	0.69	0.48	0.49
14	0.73	0.25	0.66	0.42	0.20	0.24	0.61	0.72	0.49	0.48
16	0.75	0.25	0.64	0.39	0.18	0.23	0.67	0.69	0.51	0.48
18	0.77	0.29	0.65	0.40	0.25	0.23	0.64	0.65	0.49	0.49
20	0.73	0.30	0.67	0.37	0.15	0.27	0.65	0.68	0.44	0.47

Table 6-6: results of using IIR filter combined with 5 different wavelet coif mother functions

Sub **Coif**	**Sub1**	**Sub2**	**Sub3**	**Sub4**	**Sub5**	**Sub6**	**Sub7**	**Sub8**	**Sub9**	**Avg.**
1	0.77	0.24	0.70	0.44	0.19	0.25	0.70	0.79	0.58	**0.52**
2	0.74	0.22	0.65	0.44	0.16	0.30	0.65	0.76	0.53	0.49
3	0.75	0.26	0.68	0.40	0.21	0.23	0.69	0.78	0.48	0.50
4	0.68	0.29	0.65	0.40	0.19	0.30	0.66	0.71	0.48	0.48
5	0.70	0.32	0.61	0.35	0.17	0.19	0.63	0.73	0.47	0.46

The temporal filtering operation increases the kappa coefficient as it is seen inTable 6-5. By comparing the results of this system, it is obvious that the optimum average kappa value that can be achieved is 0.53. This performance has been done by inputting the db2 wavelet coefficients to calculate CSP features. Using Coif1 also ranking an overall 0.52 average kappa value in Table 6-6.

Tables6-7 and 6-8 show the results of the spatial filtering method ICAwhich decreases the performanceslightly. If all channels are considered, the kappa coefficient value for subject 1 is0.51using db2 while it is 0.47 for the same subject using coif 1. As it is demonstrated, the classificationsuccess rates werefluctuating along with the change of thesignal quality. Researches show that the artifacts and themotor imagery signals have frequency overlap.Therefore, the ICA method can retain the integral of themotor-related signals better through separating the motorimagery signals from the raw EEG to remove the artifacts.

Table 6-7: Results of using ICAcombined with 10 different wavelet db mother functions

Sub Db	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Avg.
2	0.73	0.40	0.66	0.57	0.11	0.20	0.54	0.65	0.69	**0.51**
4	0.75	0.39	0.68	0.61	0.10	0.15	0.57	0.67	0.62	0.50
6	0.72	0.38	0.68	0.52	0.19	0.15	0.52	0.64	0.57	0.49
8	0.72	0.35	0.66	0.54	0.18	0.24	0.55	0.54	0.54	0.48
10	0.70	0.33	0.68	0.54	0.08	0.21	0.55	0.58	0.61	0.48
12	0.64	0.39	0.64	0.52	0.12	0.27	0.55	0.61	0.55	0.48
14	0.72	0.32	0.61	0.51	0.14	0.19	0.59	0.58	0.50	0.46
16	0.68	0.37	0.61	0.57	0.11	0.14	0.54	0.61	0.55	0.46
18	0.63	0.36	0.60	0.59	0.10	0.16	0.58	0.53	0.61	0.46
20	0.66	0.38	0.65	0.50	0.12	0.18	0.62	0.63	0.52	0.47

Table 6 - 8: Results of using ICA combined with 5 different wavelet coif mother functions

Sub Coif	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Avg.
1	0.74	0.42	0.63	0.53	0.12	0.19	0.59	0.53	0.46	0.47
2	0.72	0.40	0.67	0.56	0.13	0.17	0.53	0.69	0.50	0.49
3	0.72	0.33	0.61	0.57	0.03	0.19	0.61	0.50	0.50	0.45
4	0.74	0.36	0.63	0.49	0.09	0.16	0.52	0.54	0.53	0.45
5	0.70	0.38	0.62	0.49	0.13	0.14	0.60	0.50	0.50	0.45

To remove the artifacts and to improvesignal quality, 18 selected channels around the C3, Cz, C4 in the motorsomatosensory area were chosen from each data set and thealgorithm was applied on these channels.Tables 6-9 and 6-10 show the results of the spatial filtering method for selected18 ICs combined with different wavelet mothers.

Table 6-9: Results of using 18 ICs combined with 10 different wavelet db mother functions

Sub Db	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Avg.
2	0.74	0.47	0.72	0.49	0.19	0.27	0.65	0.72	0.71	**0.55**
4	0.74	0.40	0.63	0.43	0.09	0.23	0.62	0.63	0.67	0.49
6	0.73	0.41	0.67	0.36	0.17	0.21	0.58	0.64	0.65	0.49
8	0.73	0.39	0.65	0.40	0.13	0.18	0.60	0.62	0.65	0.48
10	0.70	0.40	0.66	0.35	0.08	0.19	0.65	0.60	0.67	0.48
12	0.75	0.37	0.66	0.36	0.08	0.19	0.60	0.64	0.61	0.47
14	0.74	0.40	0.63	0.36	0.15	0.16	0.67	0.61	0.69	0.49
16	0.72	0.40	0.63	0.34	0.20	0.22	0.68	0.51	0.59	0.47
18	0.71	0.42	0.63	0.35	0.13	0.24	0.67	0.54	0.66	0.48
20	0.66	0.36	0.65	0.38	0.17	0.15	0.72	0.52	0.63	0.47

Table 6 - 10: Results of using selected18 ICs combined with 5 different wavelet coif mother functions

Sub Coif	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Avg.
1	0.73	0.41	0.64	0.40	0.11	0.24	0.63	0.60	0.57	**0.48**
2	0.73	0.44	0.67	0.38	0.15	0.19	0.57	0.60	0.47	0.47
3	0.72	0.38	0.61	0.33	0.09	0.17	0.62	0.54	0.66	0.46
4	0.69	0.40	0.63	0.40	0.09	0.14	0.50	0.65	0.63	0.46
5	0.72	0.39	0.60	0.36	0.15	0.17	0.63	0.60	0.52	0.46

CHAPTER SEVEN

Analysis, Discussionsand Interpretation of Results

7.1. Frequency Filtering

As mentioned the μ (8-14) Hz and β (14–30) Hz waves which are associated with relaxed awareness and raising metal activity and the level of attention respectively. Fig.s (6-1, 6-2 and 6-3) confirmed that the frequency of the motor imagery-related signals are mainly in $\mu$and βrhythms around (10-20) Hz, and it is clear that IIR filter removed high power spectral frequencies related with artifacts from the raw data.

7.2. Results of ICA algorithms running time

Each subject’s results have been plotted to carefully observe the behavior of the algorithm given the nine different tests. Results achieved by each algorithm had shown few differences between them suggesting that there is no remarkable difference on all subjects’ brain signals nature. Moreover SOBI algorithm proved to be faster in this system in all cases.

In Table 6-3 the correlation of the two algorithms was -0.44. The negative sign indicates the reverse correlation between them. And the level of significance (0.24) reveals that there is no significant correlation between the two algorithms.

T-value of 12.36 and p-value of 0.00 in Table 6-4 confirms the highly statistical significant difference between using FastICA and SOBI algorithm.

Results of differences and correlation between running times of algorithms are agreed with the results cited in literature review (Guillermo S.*et al.*, 2017).

7.3. Classification Results

From tables 6-5 to 6-10 it can be noted that the algorithm based on 18ICA+db2–CSP scored higher overall average kcompared to other algorithms.

