PHASE SYNCHRONIZATION FOR CLASSIFICATION OF SPONTANEOUS EEG SIGNALS IN BRAIN-COMPUTER INTERFACES

THÈSE N^{O} 3397 (2005)

PRÉSENTÉE À LA FACULTÉ SCIENCES ET TECHNIQUES DE L'INGÉNIEUR

Institut de traitement des signaux

SECTION DE GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Elly GYSELS

ingénieur en physique, Université de Gent, Belgique
et de nationalité belge

acceptée sur proposition du jury:

Prof. M. Kunt, Dr P. Celka, directeurs de thèse
Prof. J.-D. Decotignie, rapporteur
Dr M. Le Van Quyen, rapporteur
Dr M. van Putten, rapporteur

Lausanne, EPFL
2006

Acknowledgments

I would like to thank everybody who directly or indirectly contributed to the successful completion of this PhD thesis.

Many thanks go to the directors of my thesis, Prof. Murat Kunt and Dr. Patrick Celka for the supervision of this work and feedback; and Dr. Philippe Renevey for the fruitful discussions and feedback. The members of the jury, Prof. Jean-Dominique Decotignie, Dr. Michel Le Van Quyen and Dr. Michel van Putten are gratefully acknowledged for the attentive and critical reading of the manuscript. I would like to thank Beat Fasel for his valuable feedback on the manuscript and for his help with the figures; and Le Song and Maria Knyazeva for the interesting discussions and collaboration. They were a great source of inspiration and motivation to me.

I would like to thank my colleagues at CSEM for the 3 pleasant years I spent in their company; especially to my office mate Olivier for the interesting discussions and for showing me the many possibilities of LaTeX; to Stephan, Josep, Patrick, Manu, Philippe, Mattia, Jens, Christophe, Victor, Olivier, Olivier and Rolf, the colleagues from the control and signal processing section; to Mario, head of the systems engineering division; to the enthusiastic always-in-good-mood secretaries Michèle, Doris and Raquel; and to the colleagues with whom I went jogging, swimming, playing basketball and beach volley: Anna, Evelyn, Myriam, Fanny, Nathalie, Aline, Véronique, Virginie, Stephan, Rolf, Antoine, Patrick, Jean-Luc, Jacques, Pascal, Alex, Philippe, Andreas, AJ and all others I forget to mention.

I heartily thank Beat, his love and support guided me through difficult moments; and my parents and family for their encouragement and for spoiling me with Belgian sweeties and drinks.
Finally, many hearty thanks go to my brother Jan, Lennie and my friends Tina, Liesbeth, Stijn and Agnieszka for the wise conversations and cheerful moments spent together; to Filip, Frederik, Tilly, Lieven, Tim, Frederika, Kristel, Erika, Peter, Pim, Dimitri, Heide, Heidy, René, Jacqueline, Tinu, Urs and Andreas for the shared laughs and discussions; and to Matteo and Joren for their sorrowless smiles.

This work was supported by the Swiss National Science Foundation through the National Center of Competence in Research on Interactive Multimodal Information Management (IM2).

Abstract

By directly analyzing brain activity, Brain-Computer Interfaces (BCIs) allow for communication that does not rely on any muscular control and therefore constitute a possible communication channel for the completely paralyzed. Typically, the user performs different mental tasks, that correspond to different output commands as recognized by the system. From the recorded brain signals (Electroencephalogram, EEG), features that characterize the mental tasks and allow their discrimination by a classifier have to be extracted.

This dissertation addresses the extraction of features in the framework of BCIs. On the one hand, new features are proposed. On the other hand, feature selection algorithms are investigated in order to select relevant features.

Currently existing BCIs mostly use power estimates in some pre-defined frequency bands, which are single-channel features. Some authors report on the use of multichannel features, but interactions between specific brain regions have not yet been studied. We propose to use the synchronization feature Phase Locking Value (PLV) for the classification of spontaneous EEG recorded during different mental tasks. It is fast to compute and can be applied to relatively short time windows, two important assets for BCI applications.
In a first instance, average synchronization values are considered. Tests on offline data show that significant classification accuracies can be obtained by the sole use of PLV. This demonstrates the relevance of synchronization features for the classification of EEG in this context.
We found that PLV and power features do not clearly outperform each other, but their combination often leads to significantly improved results and never significantly deteriorates the classification accuracies obtained by the separate subsets.

In the next step, feature selection algorithms are investigated in order to select the most interesting features. We show that Genetic Algorithms (GAs) as well as SVM-based recursive feature elimination (SVM-rfe) select physiologically meaningful features. As they are slow (computation times on the order of days and hours respectively) and thus cannot practically be used for BCIs, a modified version of the Fast Correlation-Based Filter (FCBF) is proposed. In this study, FCBF generalizes well and achieves good classification accuracies with very few features. The correspondence of the selected EEG signals with neurophysiological evidence is even stronger than for GAs and SVM-rfe. In addition, this algorithm is fast (computation time on the order of minutes) and so it can be applied between two recording sessions.

Comparing the classification results obtained with broadband and narrowband power and PLV features selected by FCBF learns that power features are preferably computed in the narrower 8-12Hz frequency band and that for PLV features, the 8-30Hz frequency band is the better one.
Furthermore, FCBF is used to evaluate a set of features comprising, on the one hand, the PLV for all possible electrode pairs, PLV averages, the cluster participation index and power features (band power and statistical mean frequency (SMF)), computed from the EEG signals. On the other hand, power (total power and SMF) and synchronization features derived from the empirical mode decomposition of these signals were included. They all proved useful. It is recommended to initially consider power features and only the PLV averages, because the resulting set of features is substantially smaller than when using PLV for all possible electrode pairs.

Automatic detection of artifacts and rejection of the corresponding data window for selecting the features and training the classifier, results in, if any, an increase of classification accuracy that is generally not

greater than 1%, for our data.
Application of global and local PCA-based denoising techniques never yielded improved results for two of the five subjects we analyzed. If it yields better results for the other 3 subjects, the increase never exceeds 1.84%.

Application of the proposed methods to blind data of the third international BCI competition held in 2005 gave significant classification accuracies. This further illustrates the relevance and the potential of the investigated techniques.

We conclude that PLV may complement currently used features and improve future BCI systems. PLV is well suited for BCI applications, because of its fast computation, needed for online feedback systems. The Fast Correlation-Based Filter is a valuable tool for evaluating features and selecting a subset of relevant features.

Version Abrégée

Les Interfaces Cerveau-Ordinateur (BCIs) analysent directement l'activité du cerveau et permettent ainsi une communication qui ne se fonde sur aucun contrôle musculaire. Cela constitue donc une voie de communication possible pour les tetraplégiques. La plupart du temps, l'utilisateur exécute différentes tâches mentales correspondant à différentes commandes reconnues par le système. Les spécificités qui caractérisent les tâches mentales et qui permettent leur distinction par un classificateur doivent être extraites des signaux cérébraux (électroencéphalogramme, EEG).

Cette thèse traite de l'extraction des caractéristiques dans le cadre de BCIs. D'une part, des nouvelles caractéristiques sont proposées. D'autre part, des algorithmes de sélection de caractéristiques sont étudiés afin de choisir les caractéristiques appropriées.

La plupart des BCIs actuels utilisent la puissance dans des bandes de fréquence prédéfinies. Ce sont des caractéristiques à canal unique. Quelques auteurs proposent l'utilisation de caractéristiques multicanaux, mais les interactions entre des régions spécifiques du cerveau n'ont pas encore été étudiées. Nous proposons d'utiliser la caractéristique de synchronisation PLV (Phase Locking Value) pour la classification d'EEG spontané enregistré pendant différentes tâches mentales. Il est rapide à calculer et peut être appliqué à des fenêtres d'analyse relativement courtes. Deux atouts importants pour des applications BCI. Dans un premier temps, des valeurs moyennes de synchronisation sont considérées. Les essais sur des données analysées en temps différé montrent que des taux de classification significatifs peuvent être obtenus par l'utilisation unique de PLV. Ceci démontre la pertinence des caractéristiques de synchronisation pour la classification d'EEG dans ce contexte.
Nous avons constaté que PLV et les caractéristiques de puissance ne se surpassent pas clairement, mais leur combinaison mène souvent à des résultats significativement améliorés et ne détériore jamais de manière significative les taux de classification obtenus par les sous-ensembles séparés.

Dans une prochaine étape, des algorithmes de sélection de caratéristiques sont étudiés afin de choisir les caractéristiques les plus intéressantes. Nous montrons que des algorithmes génétiques (GAs) aussi bien que l'élimination récursive de caractéristiques basée sur "Support Vector Machines" (SVM-rfe) sélectionnent des caractéristiques physiologiquement significatives. Ils ne peuvent pourtant pas être utilisés de manière pratique pour BCIs car ils sont lents (temps de calcul de l'ordre de quelques jours, respectivement quelques heures). Nous proposons une version modifiée du filtre à sélection de caratéristiques par corrélation rapide (FCBF). Dans cette étude, FCBF généralise bien et réalise des bons taux de classification avec très peu de caractéristiques. La correspondance des signaux EEG sélectionnés avec l'évidence neurophysiologique est encore plus forte que pour les GAs et le SVM-rfe. En outre, cet algorithme est rapide (temps de calcul de l'ordre de quelques minutes) et peut être ainsi appliqué entre deux sessions d'enregistrement.

En comparant les résultats de classification obtenus avec les caractéristiques de PLV et de puissance à bande large et à bande étroite sélectionnées par FCBF, on apprend que les caractéristiques de puissance sont de préférence calculées dans la bande de fréquence plus étroite 8-12Hz alors que pour des caractéristiques de PLV, la bande de fréquence 8-30Hz est la meilleure.
En outre, FCBF est employé pour évaluer un ensemble de caractéristiques contenant d'une part le PLV pour toutes les pairs d'électrodes possibles, les moyennes de PLV, le "Cluster Participation Index" et les caractéristiques de puissance (puissance de bande et fréquence moyenne statistique (SMF)) calculés des signaux EEG et d'autre part la puissance (puissance totale et SMF) et les caractéristiques de synchronisation dérivés de la décomposition en modes empiriques de ces signaux. Ils se sont tous avérés utiles. Nous

recommandons de considérer en premier lieu les caractéristiques de puissance avec les moyennes de PLV uniquement. En effet, l'ensemble de caractéristiques résultant est sensiblement plus petit en utilisant les moyennes qu'en utilisant PLV pour toutes les pairs d'électrodes possibles.

Emsemble, la détection automatique des artéfacts et le rejet de la fenêtre de données correspondante pour sélectionner les caractéristiques et entrainer le classificateur, résultent, le cas échéant, par une augmentation du taux de classification qui n'est généralement pas plus grand que 1% pour nos données. L'application des techniques de debruitage globales et locales basées sur PCA n'a jamais donné de résultats améliorés pour deux des cinq sujets que nous avons analysés. Si elle conduit à des meilleurs résultats pour les 3 autres sujets, l'augmentation n'excède jamais 1.84%.

L'application des méthodes proposées à des données aveugle remise lors de la troisième compétition internationale de BCI (2005) a produit des taux de classification significatifs. Ceci illustre davantage la pertinence et le potentiel des techniques étudiées.

Nous concluons que PLV peut compléter les caractéristiques actuellement utilisées et améliorer de futurs systèmes de BCI. PLV est bien approprié aux applications de BCI en raison de son calcul rapide, requis pour les systèmes d'analyse en temps réel, boucle fermée. Le FCBF est un outil valable pour évaluer des caractéristiques et sélectionner un sous-ensemble de caractéristiques appropriés.

Contents

List of Figures

List of Tables

Abbreviations

3D	3-dimensional
AAR	Adaptive Autoregressive
ALN	Adaptive Logic Network
ALS	Amyotrophic Lateral Sclerosis
AR	Autoregressive
BCI	Brain-Computer Interface
BP	Band Power
BRI	Brain Response Interface
CAT	Command Activation Term
Coh	Averaged coherence
Coh_{max}	Maximum coherence
CR	Classification Rate
CRT	Command Relaxation Term
CSP	Common Spatial Patterns
CT	Computed Tomography
CTFR	Correlative Time-Frequency
CV	Cross Validation
DA	Discriminant Analysis
DC	Direct Current
DNA	Deoxyribonucleic Acid
DSP	Digital Signal Processing
ECoG	Electrocorticogram
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EP	Evoked Potential
ER	Error Rate
ERD	Event-related Desynchronization
ERP	Event-related Potential
ERS	Event-related Synchronization
FB	Frequency Band
FCBF	Fast Correlation-Based Filter
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FS	Feature Selection
GA	Genetic Algorithm
HCI	Human-Computer Interface
HMM	Hidden Markov Models
IG	Iso-oriented Grating
ITR	Information Transfer Rate
'L'	Left (referring to left hand movement imagination)
LAR	Lagged Autoregressive
LDA	Linear Discriminant Analysis
LF-ASD	Low-Frequency Asynchronous Signal Detector

LPC	Linear Predictive Coding
LVQ	Learning Vector Quantization
MEG	Magnetoencephalogram
MRP	Movement Related Potentials
MVAR	Multivariate Autoregressive
NIR	Near Infrared
OBCI	Optical Brain-Computer Interface
OG	Orthogonal-oriented Grating
OPM	Outlier Processing Method
PC	Personal Computer
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PLV	Phase Locking Value
PSD	Power Spectral Density
'R'	Right (referring to right hand movement imagination)
RP	Readiness Potential
SCP	Slow Cortical Potential
SMF	Statistical Mean Frequency
SMR	Sensory Motor Rhythm
SPECT	Single Photon Emission Computed Tomography
SUS	Stochastic Universal Sampling
SVM	Support Vector Machine
SVM-rfe	SVM-based Recursive Feature Elimination
TFSC	Time-Frequency-Spatial Analysis of Correlations
TTD	Thought Translation Device
'U'	Unknown
UR	Unknown Rate
VCR	Videocassette Recorder
'W'	Word (referring to word generation)

Chapter 1

Introduction

1.1 Problem Statement

Brain-Computer Interfaces (BCIs) allow for communication and control that does not depend on the brain's normal output channels of peripheral nerves and muscles [156]. A BCI enables people to control their environment by generating specific states of the brain, which leave their signature in the electroencephalogram (EEG). By means of the EEG, a BCI transforms these thoughts or brain states, reflections of mental activity, into control signals. These control signals are associated to different output commands as recognized by the system. Users of BCIs [155] can actually control cursor movement [82], direct a wheelchair [38], select letters or icons on a computer screen [13], play video games, command a television or operate a neuroprosthesis [105]. The main interest of this research [147] is to develop a communication channel for persons with severe motor disabilities, being paralyzed or having lost control over their physical movements. These individuals include many with advanced amyotrophic lateral sclerosis (ALS), head trauma, brainstem stroke, severe cerebral palsy and spinal injuries. They lack accurate muscle control, but have intact brain capabilities. The practical use of BCIs will depend on the identification of different user groups, their needs and desires and the development of appropriate applications.

Reported Information Transfer Rates (ITRs), the speeds at which digital information can be transmitted from one device to another (here from the user to the application), of current BCIs are very low. They may be improved by studying cognitive tasks that generate more distinct EEG characteristics, working towards a better preprocessing and feature extraction and developing better classifiers. The IM2.BMI[1] project aims at addressing these problems. In the framework of this project, this thesis was carried out to study new feature extraction methods for BCIs. This includes both the exploration of new features and the study of feature selection techniques. The latter are necessary to find the most discriminative features amongst the ones studied and allow for personalization of the BCI.

Today's research on brain functioning and brain pathologies often assesses the synchronization between recorded brain signals to characterize interaction between different brain regions. It is believed that neurons communicate with each other by synchronous firing to transfer information amongst them. In the field of cognitive neuroscience, e.g., phase synchronization is studied to investigate, amongst others, memory, emotion, motor tasks and visual perception [35].

Where a certain degree of phase synchronization seems necessary for normal brain functioning, neurological diseases such as epilepsy manifest a pathological form of the synchronization process. For the anticipation of epileptic seizures, a research topic that received considerable attention last years, recently the degree of phase synchronization between signals recorded at different electrodes is evaluated, in order to conclude for possible pathological activity [101].

To the best of our knowledge, phase synchronization has not yet been applied to discriminate spontaneous EEG recorded during different mental tasks in the framework of BCIs. Therefore, we study whether synchronization features allow for automated classification of spontaneous EEG. Then, features

[1] Brain-Machine Interfaces project in the framework of the IM2 project (Interactive Multimodal Information Management), funded by the Swiss National Science Foundation

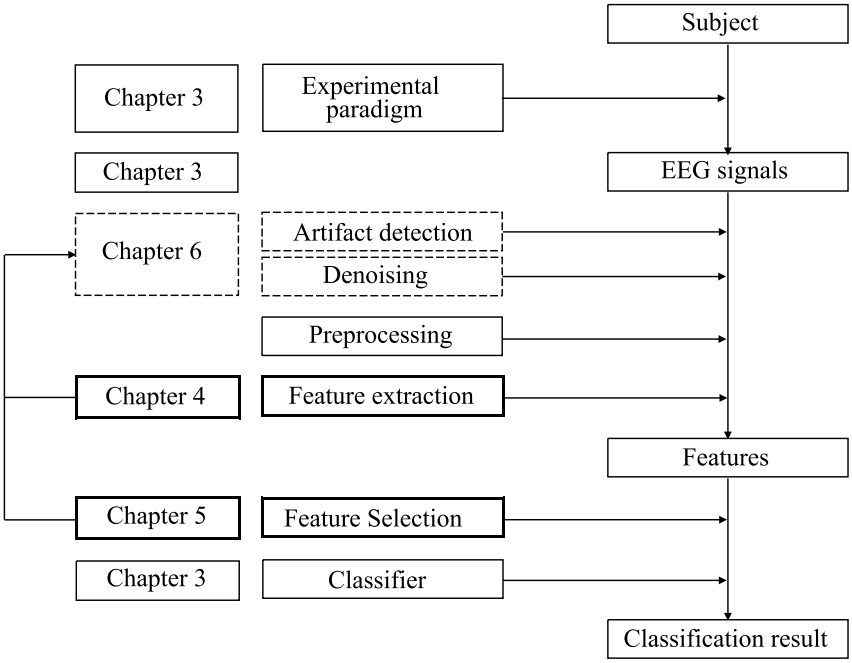

Figure 1.1: Structure of the dissertation

are evaluated with feature selection algorithms. Furthermore, we examine whether artifact removal and noise reduction algorithms allow for improvement of the classification accuracy.

1.2 Thesis Outline

This dissertation comprises 7 chapters. Chapter 2 introduces BCIs. It first explains the principle of BCI communication. Afterwards, the different types of BCIs are discussed. After reviewing BCIs, an overview of features currently used in BCIs is given. This chapter can be viewed as the state of the art.

The structure of the core of this dissertation is depicted in Figure 1.1. Chapter 3 describes the part of the employed BCI system not subject to investigation in this thesis. After an introduction to the EEG, which captures the brain activity on which BCI communication relies, we explain the EEG recording protocol, the choice of mental tasks and the classifiers used throughout this work. Feature extraction and feature selection, as well as investigation of artifact detection and denoising algorithms, are subject of this dissertation and will be treated in separate chapters (4, 5 and 6 respectively).

Chapter 4 deals with feature extraction. After motivating the idea to investigate synchronization features for the discrimination of spontaneous EEG in BCIs, we introduce the synchronization measures Phase Locking Value (PLV) and spectral coherence and discuss their properties. Finally, features derived from synchronization measures are presented, together with classification results obtained by using them to discriminate the mental tasks from the spontaneous EEG in BCIs.

Chapter 5 is dedicated to feature selection (FS), necessary to find discriminative features and to personalize the BCI. We first introduce the different types of feature selection methods and the techniques applied in BCI-research. Then, we present the FS techniques considered in this work: Genetic Algorithms (GAs), SVM(Support Vector Machine)-based recursive feature elimination (SVM-rfe) and the Fast Correlation-Based Filter (FCBF). Classification results obtained with these methods are presented. Next, we describe the application of the investigated features and feature selection algorithms to BCI competition data. Finally, we compare the classification results achieved with features selected from different sets of features with the same feature selection method. The types of features and electrode signals selected are shown.

Chapter 6 discusses artifact rejection and noise reduction techniques and studies their influence on the classification accuracy by employing features and feature selection techniques presented in Chapters 4 and 5 respectively. Independent Component Analysis (ICA) and global and local Principal Component Analysis (PCA) based methods are considered for denoising.

Chapter 7 concludes this dissertation with a summary, including the proposed strategy for analyzing EEG signals in the framework of BCIs derived from the actual results. This is followed by a brief discussion of possible issues to explore for BCIs and possible applications other than BCI of the proposed methods. The chapter is concluded with some remarks about ethical issues.

1.3 Original Contributions

These are the original contributions of this thesis:

- Demonstration of the usefulness of synchronization measures for the discrimination of spontaneous EEG recorded during different mental tasks [61, 59],
- Comparison of the classification performances of Phase Locking Value (PLV) and spectral coherence, two different synchronization measures [59],
- Comparison of the classification performances of synchronization measures and power spectral density (PSD) estimates [59],
- Demonstration that PLV computed from broadband filtered EEG signals is better for EEG-based discrimination of mental tasks than PLV computed from narrowband filtered EEG signals [66, 65],
- Computation of phase synchrony rates from the EEG for the recognition of imagined movements [138],
- Comparison of PLV and Common Spatial Patterns [138],
- Extraction of features derived from the Empirical Mode Decomposition (EMD) of the EEG signals for the classification of spontaneous EEG,
- Selection of synchronization and PSD features with different feature selection algorithms (Genetic Algorithms, SVM-based recursive feature elimination, Fast Correlation Based Filter (FCBF)) in the framework of BCIs [60, 66, 65],
- Proposal of a modified version of the FCBF algorithm to select features computed from noisy EEG [65],
- Application of an automatic artifact rejection algorithm to training data - artifact rejection did not result in a significant increase of the classification accuracy,
- Implementation of denoising algorithm based on local PCA that allows for overlapping subspaces,
- Application of denoising algorithms (based on ICA, global PCA and local PCA) to the EEG - denoising did not yield an improved discrimination of mental tasks.

Chapter 2

Brain-Computer Interfaces - State of the Art

2.1 Introduction

Brain-Computer Interfaces aim to offer means of communication to the motor disabled. The difference between BCI techniques and the more common Human-Computer Interfaces (HCI) lies in not relying on any sort of muscular response, but only on detectable signals representing responsive or intentional brain activity.
A BCI records the *brain activity* and extracts *features* that reflect the 'intentions' of the user and allow for discrimination by the *classifier*. Thus obtained, the control signal is then used to command the *external device*.
A schematic of a BCI is shown in figure 2.1 (Picture obtained from author, appeared in [155]). The BCI comprises a user, a signal capturing system, a signal processing system, a classifier and a feedback and device control unit. The signal capturing unit consists of electrodes, implanted in the brain or applied on the scalp, and a signal amplifier. After some preprocessing, features are extracted from the acquired data. These features are passed to the classifier, possibly after feature selection stage. Finally, recognition of a certain brain pattern is associated to a predefined output command.

There are 3 levels of adaptation between the user and the BCI:

- The BCI will adapt to the user's signal characteristics
- Periodic adjustments of the BCI are necessary because of the user's varying mental states (tiredness, being bored, more experience with the performance of the mental tasks, ...)
- Through the feedback the brain activity will also modify in order to produce EEG patterns that are better distinguished by the classifier

In Section 2.2, we describe different types of Brain-Computer Interfaces. Section 2.3 reviews current BCIs. As feature extraction is subject of this dissertation, Section 2.4 gives an overview of currently used features in BCIs.

2.2 Different Types of BCIs

2.2.1 Introduction

BCIs can be categorized according to different criteria. We distinguish output from input BCIs, indirect from direct BCIs, and synchronous from asynchronous BCIs. Furthermore, BCIs are distinguished according to the type of input signal.

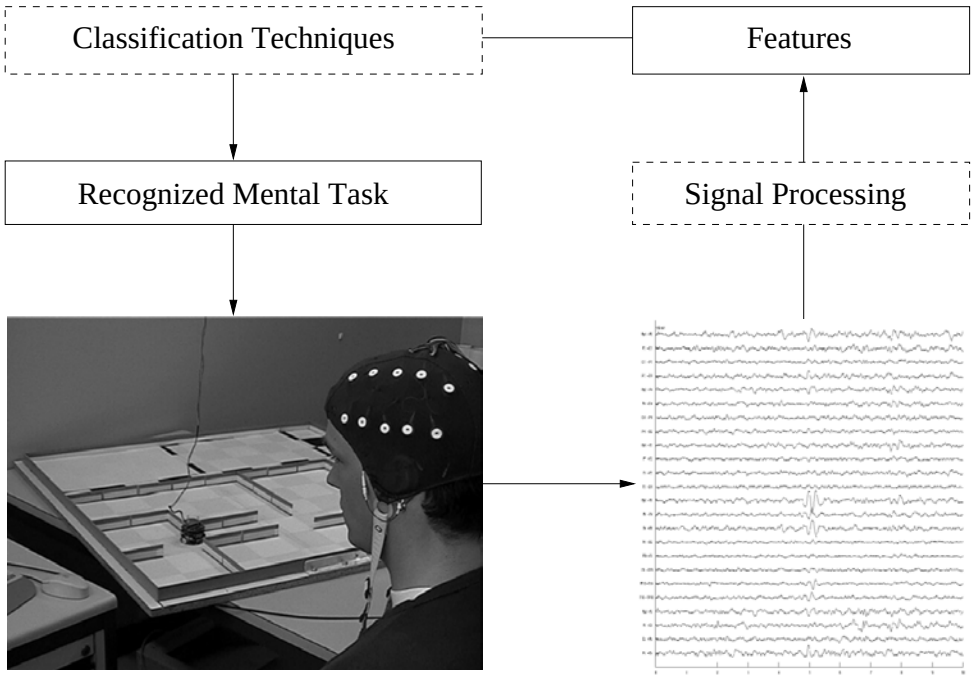

Figure 2.1: Example of a BCI. (Photo by J. del R. Millàn)

2.2.2 Input vs Output BCIs

The distinction between input and output BCIs corresponds to the direction of information flow, towards or from the brain.
Cortical input BCIs are applied in case of sensory loss [41]. E.g., visual cortex implants are designed to restore sight by direct stimulation of the visual cortex. Auditory cortex implants were developed to restore hearing in case when the hearing nerve is damaged.
A major goal of an output BCI is to provide a command signal from the cortex. This would constitute a communication interface for patients lacking intact somatic sensory pathways.
Generally, BCI refers to output BCIs. We use BCI in this sense throughout this manuscript.

2.2.3 Direct vs Indirect BCIs

Direct BCIs use neural signals recorded from neurons within the cortex, whereas indirect BCIs use neural signals from outside the cortex [41].
Currently developed BCIs use electroencephalographic activity recorded at the scalp [8, 17, 117, 157] or single-unit activity of many individual cortical neurons [88, 153, 105, 106], especially those that code for movement or its intent, recorded from implanted electrodes. Recently, invasive BCIs, recording brain activity from the cortex, have been reported [92, 86].