It is also noted that the overall k value of db2 scored greater than other db and coif mothers' wavelet for all individuals this can be attributed to better elimination of EOG from EEG signal and compatibility of db2 with data.

It is clear that the fluctuation of the k values for using same wavelet mother in different subjects; for example in table 6-10 the k value forsubject 1 using coif1 has high k value equal to 0.73 but at the same time it is 0.11 for subject 5; this can be refers to subject specific time segment or lack of concentration of subject in a group of trials.

Table 7-1showsthe values of k whenusing IIR+db2 wavelet coefficients to calculate CSP features and the values of k ofthe second winner of BCI competition IV results (Pfurtscheller G.,http://www.bbci.de/competition-/iv/results/#winners), the competitor used the conventional CSP based on the (8-30) *Hz* IIRband-pass filters. If we compare resultsof the two methodsfor each subject independently, it is clear that kvalues of our system were higher in subjects (1, 5, 6, and 8), while it were lower than the competitor in subjects (2, 3, 4, and 7), and equally for subject 9.

Table 7- 1: The Kappa Coefficient results of the proposed IIR+ db2-CSP based system compared to IIR+CSP system

The subjects	**IIR +CSP**	**IIR+db2-CSP**

Sub1	0.69	0.72
Sub2	0.34	0.30
Sub3	0.71	0.69
Sub4	0.44	0.42
Sub5	0.16	0.26
Sub6	0.21	0.26
Sub7	0.66	0.64
Sub8	0.73	0.78
Sub9	0.69	0.69
Average	**0.52**	**0.53**

As overall average, the results yield the performance of 0.53 produced by using db2 while for the competitor it is 0.52,which means the performance were improved.

Also if the result of (IIR+db2-CSP) system is compared with the system proposed by (Siavash *et al.*, 2018), which implemented a FBCSP Combined with a linear C-SVM classifier as base line of the system, a result of 0.61 kappa coefficient was obtained as overall average. Their system achieved much better performance than this system because FBCSP performs autonomous selection of key temporalspatialdiscriminative EEG characteristics.

Table 7-2 shows the values of kappa coefficients of obtained by using 18ICA+db2 wavelet coefficients to calculate CSP features and the values of k of kappa coefficients of (Bai X. *et al.*, 2014); the authors used Wavelet-CSP with ICA-filter method. The wavelet mother family used at that work didn’t mentioned. Moreover that system was built offline based.So if the results of the two methods compared for each subject independently, it is clear that k values of our system were lower than the competitor in all subjects. This is due to the signal's quality, as it is known the offline based system also offers manual rejection of the artifacted segments to the user and to feed the classifier a selected data.

Table 7- 2: The Kappa Coefficient results of the proposed 18 ICs+ db2-CSP based system compared to 18 ICs+Wavelet -CSP system

The subjects	**18 ICs + Wavelet -CSP**	**18 ICs +db2 -CSP**
Sub1	0.75	0.74
Sub2	0.61	0.47
Sub3	0.80	0.72
Sub4	0.63	0.49
Sub5	0.57	0.19
Sub6	0.52	0.27
Sub7	0.77	0.65
Sub8	0.74	0.72
Sub9	0.72	0.71
Average	**0.68**	**0.55**

The result of (18 ICs + db2 –CSP) system has better performance if it is compared with the system proposed by (Loannis*et al.*, 2018), which implement a developed multiclass BCI decodingalgorithm that uses electroencephalography (EEG) source imaging combined with CSP-ROIs and a classifier based on individual ROI classification models, a result of 0.46 kappa coefficient was obtained as overall average.

Also (18 ICs + db2 –CSP) system has the same performance if it is compared with the system proposed by (Thanh*et al.*, 2018), which implementCSP and fuzzy classification, a result of average kappa score is 0.533 was obtained.

CHAPTER EIGHT

Conclusions and Recommendations

8.1. Conclusions

In this study, different motor imagery tasks are classified in EEG signal using several signal processing techniques.In the first stage, frequency and spatial filtering,widely used to purify the raw EEG from EOG signals, were compared. Frequency filtering with an (8-30) Hz band-pass filters and extract features by common spatialpatterns (CSP)resultedin an overall average of 0.52 kappa coefficientin BCI Competition IV. To improve the classification success rate,this thesis proposed an (8-30) Hz band-pass filterwith Wavelet-CSPmethod. For the data sets from BCI Competition IV, the features of thefour-class motor imagery were trained and tested using the Support Vector Machines (SVM).As an overall average, the results yield the performance of 0.53 kappa coefficient produced by using db2 which gives abetter performance.

This thesis also proposed a spatial filtering using ICA to eliminate EOG signal with different Wavelet mother function combined with an ordinary CSP method. For the data sets from BCI Competition IV, the features of the four-class motor imagery were trained and tested using the Support Vector Machines (SVM). As overall average, the results yield the performance of 0.55 produced by using db2 for selected ICs which improve the performance.

The coefficients of different Wavelet Daubechies (2 -20) and Coiflets (1 – 5) were examinedin calculating common spatial filtersto discriminate between four imagined classes, where four detail levels of the wavelet coefficients were calculated. Form the result it is obvious that db2 has best performance in all cases.

Another observation in the experiment is the time consumed during ICA-filtering of EEG data with SOBI software is less if it is compared with the one using FastICA software.This leads to the adoption of this system to be used in real time.

From the study, it can be concluded that on one hand, theICA-filter is better to remove the artifacts in EEG than theband-pass filter; on the other hand, theglobal projectionmatrix of CSP is easier to be trained by the waveletcoefficients than by the sampling values. Therefore, theresults gained by the proposed feature extraction approachcalled Wavelet-CSP based on an ICA-filter had a higher andmore stable classification success rate.

These systems results show the effectiveness of these algorithms against its competing techniquesand the proposed models path the way for paralyzed or disordered patients touse movement imagery to control different assistive devices in their daily activities.

8.2. Recommendations

1. Increasing training data sets and using more subjects will be more representative of the problem and increase the effectiveness of the system.
2. Time windowingmethod or dividing EEG signals into a number of time segments then extract a feature vector from each time segment using Wavelet-CSP may reduce the effect of noise and outliers on extracted features.
3. Adoption of this system to be used in real time.

8.3.Limitations and Constraints

One of the major research limitationsin this field in Sudan are the very meager and scarce previous research data.

Another limitation is the unavailability of specializedBCIlaboratoriesin Sudan.

8.4. Future trendsand Recommendations for Future Research

1. It is expected that using different Wavelet mother functions may improve features. Therefore, may further improve the performance.
2. The same methodologies can also be used to classify more than four motor imagery classes.

References

AlomariM., Awada EA., Samaha A., Alkamha K. (2014)“Wavelet-based feature extraction for the analysis of EEG signals associated with imagined fists and feet movements”. Canadian Center of Science and Education, 7(2), 18-24.

AngK. K., ChinZ. Y., WangC., GuanC., and ZhangH. (2012)“Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b”, Frontiers Neurosci., 6, 39.

AngK., ChinZ.Y., ZhangH., and Guan, C. (2008) “Filterbank common spatial pattern (FBCSP) in brain-computerinterface,” in Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN’08), HongKong, 2391–2398.

Aydemir Ö.,Dan Kayikçioğlu T. (2016) “Investigation of the most appropriate mother wavelet for characterizing imaginary EEG signals used in BCI systems”, Turkish Journal of Electrical Engineering & Computer Scineces, 24, 38–49.

BAI X. , WANG Xiangzhou , ZHENG Shuhua, YU Mingxin (2014) " The Offline Feature Extraction of Four-class Motor Imagery EEG Based on ICA and Wavelet-CSP",Chinese Control Conference, 7189-7194.

BashashatiA., M. Fatourechi, Ward R.and BirchG. (2007) "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals" Journal of Neural Engineering, J. Neural Eng, 4, 1–13.

Bashashati, A., Fatourechi, M., Ward, R. K., and Birch, G. E.(2007) “A survey of signal processing algorithms in braincomputer interfaces based on electrical brain signals”, Journal of Neural Engineering, 4(2), 32.

BaylissJ. (2003) "Use of the Evoked Potential P3 Component for Control in a Virtual Apartment", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2), 113 - 116.

BishopC. M. (2006) "Pattern Recognition and Machine Learning", Springer, Berlin.

BlankertzB., (2008) BCI Competitions Webpage. Available:http://www.bbci.de/competition/iv , (Accessed: 15/06/2020).

BlankertzB., DornhegeG., LemmS., Curio G.and MüllerK."Enhancing Brain-Computer Interfaces by Machine Learning Techniques", Available: http://www.audentia-gestion.fr/research.microsoft/blankertz_bbci_print.pdf(Accessed: 15/06/2020).

BlankertzB., MüllerK., KrusienskiD., SchalkG., Wolpaw J.and BaylissJ. (2006) "The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2), 153-159.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and robert Mller, K. (2008) "Optimizing spatial filters for robust EEG single-trial analysis". In IEEE Signal Proc. Magazine, 2008b, 581–607.

BrunnerC., R. Leeb, MüllerG. R., SchlöglA., and PfurtschellerG. (2008) BCI Competition IV data set 2a Webpage. Available:http://www.bbci.de/competition/iv/desc_2a.pdf, (Accessed: 15/06/2020).

Chuong H. Nguyen , Panagiotis A. (2017) " EEG feature descriptors and discriminant analysis under Riemannian Manifold perspective", Neurocomputing, 1871–1883.

Cichocki A.and Shun-ichi A. (2002) “Adaptive Blind Signal and Image Processing”. John Wiley & Sons.

Curran E.A., Stokes M.J. (2003) "Learning to control brain activity": A review of the production and control of EEG components for driving braincomputer interface (BCI) systems. Brain and Cognition 51, 326 – 336.

Dan Zhang, Yijun Wang, Xiaorong Gao, Bo Hong, and Shangkai Gao (2007) ”An algorithm for Idle-State Detection in Motor-Imagery-Based brain-Computer Interface” Computational Intelligence and Neuroscience, 39714.

DornhegeG., BlankertzB., Curio G.and MüllerK. (2004) "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms". IEEE Transactions on Biomedical Engineering, 51, 993–1002.

EbrahimiT. (2007) "Recent Advance in Brain-computer Interfaces", IEEE Multimedia Signal Processing 9th Workshop, Digital Object Identifier 10, 117 - 117.

Eltaf A. Mohamed, Mohd Zuki B. Y., Nidal K. S., and Aamir S, M. (2014) “Enhancing EEG Signals in Brain Computer Interface Using Wavelet Transform” ,International Journal of Information and Electronics Engineering, 4(3).

Erfanian A. and. ErfaniA. (2004) " ICA-based classification scheme for EEG-based brain-computer interface: the role of mental practice and concentration skills" Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco, CA, USA, 1(Sept. 1-5), 235-238.

Erman A.(2011) “Classification of Motor Imagery Tasks in EEG Signal and its Application to a Brain Computer Interface for Controlling Assistive Environmental Devices”, MSc thesis in F. H. Electrical and Electronics Engineering, Middle East Technical University.

Farwell L.and DonchinE. (1988) “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials”, Electroencephalography and Clinical Neurophysiology, 70(6), 510–523.

Fatourechi M., Bashashati A., Ward R. K., and Birch G. E. (2007) “EMG and EOG artifacts in brain computer interface systems”, A survey. Clinical neurophysiology, 118(3), 480-494.

Fazel R.R. (2007) "Human Error in P300 Speller Paradigm for Brain-Computer Interface", Engineering in Medicine and Biology Society, EMBS 2007, 29th Annual International Conference of the IEEE, 2516 - 2519.

Graimann B., Allison B., and Pfurtscheller G. (2010) "Brain-Computer Interfaces": A Gentle Introduction. In: Graimann B., Allison B., Pfurtscheller G. (eds) Brain-Computer Interfaces, Springer, Berlin.

Guillermo S., Humberto C. (2017) “A Comparison of SOBI, FastICA, JADE and Infomax Algorithms”, Proceedings of the 8th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC), 17-22.

Gurkok, H. and Nijholt, A. (2012) "Brain-computer interfaces for multimodal interaction" A survey and principles. Int. J. Hum. Comput. Interaction, 28(5), 292–307.

Habibeh G. and Alireza A. (2012) "Temporal windowing in CSP method for multi-class Motor Imagery Classification ", Iranian Conference of Electrical Engineering, 1602-1607.

Hardik M., Nagaraj Ramraoa , Suman K. M. (2018) "Multiclass Common Spatial Pattern for EEG based Brain Computer Interface with Adaptive Learning Classifier", Pattern Recognition Letters Elsevier, 1-6.

HoffmannU., VesinJ-M., EbrahimiT., and DiserensK. (2008) "An efficient P300-based brain-computer interface for disabled subjects ", Journal of Neuroscience Methods, 167(1), 115-125.

Hsu C.-wei, Chung ChangC., and Jen LinC.(2010) “A Practical Guide to Support Vector Classification”, Bioinformatics, 1, 1-16.

Hsu W. Y. , LinCH., KuoC. Y., LiouW. F., , SunM., TsaiY. N. (2011) “Unsupervised fuzzy c-means clustering for motor imagery EEG recognition”, Int. Journal Innovative Comput Inform Control 7, 4965-4976.

HsuW. Y., LinCH., HsuHJ., ChenPH., ChenIR. (2012) “Wavelet-based envelope features with automatic EOG artifact removal: Application to single-trial EEG data”, Expert Systems with Applications 39 (3), 2743-2749.

HughesJ. R. (2008) “Gamma fast and ultrafast waves of the brain: Their relationships withepilepsy and behavior”, Epilepsy & Behavior, 13(1), 25-31.

HyvarinenA., KarhunenJ., and OjaE. (2001) "Independent Component Analysis", JOHN Wiley& Sons Inc.

Ioannis X. , Alkinoos A. , Niki P. , Dimitris K. , and Panagiotis D. B. (2018) "Decoding Motor Imagery through Common Spatial Pattern Filters at the EEG Source Space", Computational Intelligence and Neuroscience, 1-8.

Jorge B. O.(2002) “EEG Signal Classification for Brain Computer Interface Applications”, M.Sc. dissertation,Federal Institute of Technology in Lausanne, Lausanne, Switzerland.

JoyceC. A., GorodnitskyI. F., and KutasM. (2003) “Automatic removal of eye movement and blink artefacts from EEG data using blind component separation”, Psychophysiology, 41, 313 -325.

KandelE. R., SchwartzJ. H., and JessellT.M. (1991)“Principles of Neural Science”, 3rded. New York: Elsevier/North-Holland.

Kavitha P., Cuntai G., Chiew T., VinodA. P., Kai K. (2009) " A new discriminative common spatial pattern method for motor imagery brain-computer interfaces",*IEEE Transaction Biomedical Engineering*, 56(11), 2730-2733.

Koles Z.J. (1991)“The quantitative extraction and topographic mapping of theabnormal components in the clinical EEG”, Electroencephalography andClinical Neurophysiology, 79, 440–447.