2.2.4 Synchronous vs Asynchronous BCIs

Most BCIs follow a synchronous protocol [113, 13, 116], which generates predetermined control periods during which the user can respond. Synchronous protocols involve a fixed and repetitive scheme. A stimulus sequence usually comprises a warning tone, followed a few seconds later by a stimulus. Subjects are required to do the mental task only after hearing and/or seeing the stimulus.
There is some controversy about the exact definition of an asynchronous BCI. Millán and Scherer [37, 132] see it as a BCI system where the decisions are spontaneous and self-paced. In these protocols not only which mental task is being performed must be determined, but also when it is being performed. The advantage, however, is to be independent of external triggers and to have a potentially faster system as

the end of the cue does not have to be awaited. It is also a more 'natural' way of communication.
Other authors [14, 158, 21] define an asynchronous BCI as a system where not only the decisions are spontaneous and self-paced, but that also includes a preceding stage of detecting the asynchronous signal in order to switch on the BCI automatically.
We will study a surface EEG-based BCI system based on spontaneous and self-paced decisions (asynchronous according to Millàn and Scherer).

2.2.5 Different Types of Input Signals

Different brain mapping and imaging techniques exist: computed tomography (CT), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetoencephalography (MEG), near-infrared (NIR) spectroscopy, and electrocorticogram (ECoG) and surface EEG.
CT, PET and SPECT do not allow for rapid imaging of brain activity and expose the subject to radiation. Therefore, they are not suited for BCI applications. MEG is very costly and cannot be used outside the lab. FMRI, NIR spectroscopy, and intracranially recorded and surface EEG have been studied for Brain-Computer Interfacing. They are described in Paragraphs 2.2.6, 2.2.7, 2.2.8 and 2.2.9 respectively.

2.2.6 Functional magnetic resonance imaging

fMRI has been applied for BCI purposes only recently . In 2004, Weiskopf *et al* [152] described the principles of an fMRI-based BCI, which records the activity of the entire brain with a high spatial resolution. It performs data processing and feedback of the haemodynamic brain activity within 1.3s.

2.2.7 Near-infrared spectroscopy

Also in 2004, Coyle *et al* [31] presented a novel approach to Brain-Computer Interfacing that uses optical analysis to provide physiological measures of brain function (OBCI). They use near-infrared spectroscopy to detect the cerebral haemodynamic response resulting from an autoregulatory process responding to the metabolic demands of neuronal firings taking 5-8s. They distinguish an imagined motor task from the rest condition. Physiological noise such as cardiac cycle, respiratory effects and the Mayer wave are the greatest noise sources in their system. Another problem is the nature of the slow optical response.
The advantages of such a system are safety benefits - the user is isolated from electrical signals, NIR is non-ionizing and is therefore suitable for long-term use, and the technique is non-invasive. No gel is needed to make contact with the skin. The mental tasks required to operate the OBCI are straightforward and do not necessitate the learning of a new thought process.

2.2.8 Single Neuron Recordings and Electrocorticogram (ECoG)

Invasive BCIs used to be investigated on monkeys, but during the last years, several invasive ECoG-based human BCIs have been reported in literature.

Nicolelis and his co-workers experimented with invasive BCIs on rats and monkeys [88, 153, 105, 106]. In the experiments with monkeys, brain signals were picked up by arrays of fine microwires while the monkey was moving her arm. The electrical patterns were converted into instructions that direct a robotic arm. The robotic arm copied the movements of the monkey's arm. In later experiments, a monkey learned to manipulate a joystick to drive a cursor on the computer screen to hit a target. The BCI was trained to decipher the brain signals. In a later stage of the experiment, the monkey could move the cursor on the screen, by just thinking about the trajectory it should take. These results indicate a possibility for paralyzed people to control a limb prosthesis by their thoughts.

Taylor and colleagues [142] did experiments with monkeys where they had real-time visual feedback of their brain-controlled trajectories. Cell tuning properties changed. Daily practice improved movement accuracy and the directional tuning while making long sequences of 3D movements. Far fewer units than

expected were used. They showed that visual feedback combined with the tracking of changes in cortical tuning parameters improved the efficiency of cortical activity as a control signal for both fast and slow brain-controlled movements. They conclude that neural activity can be reorganized within minutes and can be used to achieve brain-controlled virtual movements with nearly the same accuracy, robustness and speed as normal arm movements.

Musallam *et al* [104], also working with monkeys, instead of decoding intended hand trajectories from motor cortical neurons and using this signal to control external devices, use higher level signals related to the goals of movements to position cursors on a computer screen without the monkeys emitting any behavior. They concluded that the goal signals can be used to operate computers, robots and vehicles.

The first one to report on a human BCI based on intracranially recorded EEG was Sutter [140], in 1992. To avoid muscle artefacts resulting from neck musculature and the problems related to placement and maintenance of scalp electrodes by relatively inexperienced nursing staff, he describes a small strip with 4 electrodes implanted in the epidural space in the head of a volunteering ALS patient. Together with a complete BRI (Brain Response Interface) communication system it permitted him to communicate with high quality synthetic speech, operate word processing software on the host computer, and control his television set and VCR (videocassette recorder). Towards the end of the evaluation, the subject reached communication rates of 10 to 12 words/minute.

Bakay and Kennedy at Emory university allowed a paralyzed man to control a computer using only his thoughts. This man received a tiny brain implant, consisting of 2 tiny hollow glass cones coated with neurotropic chemicals extracted from the recipients' peripheral nerves. These chemicals encourage nerves to grow into the cones, penetrating the glass. When thinking about movement, electrical activity in his brain increases and a signal is sent to the receiving unit in his scalp, which sends a message to control the cursor.
Kennedy *et al* [76] furthermore report on a system using conductive skull screws for more reliable access to cortical local field potentials without entering the brain itself. This system allowed an almost locked-in human subject with ALS to activate a switch using online time domain detection techniques.

Doctors of the University of Wisconsin-Madison implanted in June and November 2004 electrodes on the surfaces of the brains of two patients. The two patients were able to control a computer cursor and play a basic video game [3].

In November 2004, researchers at Brown University reported on an invasive BCI in a 25-year-old quadriplegic. After the electrodes were implanted in his brain, the man was able to read e-mail, play video games, turn on lights and change channels or adjust the volume on a TV [3].

Leuthardt and his colleagues [92] demonstrated that ECoG activity recorded from the surface of the brain in epileptic patients implanted with electrodes for presurgical evaluation can enable users to control a one-dimensional computer cursor rapidly and accurately. Success rates ranging between 74 and 100% were achieved in a closed-loop motor and speech imagery paradigm.

Lal and his colleagues [86] investigate BCIs based on ECoG as an alternative to surface EEG based BCIs, having a low signal-to-noise ratio. Intracranial EEG recordings were obtained from 3 epilepsy patients, implanted with electrode grids placed on the motor cortex, during a motor imagery experiment. Error rates range from 17.5 to 23.3%, which is higher than the best error rates obtained from EEG-based BCI, but only 1.5s data were used and very few training points were available.

2.2.9 Surface EEG

Most of the current BCIs, however, use the EEG signals recorded non-invasively from the scalp. Although the low spatial resolution, EEG has the necessary temporal resolution for a BCI, and could be imagined in a portable recording device.
Below we will briefly describe BCIs based on surface EEG. We distinguish spontaneous EEG, evoked potentials (EPs), event-related potentials (ERPs), readiness potentials (RPs) and slow cortical potentials

(SCPs). These different types of EEG responses define different types of BCIs. More details about surface EEG-based BCIs will be given in Section 2.3.

Generally, we can distinguish two approaches to the development of BCIs. A first approach uses the continuous EEG output, which the user can modify in some reliable way. The second is to evoke an EEG response with an external stimulus.

Spontaneous EEG

Spontaneous, free running EEG is the result of the rhythmic activities of neurons and is categorized according to frequency and location. We distinguish δ (1-3.5Hz), θ (4-7.5Hz), α (8-12.5Hz), β_1 (13-18Hz), β_2 (18.5-31.5Hz) and γ (30-40Hz) rhythms. E.g. moving one hand repetitively up and down, will suppress the α rhythm in the motor cortex of the contralateral hemisphere (called μ rhythm). This is true also for imagined hand movements. Many BCIs are based on the detection of this kind of rhythms, e.g. [40, 6, 103, 116, 56, 157, 81, 125, 141, 37, 52]. This is also the type of BCI that will be studied in this work.
Generally, these BCIs do not require long periods of training. Pfurtscheller and Neuper report on the participation of 324 volunteers in a BCI experiment based on a synchronous protocol and imagery left and right hand movement or imagery hand and foot movement [116]. After only about 10 minutes of training, the two brain states related to motor imagery were discriminated with an accuracy better than 80% in about 12% of the subjects. 78% of the subjects achieved classification accuracies between 60% and 80%. They are expected to further improve in additional training sessions. For 10% of the subjects no discrimination between brain states was achieved.

Readiness Potentials (RPs)

Readiness potentials, or Bereitschaftspotentials (BPs), are changes in potential reflecting a subject's readiness to move. It appears as a negative voltage deflection known as event-related synchronization/desynchronization (ERS/ERD), occurring half a second before the start of the movement. BCI systems based on RP do not require user training, but ERDs are harder to detect. The Berlin Brain-Computer Interface (BBCI) [82] is an example of such a BCI.

Slow Cortical Potentials (SCPs)

Slow cortical potentials reflect a slow change (lasting from a few hundred milliseconds up to several seconds) in potential shifts, which can be controlled by trained subjects. Scientists at the university of Tübingen have developed an electronic spelling device based on self-regulated SCP for a completely paralyzed person, called Thought Translation Device (TTD) [13]. The training period is of the order of weeks or months, depending on the subject.

Evoked Potentials (EPs)

Evoked potentials are reactions to visual or auditive stimuli in the sensory cortex. This technique can be used to train subjects to control the strength of their steady state visual EP with the use of biofeedback. These BCI systems [150, 140] have the advantage of being simple, but require an external stimulus and a long period of training.

Event-related Potentials (ERPs)

ERPs are transient signals which are characterized by a voltage deviation in the EEG and occur in response to or in advance of particular events. They are caused by external stimuli or cognitive processes triggered by external events [53].
Focusing on particular stimuli will induce changes in the EEG that can be detected by the BCI and allow to command the BCI. The P300 ERP, e.g., is a positive deflection occurring 300ms after an event to which the subject has been told to respond. The event must be one in a series of Bernouilli events and have a low probability of occurring. The amplitude of the P300 varies directly with the relevance of the eliciting events and inversely with the probability of the stimuli [46].
Many existing BCI systems use the P300 ERP for compilation of words and sentences [46, 98, 94, 15]. Subjects focus on the desired character in the matrix displayed on the screen. Each of the rows is

intensified, in random order, and subsequently the columns are intensified in a similar manner. When the row or column containing the desired character is intensified, a P300 is elicited.
ERP-based BCIs have the advantage to need little training for a new user to control the BCI. These protocols, however, are very slow as different stimuli have to be presented to the subject in a repetitive scheme.

2.3 Overview BCI Research

2.3.1 Introduction

In this section, we give a short historical introduction and present an overview of current surface-EEG based BCI research. As different research groups use different experimental paradigms, subjects, classification protocols etc., it is difficult compare the different BCI studies. Therefore, we rather present the ideas investigated.

We can distinguish off-line research from online research. In the former case types of EEG signals and processing methods are reported which could potentially be used in a BCI system. The latter reports on actual characteristics of existing BCI systems.

The foundation of current BCI work was formed in the late 60's, 70's and 80's. In recent years, BCI research has expanded significantly.

2.3.2 Early BCI Research

Early BCI research focused on the use of spontaneous EEG signals. Users were trained to control different aspects of the α rhythm to actuate various devices. In 1967, Dewan [40] used a biofeedback method to train subjects to modulate their occipital α rhythm in order to transfer Morse code messages. One subject increased the α activity by turning his closed eyes to the extreme upward position consistent with comfort and avoiding fixation or convergence and accommodation. For minimum activity he kept his eyes closed and fixated them on an imaginary point located almost immediately in front of his lids. Intervals of increased α activity were detected. On the basis of the pause duration, it was decided when a character had been completed and the Morse character was translated into its alphabetical equivalent. The minimum average time needed for each correct letter was about 35 seconds.

2.3.3 BCI Research in the Seventies

Vidal reports on brain-computer communication in 1973. In [150], he describes the detection and classification of individual evoked responses or single epochs. The experimental strategy needed to be altered. Stimulus parameters included flash intensity and color, background intensity and color and pattern shape. 4 electrodes were placed above the occipital-parietal area. Depending on which of the 4 fixation points had been in use at the time of each flash, one of 4 nonoverlapping parafoveal areas received the stimulus. Data processing included rejection of artifacts, Wiener filtering, transformation of the filtered amplitude vector, step-wise selection of the best samples, epoch class and recursive outlier rejection, real-time defaulting and decision rule updating.

2.3.4 BCI Research in the Eighties

In 1988, Farwell and Donchin [46] developed the first P300 ERP-based BCI system to operate a virtual typewriter. The letters of the alphabet are contained in a 6X6 letter matrix, visualized on a computer screen. The different rows and columns are highlighted subsequently. When the row and column in which the target letter is contained is highlighted, a P300 is elicited. This technique requires almost no training.

Meinicke and his colleagues [98] tried to improve the transfer rates obtained by Farwell and Donchin within a setup closer to an online realization than in the original studies. Improved transfer rates were

achieved using SVMs, a state-of-the-art machine learning technique, for signal classification and an augmented data space using more electrodes for the interface.

2.3.5 BCI Research in the Nineties

In 1990, Keirn and Aunon [75] identified cognitive tasks that could be differentiated from EEG signals. With spectral estimates and asymmetry coefficients, based on the Wiener-Khinchine method or on autoregressive (AR) parameters, and the AR coefficients themselves in combination with a Bayes quadratic classifier, the system could differentiate between the pairs of tasks taken from 'resting state', 'geometrical figure rotation', 'mathematical multiplication', 'mental letter composing' and 'visual counting'.
Results for features computed from 1/4-second windows were comparable to those obtained for 2-second windows. Furthermore, they found that no particular tasks stood out as being easy or difficult to classify, that the 5 subjects performed equally well and that there was no difference in classification accuracies between the opened and closed eye cases.

Sutter [140] discusses an experimental communication system for severely disabled persons that utilizes electrical responses from the brain. It determines the fixated target from weak electrical signals originating from the visual cortex of the brain and is called Brain Response Interface (BRI). The most important advantages of this approach are that it leaves the user with complete freedom of head movement and that it can work over a wide range of environmental conditions. It requires little or no calibration or adjustment during operation and does not interfere with the use of eyeglasses.

In 1992, Pfurtscheller and his colleagues at Graz University looked at μ rhythm (see Section A.2) ERDs associated with moving either the left or the right index finger. They analyzed the one-second window preceding the actual movement. This off-line experiment showed that two electrodes, C3 and C4, would be sufficient for two-dimensional cursor movement in a BCI system.
The on-line, EEG-based, Graz BCI I system, reported in 1993, detects readiness potentials from the electrodes C3 and C4. Power spectral values were computed from the second in which the planning of the movement took place. A learning vector quantizer learnt to associate planned movements with actual movements. It was demonstrated that the subject in some way modifies his ERDs by biofeedback.
Left and right index finger movement and toe and tongue movement could be differentiated by their ERDs, as demonstrated by Pfurtscheller in 1994. The use of γ ERS as well as μ ERD increased 4 way classification accuracy.

In 1996, the Graz group reported on the BCI II, discriminating between 3 brainstates. During the first 3 sessions, the subject executed right hand, left hand and right foot movement. In the next 1-2 sessions, these movements were imagined. Band power in the band 5-35Hz was estimated in intervals of 250ms from 3 bipolar EEG channels. 4 power estimates for each of the 3 channels were considered. The nonlinear classifier was trained with examples from the first sessions without feedback [116].
Furthermore, they report on the use of Hidden Markov models (HMMs) for online classification of EEG patterns related to left versus right hand imagery [116].

Müller-Gerking [103] devised spatial filters for multi-channel EEG that lead to signals which can discriminate optimally between two conditions. The method was used to classify single trial EEGs recorded during preparation for movements of left or right index finger or right foot. This method of Common Spatial Patterns (CSP) reflects the selective activation of cortical areas.

In experiments with delayed feedback, band power estimates were combined with learning vector quantization (LVQ). In experiments with continuous feedback, either the CSP or the AAR (Adaptive AutoRegressive) method was combined with the LDA (Linear Discriminant Analysis) approach. The average lowest on-line error for 3 subjects was smaller for the CSP method than for the AAR method, for the distinction of 2 classes [117, 56]. The CSP method requires a larger amount of electrodes than the other procedures, and shows some sensitivity to electrode montage.

A tetraplegic patient makes practical use of the Graz BCI system, consisting of a two-channel amplifier and a notebook, to open and close a hand orthosis by imagination of feet and right hand movement.

After 62 training sessions over a 5 months period, the patient can perform the task with an accuracy approaching 100%.

Wolpaw and his co-workers at the Wadsworth center studied motor imagery and the classification of sensorimotor EEG rhythms. An experiment reported in 1994 used two electrodes on the right and two on the left side of the sensorimotor cortex. These were used to generate a right and a left ERD signal. Power in a 5Hz wide frequency band, centered at 10Hz, was computed by an FFT every 200ms. These two power values were then converted into vertical and horizontal cursor movements by linear equations. Subjects directed the cursor from the center of a computer screen to one of four large targets at the corners of the computer screen.
They analyzed changes in μ and β rhythms associated to motor tasks [157]. An FFT (Fast Fourier Transform) implementation in a DSP (Digital Signal Processing) chip was used to calculate the band power centered at 9Hz. EEG was recorded bipolarly from the left sensorimotor area and the subject was instructed to move a cursor up or down, whereby upward movement required large and downward movement low μ rhythm amplitude. This amplitude was directly translated into cursor movement. The persons learned to change the amplitudes of the μ and β rhythms by imagining moving a hand or a body part up or down.
The scientists at the Wadsworth center try to realize more precise control over the mouse by homing in on the desired EEG frequencies and tuning the BCI to individual users. Band power is estimated from the AR spectrum (computed from short overlapping EEG segments) in 4-5Hz wide frequency bands. The cursor moves upward or downward as a linear function of the EEG magnitude in a specific frequency band. Again, this technique requires weeks or months of training to teach a person how to control their brain waves. Furthermore, subjects were selected on the basis of having a strong μ rhythm.

McMillan and Calhoun [1] used Steady State Visual Evoked Potentials (VEP) to control the roll of a flight simulator and also knee flexion.

Anderson *et al* analyzed changes in EEG rhythms associated with motor tasks and cognitive tasks. In [8], single-channel and multi-channel autoregressive techniques are compared. The MVAR method performs slightly better, distinguishing the baseline from the math task. A neural network was used for classification and applied to the data collected by Keirn and Aunon.
In [6], EEG signals recorded from one subject performing 3 mental tasks (baseline task, letter task and math task) were discriminated using an AR representation of the EEG and a multilayer neural network classifier. The output of the classifier was averaged over 10 consecutive half-second windows of data.
Furthermore, a transformation of the signals via the Karhunen-Loéve or maximum noise fraction transformations in combination with a quadratic discriminant analysis were investigated [7].

Kostov and Polak present in [80] a BCI with features based on Power Spectral Density (PSD) and the amplitude frequency distribution of recorded EEG signals in the sensory-motor rhythm (SMR). Thresholds were used to decide on real-time control actions. A human operator adjusted the feedback parameters to optimize the performance during the training stage. Off-line analysis consisted of classification of Linear Predictive Coding (LPC) and FFT features by an Adaptive Logic Network (ALN).
In [81], they trained users to control an animated cursor on the computer screen by voluntary EEG modulation. The BCI system required only 2 to 4 electrodes, and had a relatively short training time for both the user and the machine. Subjects moved the cursor in one or two dimensions to hit the targets. This BCI relies on extraction of AR features, which are subsequently classified by an ALN. They concluded that even only 4 AR coefficients can represent each voluntary modulated EEG-signal.

Roberts and Penny [124] and their colleagues study ERD and ERS in the framework of their BCI project. Instead of analyzing changes in absolute frequency content, they study temporal and spatial complexity measures. 8 subjects took part in the experiment and were asked to imagine opening and closing the fingers of their left and right hand respectively. 4 subjects were excluded because they presented insignificant control blocks.
In [125] they present an autoregressive modelling of a single EEG channel coupled with classification and temporal smoothing under a Bayesian paradigm. Uncertainty in decisions is used to reject uncertain samples. This was applied in two-way cursor movement experiments.

Sycacek *et al* [141] propose a probabilistic treatment of the signal processing part of a BCI. They obtain predictions that implicitly weight information according its certainty. Offline experiments reveal that this results in significantly higher bit rates. Probabilistic methods are also very useful to obtain adaptive learning algorithms that can cope with nonstationary problems.
4 mental tasks and a representation of the data by reflection coefficients were explored in [33]. It was found that the spatial navigation and auditory imagery tasks were significantly better discriminated than other pairs of tasks.

2.3.6 BCI Research from 2000 on

Birbaumer [13] reports on the Thought Translation Device (TTD), which trains locked-in patients to regulate slow cortical potentials of their EEG. Positive and negative slow waves are used to choose between two banks of letters. Once selected, a bank splits in two. This process is continued till the target letter is revealed. 3 ALS (Amyotrophic lateral sclerosis) patients could operate the device successfully. They needed about 2 minutes for the selection of one letter. The drawback of this method is that it requires a period of training of several months.

del R. Millán *et al* [37] do not analyze a particular EEG component, but analyze overall EEG signals at different scalp locations. The recognition of 3 different mental tasks corresponds to 3 output commands, e.g. associated to wheelchair control. Their approach is based on a protocol in which the subject is trained with the help of feedback to perform 3 different motor-related and cognitive tasks. Spectral features are computed from the EEG in sliding windows and fed into a local neural classifier. A rejection criterion is introduced to avoid risky decisions. Time of response is 0.5 seconds.
Renkens and del R. Millán [121, 38] present a brain-actuated control of a mobile platform. A portable non-invasive BCI makes it possible to continuously control a mobile in a house-like environment. To obtain control of the robot with just 3 mental tasks, the user's mental states are associated to high-level commands that the robot executes autonomously using information from its on-board sensors. After a few days of training, two human subjects successfully moved a robot between several rooms by mental control only.

Birch and Mason [14] utilize voluntary motor-related potentials recorded from the scalp for asynchronous control applications. They developed the outlier processing method (OPM) and the low-frequency asynchronous signal detector (LF-ASD) to recognize single-trial, voluntary motor-related potentials from scalp-recorded EEG signals (asynchronous signal detection task). The obtained error rates were significantly lower than with the OPM and a μ-rhythm power classification algorithm.

Müller and his co-workers [15] report on their BCI, detecting in a pseudo-online simulation upcoming finger movements in a natural keyboard typing condition and predicting their laterality. This can be done on average 100-230ms before the respective key is actually pressed, i.e., long before the onset of EMG. The approach has the advantages of having a short response time, a high classification accuracy and requiring almost no human training.
In [43] they present experiments done with imagined left and right hand finger movements. Features based on Movement Related Potentials (MRPs) were combined with features based on ERDs, assumed to be independent.
[44] presents a study on the implications of using more classes and concludes that it is practically not useful to employ more than three or four classes.
The Berlin Brain-Computer Interface (BBCI) is introduced in [82]. Using multi-channel EEG recordings, single-trial differential potential distributions of the BP preceding voluntary finger movements over the corresponding primary motor cortex are analyzed. Their neurophysiological approach aims to capture EEG indices of preparation for an immediately upcoming motor action. Possible transfer to BCI control by paralyzed patients appears worthwhile to be studied further because these patients were shown to retain the capability to generate BPs with partially modified scalp topographies. EEG-signals were multiplied by a one-sided cosine function in order to emphasize the late signal content and were filtered subsequently. Fisher's Discriminant was applied for classification. Furthermore a Command Activation Term (CAT) and a Command Relaxation Term (CRT) were introduced. This technology was applied in a Feedback 'Brain-Pacman'. A variant of BBCI designed to achieve fast classifications in normally

behaving subjects is presented in [16].
Dornhege *et al* [42] propose a method that offers a substantial speed-up for classification of SCP features as used in [15]. The technique of Common Spatial Patterns (CSP) is used to extract ERD features from EEG.

Yom-Tov and Inbar [158] developed a BCI based on the classification of movement-related potentials recorded from 8 scalp electrodes. The extracted features include AR coefficients, PSD, the mean frequency and mean amplitude in two spectral bands, and the mean and the standard deviation of the amplitude difference between every pair of recorded electrodes. Features were extracted for each microswitch press in three time intervals, containing data before and after the actual press. Feature selection was applied to this set of 1092 features. Differentiation between movements of three limbs was done with a classification accuracy 24% lower then when attempting to distinguish between two limbs (3 subjects).

Ebrahimi and his colleagues described in [45] a BCI for multimedia communication. They eliminated EEG segments contaminated with ocular artifacts. Two approaches were considered for feature extraction: a multivariate autoregressive analysis of EEG trials and a joint time-frequency-spatial analysis of the correlations (TFSC) between the univariate components of EEG signals. Neural networks were used for classification purposes. In case of MVAR, classification errors of 7, 6 and 3% higher than when using TFSC were obtained for the last session for the 3 respective subjects. The latter method was an extension of their earlier work on time-frequency representations of a signal [51]. The classification procedure consisted of calculating the Mahalanobis distance in the ambiguity domain at a given number of points of maximum Fisher contrast. Then the most likely class for the epoch under investigation was chosen.
In [52] SVMs are used to classify EEG signals recorded during imagined left and right index finger movements. The Kernel parameters are optimized for minimizing a theoretical error bound. Fourier features and correlative time-frequency (CTFR) features are compared with respect to their discriminatory power. Assuming nonlinearly separable classes, CTFR results in a 7.5% lower error as compared to Fourier features (2 subjects average).
Hoffmann *et al* [68] describe the evidence framework to develop a variant of LDA for the use in a BCI based on EEG measurements. The resulting algorithm has the advantage of giving a continuous probabilistic output, estimating fastly regularization constants and selecting amongst different feature sets.

In response to the fact that typical BCI systems are designed specifically for one particular BCI method and is therefore not suited to the systematic studies that are essential for continued progress, Schalk and his colleagues [131] developed a documented general-purpose BCI research and development platform that facilitates the evaluation, and comparison and combination of alternative brain signals, processing methods, applications and operating protocols, called BCI2000. To achieve this purpose, the BCI2000 was designed according to two principles. First, the system model was composed of 4 modules that encompass the 4 essential functions of any BCI system: signal acquisition, signal processing, output control, and operating control. Secondly, the independence, interchangeability and scalability of each module and the components of each module were maximized.

2.4 Currently Employed Features

2.4.1 Introduction

As reported in [7], there seems to be little significant difference in classification accuracy between different classifiers such as neural networks, SVMs and LDA . Improvement of BCIs is thus most likely to result from investigation of different mental tasks and examination of new features.

This section presents an overview of the features currently used in BCI research. The aim is rather to give an idea of the characteristics extracted from EEG for the purpose of classifying mental tasks, rather than comparing different features. This is difficult because different data are used, as well as different protocols, different classifiers, different mental tasks to be distinguished and different numbers of mental tasks.

2.4.2 Amplitude Samples

Vidal [150] used the (transformed) filtered amplitude samples as features, because of the nature of ERPs, demanding the preservation of time ordering.
Yom-Tov and Inbar [158] considered the mean amplitude difference between every pair of recorded electrodes and its standard deviation.