KüblerA., NijboerF., MellingerJ., VaughanT. M., PawelzikH., SchalkG., McFarlandD. J., BirbaumerN., and WolpawJ. R. (2005)“Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface,”*Neurology*, 64(10), 1775–1777.

LalorE., KellyS., FinucaneC., BurkeR., SmithR., Reilly R.and McDarbyG. (2005) " Steady-State VEP-Based Brain-Computer Interface Control in an Immersive 3D Gaming Environment", Journal on Applied Signal Processing, 3156–3164.

LeebR., KeinrathC., FriedmanD., GugerC., NeuperC., M. Garau, A. Antley, A. Steed, M. Slater and PfurtschellerG. "Walking from thoughts: Not the muscles are crucial, but the brain waves!", Paper Available: http://www.temple.edu/ispr/prev_ conferences/proceedings/2005/Leeb,%20Keinrath,%20Friedman,%20Guger,%20Neuper,%20Garau,%20et%20al.pdf, (Accessed: 15/06/2020).

LIBSVM - A Library for Support Vector Machines, http://www.csie.ntu.edu.t-w/~cjlin/libsvm/ , (Accessed: 15/06/2020).

Lopes da Silva and A. van Rotterdam. (1982) “Biophysical aspects of EEG and MEG generation”. In E. Niedermeyer and F. H. Lopes da Silva, editors, Electroencephalography, Urban & Schwarzenberg, München-Wien- Baltimore,15-26.

Lotte F. (2008) "Study of electroencephalographic signal processing and classification techniques towards the use of brain computer interfaces in virtual reality applications", PhD Thesis from the National Institute of Applied Sciences (INSA) Rennes.

Lotte F., Congedo M., L´ecuyer A., Lamarche F., and Arnaldi B. (2007) "A review of classification algorithms for EEG-based brain-computer interfaces". Journal of neural engineering, 4(2).

Magnetoencephalography, http://en.wikipedia.org/wiki/Magnetoencephalography, (Accessed: 15/06/2020).

Mahnaz A., Cuntai G., Kai K. A., and Chai Q.(2011) "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI", IEEE Transactions on Biomedical Engineering, 58(6), 1865-1872.

McFarland D.and WolpawJ. (2005) "Sensorimotor rhythm-based brain–computer interface (BCI): feature selection by regression improves performance", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13(3).

MedGadget - internet journal of emerging medical technologies http://medgadget.com/archives/2007/03/the_first_comme_1.html, (Accessed: 15/06/2020).

Misiti Y., Misiti M., Oppenheim G., MicrondeJM.P. (1995) “a Matlab Wavelet Toolbox for Signals and Images”, In: Antoniadis A., Oppenheim G. (eds) Wavelets and Statistics. Lecture Notes in Statistics, Springer, New York, NY, 103.

Moritz G., Martin B. (2008)“Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction Medicine”, Computer Science IEEE Transactions on Biomedical Engineering, 55, 8.

Muller J.G., PfurtschellerG., and FlyvbjergH. (1999)“Designing optimal spatial filtering of single trial EEG classification in a movement task”,*Clin. Neurophysiol.,* 110(5), 787–798.

Nguyen Th., Hettiarachchi I., Khatami A., Gordon-brown L., LIM C. P., and NahavandiS.(2018) "Classification of Multi-Class BCI Data by Common Spatial Pattern and Fuzzy System", IEEE, 6, 27873-27883.

Nicolas L. F. and Gomez-Gil J. (2012) "Brain computer interfaces", a review. Sensors, 12(2), 1211–1279.

Oana E., Anca L. (2017)” Feature Extraction and Classification Methods for a Motor Task Brain Computer Interface: A Comparative Evaluation for Two Databases”, International Journal of Advanced Computer Science and Applications, 8.

Oostenveld R.and PraamstraP.(2001) “The five percent electrode system for high-resolutionEEG and ERP measurements”, Clinical Neurophysiology, 112(4), 713–719.
Open-source toolbox BioSig, available at: http://biosig.source-forge.net/, (Accessed: 15/06/2020).

Orrison Jr., W. W., J. D. Lewine, J. A. Sanders, and HartshorneM. F. (1995) “Functions Brain Imaging”, St Louis: Mosby-Year Book, Inc.

PfurtschellerG."Brain-Computer Interface - State of the art and future prospects," Institute of Human-Computer Interfaces, Graz University of Technology, Paper. Available: http://www.eurasip.org/Proceedings/Eusipco/Eusipco2004/defevent/papers/cr1940.pdf, (Accessed: 15/06/2020).

Pfurtscheller G.and SilvaF. H. L. (1999) “Event-related EEG/MEG synchronization and desynchronization: Basic principles”, Clinical Neurophysiology, 110(11), 1842–1857.
Pfurtscheller G., Final results of BCI Competition IV. Available: http://www.bbci.de/competition/iv/results/#winners ,(Accessed: 15/06/2020).

Pfurtscheller G., Neuper C.(1994) "Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man", Neurosci Lett, 174, 93-96.

PfurtschellerG., NeuperC., FlotzingerD., and PregenzerM. (1997)“EEG based discrimination between imagination of right and left hand movement”*Electroencephalogh. Clin. Neurophysiol*., 103(6), 642–651.

PolikoffJ., BunnellH., BorkowskiW. (1995) "Toward a P300-based computer interface", In Proceedings of the RESNA 95 Annual Conference, Arlington Va.Precision Design Laboratory, http://www.mech.utah.edu/~bamberg/, (Accessed: 15/06/2020).

Ramoser H., Muller G. J., and Pfurtscheller G. (2000) "Optimal spatial filtering of single trial EEG during imagined hand movement". IEEE Transactions on Rehabilitation Engineering, 8(4):441–446.

RebsamenB., BurdetE., GuanC., ZhangH., TeoC., ZengQ., Ang M.and LaugierC. (2006) "A Brain-Controlled Wheelchair Based on P300 and Path Guidance," Biomedical Robotics and Biomechatronics, BioRob 2006. The First IEEE/RAS-EMBS International Conference, 1101 – 1106.

Sanei S., and Chambers J. A. (2007) "EEG Signal Processing". Wiley-Interscience.
SchalkG., McFarlandD., HinterbergerT., BirbaumerN., and WolpawJ.(2004) “BCI2000: a general-purpose brain-computer interface (BCI) system,” IEEE Transactions on Biomedical Engineering, 51(6), 1034–1043.

SchlA., HugginsJ.E., and MasonS.G.“Evaluation Criteria for BCI Research”, MIT Press Online Computer and Information Sciences Library Available at: http://www.citeulike.org/user/jmetzen/article/3211329and blink artifacts from EEG data using blind component separation," Psychophysiology,(Accessed: 15/06/2020).

Siavash Sakhavi , Cuntai Guan, and Shuicheng Y.(2018) "Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks", IEEE Transactions on Neural Networks and Learning Systems, 1-12.

Smith S. W. (1997) “The scientist and engineer's guide to digital signalprocessing”. California Technical Publishing, San Diego, CA, USA.

Soroosh S., and MohammadiA.(2017) "Error Correction Output Codding Coupled with the CSP for Motor Imagery BCI Systems ", 25th European Signal Processing Conference (EUSIPCO), 2125-2129.

Tangermann M, Müller K-R, Aertsen A, Birbaumer N, Braun C, Brunner C, Leeb R, Mehring C, Miller KJ, Müller-Putz GR, Nolte G, Pfurtscheller G, Preissl H, Schalk G, Schlögl A, Vidaurre C, Waldert S and Blankertz B (2012) “Review of the BCI competition IV”. Front. Neurosci., 6(55).