2.4.3 Band Power Estimates

The recorded EEG is the result of the oscillatory activity of neurons. As particular brain rhythms are associated with different mental tasks or processes, it is natural to study spectral characteristics of the recorded brain signals. Pfurtscheller *et al* [116] obtained an average accuracy of 78% for online classification of two motor imagery tasks by band power estimates. [36, 158] compute Power Spectral Densities (PSDs) in a broad frequency range and use the values at different frequencies (bin size 1 and 2 Hz respectively) as features. Variants include the relative changes in band-power [113], the relative power in some frequency band compared to the total power or the mean amplitude in predefined spectral bands [158].
Wang and He [151] extracted envelopes of filtered EEG signals as a measure of instantaneous power at several narrow frequency bands. After applying a PCA to the feature signals, they were classified by LDA. Two classes were discriminated with 77% accuracy.

2.4.4 Autoregressive Coefficients

Given a signal $x(t)$, the autoregressive (AR) model of an order p is defined as

$$x(t) = a_1 x(t-1) + \cdots + a_p x(t-p) + z(t) \tag{2.1}$$

where, in the ideal case, $z(t)$ is a purely random process and $\mathrm{E}(z(t)) = 0$, $\mathrm{Var}(z(t)) = \sigma_z^2$. The parameters $a_1, \ldots, a_p$ are called the AR coefficients.

AR coefficients and some variants are widely used as features in BCI research [6, 112, 158]. AAR algorithms can perform calculation of autoregressive parameters concurrent to the data acquisition [134, 116]. LDA allows for classification rates of 80% and higher.
Other variants of AR models that have been tested are lagged-autoregressive (LAR) coefficients [112].

MVAR coefficients reflect auto and cross-spectral relationships between the components of a multivariate signal. Anderson *et al* [8] compared scalar AR coefficients with MVAR coefficients. No significant difference in classification rate was found.

Curran *et al* [33] parameterized the EEG signals using a lattice filter represention of an AR process that models each EEG channel separately. The model parameters of the lattice filter AR process are the so-called reflection coefficients. They argue that the AR coefficients themselves, often used as features in BCI applications, are highly dependent upon one another and therefore they are not suited to represent the EEG characteristics.
The reflection coefficients, a linear transformation of the AR coefficients, have the advantage that an increase in model order does not effect the reflection coefficients from previous orders. Therefore, there is little interdependency between the coefficients. Using a generative classifier, they obtained a 73.89% correct classification rate, distinguishing the spatial navigation from an auditory task. EEG was recorded from electrode positions T4 and P4.

2.4.5 Fractal Dimension

Boostani and Moradi [20] propose the Hurst fractal dimension (FD) computed with Katz' method [74] as feature for BCIs. The fractal dimension of a curve can be defined as

$$FD = \frac{\log_{10} L}{\log_{10} d}, \tag{2.2}$$

where L is the total length of the curve and d is the diameter estimated as the distance between the first point of the sequence and the point of the sequence that corresponds to the farthest distance.
Results obtained with Adaboost [2] and LDA classifiers are presented for 5 subjects. Adaboost is a machine learning technique to enhance the performance of a weak learning algorithm. The principle of adaboost is that a committee machine can adaptively adjust to the errors of its components, the weak learners. Results obtained with FD are compared to those obtained with band power features and Hjorth parameters [67]. Fractal dimension did not outperform band power features. Generally, the best results were obtained using band power features together with an LDA classifier.

2.4.6 Temporal and Spatial Complexity Measures

Roberts *et al* [124] suggested to use temporal and spatial complexity measures. Embedding-Space Decomposition (ESD) has been applied to EEG data in a BCI project. The EEG signal is reconstructed in the embedding space, either by time-delay embedding, or by multichannel reconstruction. From the singular values resulting from a singular value decomposition, an entropy function is calculated. The time course of the latter is related to the imagined finger movements. They report for 4 subjects an average classification accuracy of 80% for the distinction of imagined left and right hand finger movements.

2.4.7 Common Spatial Patterns (CSP)

CSP constructs spatial filters that are optimal for the discrimination of two populations of EEG [103, 56, 116]. The method is based on a simultaneous diagonalization of two matrices. The online classification accuracy for three healthy subjects in a two-imagery task experiment ranged from 87% - 98%[116].
Dornhege and his colleagues [44] present a multi-class extension of the binary CSP algorithm. Off-line analysis demonstrated the beneficial role of a 3rd class. Furthermore, it results in a significant speed up in real time feedback experiments.

2.4.8 Joint Time-Frequency-Spatial Analysis

Joint time-frequency-space analysis of the correlations (TFSC) between the univariate components of EEG trials was reported in [45]. Correlative time-frequency representation (CTFR) is commonly characterized by the ambiguity function. The characteristic function of a multivariate signal provides information on its joint time, frequency and space correlations. For 3 subjects, when using TFSC features, classification errors of 25, 30 and 27% for the distinction of 3 mental tasks were obtained. For comparison, MVAR features yielded errors of 32, 36 and 30 % respectively.

2.5 Conclusions

This chapter explained the principle of BCI communication. BCIs aim to offer a means of communication to the motor disabled by directly analyzing recorded brain activity and not relying on any sort of muscular response.

Different types of BCIs were described and an overview of BCI research was given. Currently, the number of researchers and laboratories involved in BCI research grow rapidly. Therefore, it is difficult to give a complete overview of BCI research. In this chapter, we highlighted the most important work done in this field. Further discussions can be found in the articles [156] and [147], summarizing the first and second international meeting on Brain-Computer Interface technology.

BCIs using continuous EEG output, modified by the user, are more promising than BCIs using an EEG response evoked with an external stimulus. The latter require several repetitions for the evoked response to be detected and are thus intrinsically slower.

Users can modify their EEG by executing different mental tasks. Motor imagery tasks alter the EEG in a reliable way and are most commonly used. However, BCI research may benefit from investigation of

mental tasks with different mental loads, and involving distinct brain regions.

Current BCIs use frequency-domain features, such as EEG μ or β rhythms occurring in specific cortical areas, or time-domain features, such as slow cortical potentials or P300 potentials. Features based on the frequency content of the signals and AR models are most often used. Besides extracting features from single EEG channels, also multivariate features such as CSPs are employed.
Band power features and CSPs should be preferred. Band power features directly reflect the change in EEG rhythm associated to different mental tasks. CSPs are spatial filters optimal for discriminating two populations of EEG. The best ranked contributions of the BCI Competition 2005 also used these types of features.

In the BCI Competition 2003 [17] and 2005, the state-of-the-art of available signal processing and classification tools were evaluated. Overfitting seemed to be the main problem. Some of the results proved that it is possible to adapt complex models to intricate data like EEG with good generalizability.

The state-of-the-art BCI generates EEG patterns resulting from limb movement or imagined limb movement that can be reliably classified given a long enough time window of data.

Chapter 3

Brain-Computer Interface under Investigation

3.1 Introduction

Figure 3.1 shows a schematic of the investigated BCI. EEG signals are acquired from several subjects according to a predefined recording protocol and stored for offline analysis. From these EEG signals features are extracted, possibly after artifact detection, denoising and preprocessing. Finally, feature selection is applied to the extracted features and the selected features are passed to a classifier, which returns a classification result. The percentage of correctly classified data windows indicates how well the combination of artifact detection, denoising, preprocessing, feature extraction, feature selection and classification discriminates the spontaneous EEG recorded during performance of different mental tasks.

Feature extraction, here consisting of the study of synchronization features, and feature selection are the main subjects of this thesis and will be treated in separate chapters, Chapters 4 and 5 respectively. In Chapter 6, features and feature selection algorithms presented in Chapters 4 and 5 are used to study the influence of artifact detection and rejection and denoising on the classification accuracy.
The experimental paradigm and classification algorithms will be detailed in this chapter. First, EEG will be introduced in Section 3.2. Section 3.3 describes the protocol used to record the EEG. The choice of mental tasks will be explained in Section 3.4. Finally, in Sections 3.5, 3.6 and 3.7, the employed classifiers will be presented.

3.2 Electroencephalogram (EEG)

Monitoring of the brain activity is necessary to understand what the brain intends to communicate. The scalp recorded electroencephalogram (EEG) has the advantages of having a good time resolution, and being noninvasive and low-cost compared to NIR spectroscopy, fMRI, ECoG and MEG. In addition, EEG could be recorded with portable devices.
In this work, a surface-EEG based BCI is studied. Therefore, this Section briefly describes EEG. More information about EEG can be found in Appendix A. It discusses the recording of EEG, the different brain rhythms and the entrainment of brain rhythms.

The EEG [107] is a recording of the brain's electrical activity. It yields information about pathologic phenomena such as, e.g., epilepsy, and is very often used to study mental activity, visual processing, motor tasks, auditive processing, emotions etc in the field of cognitive neuroscience.

EEG is recorded from electrodes placed on the scalp or implanted in the brain. Intracranially recorded signals have amplitudes on the order of 1-2 mV. The signals recorded at the scalp have a very low amplitude, on the order of 10-100 μV, because of the low conductivity of the cerebrospinal fluid, the brain tissue and the skull. While the former usually result in cleaner signals, having higher amplitudes and containing

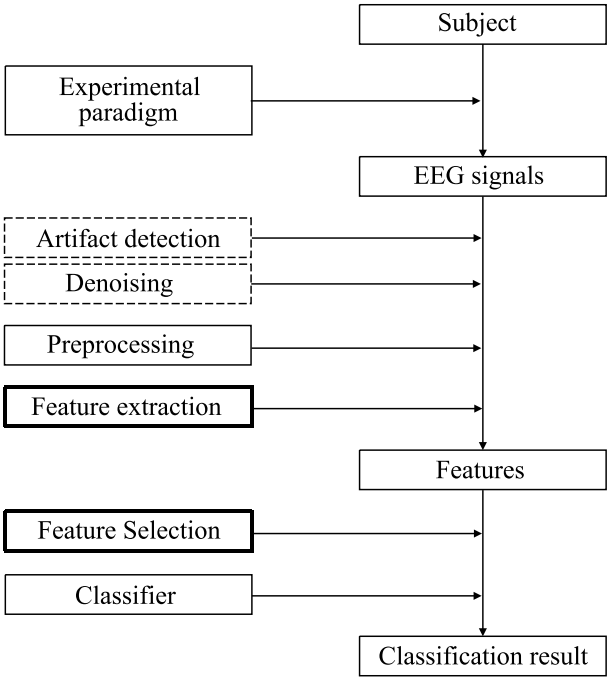

Figure 3.1: Schematic of the BCI under study

no eyeblinking and muscle artifacts, the latter method has the great advantage of being non-invasive.

The recorded scalp EEG signals are potential variations detected at the scalp, resulting from the sum of post-synaptic inhibitory and excitatory potentials affecting pyramidal neurons in the region of the cortex below the superficial electrodes. The electromagnetic fields are affected by dispersion and inhomogeneity.

The way to place electrodes on the scalp has been standardized and is referred to as the 10-20 electrode placement system [71]. The recording sites used in our experiments, an extension of the 10-20 electrode placement system, are shown in the left panel of Figure 3.2. The names of the electrodes are according to the brain lobe from which activity is recorded. They are presented in the right panel of Figure 3.2. The frontal lobe, which includes the motor cortex, is the center of personality and emotion, and performs tasks that involve logic and reasoning. The parietal lobe, is responsible for processing and storing sensory information from the body. The occipital lobe receives and processes visual information and maps it onto the cerebral cortex in a complex network. The temporal lobes are responsible for auditory processing as well as some language processing in the left temporal lobe.

Different EEG rhythms can be distinguished and are associated with different mental activities. The ones relevant for this study are:

- Alpha is the frequency range from 8.5 Hz to 12 Hz. It is characteristic of a relaxed, alert state of consciousness and is present by the age of two years. Alpha rhythms are best detected with the eyes closed. Alpha attenuates with drowsiness and open eyes, and is best seen over the occipital (visual) cortex. An alpha-like normal variant called μ is sometimes seen over the motor cortex (central scalp) and attenuates with movement, with imagined movement or with the intention to move.

- Beta is the frequency range from 12 Hz to 30Hz. Disorganized, low amplitude beta is often associated with active, busy or anxious thinking and active concentration.

High voltage, regular, synchronized EEG rhythms reflect coordinated activity of large groups of neurons [139] and are probably correlated with an idling state of the brain. Low voltage, irregular desynchronized EEG activity reflects a state of increased cortical excitability and corresponds to a state of information processing and mental activity.
The μ rhythm [83] appears maximally over the central Rolandic or sensorimotor area and is attenuated

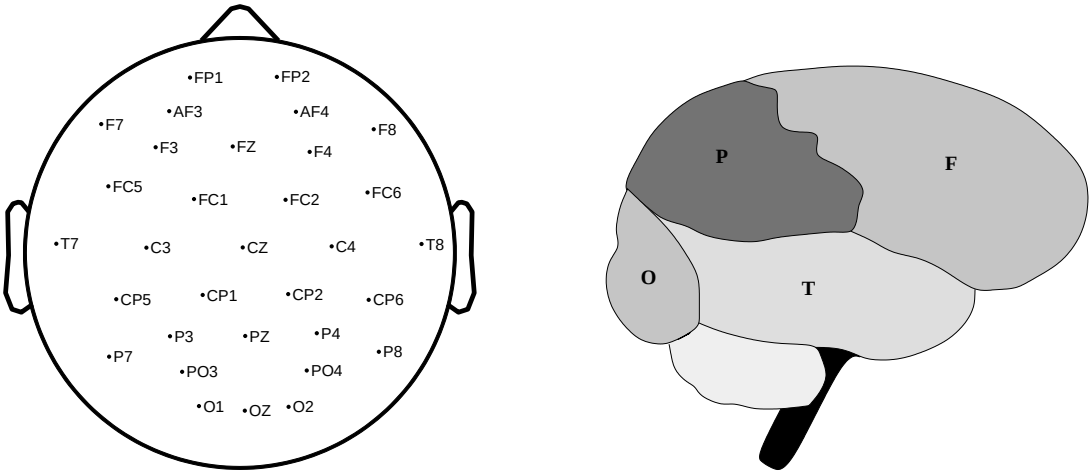

Figure 3.2: Left: Electrode positions used in our experiments; Right: The different lobes of the cortex: F: frontal lobe; P: parietal lobe; O: occipital lobe; T: temporal lobe.

or blocked by movement. This rhythm is minimally affected by visual stimulation and can thus be distinguished both spatially and functionally from the occipital α rhythm. Kuhlman studied the reaction of EEG to movement (self-paced continuous opening and closing of the hand) and visual stimulation (open and close eyes). The frequencies of α and μ rhythm are not necessarily the same. He found that for some subjects the μ activity was higher in frequency than the α, in other subjects it was lower. The dominant frequency of these rhythms, when present, rarely varied more than 0.5Hz within each individual.

3.3 Recording Protocol

For the purpose of introducing and studying new features, we performed an offline study of data recorded without feedback. In this condition the subjects could not adapt to any features or classification algorithms used to provide the feedback.

Five healthy right-handed volunteers (2 women and 3 men), aged between 22 and 33, took part in the experiments. No selection criteria were used. None of the subjects had participated in BCI experiments before. For each subject, a total of 60 minutes of EEG was recorded in 5 sessions of 4 minutes per day, on 3 consecutive days. The sessions were separated by breaks of 5 to 10 minutes. Subjects were seated comfortably in front of a computer. They were asked not to move during the sessions and look at the computer screen.
About every 20 seconds the operator informed the subject which new task they have to perform. The 3 tasks were alternated randomly and such that the incidence of the 3 tasks was equal.

In our protocol, no cue was used. Subjects switched about every 20s randomly between the 3 mental tasks. In such a protocol not only must it be determined which mental task is being performed, but also when it is being performed. The communication, however, is more 'natural' as the users can spontaneously switch between tasks. The advantage is not to be dependent on external triggers and to have thus a potentially faster system. In this study, we deal with resolving the mental task corresponding to an EEG window.

EEG signals are recorded with the 32-channel Biosemi ActiveTwo system[1]. Electrodes were placed on the scalp according to the 10-20 international electrode placement system. The 32 electrodes are shown in Figure 3.2. The ground electrode is replaced by two separate electrodes, located between C3 and Cz, and Cz and C4 respectively. They form a feedback loop, decreasing the effective impedance with a factor 100 at $50Hz$. EEG signals were digitized at a sampling rate $f_s = 2048Hz$, subsampled to $f_s = 512Hz$

[1] http://www.biosemi.com

and stored for offline analysis. Together with the EEG signals, we recorded the 'task signal', indicating which task the subject is doing at every moment.

3.4 Choice of Mental Tasks

The goal is to choose mental tasks that are spatially as distinct as possible and that have a different mental load, in order to better distinguish between them. Many BCIs use motor imagery paradigms.

Beisteiner *et al* [9] assessed the existence of systematic changes of the topography of DC potentials for unilateral left and right hand movements and for bilateral movements. Subjects either imagined or executed a sequence of unilateral or bilateral hand movements. Changes of DC potentials between task execution and imagination were localized in central recordings (C3, Cz, C4) with larger amplitudes when executing the task than when imagining to do so. With unilateral performance, the side of the performing hand (right, left) had localized effects in the recordings over the sensorimotor hand area (C3, C4). They were qualitatively the same with imagination and execution and quantitatively similar. There were no significant differences. They also found that DC potentials were always more negative with motor imagery as compared to visual imagery with differences being most pronounced in central recordings.

Penny *et al* [113] found that imagined left and right hand movements (opening and closing of the hand) could be identified in 6 out of 7 subjects with an average accuracy of 70%. Real movements could be identified from the μ-rhythm in all subjects with an average accuracy of 76 %. These classification results were based on μ-rhythm power features.

Pfurtscheller and Neuper [116] chose motor imagery tasks because the suppression of μ and central β rhythms is more pronounced at the contralateral hemisphere when subjects imagine one-sided hand movements than when they actually perform such movements. Motor imagery could thus be used to achieve asymmetrical electrocortical responses.

Curran and Stokes [34] review research on cognitive tasks and other methods of generating and controlling specific changes in EEG activity that can be used to drive BCI systems. Much work is required to develop a range of well-defined, easily learned cognitive tasks that are shown to enable all or most subjects to produce and control EEG signals that can be used to drive a BCI system. Identifying pairs of cognitive tasks which produce maximally discriminable EEG patterns will help improving the classification accuracy.

The mental tasks chosen in this study, were imagination of repetitive self-paced left hand movement ('L') and imagination of repetitive self-paced right hand movement ('R'). The imagined hand movement was a flexion at the wrist causing the hand moving up and down. The third task, 'W', consisted of generating words that begin with the same letter, freely chosen by the subject. The words were not spoken. All tasks were executed with opened eyes.
The motor imagery tasks activate the motor cortex, indicated in the right panel of Figure 3.3. The "word" task involves the language center, located in the left temporal lobe. Electrodes overlying these brain regions are indicated in the left panel of Figure 3.3.

3.5 Classifiers

In order to evaluate the features with regard to their ability to distinguish spontaneous EEG recorded during different mental tasks, it was necessary to apply a classifier. Two different classifiers have been used: Support Vector Machines (SVMs) [22] and Discriminant Analysis (DA) [32].

In the framework of BCIs, it was discussed that non-linear methods in some applications can provide better results than linear methods, particularly with complex and/or very large data sets, but it was agreed that simplicity is generally best [102]. Therefore, the use of linear classification methods is recommended

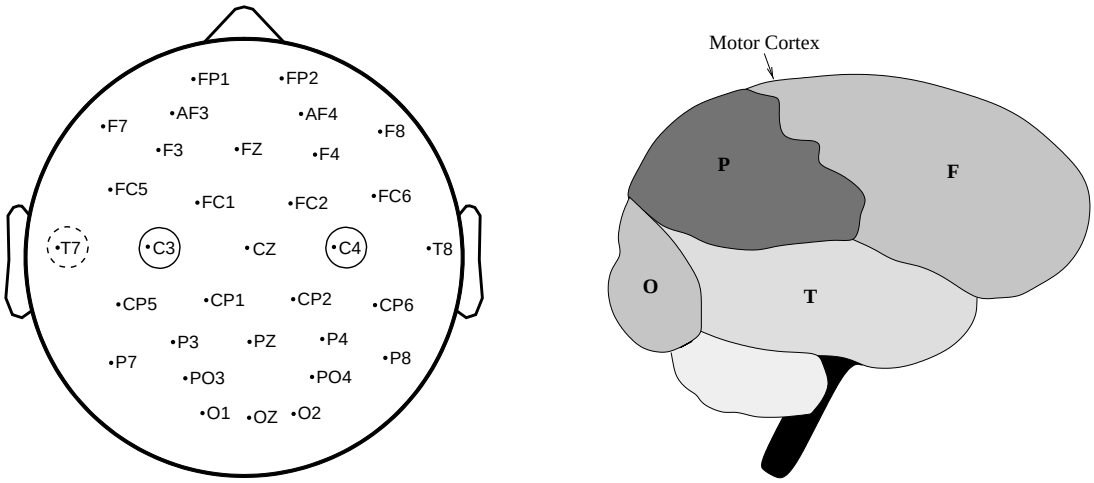

Figure 3.3: Left: Electrodes overlying brain regions involved in the mental tasks considered in this study; Right: Location of the motor cortex.

wherever possible. This motivates the use of linear SVMs and linear DA in our study.

Linear classifiers are more robust than their nonlinear counterparts, because they have only limited flexibility and are less prone to overfitting. However, the influence of single data points on learning is not limited. In the presence of strong noise and outliers the decision surface can change drastically. Both SVM and DA try to deal with outliers, by regularization and robust estimation of parameters respectively.

3.6 Support Vector Machines

3.6.1 Generalities

SVMs belong to the state-of-the-art machine learning techniques and have been successfully applied in BCI research [98]. Regularization helps to limit the influence of outliers or strong noise, the complexity of the classifier and the raggedness of the decision surface. Theoretical results guarantee a high generalization performance when the margin is large or when the number of support vectors is small. In addition, SVMs find a global solution [22].

SVMs generally distinguish between two classes only. To resolve the 3-class classification problem, we used a combination of SVMs. In this paragraph, we will discuss some generalities about SVMs. The combination of classifiers used for the 3-class discrimination problem is explained in the next paragraph.

An introduction to SVMs can be found in [22]. For several applications (e.g., handwritten digit recognition, object recognition, speaker identification, face detection in images, text categorization), SVM generalization performance either matches or is significantly better than that of competing methods. The best generalization performance will be achieved if the right balance is found between the accuracy attained on a particular training set and the ability of the machine to learn any training set without error.

The training data are labeled $\{\mathbf{x}_i, y_i\}$, $i = 1, \ldots, l$, $\mathbf{x}_i \in \mathbf{R}^d$, $y_i \in \{-1, 1\}$. Suppose we have a hyperplane which separates the positive from the negative examples, a separating hyperplane (see Figure 3.4). The points $\mathbf{x}$ which lie on the hyperplane satisfy $\mathbf{w}.\mathbf{x} + b = 0$, where $\mathbf{w}$ is the normal to the hyperplane, $|b|/\|\mathbf{w}\|$ is the perpendicular distance from the hyperplane to the origin and $\|\mathbf{w}\|$ is the Euclidean norm of $\mathbf{w}$. The margin of the separating hyperplane is defined as $d_+ + d_-$, d_+ (d_-) being the shortest distance from the separating hyperplane to the closest positive (negative) example. For the linearly separable case, the support vector algorithm looks for the separating hyperplane with the largest margin. The training

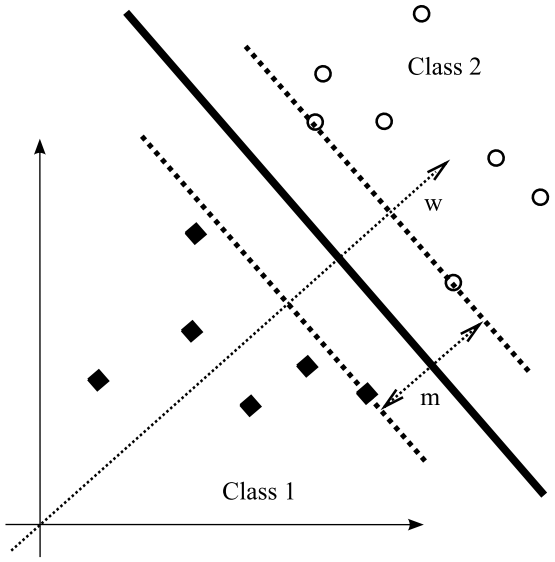

Figure 3.4: Linear separating hyperplane for the separable case; m: margin

data will satisfy the following constraints:

$$\mathbf{x}_i.\mathbf{w} + b \geq +1 \quad \text{for } y_i = +1 \tag{3.1}$$

$$\mathbf{x}_i.\mathbf{w} + b \leq -1 \quad \text{for } y_i = -1 \tag{3.2}$$

This defines two hyperplanes H_1 and H_2. In that case the margin equals $\frac{2}{\|\mathbf{w}\|}$. The training points for which Equations (3.1) or (3.2) holds, and whose removal would change the solution found are called support vectors.

In the Lagrangian formulation of the problem, the training data will appear in the form of dot products between vectors, which allows generalization of the procedure to the nonlinear case. The Lagrangian is as follows:

$$L_P = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{x}_i \mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i \tag{3.3}$$

L_P must be minimized with respect to $\mathbf{w}$ and b, and at the same time the derivatives of L_P with respect to all α_i vanish, all subject to the constraint $\alpha_i \geq 0$. Substituting equalities resulting from the dual formulation of the problem we obtain

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j \mathbf{x}_i.\mathbf{x}_j, \tag{3.4}$$

which has to be maximized. In the solution, those points for which $\alpha_i > 0$ are called support vectors. Solving the SVM problem is equivalent to finding a solution to the Karush-Kuhn-Tucker conditions, playing a central role in the theory and practice of constrained optimization.

In the non-separable case, a further cost is introduced to relax the constraints (3.1) and (3.2). Positive slack variables $\xi_i, i = 1, \ldots, l$ are introduced in the constraints.

$$\mathbf{x}_i.\mathbf{w} + b \geq +1 - \xi_i \quad \text{for } y_i = +1 \tag{3.5}$$

$$\mathbf{x}_i.\mathbf{w} + b \leq -1 + \xi_i \quad \text{for } y_i = -1 \tag{3.6}$$

$\xi_i \geq 0, \forall i$ An extra cost for errors can be assigned by changing the objective function to be minimized from $\|\mathbf{w}\|^2/2$ to $\|\mathbf{w}\|^2/2 + C(\sum_i \xi_i)$. The larger the parameter C (cost of the constrain violation), the higher the penalty to errors. The α_i now have an upper bound of C. The most important data points are the support vectors with highest values of α. For the non-separable case, the upper bound $\alpha_i \leq C$ corresponds to an upper bound on the force any given point is allowed to exert on the sheet.

3.6.2 SVM for 3-class Discrimination

We fixed the values of the parameters C and ϵ (tolerance of the termination criterion) of the SVM.

The classifier we used to distinguish between 3 classes consisted of a combination of 3 Support Vector Machines (SVMs) with a linear Kernel. Each individual SVM was trained with data from two tasks only: 'left' and 'right', 'left' and 'word', and 'right' and 'word'. To discriminate the 3 mental tasks of the test set, the data were passed to the 3 individual SVMs and the classification results of the 3 SVMs were combined to obtain the final class ('left', 'right' or 'word') or a label 'unknown'.

The first classifier ('L'-'R') results in output '1' when class 'left' is predicted and '2' when 'right' is predicted. Similarly the second ('L'-'W') and third classifier ('R'-'W') have output '1' and '4', and '2' and '4' respectively. The sum of the 3 outputs indicates whether one class has been recognized twice or whether 3 times a different class was recognized. In the former case, the window is associated to the class recognized twice, 'left', 'right' or 'word'; in the latter case, the window is categorized as 'unknown' ('U'). The numbers '1', '2' and '4' were chosen because they result in different sums for the classes 'left', 'right', 'word' and 'unknown'. Table 3.1 gives an overview.

Classifier				
1st	2nd	3rd	Sum	Predicted Class
1	1	2	4	'left'
1	1	4	6	'left'
1	4	2	7	'unknown'
1	4	4	9	'word'
2	1	2	5	'right'
2	1	4	7	'unknown'
2	4	2	8	'right'
2	4	4	10	'word'

Table 3.1: Output of the 3 SVM classifiers and attributed class label

Features were computed from 1s sliding windows with $\frac{7}{8}$s overlap and classified. The length of 1s was chosen arbitrarily. The shorter the windows, the less accurate the estimation of the features, while longer windows make the system react slowlier.
Every second, strict majority voting was applied to the 8 classification results to determine the class. If no strict majority was attained, the window was categorized as 'unknown'. Otherwise the label of the class occurring strictly more than the others was associated to the data window.

3.7 Discriminant Analysis

3.7.1 Generalities

We applied robust linear DA, presented in [32].
In Discriminant Analysis (DA) a training sample composed of two groups $\mathcal{P}_1$ and $\mathcal{P}_2$, corresponding to two classes (mental tasks), of p-variate observations is considered. The observations $x_{11}, \ldots, x_{1n_1}$ and $x_{21}, \ldots, x_{2n_1}$ come from populations $\mathcal{P}_1 \sim H1 = N_p(\mu_1, \Sigma_1)$ and $\mathcal{P}_2 \sim H2 = N_p(\mu_2, \Sigma_2)$ respectively. Σ and μ are the covariance matrix and the mean of the two multivariate normal distributions. From these data, a linear discriminant is determined, which is used to classify new observations into one of the two groups. Assuming equal covariance matrices, Fisher's linear discriminant rule is as follows:

$$L(x) = (\mu_1 - \mu_2)^t \Sigma^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)^t \Sigma^{-1}(\mu_1 + \mu_2). \tag{3.7}$$

A new observation x is assigned to class $\mathcal{P}_1$ if $L(x) > log(\frac{c_2 \pi_2}{c_1 \pi_1}) = \tau$, where c_1 and c_2 are the costs of misclassifying a unit of, respectively, $\mathcal{P}_1$ and $\mathcal{P}_2$ and π_1 and π_2 are the prior probabilities that x will belong to $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively. In practice these parameters are unknown and therefore $\tau = 0$.

Outliers and atypical observations may shift the estimated means and they may blow up the dispersion matrices, and thus change the discriminant rules, based on estimates of the population parameters. This may have an influence on the results of classical DA. To prevent this, robust estimators of the population parameters were proposed [32]. We used the MCD-estimator (Minimum Covariance Discriminant), because this estimator has good statistical properties. The estimator is given by the subset of observations of size h for which the determinant of its covariance matrix is minimal. The location and covariance are then given by the mean and covariance matrix of these h observations.

3.7.2 DA for 3-class discrimination

The DA classifier used to distinguish 3 classes is analogous to the 3-class SVM classifier described in Paragraph 3.6.2. In this case, instead of using 3 SVM-classifiers, 3 DA classifiers are used to distinguish the 3 respective pairs of tasks and produce the outputs '1', '2' and '4'. Otherwise the classifier is identical to the one in Paragraph 3.6.2.

3.8 Information Transfer Rate

The performance of a BCI is measured not only by the classification accuracy, CR, but also by the rate of information transfer, measured by the Information Transfer Rate (ITR). Wolpaw *et al* propose the following ITR definition:

$$ITR = \frac{\log_2 N + (CR)\log_2(CR) + (1-CR)\log_2(\frac{1-CR}{N-1})}{T}. \tag{3.8}$$

N is the number of classes (mental tasks) distinguished, CR the fraction of correctly classified samples and $\frac{1}{T}$ is the classification speed.
Some simplifying assumptions were made. First, no "not recognized" class was considered. Secondly, it was assumed that all classes have the same a priori occurance and that, thirdly, the classification accuracies are the same for all classes. The fourth assumption is that the classification error (1-CR) is equally distributed amongst the remaining classes.
Figure 3.5 exemplifies the ITR for a 3-class distinction problem where the classifier returns a result every second. The maximum bit rate 95 bits/minute.

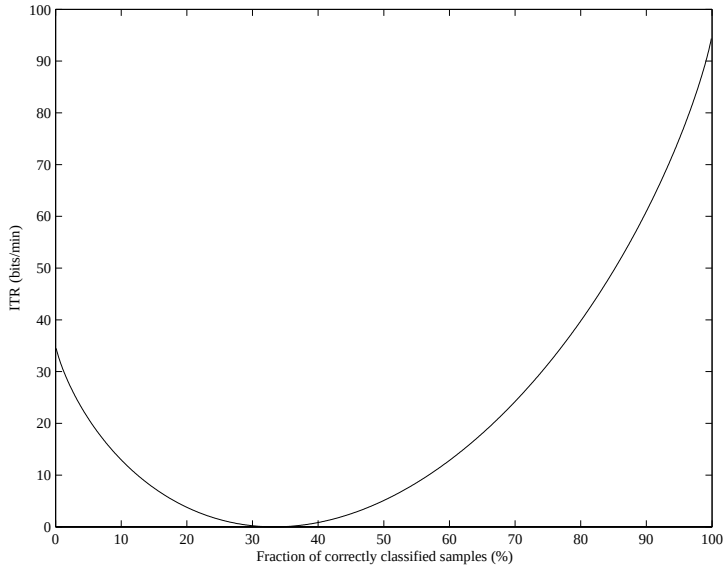

Figure 3.5: Information Transfer Rate (bits/minute) in function the classification accuracy (%) for 3-class distinction every 1s

3.9 Conclusions

Scalp recorded EEG has a good time resolution, is noninvasive and can be recorded from portable devices. Therefore, it is well suited for recording brain activity in BCI applications. In the EEG different rhythms can be distinguished and are associated with different mental activities.
We detailed the recording protocol of the EEG data used throughout this thesis. 32-channel EEG was recorded from 5 subjects during 3 different mental tasks. Mental tasks were chosen according to the brain regions they activate. The goal was to activate spatially distinct brain regions. A combination of two motor imagery tasks (left and right hand movement) and a word generation task was chosen.
Finally, it was argued that in the framework of BCIs, it is recommended to use linear classification methods wherever possible. Therefore, we chose linear SVMs and a DA classifier in this thesis.

Chapter 4

Feature Extraction: Synchronization

4.1 Introduction

As detailed in Section 2.4, in surface EEG based interfaces, very often spectral estimates [116, 113] and autoregressive (AR) coefficients [8, 134] and measures derived from them [33] are used to characterize the EEG . Some authors exploit spatial information by considering multivariate autoregressive models (MVAR) [8], temporal and spatial complexity measures [124], common spatial patterns (CSP) [56], or joint Time-Frequency-Space Correlation (TFSC) [45]. In this work, we examine yet another measure that quantifies interaction between EEG signals and has a more direct neurophysiological interpretation: the Phase Locking Value (PLV). We compare it to spectral coherence, another widely-used measure of synchronization.

Different regions widely distributed over the brain must communicate to provide the basis for integration of sensory information and for many functions that are critical for learning, memory, information processing, perception and behavior of organisms. Transient periods of synchronization of oscillating neuronal discharges have been proposed to act as an integrative mechanism that may bring a widely distributed set of neurons together into a coherent ensemble that underlies a cognitive act [99] . Rodriguez *et al* [126] demonstrated a direct participation of synchrony in a cognitive task.

Many studies employ spectral coherence to quantify synchronization processes in the brain, e.g. [108, 54, 129, 130]. Recently, Delorme and Makeig [39] studied power and coherence changes in EEG data recorded from a highly trained subject during sessions in which he attempted to regulate power at 12Hz over his left- and right-central scalp. They observed changes in partial phase coherence between maximally independent components.

The PLV is a time-domain synchronization measure frequently applied in recent EEG analysis. It was described by Mardia [96] and by Hoke *et al* [69] for magnetoencephalography as Phase Coherence (R). In 1999, it was introduced in electroencephalography by Lachaux *et al* [85] as PLV and in 2000 by Mormann *et al* [101] as R in the fields of cognitive neuroscience and epilepsy research respectively. Since then, this measure has been used in a variety of studies, e.g. [84, 89, 90, 146, 47, 10, 12, 11, 64, 100, 145]. Lachaux *et al* introduced the Phase Locking Value (PLV) to quantify the degree of phase locking between two signals. To detect significant deviations from background fluctuations, Phase Locking Statistics (PLS) is used.
We prefer the notation PLV to R in order to avoid confusion with spectral coherence.

A very interesting study related to functional coupling during simple paced finger movements is the one by Gerloff *et al* [54]. They suggest that important aspects of information processing in the human motor system could be based on network-like oscillatory cortical activity and might be modulated on at least two levels, which to some extent can operate independently from each other: regional activation (task-related power) and inter-regional functional coupling. In general, task-related coherence showed the most prominent task-related changes, and these were typically restricted to the central region. Internal pacing of movement is associated with greater functional coupling of cortical premotor and sensorimotor areas than external pacing. Disruption of the orderly idling rhythm would increase the entropy of the system

and thus the information content. If, in this model, the burst were then synchronized with concomitant bursts in other activated areas, one would observe increased inter-regional coherence together with decreased power, as observed by Gerloff *et al.*

Also Ohara *et al* [109] tested for inter-regional functional coupling modulated by motor tasks accompanying local cortical activation and therefore analyzed movement-related changes of ECoG coherence among different motor areas. They found an increase in functional coupling starting in the preparatory phase of voluntary hand movements and a power suppression in the same region during the preparatory phase of the movement. The power suppression preceded the coherence change by a significant length of time. Spatial distribution of coherence and power was not significantly correlated.

Another reason why synchronization could be a useful feature in BCIs, is that it could reflect the effects of learning [136], an important aspect of BCI. Sommerfeld *et al* report that a training-dependent reduction of mental effort for control processes is accompanied by a reduction of the strength of the functional coupling in the beta band between the specific frontal and parietal brain regions during a linear ordering problem. However, to fully exploit this matter to improve BCIs, synchronization features should be used together with adaptive classifiers (online learning), which is not subject of this work.

Quiroga *et al* [119] applied several linear and nonlinear measures of synchronization to three typical EEG signals. Mutual information was not robust due to the low number of data points. The other measures, namely nonlinear interdependences, phase synchronizations, cross correlation and the coherence function, all resulted in a useful quantification of synchronization. Separation between synchronization levels was more pronounced with nonlinear measures, but qualitative results were the same. It was concluded that these measures are valuable for the study of synchronization in real data.

To our knowledge, synchronization measures in general and PLV and spectral coherence in particular have not yet been applied to classify EEG in the framework of BCIs. Here, we study PLV for EEG-based discrimination of 3 mental tasks and compare it to spectral coherence.

4.2 Phase Locking Value

4.2.1 Calculating PLV

The PLV characterizes the stability of the phase differences between the phases $\varphi_x(t)$ and $\varphi_y(t)$ of 2 signals $s_x(t)$ and $s_y(t)$ across subsequent time samples of one window as follows:

$$PLV = \left| \left\langle e^{j\Delta\varphi(t)} \right\rangle \right|, \tag{4.1}$$

where $\Delta\varphi(t) = \varphi_y(t) - \varphi_x(t)$ and $\langle . \rangle$ is the averaging operator. For the discrete case, it suffices to replace t by $\frac{k}{f_s}$, $k = 1 \dots N$, N being the number of samples in the window. PLV equals the length of the sum vector of all unit vectors $e^{j\Delta\varphi(t)}$ in one window, divided by the number of samples in that window. When the phase difference is constant (phase synchronization), all phase difference vectors will be aligned, resulting in a PLV equal to 1. If the phase differences are randomly distributed over $[0, 2\pi]$, the vector sum, and thus the PLV, will be 0.

4.2.2 Calculating the Instantaneous Phase of a Signal

To compute the PLV, we first need to know the instantaneous phase $\varphi(t)$.
The phase $\varphi(t) = 2\pi ft + \phi$ of a periodic signal $s_p(t)$ at frequency f

$$s_p(t) = cos(2\pi ft + \phi) \tag{4.2}$$

indicates where in its cycle the process can be located at a given time t. For a periodically varying phenomenon, it characterizes any distinguishable instantaneous state of the phenomenon, referred to a fixed reference or another periodically varying phenomenon.

We will consider two methods frequently used to determine the instantaneous phase of an arbitrary signal: the Hilbert transform method and the method using a wavelet transform.

Hilbert transform

The Hilbert transform [110, 93] finds its origin in the fact that, for causal signals, a relation between the real and imaginary parts of the Fourier transform of the signal exists. In the ideal case the Hilbert transform is a $\frac{\pi}{2}$-degree phase shifter. Thus, e.g.,

$$s_e(t) = \cos(2\pi f t + \phi) \tag{4.3}$$

transforms into

$$\tilde{s}_e(t) = \sin(2\pi f t + \phi). \tag{4.4}$$

Because of linearity, a sum of cosines will transform in a sum of sines having the same arguments.
The Hilbert transform of $s(t)$ is given by

$$\tilde{s}(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{+\infty} \frac{s(\tau)}{t-\tau} d\tau, \tag{4.5}$$

$p.v.$ denoting the Cauchy principal value, and allows us to determine the instantaneous phase φ_s as follows:

$$\varphi_s(t) = \arctan\left(\frac{\tilde{s}(t)}{s(t)}\right). \tag{4.6}$$

The application of the Hilbert transform to nonstationary signals and its interpretation are discussed in [19].

A significant improvement in computation time, allowing for real time computation on a PC running MATLAB, was achieved with the following algorithm.

- Compute the Hilbert transform of the signals $s_x(t)$ and $s_y(t)$ and construct the analytical signals:

$$S_x(t) = s_x(t) + j\tilde{s_x}(t) \tag{4.7}$$

$$S_y(t) = s_y(t) + j\tilde{s_y}(t) \tag{4.8}$$

- Instead of explicitly computing the phase of both signals as in equation (4.6), we normalize the analytical signals:

$$S_{xn}(t) = \frac{s_x(t) + j\tilde{s_x}(t)}{s_x^2(t) + \tilde{s_x}^2(t)} \tag{4.9}$$

$$S_{yn}(t) = \frac{s_y(t) + j\tilde{s_y}(t)}{s_y^2(t) + \tilde{s_y}^2(t)}. \tag{4.10}$$

From this we can directly derive $e^{j\Delta\varphi(t)}$:

$$\begin{aligned} e^{j\Delta\varphi(t)} &= \cos(\varphi_y(t) - \varphi_x(t)) + j\sin(\varphi_y(t) - \varphi_x(t)) && (4.11)\\ &= Re(S_{yn}(t))Re(S_{xn}(t)) + Im(S_{yn}(t))Im(S_{xn}(t)) + \\ &\quad j(Im(S_{yn}(t))Re(S_{xn}(t)) - Re(S_{yn}(t))Im(S_{xn}(t))). && (4.12) \end{aligned}$$

In the discrete case, the Hilbert transform [110] is computed by the so-called discrete-time analytic signal. Integrals are replaced by sums.

Wavelet transform

Another approach to compute the instantaneous phase applies wavelets. The original signal is convolved with a complex Gabor wavelet. The phase is then determined from the coefficients of their wavelet transform at the target frequency. For a signal $s(t)$, the wavelet coefficients as a function of time τ and frequency f are defined as:

$$W_s(\tau, f) = \int_{-\infty}^{+\infty} s(t).\Psi^*_{\tau.f}(t)dt, \tag{4.13}$$

where the Gabor function $\Psi_{\tau.f}(t)$ at frequency f and time τ is defined by

$$\Psi_{\tau.f}(t) = \sqrt{f}exp(i2\pi f(t-\tau))exp\bigg(-\frac{(t-\tau)^2}{2\sigma^2}\bigg) \tag{4.14}$$

The instantaneous phase can be evaluated from

$$\varphi(\tau, f) = arctan\bigg(\frac{Im(W_s(\tau, f))}{Re(W_s(\tau, f))}\bigg) \tag{4.15}$$

Hilbert transform vs Wavelet transform

Lachaux computed the instantaneous phase by means of a wavelet transform, whereas Mormann *et al* used the Hilbert transform method. Le Van Quyen *et al* [89] investigated the difference between the Hilbert and wavelet transform method by means of neural models, intracranial signals from epileptic patients and scalp EEG recordings. They demonstrated that the differences between the two methods were minor and concluded the equivalence of the two approaches for the study of neuroelectrical signals. In this study, the Hilbert transform method is adopted.

4.3 Spectral Coherence

The spectral coherence of 2 signals $s_x(t)$ and $s_y(t)$ is the squared normalized cross-power spectrum:

$$K^2_{XY}(f) = \frac{|\langle C_{XY}(f)\rangle|^2}{\langle C_{XX}(f)\rangle\langle C_{YY}(f)\rangle}, \tag{4.16}$$

where $C_{XY}(f) = X(f)Y^*(f)$, $C_{XX}(f) = X(f)X^*(f)$, $C_{YY}(f) = Y(f)Y^*(f)$, $X(f)$ and $Y(f)$ are the Fourier transforms of $s_x(t)$ and $s_y(t)$, and $\langle\rangle$ is the averaging operator.
The averaging in equation (4.16) is necessary in order to make cross terms, that are due to the correlated and uncorrelated parts of the signals, disappear. At the same time, this has the advantage that epochs with artifacts or distorted epochs can be easily rejected.
$C_{XY}(f)$ is the Fourier transform of the correlation function:

$$C_{XY}(f) = X(f)Y^*(f) = \mathcal{F}\bigg(\int_{-\infty}^{+\infty} \bar{x}(t+\tau)y(\tau)d\tau\bigg). \tag{4.17}$$

The spectral coherence as defined in equation (4.16) has, like the PLV, a value between 0 and 1. A coherence value of 1 means that the corresponding frequency components of the signals $s_x(t)$ and $s_y(t)$ are identical, except for a multiplicative amplitude difference and a constant time relation (phase delay). Practically, for finite length signals it means that both signals are related by a linear transform. A zero coherence indicates that the corresponding frequency components of both signals are not correlated.

The coherence spectrum is as broad as half of the sampling frequency. To obtain a scalar value from the coherence $K^2_{XY}(f)$, this spectrum is usually averaged, over the whole spectrum or over a specific frequency band ΔB, i.e. $\langle K^2_{XY}(f)\rangle_{\Delta B}$. Also maximum or minimum coherence values can be considered.

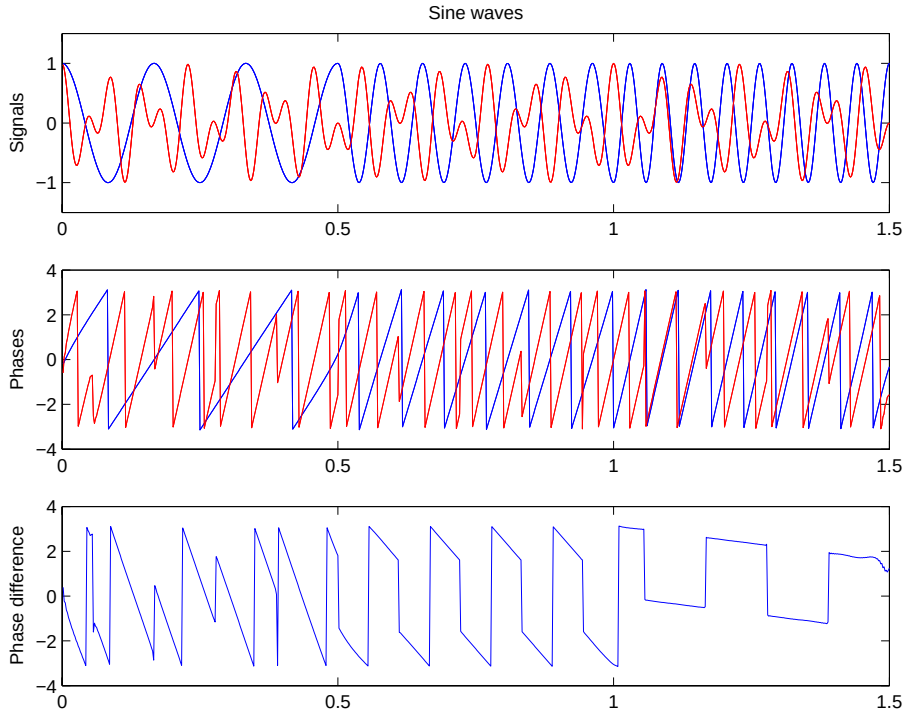

Figure 4.1: Sine waves without noise, their respective phases and the phase difference. Only 0.5s is shown for each regime. Horizontal axis: t(s).

4.4 Properties of PLV and Spectral Coherence

We will examine the common features of and the differences between PLV and spectral coherence in the following paragraphs. To illustrate PLV and coherence, we use 3 pairs of signals: sine waves and transients of weakly and strongly coupled Rössler oscillator systems [101]. The sine waves share a common frequency, while for the transients of the Rössler oscillator systems the amplitude of the envelope as well as the frequency vary in time.

First, we consider 2 sine waves, one of which is the sum of two frequency components and the other one consisting of one frequency component, changing in 3 subsequent time intervals. Only in the second time interval both signals have a frequency component in common. 60s signals are considered, but for visualization purposes, only 0.5s of each regime is shown (Figure 4.1). The phases of both signals as well as the phase differences are presented in the same figure.
The considered sinusoidal signals are given by:

$$s_x(t) = (\cos(2\pi 13t) + \cos(2\pi 22t))/2 + \sigma\zeta(t) \quad \forall t \tag{4.18}$$

$$s_y(t) = \begin{cases} \cos(2\pi 6t) + \sigma\zeta(t) & \forall t = \frac{1}{f_s} ... \frac{t_{tot}}{3} \\ \cos(2\pi 13t) + \sigma\zeta(t) & \forall t = \frac{t_{tot}}{3} + \frac{1}{f_s} ... \frac{2t_{tot}}{3} \\ \cos(2\pi 17t) + \sigma\zeta(t) & \forall t = \frac{2t_{tot}}{3} + \frac{1}{f_s} ... t_{tot} \end{cases} \tag{4.19}$$

$\sigma\zeta(t)$ is a Gaussian white noise with variance σ^2 and t_{tot} is the total time (60s). $f_s = 512Hz$ in these simulations.

Further illustrations are on the x-components of a 3-dimensional weakly coupled and strongly coupled Rössler oscillator respectively, in transient regime. These signals are presented in the Figures 4.2 and 4.3 respectively, together with their phases and phase difference. The equations of the considered Rössler

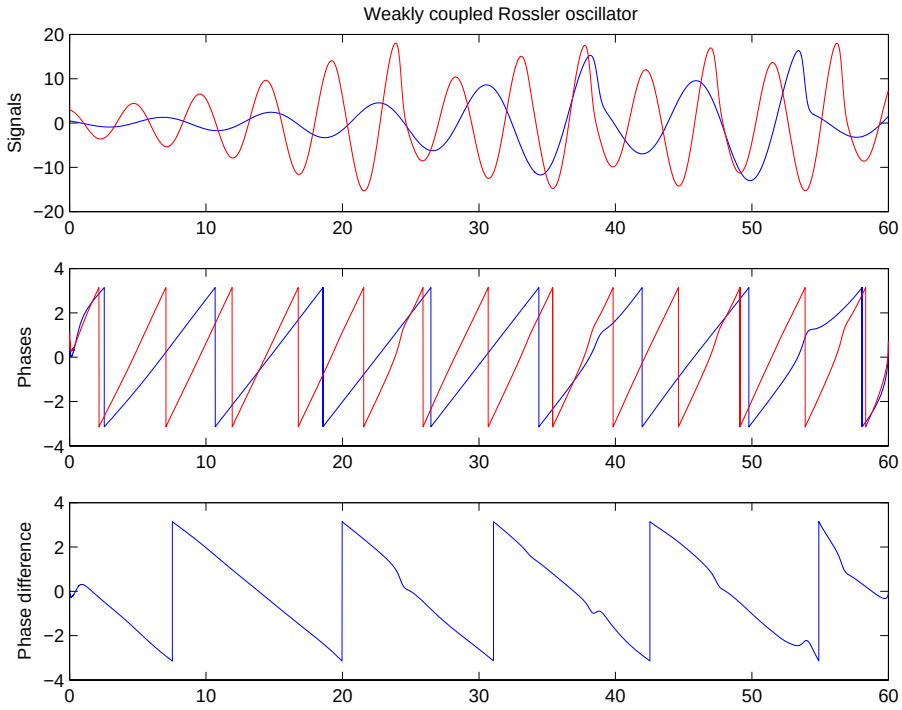

Figure 4.2: X-components of a weakly coupled Rössler oscillator system without noise, their respective phases and the phase difference. Horizontal axis: t(s).

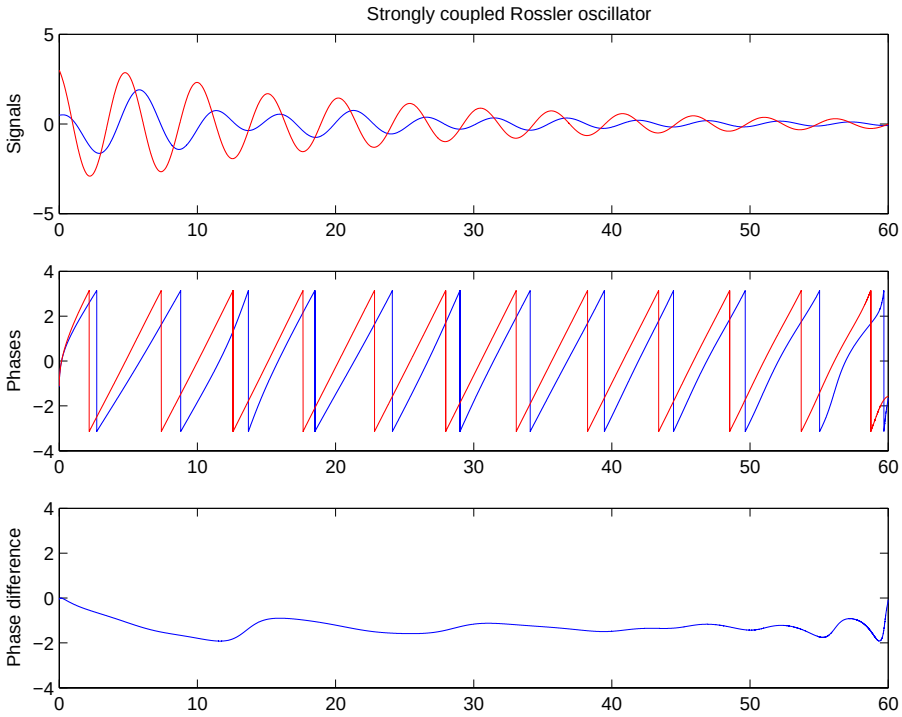

Figure 4.3: X-components of a strongly coupled Rössler oscillator system without noise, their respective phases and the phase difference. Horizontal axis: t(s).

oscillator system are as follows.

$$\begin{cases} \dot{x_1} = -w_1 y_1 - z_1 + \epsilon(x_2 - x_1) \\ \dot{y_1} = w_1 x_1 + 0.165 y_1 \\ \dot{z_1} = 0.2 + z_1(x_1 - 10) \\ \dot{x_2} = -w_2 y_2 - z_2 + \epsilon(x_1 - x_2) \\ \dot{y_2} = w_2 x_2 + 0.165 y_2 \\ \dot{z_2} = 0.2 + z_2(x_2 - 10) \end{cases} \tag{4.20}$$

Where w_1 was set to 0.8 and w_2 to 1.3. The coupling parameter ϵ was 0.001 in the weakly coupled case and 0.3 in the strongly coupled case. The signals used here for illustration are the x-components of this Rössler oscillator system, x_1 and x_2. To study the influence of noise, $a_1\sigma\zeta(t)$ and $a_2\sigma\zeta(t)$ were added to the computed signals x_1 and x_2. These are Gaussian white noises with variance $a_1^2\sigma^2$ and $a_2^2\sigma^2$ respectively. a_1 and a_2 are the maxima of the absolute value of the amplitudes of x_1 and x_2 respectively.