ThanT. A. (2014) "Multivariate Features for Multi-class Brain Computer Interface Systems", PhD thesis for Information Sciences and Engineering of the University of Canberra.

The "10-20 System" of Electrode Placement. http://faculty.washington.edu/chudler/1020.html, (Accessed: 15/06/2020).

TilleryS., Taylor D.and SchwartzA. (2003) "The general utility of a neuroprosthetic device under direct cortical control," Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE, 3, 2043 – 2046.

VargheseJ. P. (2009) “Analysis of EEG Signals For EEG-based Brain-Computer Interface”, School of Innovation, Design and Technology, Malardalen University, Vasteras, Sweden.

Vigrio R. N. (1997) “Extraction of ocular artefacts from fEEGg usingindependent component analysis”. Electroencephalography and Clinical Neurophysiology, 103(3), 395 - 404.

Vigrio, R., Srel, J., Jousmki, V., Hmlinen, M., and Oja, E.(2000). “Independent component approach to the analysis of eeg and meg recordings”.IEEE Transactions on Biomedical Engineering, 47, 589-593.

W. Penny, S. Roberts, A. Curran and J. Stokes (2000) "EEG-based communication: a pattern recognition approach", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 8(2), 214–5.

Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., and Fox, N. A. (2004) "Automatic correction of ocular artifacts in the eeg: a comparison of regression-based and component-based methods". International Journal of Psychophysiology, 53(2), 105 – 119.

What is Magnetoencephalography (MEG)?, http://ilabs.washington.edu/what-magnetoencephalography-meg, (Accessed: 15/06/2020).

Wikipedia Webpage. Available: http://en.wikipedia.org/ , (Accessed: 15/06/2020).

Wolpaw J. and McFarlandD. (2004) "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans", Proceedings of the National Academy of Sciences (PNAS), 101(51), 17849–17854.

Wolpaw J. R., Birbaumer, McFarlandN. D.,Pfurtscheller G.,and Vaughan T. M.(2002) "Brain computer interfaces for communication and control", Clinical Neurophysiology, 113(6), 767–791.

Wu W., Gao X., Hong B., and Gao S.(2008) “Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)”, IEEE Trans. Biomed. Eng. 55, 1733–1743.

XuN., GaoX., HongB., MiaoX., GaoS., and YangF. (2004) " BCI Competition 2003—Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications", IEEE Transactions on Biomedical Engineering, 51(6), 1067 – 1072.

Yang C.Z., Keng K.A., WangC., Guan C.and ZhangH. (2009) “Multiclass Filter Bank Common Spatial Pattern for Four-Class Motor Imagery BCI’’, 31th Annual International Conference of the IEEE EMBS, 571-574.

ZhangH., ChinZ. Y., AngK. K., GuanC., WangC.(2011) "Optimum spatio-spectral filtering network for brain-computer interface", *IEEE Transaction Neural Networks,* 22(1), 52-63.

Appendix (A)

Appendix (A)

Equations and Implementation Issues

A1. Second Order Blind Identification (SOBI):

The standard mixing model:

$$x=Hs\ (k)\text{(A.1)}$$

Where$(k)$is mixed signals and$H$is mixing matrix.

Before setting time delayed covariance matrices of mixed signals, formulating the robust orthogonalization must be done first as:

$$\bar{x}(k) = Q\ x\ (k)\text{(A.2)}$$

By using several time lags, up to 100 number of time lags, the time delayed covariance matrices of mixed signal for preselected time delays$(p_1, p_2,\ldots\ldots,)$are defined as:

$$R_{\bar{x}}(p_i) = \frac{1}{N}\sum_{k=1}^{N}\bar{x}(k)\bar{x}^T\ (k-p_i) = QR_{\bar{x}}(p_i)Q^T\text{(A.3)}$$

And then, the orthogonalized mixing matrix$A=QH,$ perform Joint Aproximation Diagonalization (JAD):

$$R_{\bar{x}}(p_i) = QR_x(p_i)Q^T = AR_S(p_i)A^T = UD_iU^T\text{(A.4)}$$

For$i=(1,2,3,\ldots.L,\ )$JAD reduces the probability of un-identifiability of a mixing matrix caused by an unfortunate choice of time delay$p$. Then orthogonal mixing matrix can be estimated as $\hat{A}=Q\hat{H}=U$and diagonal matrix(D_i). Finally, the estimated of source signals as (Cichocki A. et al., 2002):

$$\hat{S}\,(k) =(k)\quad\text{(A.5)}$$

And the mixing matrix as:

$$\hat{H}=Q+U \quad (A.6)$$

Second Order Blind Identification (SOBI) by joint diagonalization of correlation matrices. This code assumes temporally correlated signals, and uses correlations across times in performing the signal separation. Thus, estimated time delayed covariance matrices must be nonsingular for at least some time delays.

```
% Usage:
%       >> winv = sobi(data);
%       >> [winv,act] = sobi(data,n,p);
% Inputs:
%   data - data matrix of size [m,N] ELSE of size [m,N,t] where
%              m is the number of sensors,
%              N is the  number of samples,
%              t is the  number of trials (here, correlations avoid epoch boundaries)
%     n - number of sources {Default: n=m}
%     p - number of correlation matrices to be diagonalized {Default: min(100, N/3)}
%        Note that for noisy data, the authors strongly recommend using at least 100
%        time delays.
%
% Outputs:
%   winv - Matrix of size [m,n], an estimate of the *mixing* matrix. Its
%        columns are the component scalp maps. NOTE: This is the inverse
%        of the usual ICA unmixing weight matrix. Sphering (pre-whitening),
%        used in the algorithm, is incorporated into winv. i.e.,
%           >> icaweights = pinv(winv); icasphere = eye(m);
%   act  - matrix of dimension [n,N] an estimate of the source activities
%           >> data          = winv          * act;
%              [size m,N]       [size m,n]       [size n,N]
%           >> act = pinv(winv) * data;
%
% Authors:  A. Belouchrani and A. Cichocki (papers listed in function source)
%
% Note: Adapted by Arnaud Delorme and Scott Makeig to process data epochs

% REFERENCES:
% A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, ``Second-order
%  blind separation of temporally correlated sources," in Proc. Int. Conf. on
%  Digital Sig. Proc., (Cyprus), pp. 346--351, 1993.
%
%  A. Belouchrani and K. Abed-Meraim, ``Separation aveugle au second ordre de
%  sources correlees," in  Proc. Gretsi, (Juan-les-pins),
%  pp. 309--312, 1993.
%
%  A. Belouchrani, and A. Cichocki,
%  Robust whitening procedure in blind source separation context,
%  Electronics Letters, Vol. 36, No. 24, 2000, pp. 2050-2053.
%
%  A. Cichocki and S. Amari,
%  Adaptive Blind Signal and Image Processing, Wiley,  2003.

function [H,S,D]=sobi(X,n,p),

[m,N,ntrials]=size(X);
if nargin<1 | nargin > 3
help sobi
```

```
elseif nargin==1,

 n=m; % Source detection (hum...)
 p=min(100,ceil(N/3)); % Number of time delayed correlation matrices to be diagonalized
% Authors note: For noisy data, use at least p=100 the time-delayed covariance matrices.
elseif nargin==2,
 p=min(100,ceil(N/3)); % Default number of correlation matrices to be diagonalized
% Use < 100 delays if necessary for short data epochs
end;

%
% Make the data zero mean
%
X(:,:)=X(:,:)-kron(mean(X(:,:)')',ones(1,N*ntrials));

%
% Pre-whiten the data based directly on SVD
%
[UU,S,VV]=svd(X(:,:)',0);
Q= pinv(S)*VV';
X(:,:)=Q*X(:,:);

% Alternate whitening code
% Rx=(X*X')/T;
% if m<n, % assumes white noise
%   [U,D]=eig(Rx);
%   [puiss,k]=sort(diag(D));
%   ibl= sqrt(puiss(n-m+1:n)-mean(puiss(1:n-m)));
%    bl = ones(m,1) ./ ibl ;
%   BL=diag(bl)*U(1:n,k(n-m+1:n))';
%   IBL=U(1:n,k(n-m+1:n))*diag(ibl);
% else    % assumes no noise
%    IBL=sqrtm(Rx);
%    Q=inv(IBL);
% end;
% X=Q*X;

%
% Estimate the correlation matrices
%
 k=1;
pm=p*m; % for convenience
for u=1:m:pm,
   k=k+1;
for t = 1:ntrials
if t == 1
        Rxp=X(:,k:N,t)*X(:,1:N-k+1,t)'/(N-k+1)/ntrials;
else
        Rxp=Rxp+X(:,k:N,t)*X(:,1:N-k+1,t)'/(N-k+1)/ntrials;
end;
end;
M(:,u:u+m-1)=norm(Rxp,'fro')*Rxp;  % Frobenius norm =
end;                                % sqrt(sum(diag(Rxp'*Rxp)))

%
% Perform joint diagonalization
%
epsil=1/sqrt(N)/100;
encore=1;
V=eye(m);
while encore,
encore=0;
for p=1:m-1,
for q=p+1:m,
   % Perform Givens rotation
   g=[   M(p,p:m:pm)-M(q,q:m:pm)  ;
         M(p,q:m:pm)+M(q,p:m:pm)  ;
      i*(M(q,p:m:pm)-M(p,q:m:pm)) ];
      [vcp,D] = eig(real(g*g'));
```

```
        [la,K]=sort(diag(D));
angles=vcp(:,K(3));
angles=sign(angles(1))*angles;
   c=sqrt(0.5+angles(1)/2);
sr=0.5*(angles(2)-j*angles(3))/c;
sc=conj(sr);
oui = abs(sr)>epsil ;
encore=encore | oui ;
if oui , % Update the M and V matrices
colp=M(:,p:m:pm);
colq=M(:,q:m:pm);
M(:,p:m:pm)=c*colp+sr*colq;
M(:,q:m:pm)=c*colq-sc*colp;
rowp=M(p,:);
rowq=M(q,:);
M(p,:)=c*rowp+sc*rowq;
M(q,:)=c*rowq-sr*rowp;
temp=V(:,p);
    V(:,p)=c*V(:,p)+sr*V(:,q);
V(:,q)=c*V(:,q)-sc*temp;
end%% if
end%% q loop
end%% p loop
end%% while

%
% Estimate the mixing matrix
%
H = pinv(Q)*V;

%
% Estimate the source activities
%
if nargout>1
  S=V'*X(:,:); % estimated source activities
end
```

A2. FastICA - Hyvärinen’s Fixed Point Algorithm:

Hyvärinen’s algorithm is often used in ‘real time’ applicationsbecause of the possible parallel implementation. Thisalgorithm converges quickly as it seeks for a component one byFastICA uses kurtosis for the independent componentsestimation. Whitening is usually performed on data beforethe execution of the algorithm.

Assume that we have collected a sample of the sphere (or prewhitened) random vector x,which is in case of blind source separation is a collection of linear mixture of independent source signals. The basic method of Fast ICA algorithm is as follows:

1. Take a random initial vector $w(0)$ and divide it by its norm. Let k = 1.

2. Let $w(k) \ = \ E\{Z[Z^T w(k-1)]\ ^3\} - \ 3w(k-1)$(A.7)

3. Divide $w(k)$ by its norm .

4. If $|\, w^T\ (k)w(k-1)|$ is not close enough to 1, let $k \ = \ k + 1$, and go back to step 2.Otherwise the algorithm is convergent and outputs (k) .

The final vector $w(k)$ given by the algorithm equals one of the columns of the (orthogonal) demixing matrix B. In case of blind source separation, this means that $w(k)$ separates one of the non-Gaussian source signals in the sense that $w(k)^T x(t), t \ = \ 1,2$, ... equals one of the source signals.

To estimate n independent components, run these algorithm n times. To ensure that we estimate each time a different independent component, we only need to add a simple orthogonalizing projection inside the loop. The column of the demixing matrix B is orthonormal because of the sphering. Thus we can estimate the independent components one by one by projecting the current solution $w(k)$ on the space orthogonal to the columns of the demixing matrix B previously found. Define the matrix B as the matrix whose columns are the previously found columns of B.

Then adding the projection operation in the beginning of step 3.

3. $W \ = \ W - B\bar{B}^T \times \ W$(A.8)

Divide $w(k)$ by its norm.

Also the initial random vector should be projected this way before starting the iterations. To prevent estimation error in$\bar{B}$from deteriorating the estimate $w(k)$, this projection step can be omitted after the first few iterations: once the solution $w(k)$ has entered the basin of attraction of one of the fixed points, it will stay there and converge to that fixed point.

In addition to the hierarchical (or sequential) orthogonalization described above, any other method of orthogonalizing the weight vectors could also be used. In some applications, a symmetric orthogonalization might be useful. This means that the fixed point step is first performed for all the n weight vectors, and then the matrix $W(k) = (w_1(k), \ldots \ldots w_n(k))$ of the weight vector is orthogonalized, e.g., using the well-known formula:

$$W(k) = W(k) \, (W(k)^T \, W(k))^{1/2} \text{(A.9)}$$

Where $(W(k)^T \, W(k))^{1/2}$ is obtained from the eigenvalue decomposition of $W(k)^T \, W(k) = EDE^T$ as $(W(k)^T \, W(k))^{1/2} = ED^{1/2}E^T$.