4.4.1 Influence of Window Length

The phase is derived from information contained in a segment of the signal. Therefore, we examine the influence of the window length on the synchronization measures.
Simulations on sinusoidal signals suggest that the time resolution of PLV is better than the one of coherence. We refer to Figures 4.4, 4.5 and 4.6. The PLV, the averaged coherence (Coh) and maximum of the coherence spectrum (Coh_{max}) are shown respectively. The plots show synchronization values, derived from windows of 128, 256 and 512 samples respectively, as a function of time.

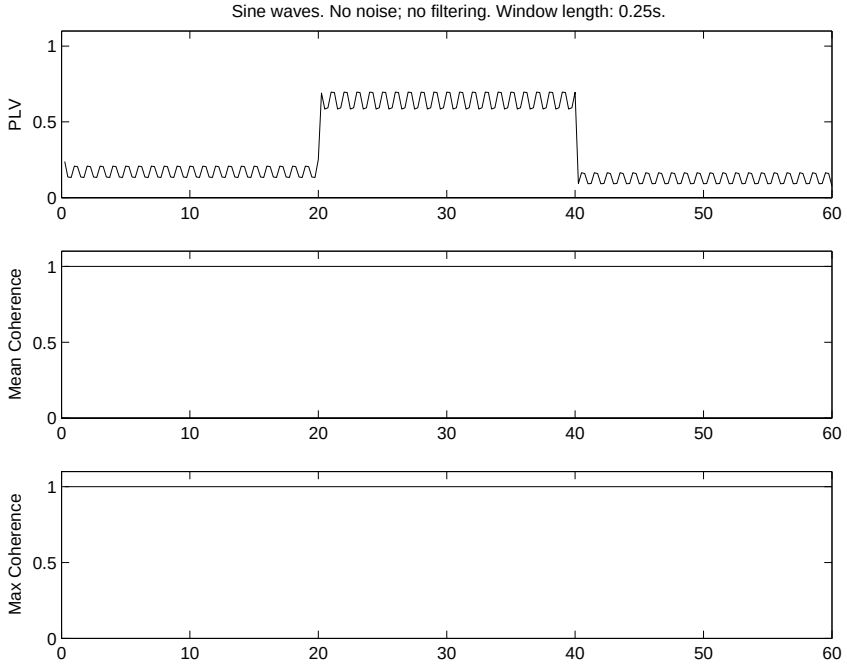

Figure 4.4: Synchronization measures PLV, Coh and Coh_{max} for sine waves. Window length: 0.25s (128 samples). Horizontal axis: t(s).

For very short time windows, Coh does not distinguish different intervals of common activity, but PLV does, although with great variability. Increasing the window length decreases the variance and increases the difference in synchronization values for the different regimes. The variability for short windows originates from the fact that Fourier transforms are performed on fractional periods of the signals, which creates phase jumps. Adding noise smooths this effect.
We also noticed that in the case of one sine wave, the phase jumps over 2π occur at the same time as the minima appear in the curve, only when an exact multiple of the period is chosen as time window.

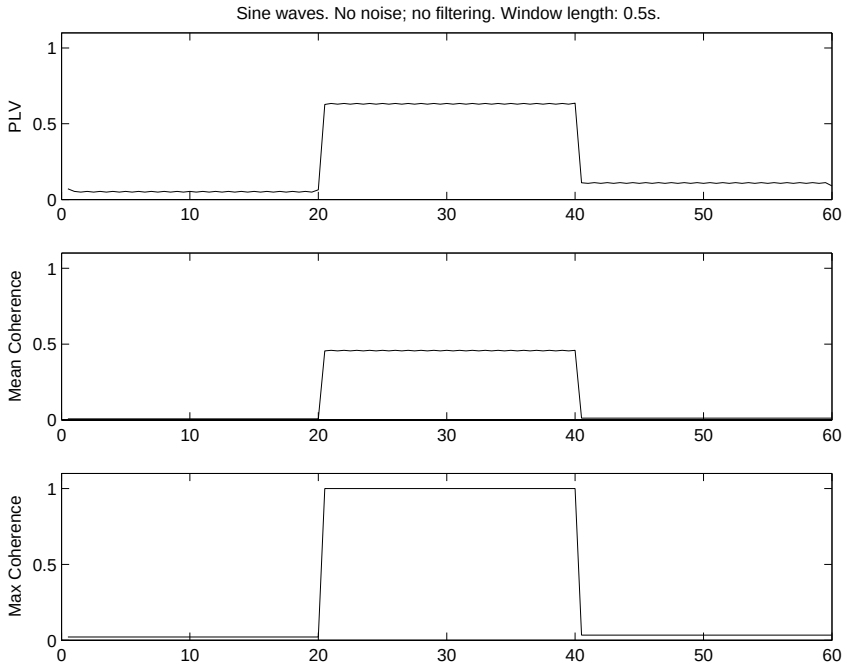

Figure 4.5: Synchronization measures PLV, Coh and Coh_{max} for sine waves. Window length: 0.5s (256 samples). Horizontal axis: t(s).

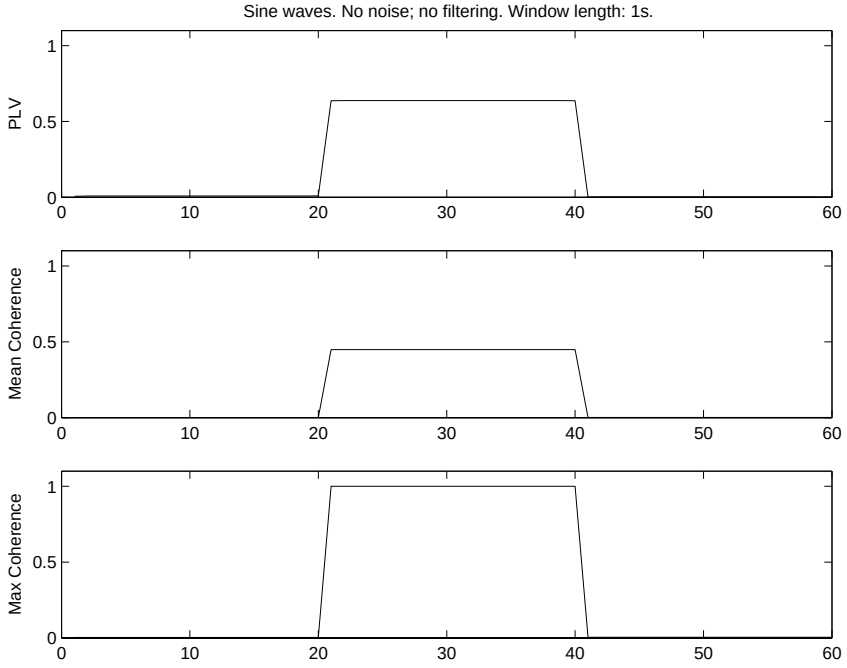

Figure 4.6: Synchronization measures PLV, Coh and Coh_{max} for sine waves. Window length: 1s (512 samples). Horizontal axis: t(s).

For the Rössler oscillator signals, windows of 1s, 5s and 10s were considered, see Figures 4.7, 4.8, 4.9. For 1s windows, the weakly and strongly coupled Rössler oscillator can hardly be distinguished. There is a great variability in PLV, Coh and Coh_{max} values. For 5s windows, approaching the period of the Rössler oscillator signal, PLV is close to 1 for the strongly coupled Rössler oscillator and has slightly smaller values for the weakly coupled case. The computed coherence values are smaller for the strongly coupled than for the weakly coupled case. The Coh_{max}, however, distinguishes the weakly coupled from the strongly coupled case correctly. Only for 10s windows, we see a very clear difference between the weakly and strongly coupled Rössler oscillator for PLV and Coh_{max}. Coh also distinguishes both cases, but less clearly. We can conclude that a window length of at least one period is needed in order to reliably compute the phase.
We also notice a similar behavior for PLV and Coh_{max} in Figures 4.8 and 4.9.

There are some edge effects to observe. As we integrate only over a finite number of samples, instead of infinity as occurs in the definition of the Hilbert transform (equation (4.5)), we better work with overlapping windows and discard some samples at each side of every window.

4.4.2 Influence of Noise

Here, we show examples of noise added to the above presented signals. For the sinusoidal signals, we also present figures for noise added to the amplitude, phase and frequency respectively. The figures display the mean and the mean $\pm$ the standard deviation for PLV, Coh and Coh_{max} as obtained after 50 Monte Carlo runs.

Noise Added to the Signal

Adding a Gaussian white noise with standard deviation equal to the maximum magnitude of the original signal to the Rössler signals results in Figures 4.10 and 4.11. For Figure 4.10 no filtering was applied and it can be seen that neither PLV, Coh nor Coh_{max} distinguishes the weakly coupled Rössler oscillator from the strongly coupled one.
When using a filter (Figure 4.11), the 3 measures present differences for the weakly and strongly coupled case. Most discriminative are PLV and Coh_{max}. The pass bands of the filters for the weakly and strongly coupled Rössler oscillator are chosen according to the frequencies of the respective signals.

Adding a gaussian white noise to the sine waves defined in equations (4.18) and (4.19) results in Figure 4.12 for $\sigma = 1$. The top panel is the non-filtered case, the lower panel presents the measures computed from the noisy signals after application of an 8-30Hz band pass filter. It can be seen that Coh averaged over the whole spectrum should not be used without filtering. PLV and Coh_{max} discriminate the regime of coupling, even without filtering.

Coh can be made more robust by not averaging over all frequencies, but just over a limited frequency band, see Figure 4.13. This requires prior knowledge about the frequencies involved in the system under investigation.

Noise Added to the Amplitude

We added Gaussian white noise to the amplitude of the above considered sinusoidal signals as indicated in the following equations:

$$s_x(t) = (1 + 1.5\zeta(t))(\cos(2\pi 13t) + \cos(2\pi 22t))/2 \quad \forall t \tag{4.21}$$

$$s_y(t) = \begin{cases} (1 + 1.5\zeta(t))\cos(2\pi 6t) & \forall t = \frac{1}{f_s}...\frac{t_{tot}}{3} \\ (1 + 1.5\zeta(t))\cos(2\pi 13t) & \forall t = \frac{t_{tot}}{3} + \frac{1}{f_s}...\frac{2t_{tot}}{3} \\ (1 + 1.5\zeta(t))\cos(2\pi 17t) & \forall t = \frac{2t_{tot}}{3} + \frac{1}{f_s}...t_{tot} \end{cases} \tag{4.22}$$

Figure 4.14 shows the considered synchronization measures computed from the signals with noise added to the amplitude, without filtering and after application of an 8-30Hz band pass filter respectively. Again,

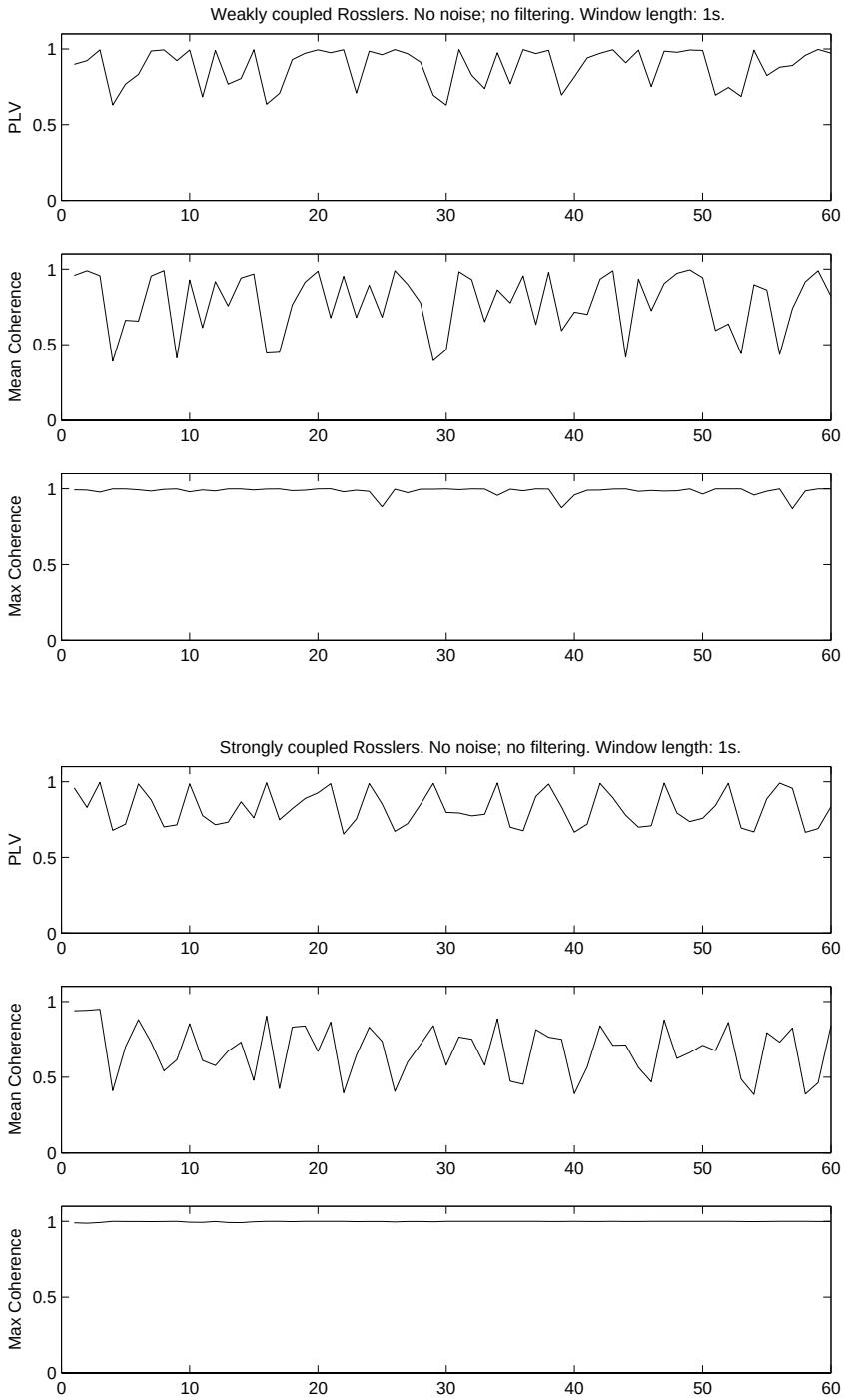

Figure 4.7: Synchronization measures PLV, Coh and Coh_{max} for weakly and strongly coupled Rössler oscillator system. Window length: 1s. Horizontal axis: t(s).

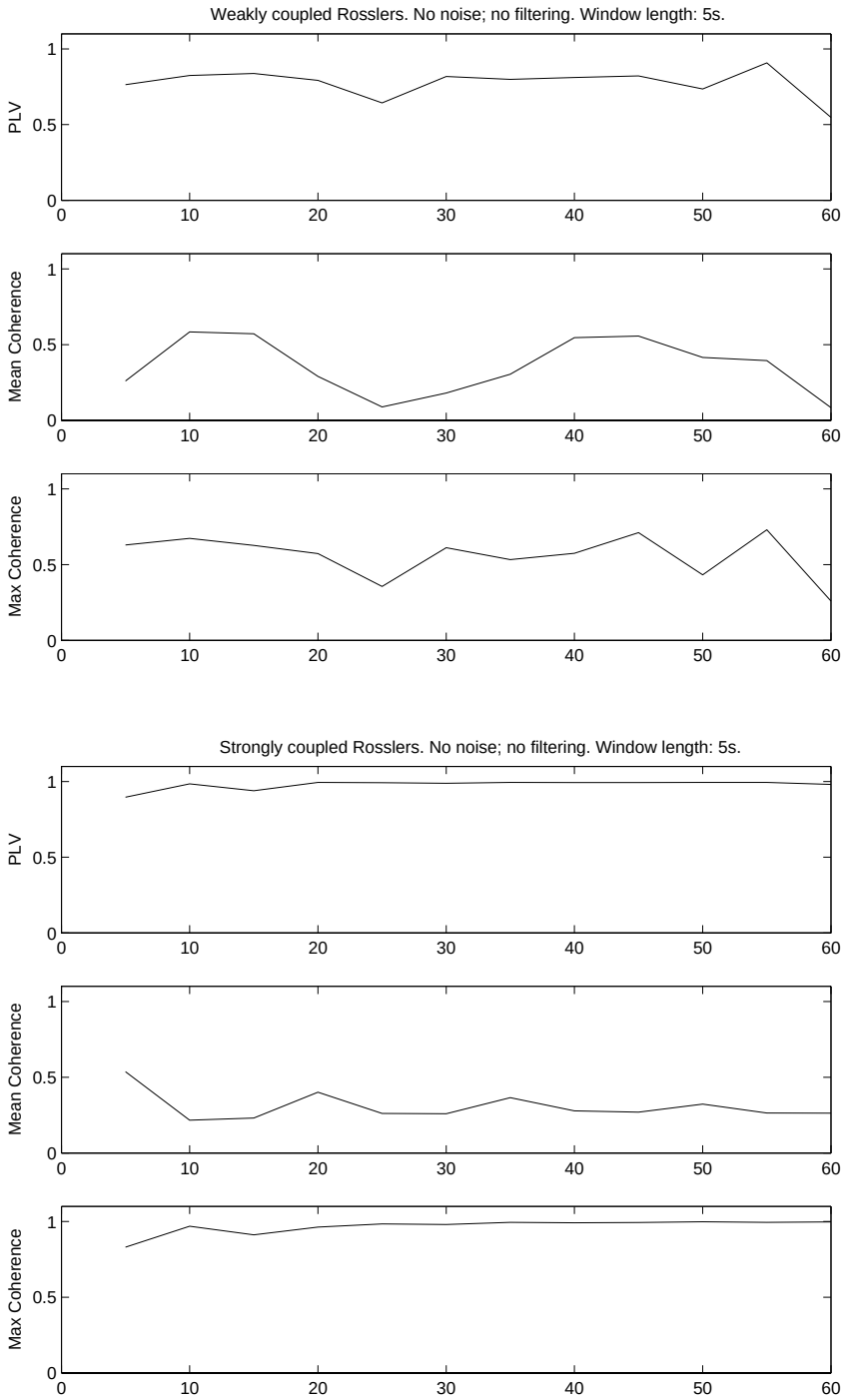

Figure 4.8: Synchronization measures PLV, Coh and Coh_{max} for weakly and strongly coupled Rössler oscillator system. Window length: 5s. Horizontal axis: t(s).

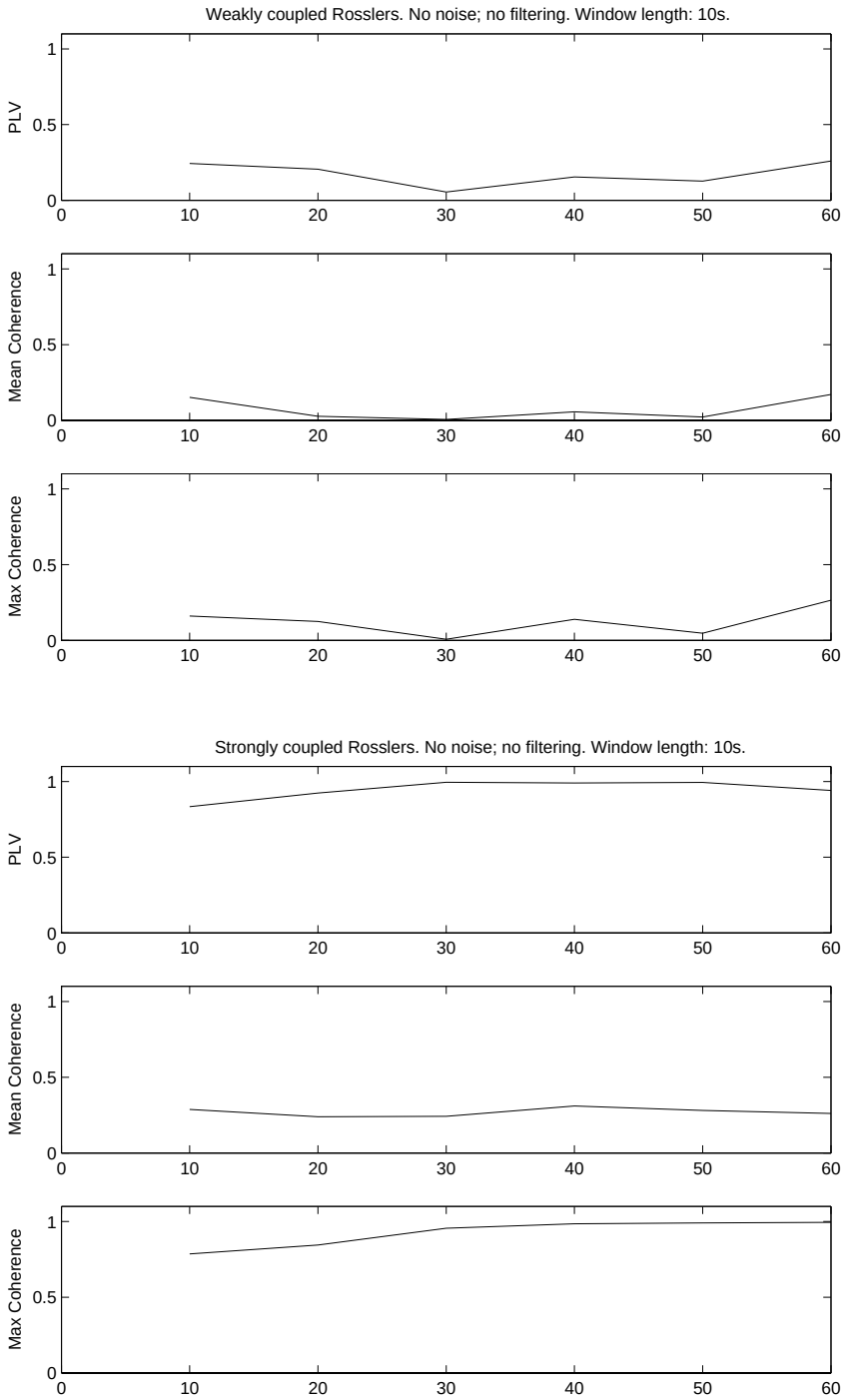

Figure 4.9: Synchronization measures PLV, Coh and Coh_{max} for weakly and strongly coupled Rössler oscillator system. Window length: 10s. Horizontal axis: t(s).

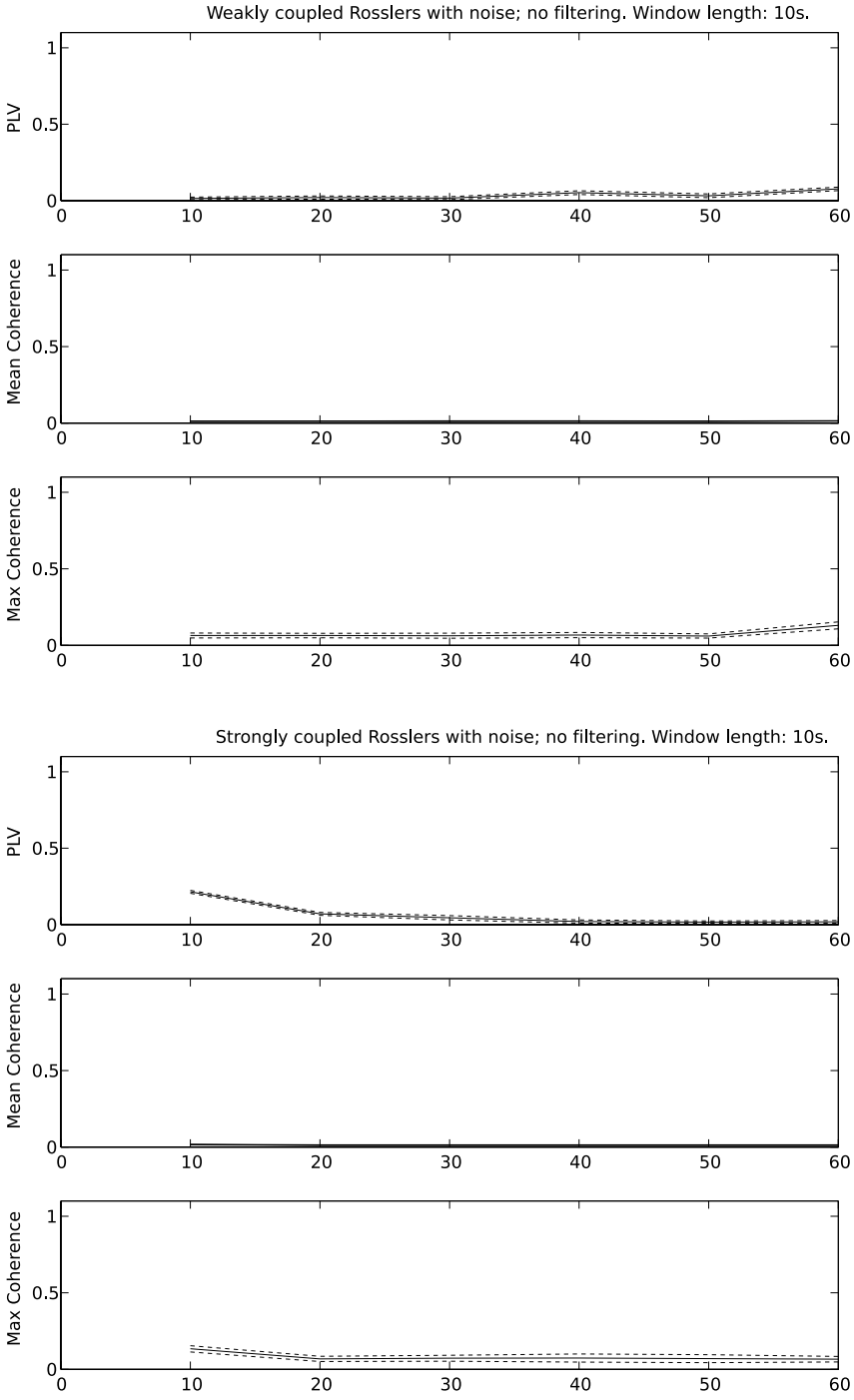

Figure 4.10: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for weakly (top) and strongly (bottom) coupled Rössler oscillator. Window length= 10s; $\sigma = 1$; no filtering. Horizontal axis: t(s).

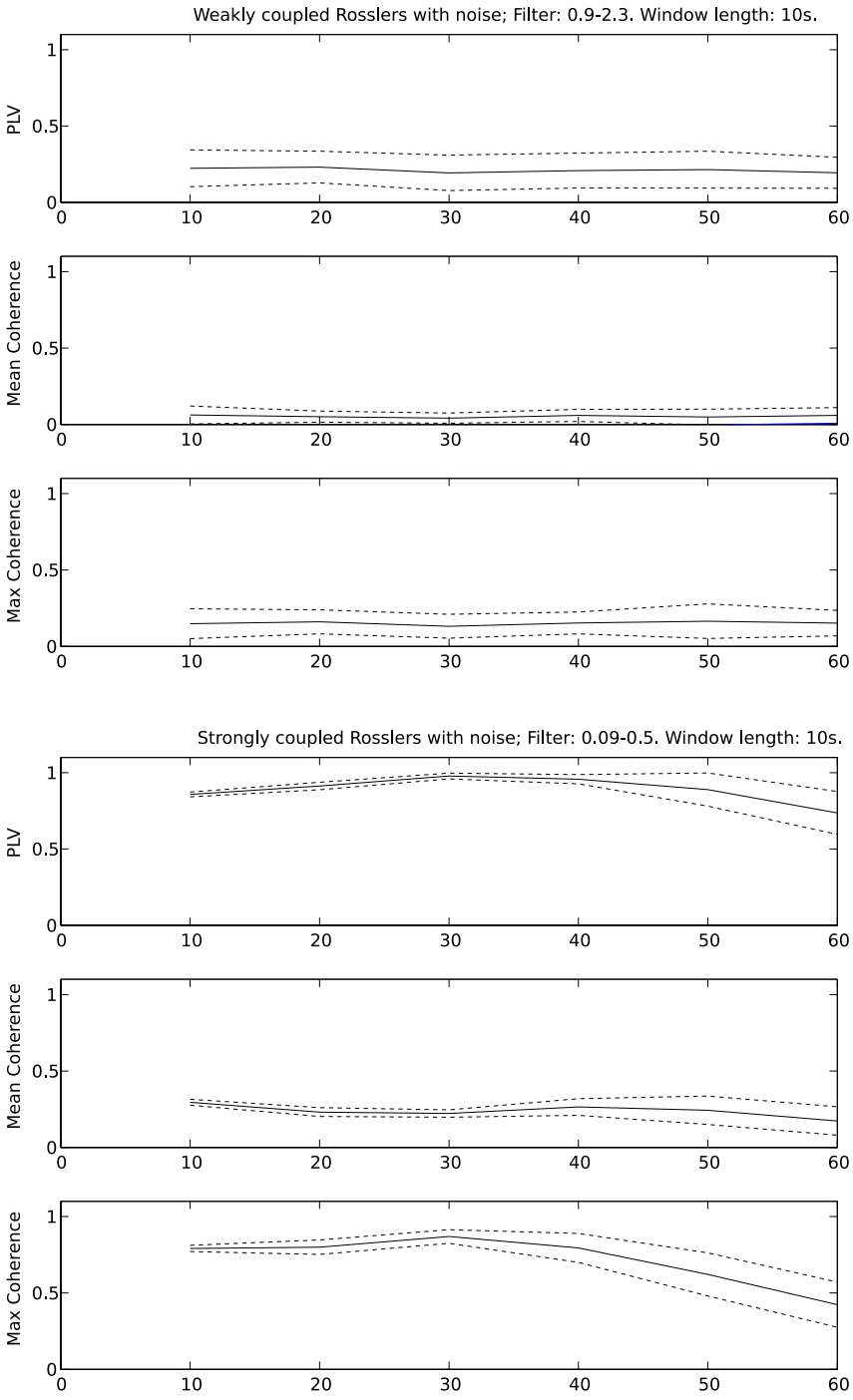

Figure 4.11: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for weakly (top) and strongly (bottom) coupled Rössler oscillator. Window length= 10s; $\sigma = 1$; 0.9-2.3Hz filter and 0.09-0.5Hz filter respectively. Horizontal axis: t(s).

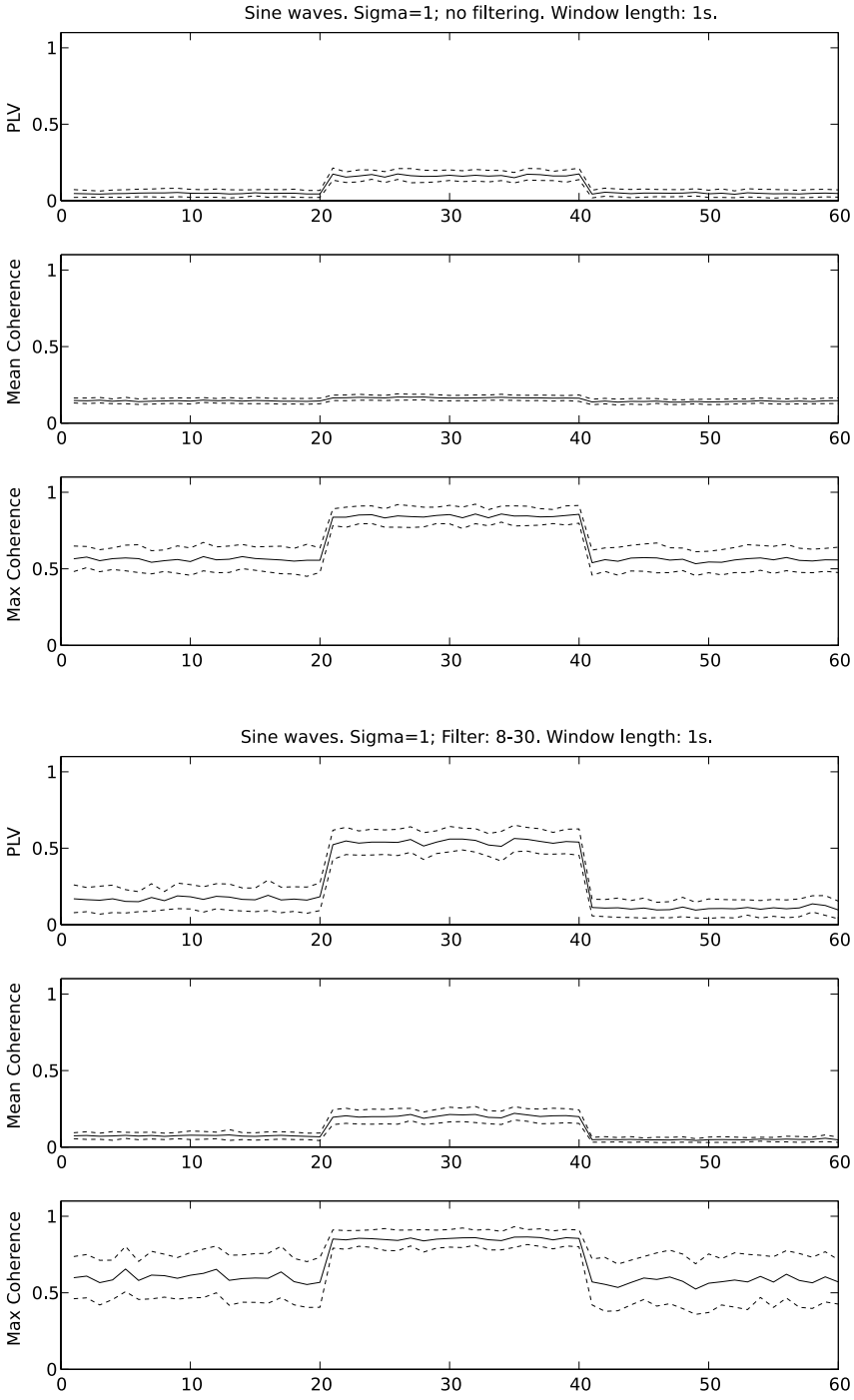

Figure 4.12: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for sinusoidal signals with noise added to the signal. Window length= 1s; $\sigma = 1$; (top): no filter applied; (bottom): 8-30Hz band pass filter applied. Horizontal axis: t(s).

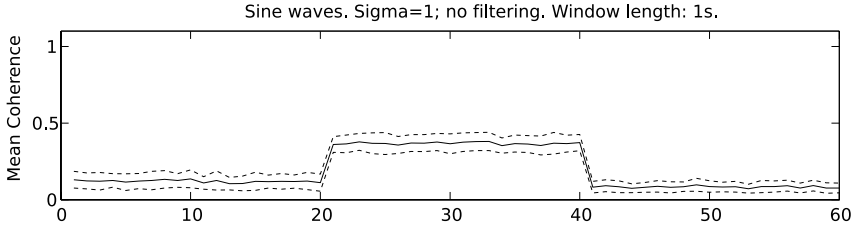

Figure 4.13: Coh (± standard deviation) for sine waves. Window length= 1s; $\sigma = 1$; no filtering, but Coh computed from values at frequencies between 8-30Hz only. Horizontal axis: t(s).

Coh distinguishes the period of coupling only after application of a filter. PLV and Coh_{max} can be used without filtering.

Noise Added to the Phase

Next, we add a Gaussian white noise to the phase of the sinusoids:

$$s_x(t) = (\cos(2\pi 13t + 0.23 * 2\pi\zeta(t)) + \cos(2\pi 22t + 0.23 * 2 * \pi\zeta(t)))/2 \quad \forall t \tag{4.23}$$

$$s_y(t) = \begin{cases} \cos(2\pi 6t + 0.23 * 2\pi\zeta(t)) & \forall t = \frac{1}{f_s}...\frac{t_{tot}}{3} \\ \cos(2\pi 13t + 0.23 * 2\pi\zeta(t)) & \forall t = \frac{t_{tot}}{3} + \frac{1}{f_s}...\frac{2t_{tot}}{3} \\ \cos(2\pi 17t + 0.23 * 2\pi\zeta(t)) & \forall t = \frac{2t_{tot}}{3} + \frac{1}{f_s}...t_{tot} \end{cases} \tag{4.24}$$

The resulting PLV, Coh and Coh_{max} are shown in Figure 4.15. Without filtering, only Coh_{max} can distinguish the different regimes of coupling. Interestingly, application of filtering makes Coh_{max} distinguish less clearly the different regimes. PLV discriminates the periods of coupling from the other periods, Coh hardly does.

Noise Added to the Frequency

Last, we study the influence of noise added to the frequency of sinusoidal signals:

$$s_x(t) = (\cos(2\pi(13 + 0.009\zeta(t))t) + \cos(2\pi 22t + 0.23 * 2 * \pi\zeta(t)))/2 \quad \forall t \tag{4.25}$$

$$s_y(t) = \begin{cases} \cos(2\pi(6 + 0.009\zeta(t))t) & \forall t = \frac{1}{f_s}...\frac{t_{tot}}{3} \\ \cos(2\pi(13 + 0.009\zeta(t))t) & \forall t = \frac{t_{tot}}{3} + \frac{1}{f_s}...\frac{2t_{tot}}{3} \\ \cos(2\pi(17 + 0.009\zeta(t))t) & \forall t = \frac{2t_{tot}}{3} + \frac{1}{f_s}...t_{tot} \end{cases} \tag{4.26}$$

As expected, the synchronization measures fail to detect the periods of coupling in this case 4.16. After filtering the signal in the 8-30Hz or 11-15Hz frequency band 4.17, the onset of the period of synchronization can be observed, but not an increased synchronization for the whole period of coupling.

Conclusions

After adding Gaussian white noise with a standard deviation equal to the maximum magnitude of the original signal to uncoupled and coupled Rössler oscillators, none of the measures PLV, Coh and Coh_{max} distinguishes uncoupled from coupled Rössler oscillators. After improving the SNR by filtering, however, PLV and Coh_{max} do. In case of sine waves, PLV and Coh_{max} discriminate coupling and non-coupling, even without application of a filter. PLV and Coh_{max} are robust to addition of noise to the amplitude. To a certain degree, they support noise on the phase and they are not robust at all to noise on the frequency. For PLV, these findings were confirmed by a mathematical analysis [25].

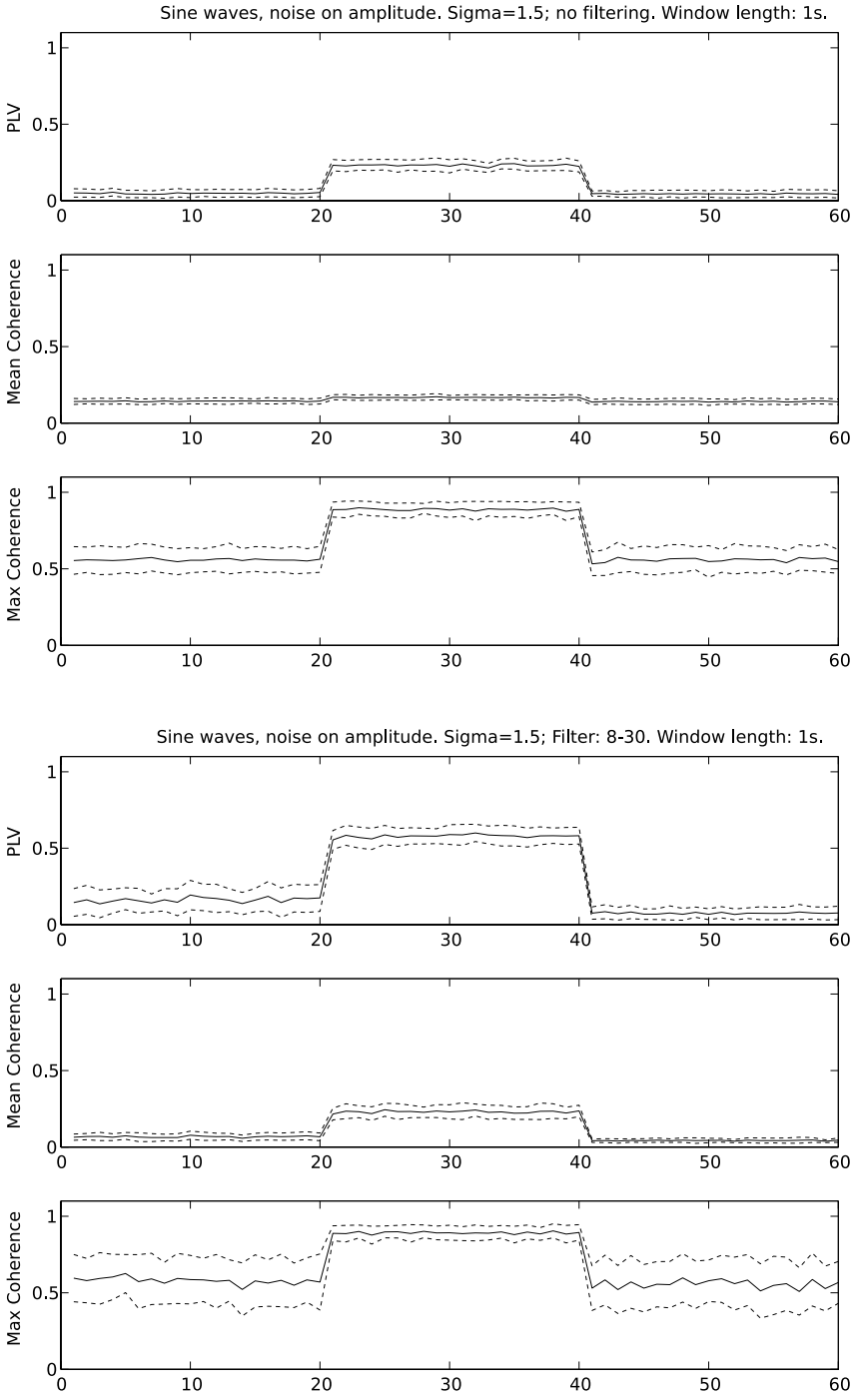

Figure 4.14: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for sinusoidal signals with noise added to the amplitude. Window length= 1s; $\sigma = 1.5$; (top): no filter applied; (bottom): 8-30Hz band pass filter applied. Horizontal axis: t(s).

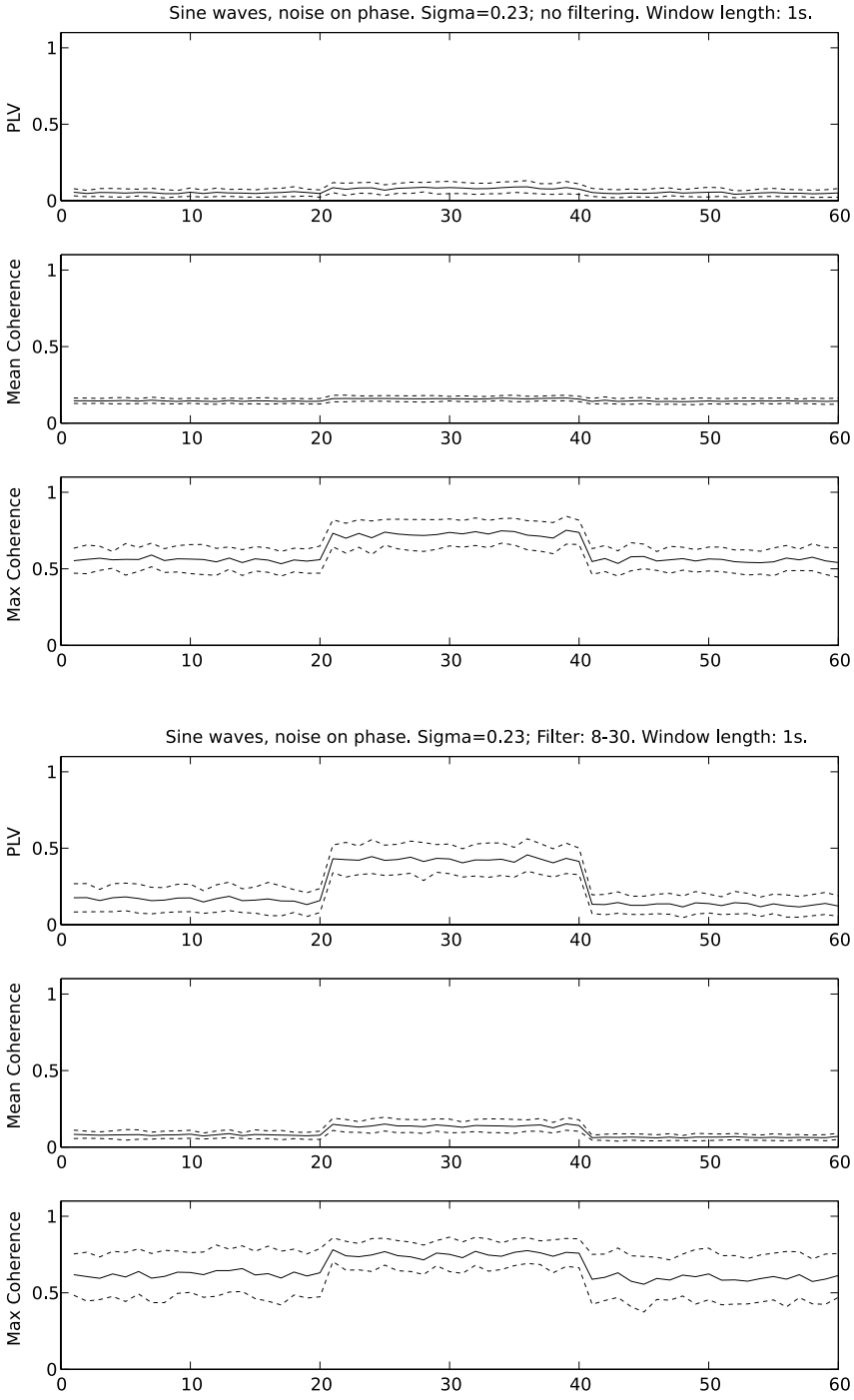

Figure 4.15: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for sinusoidal signals with noise added to the phase. Window length= 1s; $\sigma = 0.23$; (top): no filter applied; (bottom): 8-30Hz band pass filter applied. Horizontal axis: t(s).

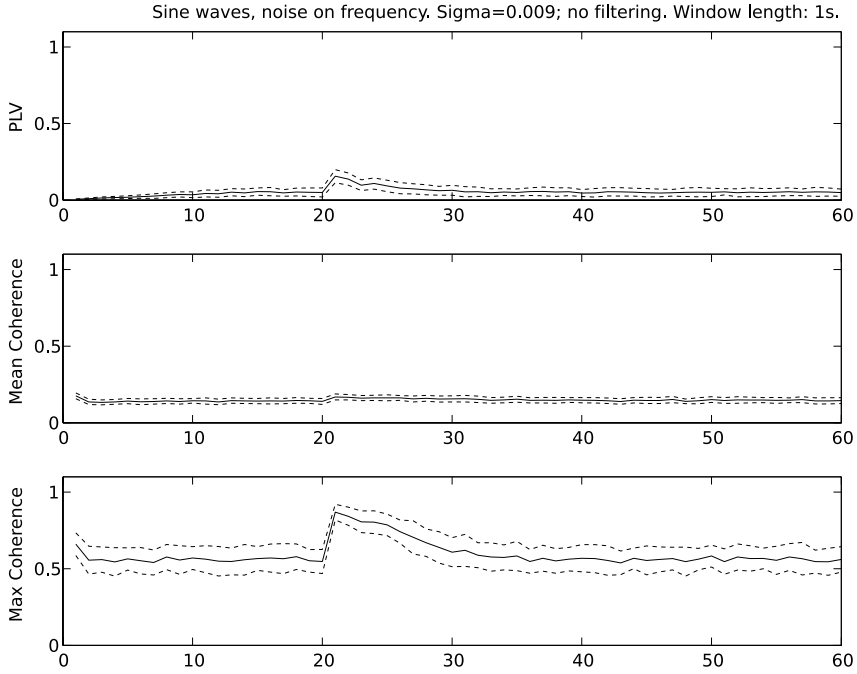

Figure 4.16: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for sinusoidal signals with noise added to the frequency. Window length= 1s; $\sigma = 0.009$; No filter applied; Horizontal axis: t(s).

4.4.3 Computation Time

The computation time is considerably smaller for PLV than for coherence, if calculation of PLV is implemented as in Section 4.2. This is an important asset for BCI applications, necessitating online computation of features.

4.4.4 Other

Coherence depends on the parameters for its computation (number of points for Fourier transform, overlap) and the measure becomes unreliable if these parameters are not chosen properly. There is a trade-off between temporal and spectral resolution on the one hand and between robustness and temporal resolution on the other hand.

4.5 PLV Computed from Broadband Signals

4.5.1 Broadband vs Narrowband: Example Epileptic Seizures

In previous research [64], we examined very long EEG recordings of epileptic patients, up to 90 hours per patient. From the 26 seizures recorded, only one seizure, from a patient with left medial temporal lobe epilepsy, showed clear changes in PLV related to the seizure [62]. The PLV was determined from 15s windows.
About two hours before the seizure, the PLV decreased significantly and remained at this level till seizure onset. Interestingly, we observed this phenomenon only in the frequency bands 1-99Hz and 1-49Hz. In an attempt to localize it in a smaller frequency band, we noticed that this loss in synchrony preceding the seizure, was not present in any of the frequency bands 51-99Hz, 1-40Hz, 40-49Hz and 20-49Hz.

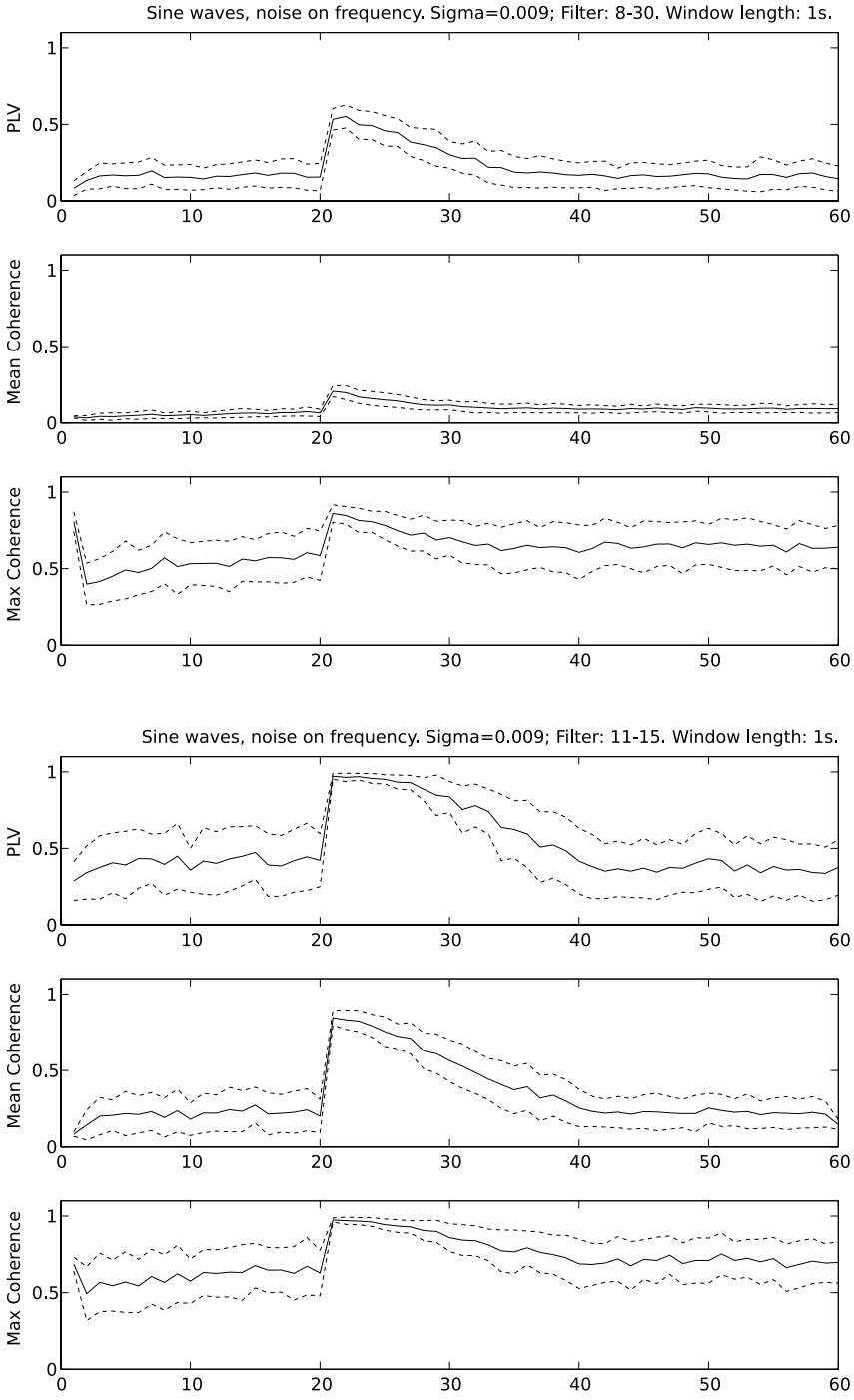

Figure 4.17: Synchronization measures PLV, Coh and Coh_{max} ($\pm$ standard deviation) for sinusoidal signals with noise added to the frequency. Window length= 1s; $\sigma = 0.009$; (top): 8-30Hz band pass filter applied; (bottom): 11-15Hz band pass filter applied. Horizontal axis: t(s).

4.5.2 Broadband vs Narrowband: Example Visual Binding Paradigm

This example is taken from EEG data obtained in an experiment based on the well-known trend of the visual system to group iso-oriented collinear patterns and to segment cross-oriented ones. Nine subjects viewed centrally presented bilateral moving gratings (Figure 4.18), which either obeyed Gestalt grouping rules (iso-oriented, IG) or violated them (orthogonally-oriented, OG). IG stimuli reliably increase interhemispheric synchronization compared to OG stimuli.
In [77], electrode 70 in the left occipital region was identified as an electrode of interest and an increased interhemispheric coherence was detected at 22Hz. The left panel of figure 4.19 visualizes the difference in PLV between the IG and OG stimuli conditions. PLV between electrode 70, in the left occipital region, and all other electrodes is shown, computed from both narrow (21-23Hz) and broad (3-70Hz) band filtered EEG signals. The right panel indicates for which electrodes this difference was significant (Wilcoxon test).