FASTICA(mixedsig) estimates the independent components from givenmultidimensional signals. Each row of matrix mixedsig is oneobserved signal. FASTICA uses Hyvarinen's fixed-point algorithm:

```
function [Out1, Out2, Out3] = fastica(mixedsig, varargin)
%FASTICA - Fast Independent Component Analysis
%
% FastICA for Matlab 7.x and 6.x
% Version 2.5, October 19 2005
% Copyright (c) Hugo Gنvert, Jarmo Hurri, Jaakko Sنrelن, and Aapo Hyvنrinen.
% [icasig] = FASTICA (mixedsig); the rows of icasig contain the
% estimated independent components.
%
% [icasig, A, W] = FASTICA (mixedsig); outputs the estimated separating
% matrix W and the corresponding mixing matrix A.
%
% [A, W] = FASTICA (mixedsig); gives only the estimated mixing matrix
% A and the separating matrix W.
%
% Some optional arguments induce other output formats, see below.
%
% A graphical user interface for FASTICA can be launched by the
% command FASTICAG
%
% FASTICA can be called with numerous optional arguments. Optional
% arguments are given in parameter pairs, so that first argument is
% the name of the parameter and the next argument is the value for
% that parameter. Optional parameter pairs can be given in any order.
%
% OPTIONAL PARAMETERS:
%
```

```
% Parameter name        Values and description
%
%=======================================================================
% --Basic parameters in fixed-point algorithm:
%
% 'approach'            (string) The decorrelation approach used. Can be
%                       symmetric ('symm'), i.e. estimate all the
%                       independent component in parallel, or
%                       deflation ('defl'), i.e. estimate independent
%                       component one-by-one like in projection pursuit.
%                       Default is 'defl'.
%
% 'numOfIC'             (integer) Number of independent components to
%                       be estimated. Default equals the dimension of data.
%
%=======================================================================
% --Choosing the nonlinearity:
%
% 'g'                   (string) Chooses the nonlinearity g used in
%                       the fixed-point algorithm. Possible values:
%
%                       Value of 'g':      Nonlinearity used:
%                       'pow3' (default)   g(u)=u^3
%                       'tanh'             g(u)=tanh(a1*u)
%                       'gauss             g(u)=u*exp(-a2*u^2/2)
%                       'skew'             g(u)=u^2
%
% 'finetune'            (string) Chooses the nonlinearity g used when
%                       fine-tuning. In addition to same values
%                       as for 'g', the possible value 'finetune' is:
%                       'off'              fine-tuning is disabled.
%
% 'a1'                  (number) Parameter a1 used when g='tanh'.
%                       Default is 1.
% 'a2'                  (number) Parameter a2 used when g='gaus'.
%                       Default is 1.
%
% 'mu'          (number) Step size. Default is 1.
%                       If the value of mu is other than 1, then the
%                       program will use the stabilized version of the
%                       algorithm (see also parameter 'stabilization').
%
%
% 'stabilization'       (string) Values 'on' or 'off'. Default 'off'.
%                       This parameter controls wether the program uses
%                       the stabilized version of the algorithm or
%                       not. If the stabilization is on, then the value
%                       of mu can momentarily be halved if the program
%                       senses that the algorithm is stuck between two
%                       points (this is called a stroke). Also if there
%                       is no convergence before half of the maximum
%                       number of iterations has been reached then mu
%                       will be halved for the rest of the rounds.
%
%=======================================================================
% --Controlling convergence:
%
% 'epsilon'             (number) Stopping criterion. Default is 0.0001.
%
% 'maxNumIterations'    (integer) Maximum number of iterations.
%                       Default is 1000.
%
% 'maxFinetune'         (integer) Maximum number of iterations in
%                       fine-tuning. Default 100.
%
% 'sampleSize'          (number) [0 - 1] Percentage of samples used in
%                       one iteration. Samples are chosen in random.
%                       Default is 1 (all samples).
%
```

```
% 'initGuess'         (matrix) Initial guess for A. Default is random.
%                     You can now do a "one more" like this:
%                     [ica, A, W] = fastica(mix, 'numOfIC',3);
%                     [ica2, A2, W2] = fastica(mix, 'initGuess', A, 'numOfIC', 4);
%
%=======================================================================
% --Graphics and text output:
%
% 'verbose'           (string) Either 'on' or 'off'. Default is
%                     'on': report progress of algorithm in text format.
%
% 'displayMode'       (string) Plot running estimates of independent
%                     components: 'signals', 'basis', 'filters' or
%                     'off'. Default is 'off'.
%
% 'displayInterval'   Number of iterations between plots.
%                     Default is 1 (plot after every iteration).
%
%=======================================================================
% --Controlling reduction of dimension and whitening:
%
% Reduction of dimension is controlled by 'firstEig' and 'lastEig', or
% alternatively by 'interactivePCA'.
%
% 'firstEig'          (integer) This and 'lastEig' specify the range for
%                     eigenvalues that are retained, 'firstEig' is
%                     the index of largest eigenvalue to be
%                     retained. Default is 1.
%
% 'lastEig'           (integer) This is the index of the last (smallest)
%                     eigenvalue to be retained. Default equals the
%                     dimension of data.
%
% 'interactivePCA'    (string) Either 'on' or 'off'. When set 'on', the
%                     eigenvalues are shown to the user and the
%                     range can be specified interactively. Default
%                     is 'off'. Can also be set to 'gui'. Then the user
%                     can use the same GUI that's in FASTICAG.
%
% If you already know the eigenvalue decomposition of the covariance
% matrix, you can avoid computing it again by giving it with the
% following options:
%
% 'pcaE'              (matrix) Eigenvectors
% 'pcaD'              (matrix) Eigenvalues
%
% If you already know the whitened data, you can give it directly to
% the algorithm using the following options:
%
% 'whiteSig'          (matrix) Whitened signal
% 'whiteMat'          (matrix) Whitening matrix
% 'dewhiteMat'        (matrix) dewhitening matrix
%
% If values for all the 'whiteSig', 'whiteSig' and 'dewhiteMat' are
% supplied, they will be used in computing the ICA. PCA and whitening
% are not performed. Though 'mixedsig' is not used in the main
% algorithm it still must be entered - some values are still
% calculated from it.
%
% Performing preprocessing only is possible by the option:
%
% 'only'              (string) Compute only PCA i.e. reduction of
%                     dimension ('pca') or only PCA plus whitening
%                     ('white'). Default is 'all': do ICA estimation
%                     as well.  This option changes the output
%                     format accordingly. For example:
%
%                     [whitesig, WM, DWM] = FASTICA(mixedsig,
%
```

```
    'only', 'white')
%             returns the whiteed signals, the whitening matrix
%             (WM) and the dewhitening matrix (DWM). (See also
%             WHITENV.) In FastICA the whitening matrix performs
%             whitening and the reduction of dimension. Dewhitening
%             matrix is the pseudoinverse of whitening matrix.
%
%             [E, D] = FASTICA(mixedsig, 'only', 'pca')
%             returns the eigenvector (E) and diagonal
%             eigenvalue (D) matrices  containing the
%             selected subspaces.
%
%===========================================================================
% EXAMPLES
%
%     [icasig] = FASTICA (mixedsig, 'approach', 'symm', 'g', 'tanh');
%           Do ICA with tanh nonlinearity and in parallel (like
%           maximum likelihood estimation for supergaussian data).
%
%     [icasig] = FASTICA (mixedsig, 'lastEig', 10, 'numOfIC', 3);
%           Reduce dimension to 10, and estimate only 3