It can be seen that in the narrowband case, the two stimuli conditions can be hardly distinguished [63]. When computed from broadband EEG, however, many significant differences between both stimuli conditions are detected. Most of them concern interhemispheric synchronization.

4.5.3 Discussion

These findings show that we should not only consider PLV computed from narrowband signals and suggest that the PLV compares phases referred to instantaneous frequencies from two signals. They are determined by all frequencies present in the signals in a certain time interval. Unfortunately, a mathematical expression telling us how to ascertain the local (instantaneous) frequency from the frequency content of a signal is not available.

Frequency and phase are clearly defined for sinusoidal signals, but not for, for example, signals composed of several sinusoids or biological signals, such as EEG and MEG. Instead of using the neurophysiological definition of synchrony, assuming two or many subsystems sharing specific common frequencies, we prefer the definition of synchronization used in system theory referring to a process whereby two or many subsystems adjust some of their time-varying properties to a common behavior due to coupling or common external forcing. This general definition of synchronization does not imply a relationship between systems at specific frequencies but at all of them [23].
We can also consider the difference between PLV and coherence from this point of view. Coherence reflects a coupling at specific frequencies and can be very usefull to study systems exhibiting interaction between subsystems at specific frequencies. PLV indicates an adjustment of the time-varying properties.

PLV is mostly computed from narrowband filtered signals, e.g. [84, 11], because only then phase has a clear physical meaning. However, PLV can be computed also for broadband signals and only in this case a synchronization in the sense of an adjustment of the time-varying properties can be detected. Indeed, filtering the signals in narrow frequency bands removes all components that could desynchronize the signals and as such artificially increases the computed synchrony.
If we filter the signals in a narrow frequency band in which the signal has no power, the signal will be suppressed. However, the signal will never be exactly 0. Small fluctuations will still be present, inducing a phase. Studying the PLV in a narrow frequency band may be a detection of the presence of the specific frequency in both signals.

4.5.4 Phase Defined from Minima and Maxima in Noise-free Signals

To detect phase locking, the instantaneous frequencies should be about the same. The phase is defined with regard to a period (frequency), so phase synchronization comes naturally with having a common (instantaneous) frequency.
Each frequency component will generate minima and maxima in the signal. Therefore, we propose, as an illustration, to define the phase from the minima and maxima of the signals (here assumed to be without noise). We define a local period, and from this an instantaneous frequency, as the distance between the minima (or the maxima) of the signal we are observing.

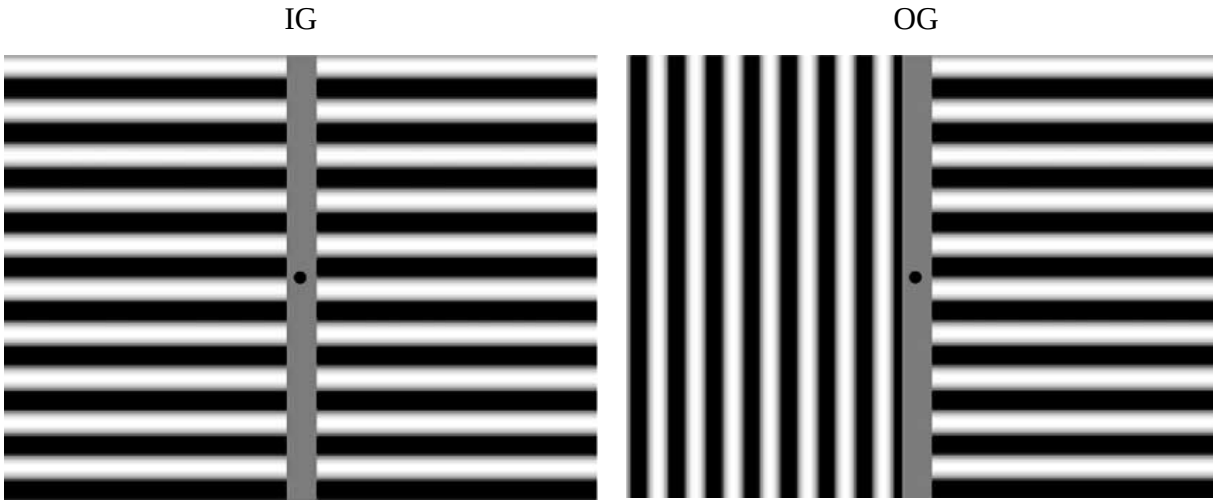

Figure 4.18: Iso-oriented and orthogonally-oriented grating

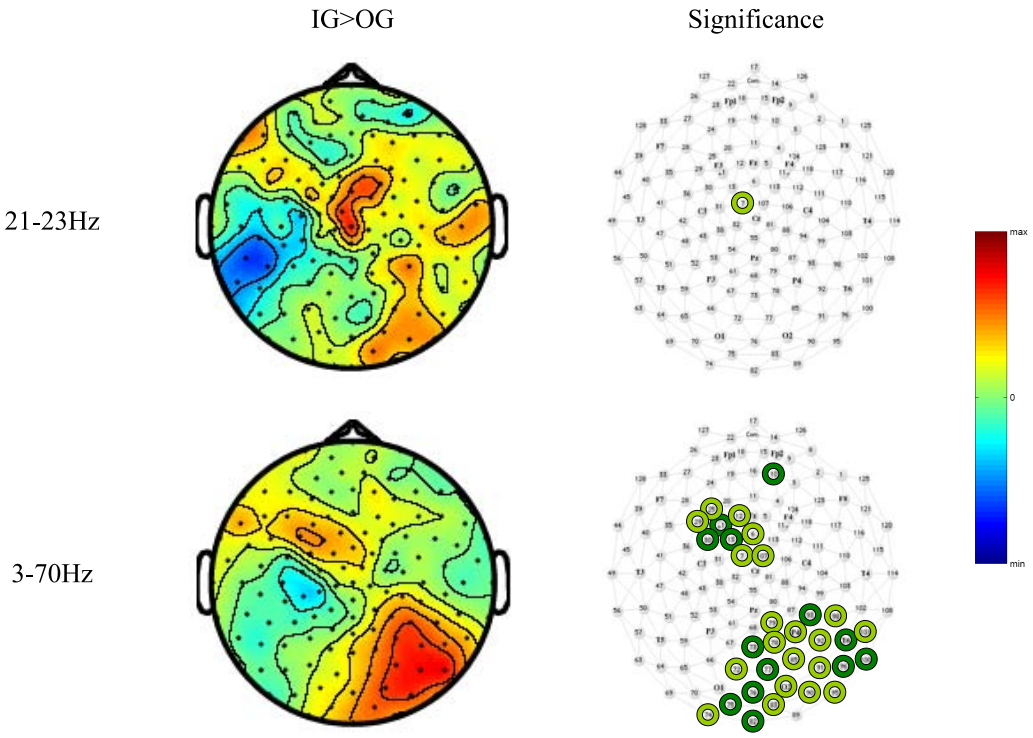

Figure 4.19: For IG vs OG, difference in PLV referred to electrode 70 (left occipital), computed from narrowband (top) and broadband (bottom) filtered EEG signals; and electrodes where this difference was significant (right)

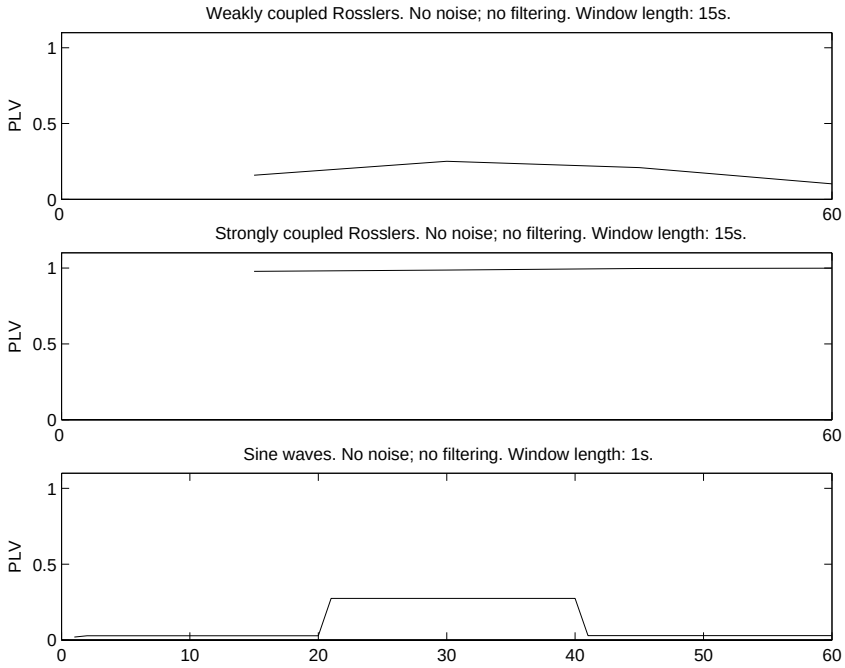

Figure 4.20: PLV_m for the weakly and strongly coupled Rössler oscillator systems, shown in figures 4.2 and 4.3 respectively, and between the sinusoids shown in figure 4.1. The respective moving windows were 15s, 15s and 1s long. Horizontal axis: t(s).

To test the idea of a local frequency, defined from the distance between minima or maxima of a signal, we calculate the phase of a signal as follows. We detect the minima and maxima, and let the phase increase linearly from $-\pi$ to 0 between a minimum and a maximum, and from 0 to π between a maximum and a minimum. Subsequently, the PLV_m is calculated from the phase difference between the two signals. Thus, the only difference with the PLV described in Section 4.2, is the way the 'instantaneous', local, phase is calculated.

We see in the third plot of figure 4.20 that for the two sinusoids, even though less significant than in the first plot of figure 4.6, the presence of common activity is detected. The distinction between the coupled and uncoupled Rössler oscillators is very clear, as we can see in the first and second plot of figure 4.20. The difference between these results, compared to those obtained in Paragraph 4.4.1, may be explained by the fact that no difference is made between the local and global minima and maxima. They are treated equally if we only consider the fact that, locally, a minimum or a maximum exists. The Hilbert transform, however, changes the slope of the phase curve or induces a small drop of phase in the plots of instantanous phase.

4.6 Synchronization for the Recognition of Mental Tasks in a BCI

4.6.1 Introduction

This section introduces synchronization features applied for the recognition of mental tasks in a BCI [59].

We use PLV and a coherence-based measure, Coh_{max}, to study synchronization for EEG-based classification of mental tasks. Coh_{max} was preferred to Coh, because its properties proved to be better (see Section 4.4).

Considering twice (for PLV and Coh_{max}) 496 electrode pairs together with power features in a few frequency bands would result in a huge amount of features, making a comparison of PLV, Coh and PSD features very difficult. Therefore, we define different features from the pair-wise synchronization mea-

sures. Additional features are defined from PSD estimates. For the actual comparison of the different types of features, 53 different subsets of the available set of features were considered.

First, Paragraph 4.6.2 describes the here applied preprocessing. Then, different synchronization features are proposed in Paragraphs 4.6.3, 4.6.4, 4.6.5 and 4.6.6. Band power estimates are considered for comparison (see Paragraph 4.6.7). From the obtained set of features, different subsets were defined (Paragraph 4.6.8).

Features were computed from sliding 1s windows, 8 times per second, in the frequency band 8-30Hz. The length of 1s was chosen arbitrarily. The shorter the windows, the less accurate the estimation of the features, while longer windows make the system react slowlier.

4.6.2 Data Preprocessing

If the mental task changed, the corresponding 1-second window and its subsequent window were removed from the recorded EEG data. For each day of recording, about 1050 windows remained available for classification.
No artifact detection or denoising techniques were applied. It was not necessary to exclude artifacts to be sure to classify brain activity (rather than artifacts).

Preprocessing of Synchronization Features

A linear phase FIR 8-30Hz band-pass filter of order 1023 was applied to the EEG before computing the synchronization measures (PLV and maximum spectral coherence). The 8-30Hz frequency band was chosen because it contains the μ and β rhythms, which have been proved to be involved in motor-tasks [116].

Preprocessing of Power Features

Before calculating power features, a Laplacian filter [97] was applied, followed by the same band-pass filter as applied to the computation of the synchronization features. The surface Laplacian was computed from a spherical spline interpolation scheme [114, 115]. Its application to the scalp surface potential was motivated by the fact that it has been shown to be an estimate of the cortical surface potential, which is a more accurate estimation of the cortical activity than the surface EEG.

4.6.3 Synchronization Based Features

We propose different features based on PLV and Coh_{max}, here representing the maximum coherence in the frequency band $\Delta B = 8 - 30Hz$, that assess the interaction between EEG signals recorded at neighboring electrodes.
To compute the discrete Hilbert transform, the number of samples for fast Fourier transform, N_{fft}, was set to $f_s{*}1s$. After visual inspection of the coherence spectra, the parameters for computing the coherence were set as follows: N_{fft} was $2.f_s.1s$; a Hanning window of length $\frac{N_{fft}}{8}$ was used; the overlap was $\frac{N_{fft}}{16} - 1$.

From the PLV and Coh_{max} for all 496 electrode pairs, we computed the following features: general levels of synchronization in different regions (Paragraph 4.6.4), the coupling between signals recorded from an electrode and its neighbors (Paragraph 4.6.5), and a few of the individual electrode pairs (Paragraph 4.6.6).

4.6.4 Synchronization Based Features: Group Averages

9 averages over respectively the PLV and Coh_{max} values of all possible electrode pairs within a group of electrodes were considered. These 9 group averages are illustrated in the left panel of Fig. 4.21. One

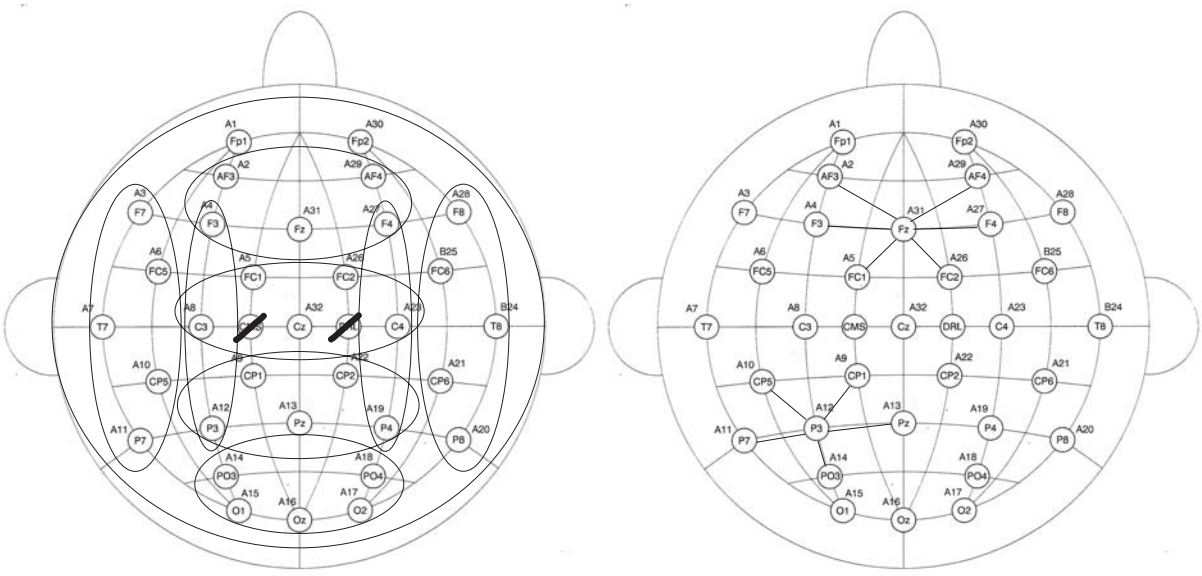

Figure 4.21: Illustration to the Synchronization Features Computed from the EEG Signals. See text for details.

value represented the average over all possible electrode pairs, indicating a general level of synchronization. Furthermore, 6 averages for a 'frontal' (AF3, AF4, F3, Fz, F4), 'central' (FC1, FC2, C3, Cz, C4), 'parietal' (CP1, CP2, P3, Pz, P4), 'occipital' (PO3, PO4, O1, Oz, O2), 'left' (F7, T7, P7, FC5, CP5) and 'right' (F8, T8, P8, FC6, CP6) group were obtained. Two more values resulted from averaging over the pairs F3-C3, F3-P3 and C3-P3 on the one hand and F4-C4, F4-P4 and C4-P4 on the other hand.

4.6.5 Synchronization Based Features: Interaction with Neighbors

Furthermore, each electrode was attributed two values corresponding to the averaged PLV and Coh_{max} values of the electrode pairs formed by this electrode and its neighbors. The right panel of Fig. 4.21 shows an example for two electrodes.

4.6.6 Synchronization Based Features: Details in the Fronto-centro-parietal and Temporal Region

At last, 40 synchronization values for individual electrode pairs in the fronto-centro-parietal region (formed with electrodes F3, C3, P3, Fz, Cz, Pz, F4, C4, P4) were considered, for both PLV and Coh_{max}, because of the involvement of this part of the cortex in the motor tasks. Unfortunately, this electrode placement has only one electrode in the temporal region, which is involved in language tasks. Therefore, we included the synchronization values between the signals recorded at electrodes T7-FC5, T7-CP5, T8-FC6 and T8-CP6.

This specific choice of averages and individual electrode pairs was arbitrary, but symmetrical. It relied on the one hand on physiological information regarding motor and language tasks, on the other hand we wanted to reduce the number of features that would result from considering all possible electrode pairs and enhance their relevance by combining certain of them. In fact, in cognitive neuroscience there are often just one or a few electrode pairs considered, or all pairs are summed.

4.6.7 Band Power Estimates

The PSDs in the α (8-12Hz), β_1 (13-18Hz) and β_2 (19-30Hz) frequency bands, as well as the PSD in the larger band 8-30Hz, which we will denote with λ, were computed for the 32 EEG signals by means of the

Welch's averaged Power Spectral Density estimate from 1s sliding windows, 8 times per second. N_{fft} was set to $\frac{f_s.1s}{2}$, the overlap to $\frac{N_{fft}}{2}$. A Hanning window of length N_{fft} was used.

As the classifier needs numbers of the same order of magnitude at its input, every PSD feature was rescaled by dividing it by the average over all PSD features of the same frequency at the same time instance. The same procedure was applied to the PLV and Coh_{max} features.

4.6.8 Subsets of Features

With the total of 162 synchronization features (81 PLV and 81 Coh_{max}) and 128 power features, 53 different groups of features, called subsets, were constructed and tested on the available data. Some subsets consisted of one type of feature only (PLV, Coh_{max}, α, β_1, β_2 and λ PSD features), other subsets combined PSDs with synchronization based features. The PSD features or the synchronization measures at single electrodes, as defined in Paragraph 4.6.5, were considered at all electrodes or only at C&P electrodes (C3, Cz, C4, P3, Pz, P4). Also a combination of synchronization values for individual electrode pairs in the fronto-centro-parietal region was considered, as well as their combination with α power features at the C&P electrodes.
A detailed description of all of the 53 subsets would lead us too far, but, as an example, the best performing subsets for the 3 best subjects will be described in detail in Paragraph 4.7.3.

4.7 Results for the Use of Synchronization features to Classify Mental Tasks

4.7.1 Introduction

In this section we present the results from our experiments on the application of synchronization for the EEG-based discrimination of mental tasks, using the features presented in Section 4.6. We will first, in Paragraph 4.7.2, discuss some general observations for the 5 subjects. Then, in Paragraph 4.7.3, we show the classification accuracies for the best subsets and describe these subsets in more detail. Subsequently, the sole use of synchronization and power features will be illustrated (see Paragraph 4.7.4). In Paragraph 4.7.5, the different types of features are compared by means of significance tests. The results are discussed in Paragraph 4.7.6.

The tables present, for each day and each subject, results from 5-fold cross-validation, using 4 sessions for training and the 5th one for testing. Each of the 5 sessions was test set once. 'CR' refers to the average percentage of correctly classified windows. 'UR' is the percentage of windows for which the system responds 'unknown' (neither correct classification, nor mistake). The remaining percentage is the error rate ('ER'). For this study, we used an SVM classifier, described in Paragraph 3.6.2.

4.7.2 Observations on the Different Subjects

From a global inspection of the performances of the 5 subjects, it appeared that subjects 2 and 4 did not achieve good classification results. For many subsets, classification accuracies were hardly better than random (33%). Subject 2, trying to avoid eye blinking during recording of the EEG, achieved best performance on day 1. In case of Subject 4, results were better for the 2nd day of recording. This subject reported getting bored with the paradigm.
Possibly these subjects did not well perform the mental tasks.

The performance of subject 1 increases over the 3 days. This subject reported difficulties with imagining the hand movements the first day and felt comfortable with the task only the third day. Subject 3 performs slightly less on days 2 and 3 than on day 1. We learned afterwards that this subject had a minor car accident between day 1 and 2. Furthermore, he reported he got used to the procedure. The performance of subject 5 was better for the first day and the last one. The subject reported that he felt

			1 subset		3 subsets	
Subject	Day	Type	CR(%)	UR(%)	CR(%)	UR(%)
1	1	PLV + Coh_{max}	57.65	6.51	57.55	6.14
	2	α	63.15	7.24	66.10	6.22
	3	PLV + α	67.81	6.77	71.29	5.20
2	1	PLV + β_1	45.05	7.95	46.82	7.67
	2	All	41.87	10.00	44.52	6.06
	3	PLV + β_2	41.25	3.50	41.45	6.14
3	1	Coh_{max} + α	64.85	7.81	66.23	8.19
	2	Coh_{max} + λ	62.21	7.64	65.39	5.77
	3	Coh_{max} + α	60.63	9.06	61.47	8.50
4	1	All	42.18	10.48	45.16	9.65
	2	Coh_{max} + all PSD	49.48	10.53	47.41	8.84
	3	PLV + all PSD	44.66	8.22	47.98	8.21
5	1	PLV	55.14	6.45	58.50	5.70
	2	PLV	46.38	8.09	48.78	7.16
	3	PLV + Coh_{max}	52.97	7.06	57.34	7.07

Table 4.1: For the respective subjects and days of recording: The types of features contained in the best feature subset and the corresponding results for 3-class distinction; Results for 3-class discrimination using for the distinction of each task pair the respective best feature subset

more attentive day 1 and part of day 3.

4.7.3 Best Feature Subsets

Table 4.1 shows, for the 3 days of recording of each subject, the types of features contained in the set of features achieving best classification accuracy and the 3-class classification results obtained with these subsets. In all but one case, synchronization-based measures are contained in the best performing feature subset. 11 out of 15 subsets contain power features.

Each window contains 32*512 = 1664 samples, and is then characterized by a few tens of feature values. Here follow, as an example, the details of the best features subsets, for the different days of recording of the 3 best subjects:

- Subject 1, day 1: PLV and Coh_{max}, values for the electrode pairs described in Paragraph 4.6.6 (80 features).
- Subject 1, day 2: PSD in the α band, all electrodes (32 features)
- Subject1, day 3: PLV, group values and values at C&P electrodes; and PSD in the α band, all electrodes (47 features)
- Subject 3, day 1: Coh_{max}, group values and values at C&P electrodes; and PSD in the α band, all electrodes (47 features)
- Subject 3, day 2: Coh_{max}, group values and values at C&P electrodes; and PSD in the λ band, all electrodes (47 features)
- Subject 3, day 3: Coh_{max}, group values and values at C&P electrodes; and PSD in the α band, at C&P electrodes (21 features)
- Subject 5, day1: PLV values for the electrode pairs described in Paragraph "Details in the fronto-centro-parietal and temporal region" (40 features).
- Subject 5, day 2: PLV for the electrode pairs C3-F3, C3-Cz, C3-P3, C4-F4, C4-Cz, C4-P4, P3-Pz, P4-Pz, T7-CP5 and T8-CP6 (10 features)
- Subject 5, day 3: PLV and Coh_{max}, group values and values at C&P electrodes (30 features)

The PLV and Coh_{max} group values and values at individual electrodes are as defined in Paragraphs 4.6.4 and 4.6.5 respectively.

Furthermore, Table 4.1 presents the results obtained when using for the 3 SVMs discriminating the different pairs of tasks the respective best subsets. In all but 2 cases, using 3 different sets of features yielded better 3-class discrimination results than using one single subset. For the respective subjects, classification accuracies of 71.29%, 46.82%, 66.23%, 47.98% and 58.50% were obtained.

4.7.4 Sole Use of Synchronization and Power Features

The best classification rate when using a subset of only PLV, Coh_{max}, α, β_1, β_2 or λ features respectively, is indicated in Table 4.2. For each day of recording of each subject, the best result is highlighted with bold font. When using only synchronization measures, 3-class distinction with accuracies up to 62.35%, 44.13%, 60.81%, 40.40% and 55.14% was achieved for subjects 1, 2, 3, 4 and 5 respectively. The achievement of significant classification accuracies using only synchronization features demonstrates their relevance for EEG-based discrimination of mental tasks.
When using only α-power estimates, the maximum accuracies were 65.05%, 40.76%, 62.64%, 38.42% and 48.97% for the 5 respective subjects. Sole use of λ-power estimates resulted in maximum classification rates of 63.32%, 41.13%, 59.80%, 42.15 and 48.12% respectively.
When using synchronization and power features separately for classifying the 3 mental tasks, PLV outperforms the power-subsets on the 3 days for Subject 5, on 2 days for Subjects 2 and 4 and on 1 day for Subjects 1 and 3. The other 2 days for subjects 1 and 3, α-power features perform better. For Subject 2 and 4, λ-power and β_2 respectively are best for the remaining day.

4.7.5 Significance Tests

A total of 25 pairs of feature subsets were compared with one another to test for significant differences between PLV and Coh, between the different power features, between power features and synchronization measures, and between the combination of synchronization measures and power features and the individual subsets.

To this end, we considered the following 20 subsets. 10 subsets consisted of the PLV, Coh, α, β_1, β_2, λ, PLV and α, PLV and λ, Coh and α, and Coh and λ features at the 32 electrodes. Another 10 subsets were constructed with these features at the C&P electrodes only. As explained in Paragraph 4.6.5, the PLV or Coh value for a single electrode is the mean of the PLV or Coh values of the electrode pairs formed by this electrode and its neighbors.
The data from the different days of recording for different subjects were analyzed separately. After computation of the CRs for the 5 sessions, by means of 5-fold cross validation, the results for the corresponding subsets with features at all electrodes and features at the 6 C&P electrodes only were concatenated. A pairwise student's t-test was then applied to these 10 results for the 2 respective types of features that we wanted to compare.

Table 4.3 gives an overview of the results. 'SB' means that the firstly mentioned type of feature is performing significantly better than the second one, with a level of significance < 0.05. 'B' means that the first type is better than the second one, but with a level of significance p: $0.05 \leq p < 0.1$. 'SW' and 'W' mean that the second type is the (significantly) better one. '-' means that the performance of the different types of features is not significantly different.

PLV vs Coh_{max}

For subjects 1, 3 and 5, PLV was significantly better than Coh_{max} for 2 of the 3, 1 of the 3 and all 3 days respectively. For subject 2, no significant differences between PLV and Coh_{max} were detected. In case of Subject 4, Coh_{max} was significantly better than PLV for the second day of recording. These results shows that both measures, PLV and Coh_{max}, extract different information from the signals.