%           independent components.
%
%     [icasig] = FASTICA (mixedsig, 'verbose', 'off', 'displayMode', 'off');
%           Don't output convergence reports and don't plot
%           independent components.
%
%
% A graphical user interface for FASTICA can be launched by the
% command FASTICAG
%
%   See also FASTICAG

% @(#)$Id: fastica.m,v 1.14 2005/10/19 13:05:34 jarmo Exp $

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%
% Check some basic requirements of the data
if nargin == 0,
error ('You must supply the mixed data as input argument.');
end

if length (size (mixedsig)) > 2,
error ('Input data can not have more than two dimensions.');
end

if any (any (isnan (mixedsig))),
error ('Input data contains NaN''s.');
end

if ~isa (mixedsig, 'double')
fprintf ('Warning: converting input data into regular (double) precision.\n');
mixedsig = double (mixedsig);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%
% Remove the mean and check the data

[mixedsig, mixedmean] = remmean(mixedsig);

[Dim, NumOfSampl] = size(mixedsig);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%
% Default values for optional parameters

% All
verbose         = 'on';
```

```
% Default values for 'pcamat' parameters
firstEig          = 1;
lastEig           = Dim;
interactivePCA    = 'off';

% Default values for 'fpica' parameters
approach          = 'defl';
numOfIC           = Dim;
g                 = 'pow3';
finetune          = 'off';
a1                = 1;
a2                = 1;
myy               = 1;
stabilization     = 'off';
epsilon           = 0.0001;
maxNumIterations  = 1000;
maxFinetune       = 5;
initState         = 'rand';
guess             = 0;
sampleSize        = 1;
displayMode       = 'off';
displayInterval   = 1;


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%
% Parameters for fastICA - i.e. this file

b_verbose = 1;
jumpPCA = 0;
jumpWhitening = 0;
only = 3;
userNumOfIC = 0;


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%
% Read the optional parameters

if (rem(length(varargin),2)==1)
error('Optional parameters should always go by pairs');
else
for i=1:2:(length(varargin)-1)
if ~ischar (varargin{i}),
error (['Unknown type of optional parameter name (parameter' ...
        ' names must be strings).']);
end
   % change the value of parameter
switch lower (varargin{i})
case 'stabilization'
stabilization = lower (varargin{i+1});
case 'maxfinetune'
maxFinetune = varargin{i+1};
case 'samplesize'
sampleSize = varargin{i+1};
case 'verbose'
verbose = lower (varargin{i+1});
     % silence this program also
if strcmp (verbose, 'off'), b_verbose = 0; end
case 'firsteig'
firstEig = varargin{i+1};
case 'lasteig'
lastEig = varargin{i+1};
case 'interactivepca'
interactivePCA = lower (varargin{i+1});
case 'approach'
approach = lower (varargin{i+1});
case 'numofic'
```

```
numOfIC = varargin{i+1};

% User has supplied new value for numOfIC.
% We'll use this information later on...
userNumOfIC = 1;
case 'g'
      g = lower (varargin{i+1});
case 'finetune'
finetune = lower (varargin{i+1});
case 'a1'
      a1 = varargin{i+1};
case 'a2'
      a2 = varargin{i+1};
case {'mu', 'myy'}
myy = varargin{i+1};
case 'epsilon'
epsilon = varargin{i+1};
case 'maxnumiterations'
maxNumIterations = varargin{i+1};
case 'initguess'
% no use setting 'guess' if the 'initState' is not set
initState = 'guess';
guess = varargin{i+1};
case 'displaymode'
displayMode = lower (varargin{i+1});
case 'displayinterval'
displayInterval = varargin{i+1};
case 'pcae'
      % calculate if there are enought parameters to skip PCA
jumpPCA = jumpPCA + 1;
      E = varargin{i+1};
case 'pcad'
      % calculate if there are enought parameters to skip PCA
jumpPCA = jumpPCA + 1;
      D = varargin{i+1};
case 'whitesig'
      % calculate if there are enought parameters to skip PCA and whitening
jumpWhitening = jumpWhitening + 1;
whitesig = varargin{i+1};
case 'whitemat'
      % calculate if there are enought parameters to skip PCA and whitening
jumpWhitening = jumpWhitening + 1;
whiteningMatrix = varargin{i+1};
case 'dewhitemat'
      % calculate if there are enought parameters to skip PCA and whitening
jumpWhitening = jumpWhitening + 1;
dewhiteningMatrix = varargin{i+1};
case 'only'
% if the user only wants to calculate PCA or...
switch lower (varargin{i+1})
case 'pca'
only = 1;
case 'white'
only = 2;
case 'all'
only = 3;
end

otherwise
      % Hmmm, something wrong with the parameter string
error(['Unrecognized parameter: ''' varargin{i} '''']);
end;
end;
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%
% print information about data
if b_verbose
```

```
fprintf('Number of signals: %d\n', Dim);

fprintf('Number of samples: %d\n', NumOfSampl);
end

% Check if the data has been entered the wrong way,
% but warn only... it may be on purpose

if Dim > NumOfSampl
if b_verbose
fprintf('Warning: ');
fprintf('The signal matrix may be oriented in the wrong way.\n');
fprintf('In that case transpose the matrix.\n\n');
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%
% Calculating PCA

% We need the results of PCA for whitening, but if we don't
% need to do whitening... then we dont need PCA...
if jumpWhitening == 3
if b_verbose,
fprintf ('Whitened signal and corresponding matrises supplied.\n');
fprintf ('PCA calculations not needed.\n');
end;
else

% OK, so first we need to calculate PCA
  % Check to see if we already have the PCA data
if jumpPCA == 2,
if b_verbose,
fprintf ('Values for PCA calculations supplied.\n');
fprintf ('PCA calculations not needed.\n');
end;
else
% display notice if the user entered one, but not both, of E and D.
if (jumpPCA > 0) & (b_verbose),
fprintf ('You must suply all of these in order to jump PCA:\n');
fprintf ('"pcaE", "pcaD".\n');
end;

   % Calculate PCA
   [E, D]=pcamat(mixedsig, firstEig, lastEig, interactivePCA, verbose);
end
end
% skip the rest if user only wanted PCA
if only > 1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%
% Whitening the data

% Check to see if the whitening is needed...
if jumpWhitening == 3,
if b_verbose,
fprintf ('Whitening not needed.\n');
end;
else

   % Whitening is needed
% display notice if the user entered some of the whitening info, but not all.
if (jumpWhitening > 0) & (b_verbose),
fprintf ('You must suply all of these in order to jump whitening:\n');
fprintf ('"whiteSig", "whiteMat", "dewhiteMat".\n');
end;

   % Calculate the whitening
   [whitesig, whiteningMatrix, dewhiteningMatrix] = whitenv ...
                         (mixedsig, E, D, verbose);
  End


end % if only > 1

% skip the rest if user only wanted PCA and whitening
if only > 2


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%
% Calculating the ICA

  % Check some parameters
% The dimension of the data may have been reduced during PCA calculations.
  % The original dimension is calculated from the data by default, and the
% number of IC is by default set to equal that dimension.

  Dim = size(whitesig, 1);

% The number of IC's must be less or equal to the dimension of data
if numOfIC > Dim
numOfIC = Dim;
    % Show warning only if verbose = 'on' and user supplied a value for 'numOfIC'
if (b_verbose & userNumOfIC)
fprintf('Warning: estimating only %d independent components\n', numOfIC);
fprintf('(Can''t estimate more independent components than dimension of data)\n');
end
end

% Calculate the ICA with fixed point algorithm.
  [A, W] = fpica (whitesig,  whiteningMatrix, dewhiteningMatrix, approach, ...
numOfIC, g, finetune, a1, a2, myy, stabilization, epsilon, ...
maxNumIterations, maxFinetune, initState, guess, sampleSize, ...
displayMode, displayInterval, verbose);

  % Check for valid return
if ~isempty(W)
% Add the mean back in.
if b_verbose
fprintf('Adding the mean back to the data.\n');
end
icasig = W * mixedsig + (W * mixedmean) * ones(1, NumOfSampl);
    %icasig = W * mixedsig;
if b_verbose & ...
      (max(abs(W * mixedmean)) > 1e-9) & ...
      (strcmp(displayMode,'signals') | strcmp(displayMode,'on'))
fprintf('Note that the plots don''t have the mean added.\n');
end
else
icasig = [];
end

end % if only > 2

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%
% The output depends on the number of output parameters
% and the 'only' parameter.

if only == 1    % only PCA
  Out1 = E;
  Out2 = D;
elseif only == 2  % only PCA & whitening
if nargout == 2
    Out1 = whiteningMatrix;
    Out2 = dewhiteningMatrix;
else
    Out1 = whitesig;
```

```
   Out2 = whiteningMatrix;

   Out3 = dewhiteningMatrix;
end
else      % ICA
if nargout == 2
   Out1 = A;
   Out2 = W;
else
   Out1 = icasig;
   Out2 = A;
   Out3 = W;
end
end
```