Subj	Day		PLV	Coh_{max}	α	β_1	β_2	λ
1	1	CR(%)	**57.55**	42.28	49.33	42.09	39.30	45.71
		UR(%)	5.40	11.83	10.99	10.51	9.60	11.70
	2	CR(%)	62.35	50.79	**63.15**	53.58	44.51	61.53
		UR(%)	5.49	8.96	7.24	8.15	6.86	6.88
	3	CR(%)	60.95	50.75	**65.05**	56.66	48.01	63.32
		UR(%)	4.58	10.71	7.24	8.06	6.86	7.96
2	1	CR(%)	**44.13**	42.26	40.11	40.48	39.26	39.27
		UR(%)	7.11	8.78	8.77	8.59	0.94	4.86
	2	CR(%)	37.08	40.31	40.76	38.01	40.68	**41.13**
		UR(%)	2.84	2.85	9.92	3.31	0.18	6.51
	3	CR(%)	**40.05**	38.58	37.65	37.92	39.31	38.94
		UR(%)	3.51	3.84	1.38	0.28	0.09	0.09
3	1	CR(%)	57.31	44.17	**62.64**	43.89	32.47	52.34
		UR(%)	6.70	9.68	8.54	10.53	2.71	9.03
	2	CR(%)	**60.81**	50.51	58.03	47.15	36.84	59.70
		UR(%)	5.02	10.59	7.36	10.51	9.39	8.19
	3	CR(%)	56.94	46.39	**60.54**	48.99	37.43	59.80
		UR(%)	4.81	8.32	8.97	9.24	6.10	7.58
4	1	CR(%)	**38.93**	37.28	35.62	37.11	35.41	38.11
		UR(%)	6.57	10.98	10.03	10.13	4.86	9.09
	2	CR(%)	39.73	38.57	36.01	37.73	**42.89**	42.15
		UR(%)	7.53	13.54	11.30	9.70	6.86	10.07
	3	CR(%)	**40.40**	37.85	38.42	33.44	37.11	35.02
		UR(%)	7.08	11.44	9.54	6.03	9.07	9.25
5	1	CR(%)	**55.14**	40.28	48.97	48.41	31.21	48.12
		UR(%)	6.45	12.15	8.69	8.70	6.17	8.04
	2	CR(%)	**46.38**	38.38	38.76	40.24	35.41	42.65
		UR(%)	8.09	12.09	10.50	6.69	8.18	8.64
	3	CR(%)	**50.91**	46.94	44.22	45.84	34.83	47.66
		UR(%)	6.41	11.53	9.95	8.00	7.72	8.47

Table 4.2: CRs and URs for 3-class distinction for sole use of synchronization (PLV and Coh_{max}) and PSD (α, β_1, β_2, λ) features

	Subject 1			Subject 2			Subject 3		
Types tested	1	2	3	1	2	3	1	2	3
PLV / $\mathrm{Coh}_{\max}$	SB	–	SB	–	–	–	SB	–	–
PLV / λ	B	SW	–	SB	–	–	–	SW	SW
PLV / α	–	SW	–	–	–	–	SW	SW	SW
PLV / β_1	SB	W	–	B	–	–	SB	–	–
PLV / β_2	SB	B	SB	–	–	–	SB	SB	SB
$\mathrm{Coh}_{\max}$ / λ	W	SW	SW	SB	–	–	SW	SW	SW
$\mathrm{Coh}_{\max}$ / α	SW	SW	SW	–	–	B	SW	SW	SW
$\mathrm{Coh}_{\max}$ / β_1	–	–	SW	–	–	–	SW	–	–
$\mathrm{Coh}_{\max}$ / β_2	–	SB	–	–	SB	–	SB	SB	SB
λ / α	W	–	–	SW	–	–	SW	–	–
λ / β_1	B	SB	SB	–	–	–	SB	SB	SB
λ / β_2	SB	SB	SB	–	–	–	SB	SB	SB
α / β_1	SB	SB	SB	–	–	–	SB	SB	SB
α / β_2	SB	SB	SB	–	–	–	SB	SB	SB
β_1 / β_2	B	SB	SB	–	–	–	SB	SB	SB
PLV+α / $\mathrm{Coh}_{\max}$+α	B	–	–	B	–	–	B	SW	–
PLV+λ / $\mathrm{Coh}_{\max}$+λ	–	W	–	–	–	–	SB	–	–
PLV+α / PLV	B	SB	SB	SB	–	–	SB	SB	SB
PLV+α / α	SB	–	B	B	–	–	–	B	–
PLV+λ / PLV	–	SB	SB	–	–	–	–	SB	SB
PLV+λ / λ	SB	–	SB	SB	–	–	SB	SB	–
$\mathrm{Coh}_{\max}$+α / $\mathrm{Coh}_{\max}$	SB	SB	SB	–	–	–	SB	SB	SB
$\mathrm{Coh}_{\max}$+α / α	SB	B	SB	–	–	–	–	SB	–
$\mathrm{Coh}_{\max}$+λ / $\mathrm{Coh}_{\max}$	SB	SB	SB	–	–	–	SB	SB	SB
$\mathrm{Coh}_{\max}$+λ / λ	SB	B	SB	B	–	–	–	SB	–

	Subject 4			Subject 5		
	1	2	3	1	2	3
PLV / $\mathrm{Coh}_{\max}$	–	SW	–	SB	SB	SB
PLV / λ	–	SW	–	–	–	–
PLV / α	–	–	–	–	B	SB
PLV / β_1	–	W	–	–	–	–
PLV / β_2	–	SW	–	SB	SB	SB
$\mathrm{Coh}_{\max}$ / λ	–	SW	SB	SW	–	W
$\mathrm{Coh}_{\max}$ / α	–	–	–	SW	–	–
$\mathrm{Coh}_{\max}$ / β_1	–	–	–	SW	B	–
$\mathrm{Coh}_{\max}$ / β_2	–	SW	–	SB	SB	SB
λ / α	–	SB	SB	–	–	SB
λ / β_1	–	SB	–	–	–	–
λ / β_2	–	–	–	SB	SB	SB
α / β_1	–	–	–	–	–	–
α / β_2	–	SW	–	SB	SB	SB
β_1 / β_2	–	SW	–	SB	SB	SB
PLV+α / $\mathrm{Coh}_{\max}$+α	SB	SW	W	–	SB	–
PLV+λ / $\mathrm{Coh}_{\max}$+λ	–	SW	–	–	SB	–
PLV+α / PLV	–	–	–	SB	SB	–
PLV+α / α	SB	–	–	–	SB	SB
PLV+λ / PLV	–	SB	SB	SB	–	SB
PLV+λ / λ	–	–	SB	–	SB	SB
$\mathrm{Coh}_{\max}$+α / $\mathrm{Coh}_{\max}$	–	–	SB	SB	–	SB
$\mathrm{Coh}_{\max}$+α / α	–	SB	SB	SB	–	SB
$\mathrm{Coh}_{\max}$+λ / $\mathrm{Coh}_{\max}$	–	SB	–	SB	SB	SB
$\mathrm{Coh}_{\max}$+λ / λ	SB	SB	SB	–	–	SB

Table 4.3: Differences between synchronization and PSD features for different subjects and days of recording. (S)B (S)W: first type of feature is (significantly) better (worse) than the one behind slash. Level of significance: 0.05

Power Estimates

For Subjects 1, 3 and 5, other PSD estimates achieve significantly better classification accuracies than the PSD for the β_2 band. α and λ power measures do not show a clear distinction regarding performance. For subjects 1 and 3, they are significantly better than β_1 PSD estimates. For Subject 5, there are no significant differences between β_1 PSD on the one hand and α and λ PSD estimates on the other hand. For the first day of recording of Subject 2, PSD in the α band proves significantly better than the PSD in the λ band. Otherwise no significant differences in performance are detected for any pair of PSD features and for any day of recording of this subject.
For Subject 4, λ power is significantly better than α power for two days of recording and β_2 was significantly better than α and β_1 for one day of recording.

Synchronization vs Power Estimates

If there were any significant differences, for Subject 2 PLV and Coh_{max} were better than the PSD estimates, and for Subject 4 power features were better than PLV and Coh_{max}.
For the other 3 subjects, λ and α PSDs perform significantly better than PLV for Subject 3 and significantly better than Coh_{max} for all 3 subjects. There was no clear difference between the performance of PLV and β_1 PSD estimates. β_1 PSD performed significantly better than Coh_{max}. Coh_{max} and PLV were significantly better than β_2 PSD for the 3 subjects.

Combination of Synchronization and Power Features

The last part of Table 4.3 shows that combining synchronization measures with PSD estimates can lead to (significantly) improved results compared to classifications using the individual subsets separately. Results obtained combining synchronization and PSD features are never worse than the results obtained using these types of features separately.
However, conclusions about reciprocal additional information should be drawn carefully. For subject 5, day 3, e.g., the fact that the subset PLV $+\alpha$ performs significantly better than α is probably due to the good performance of PLV features only. From the same table, we can see that PLV performs significantly better than α PSD estimates for that subject on that day.
Of interest are the cases where there is no significant difference in performance between a synchronization measure and α or λ PSD while their combination leads to significantly improved results. For example, we see that there is no significant difference in performance of PLV and λ features for subject 1 and 5 on day 3, while the subset combining these features performs significantly better than the individual subsets.
Table 4.3 also shows that for subject 3, day 2, α PSD features achieve significantly better results than Coh_{max} features. The combination of both types of features not only performs significantly better than Coh_{max} features, but also significantly better than the α features.

Similar conclusions can be drawn when using averaged coherence instead of maximum coherence. Classification accuracies obtained with PLV are greater than those obtained with Coh. In most cases the same significant differences have been detected (see Table 4.4). Inspecting the differences between Table 4.4 and Table 4.3 one learns that if coherence features are used alone, for most subjects it is better to use the average coherence. However, in combination with power features Coh_{max} is of advantage.

We would like to emphasize that for these significance tests only the features at individual electrodes have been considered. Group averages and synchronization values between individual electrode pairs reported in Paragraphs 4.6.4 and 4.6.6 were omitted. In addition, the two subsets considered (all electrodes and C&P electrodes) may not be optimal for a given type of feature. The different types of features contained in the whole set of 290 features will be compared by means of feature selection algorithms in Chapter 5.

4.7.6 Discussion

In the preceding sections, we presented the use of synchronization measures in the framework of BCIs. The suitability of the PLV and the maximum spectral coherence, Coh_{max}, for classifying mental tasks

	Subject 1			Subject 2			Subject 3		
Types tested	1	2	3	1	2	3	1	2	3
PLV / Coh	SB	–	B	–	–	–	SB	SW	SW
Coh / λ	–	SW	SW	B	–	–	–	–	SW
Coh / α	SW	SW	SW	–	–	–	SW	SW	SW
Coh / β_1	–	W	SW	–	–	–	SB	SB	–
Coh / β_2	SB	SB	B	–	–	–	SB	SB	SB
PLV+α / Coh+α	–	–	–	–	–	–	SB	–	–
PLV+λ / Coh+λ	–	–	–	–	–	–	SB	–	–
Coh+α / Coh	SB	SB	SB	–	–	–	SB	SB	SB
Coh+α / α	SB	–	SB	–	–	–	–	SB	–
Coh+λ / Coh	SB	SB	SB	–	–	–	SB	SB	SB
Coh+λ / λ	SB	–	SB	–	–	–	–	B	–

	Subject 4			Subject 5		
	1	2	3	1	2	3
PLV / Coh	–	SW	–	–	B	–
Coh / λ	–	–	–	–	–	–
Coh / α	–	B	–	–	–	SB
Coh / β_1	–	–	–	–	W	B
Coh / β_2	–	–	–	SB	B	SB
PLV+α / Coh+α	–	W	SW	–	SB	–
PLV+λ / Coh+λ	–	SW	–	–	SB	–
Coh+α / Coh	–	–	SB	SB	SB	–
Coh+α / α	–	SB	SB	SB	–	SB
Coh+λ / Coh	–	SB	SB	SB	B	SB
Coh+λ / λ	SB	SB	SB	B	–	SB

Table 4.4: Differences between synchronization and PSD features for different subjects and days of recording. Coh used instead of Coh_{max}. (S)B (S)W: first type of feature is (significantly) better (worse) than the one behind slash. Level of significance: 0.05

is evaluated. There are significant differences between these measures. In general, PLV yields better results than Coh_{max}. This may be because synchronization is more restrictive than coherence. The signals of two synchronized systems are correlated, but increased coherence does not necessarily imply synchronization [127].

We demonstrated that interesting features can be derived, containing relevant information for classification of spontaneous EEG recorded during different mental tasks, from PLV and Coh_{max}, computed from broadband signals, . For the sole use of synchronization features, for the 5 respective subjects, classification rates of 62.35%, 44.13%, 60.81%, 40.40% and 55.14% were obtained for distinction of the three mental tasks 'left', 'right' and 'word'.
While synchronization measures computed from narrow band signals have a clearer neurophysiological interpretation, synchronization between broadband signals may be more difficult to achieve and may therefore be more discriminative.

Subjects 2 and 4 achieved significantly lower classification accuracies than Subjects 1, 3 and 5. For the pairs of feature types tested for significant differences, the former showed less often significant differences than the latter. This may suggest that they did not well perform the mental tasks.

For 3 best subjects we observe from Table 4.3 that α and λ features outperform β_1 features in case of subject 1 and 3 and β_2 features for all 3 subjects. High frequency features are systematically worse than lower frequency or broad band features.

A comparison of synchronization and PSD features is difficult, because they are intrinsically different. While power features are computed from single EEG signals, the here considered synchronization features characterize the interaction between 2 EEG signals. We compared the performance of power measures with the one of synchronization features at individual electrodes, characterizing the interaction with neighbors, in a paired student's t- test. We found that the α and λ features are significantly better than Coh_{max} for the 3 subjects and significantly better than PLV for subject 3. Considering their combination,

however, we see that synchronization measures and PSD add information to each other. No combination of features is worse than the separate subsets.

Since coupling features are broad band, the superior performance of subsets containing synchronization features as well as PSD features may be due to the availability of more information for the classifier. However, also subsets containing the broad band synchronization features as well as broad band PSD features (λ) perform better than the separate ones. This suggests that the superior performance of sets containing features of both types is not due to the availability of more information for the classifier, but rather to the the presence of coupling features.

If one had to give feedback to the subject, offline evaluation of the performance of different feature subsets on the first day of recording will be a good indicator. In our experiments, the feature subset with the best 3-day average performance was also the best feature subset on day 1 for 4 subjects. Also, evaluated over the different days, the combination of synchronization with PSD features is generally one of the best feature subsets.

Combining synchronization and power features 3-class classification accuracies of up to 72.29%, 46.82%, 66.23%, 47.98% and 58.50% were obtained for the 5 respective subjects. Error rates were 23.51%, 45.51%, 25.58%, 43.81% and 35.80%. Results reported by other authors on 3-class differentiation are discussed in the following.
Anderson [6] investigated the 3 mental tasks 'baseline', 'letter' and 'math' for 1 subject and reports 85% classification accuracy when averaging over 4 0.5s windows, overlapping with 0.25s.
Millán and co-workers [37] distinguish between the mental tasks 'relax', 'left' and 'right' (no feedback). While left and right motor imagery is performed with eyes opened, 'relax' was done with closed eyes. This results in a much higher recognition rate of the task 'relax' (100% and 92% respectively) than for the tasks 'left' (42% and 51%) and 'right' (36% and 56%). The average true positive rates reported were 52% and 63%. The merit of this study was the very low error rate. The false positive rate was as low as 0 and 5%. None of the results, however, were presented for all 3 subjects.
Yom-Tov and Inbar [158] report 63.2% average classification accuracy (3 subjects) for the distinction of real left and right finger movements and left toe movements.
Ebrahimi *et al* [45] consider the mental tasks 'mental counting' and imagined left and right index finger movements. Using time-frequency-space correlation, they obtain error rates of 25%, 30% and 27% for the best session of 3 subjects. A session comprises 3 1-minute recordings.
Apart from the system of Anderson, achieving 85% for 1 subject, the performance of our method is in the range of the accuracies reported in the other studies. It is difficult, however, to directly compare the results from our study and these 4 studies, as the recording equipments, recording and classification protocols and the mental tasks considered are different. In addition, the amount of data and the amount of training the subjects had before participating in the BCI experiments is different for the different studies.

4.8 Conclusions

In this chapter, we introduced the synchronization measures PLV and coherence. Their properties were discussed and illustrated on sinusoidal signals as well as weakly and strongly coupled Rössler oscillators. We found that PLV and Coh_{max} were more noise-robust than Coh. Another advantage of PLV is that its computation is fast.
The use of PLV computed from broadband filtered EEG signals was illustrated and proved better than PLV computed from narrowband filtered EEG signals.
The usefulness of PLV and coherence for classifying mental tasks was demonstrated. There are significant differences between these measures. In general, PLV yielded better classification results than coherence. Combining synchronization features with PSD features never decreased the classification accuracies and often yielded (significantly) improved results when compared to the separate subsets. This suggests some complementarity between synchronization measures and PSD estimates and indicates the potential of synchronization measures to improve current BCI systems.

Chapter 5

Automatic Feature Selection

5.1 Introduction

After evaluating synchronization features for their relevance for the recognition of mental tasks, feature selection algorithms are investigated. The objective of feature selection (FS) is three-fold: improving the classification performance of the classifiers, providing faster and more cost-effective classifiers, and providing a better understanding of the underlying process that generated the data [57].

Feature selection algorithms allow us to distinguish between more and less informative EEG signals and measures derived from them. Observations from such studies could allow for direct improvement of BCI performance.

The computational complexity of many learning algorithms depends heavily on the number of features. Also therefore it is useful, especially in BCI applications, to reduce the feature set before the construction of the classifier. In addition, a smaller number of features reduces the risk for overfitting (but a too small number of features may lead to underfitting).

We are interested in automatic FS algorithms in order to study many different features and their capability of classifying spontaneous EEG recorded during mental tasks. In addition, automated procedures allow us to deal with subject to subject variability and fast algorithms would allow for dealing with variability between different days of recording and the session-to-session variability for one subject.

FS not only includes deciding which features to use in describing the concept, but also decide how to combine those features [18]. Any feature selection method can be characterized by 4 basic issues that determine the nature of the heuristic search process:

- The starting point (forward vs. backward elimination)
- Organization of the search
- Strategy used to evaluate alternative subsets of attributes
- A criterion for halting the search

The papers by Koller and Sahami [78], Blum and Langley [18] and Guyon and Elisseeff [57] give a nice introduction to feature selection.

This chapter is structured as follows. First, in Section 5.2, the different types of FS algorithms will be explained. Then, in Section 5.3, we shortly review FS methods employed in BCI research. In Section 5.4, we introduce the 3 different FS methods applied to our data. The algorithms and classification results are presented for each them, in Sections 5.5, 5.6 and 5.7 respectively. In Section 5.8, application of the best method to data from the BCI competition is presented. This chapter is concluded with an application of the best feature selection method to different sets of features (Section 5.9).

5.2 Feature Selection Methods

5.2.1 Introduction

Generally speaking, one can distinguish between *embedded*, *filter* and *wrapper* methods for feature selection, depending on the relation between the selection scheme and the basic induction algorithm [18]. We shortly describe these 3 methods and illustrate them with some examples.

5.2.2 Embedded Methods

In embedded methods for feature selection, e.g. decision trees [120, 128, 36] and SVM-based Recursive Feature Elimination (SVM-rfe)[58, 87], the classification algorithm and feature selection are indivisible. They incorporate variable selection as part of the training process and may therefore be more efficient.

An example of decision trees is the C4.5 algorithm [118]. In such a tree, each node corresponds to an attribute, and each outgoing arc to a node at the next level is associated with a possible value of the attribute. Each node should be associated with the most informative attribute among the attributes not yet considered in the path from the root. The classifier thus constructed is an embedded approach to feature subset selection because it builds a tree in which only the most relevant features appear.

Guyon *et al* introduce SVM-rfe [58] to select a small subset of genes from broad patterns of gene expression data, recorded on DNA micro-arrays. SVM-rfe is an application of rfe using the weight magnitude of the SVM being trained as ranking criterion and allows for feature subset ranking (rather than ranking features according to their individual relevance). Feature subsets are nested and contain complementary features, not necessarily individually most relevant.

5.2.3 Filter Methods

Filter methods precede the classification algorithm. They select relevant features before attempting to train a classifier. Relevant/irrelevant features can be defined in a number of ways [18].
Filter methods are independent of the classification algorithm and are therefore computationally less demanding. The classifier does not have to be retrained for selecting different features or feature subsets. However, the resulting feature subset will not be optimized for the specific classifier used.

The Relief algorithm, proposed by Kononenko [79], is an example of a filter method. It was shown to be very efficient in estimating the quality of the attributes. The idea behind this algorithm is that a feature is relevant if its values are very different for two samples of different classes (high interclass variability), and very similar for two samples of the same class (low intraclass variability). Relief is very efficient in estimating the quality of the features, but is ineffective at removing redundant features, as two predictive but highly correlated features are both likely to be given high relevance weightings.

Yu and Liu [159] presented another filter method for FS, the "Fast Correlation Based Filter" (FCBF). Not only does the algorithm determine the relevant features, it also includes a redundancy analysis. First, the relevant features are determined from the symmetrical uncertainty between each feature and class. Second, redundant features are detected by means of approximate Markov blankets and removed.

5.2.4 Wrapper Methods

Wrapper methods make explicitly use of the prediction performance of the induction algorithm to perform feature selection [57]. The classification algorithm is used as a subroutine for the FS task in order to evaluate the selected feature set. The results of the classification algorithm are the target function. The induction method that will use the feature subset should provide a better estimate of accuracy than a separate measure that may have an entirely different inductive basis. The disadvantage of wrapper methods over filter methods is their computational cost, which results from calling the induction algorithm for each feature set considered. The resulting feature subset will be optimized for the specific induction

algorithm used, but may perform badly when applied to other classification techniques.

Genetic Algorithms (GAs) can be used as wrapper methods. They have a wide range of application areas, including parametric optimization, controller structure selection, fault identification, nonlinear system identification, antenna design and medicine [111]. Applications in medicine include data mining (knowledge discovery), mainly applied to diagnosis and prognosis, medical imaging and signal processing and planning and scheduling [111].
GAs are inspired by evolution. A potential solution is encoded as a chromosome-like data structure and recombination and mutation operators are applied to these structures. The principle of survival of the fittest is applied to produce successively better approximations to a solution.

5.3 Feature Selection Methods Employed in BCI Research

Millán *et al* [36] use the C4.5 algorithm, preceded by the Relief algorithm to find the most relevant features. Classification results were obtained from 5-fold cross validation. For distinction of two classes, they obtained an average (5 subjects) error rate of 21.0%.

Yom-Tov and Inbar [158] compare a filter method with a wrapper method. They define irrelevant features to be those features that are detrimental to the classifier either because they provide redundant information or because they are corrupted by noise. They use an algorithm, proposed by Koller and Sahami [78], to estimate the cross-entropy between each pair of features and then iteratively remove those features that have low cross-entropy and thus add little extra information to the classification task.
The wrapper method they considered is a GA. For faster convergence, they use a Culling GA. As compared to other genetic algorithms, such algorithms perform a more aggressive search of the solution space, discarding at each iteration the weakest individuals. Features are partitioned into groups. The examples are classified using the feature groups. The groups are attributed a score (probability of a correct decision as determined by an SVM classifier using fivefold cross validation) and a predetermined percentage of groups is discarded. 33% of the feature groups are discarded. This is a compromise between fast convergence and good performance. This procedure is iterated for a predetermined number of iterations.
Results were obtained with SVM classifiers using 5-fold cross validation: 87.0% (3 subjects) classification accuracy for distinction of toe movements from left or right finger movements and 63.2% for the three limb classification task.

Schröder *et al* [135] apply GAs using SVMs for evaluation to select EEG channels. They report that the GA-SVM choice improved the classification accuracy by 2.08% compared to the all-channel choice. Comparison of the GA-SVM with the physiologically motivated choice reveals an average improvement of 10.19%. 9 sessions from 4 different subjects, covering runs of 50 trials each, have been considered. The final average accuracy values plotted, however, were based 20 cross validations, whereas the GA search internally only used one cross validation.

Citi *et al* [29] use a GA to perform joint optimization of features and classifier in a P300 paradigm. The classifier was a polynomial classifier. Their approach confirmed the usefulness of linear detectors, and revealed the importance of selecting certain EEG channels and using their differences to cancel non-P300 components.

Lal *et al* [87] use Zero-Norm Optimization[1] and SVM-rfe, closely related to Support Vector Machines, as well as the Fisher Criterion to select relevant channels for a BCI. Zero-Norm Optimization minimizes the zero-norm of the SVM being trained instead of the l_1-norm or l_2-norm as in standard SVMs. The solutions thus obtained are usually much sparser. The Fisher Criterion determines how strongly a feature is correlated with the labels.
Scores were defined for each channel (derived from features computed from this channel) and at each iteration step the channel with the lowest score is removed. This allows for spatial interpretation of the solution.

[1] The zero-norm of a vector v is equal to the number of nonzero entries of v

They found SVM-rfe and Zero-Norm Optimization to be capable of selecting relevant channels, whereas the Fisher Criterion failed for some subjects. Especially for small numbers of channels SVM-rfe is slightly superior over the other two methods. For larger numbers of channels the performance of Zero-Norm Optimization is comparable to SVM-rfe. SVM-rfe method was capable of estimating physiologically meaningful EEG channels for the imagined left/right hand movement paradigm. The average error rate (taken over 5 subjects) was 24% using 12 channels.

5.4 Applied Feature Selection Strategies

In Sections 4.6 and 4.7, we introduced and compared the coupling measures Phase Locking Value (PLV) and the maximum spectral coherence in some frequency band (Coh_{max}) for classifying spontaneous EEG recorded during different mental tasks in BCIs. To reduce the amount of features that would result from computing PLV and Coh_{max} between all possible electrode pairs, we considered average synchronization values within different brain regions; the average coupling of each electrode signal with the signals recorded from neighboring electrodes; and a few synchronization values for individual electrode pairs in the fronto-centro-parietal region.
Instead of manually choosing subsets of features, application of automated feature selection algorithms would be desirable. In addition, they could allow for studying larger sets of features, including, e.g., synchronization values for all possible electrode pairs.

3 FS algorithms have been investigated, a wrapper, an embedded and a filter algorithm. GAs were applied as wrapper approach (Section 5.5) to feature selection; the embedded method considered was SVM-rfe (Section 5.6) and finally the filter method Fast Correlation-Based Filter (FCBF) was implemented and tested (Section 5.7). Each of these methods will be described and the obtained classification results and selected features will be presented.

5.5 Genetic algorithms

The first algorithm we applied for FS, was a GA. We used the GA as a wrapper method, the SVM classifier described in Paragraph 3.6.2 was used for evaluation of the fitness of individuals.

5.5.1 Generalities

GAs are inspired by evolution. The principle of survival of the fittest is applied to produce successively better approximations to a solution. A general scheme of a genetic algorithm is shown in Figure 5.1. First, an initial population of potential solutions encoded in chromosome-like structures should be generated and evaluated for their fitness. A fitness value is attributed to the individuals according to a performance function. If the stop criterion is not met, individuals are randomly selected for generating offspings. Usually, the chance of picking an individual to generate offsprings is proportional to its fitness. Recombination and mutation operators are applied to the selected individuals to generate the new population and the fitness of the individuals of this new population is evaluated.

GAs are robust, global and need only little or no a priori knowledge about the process to be controlled. GAs make fewer assumptions about the search space, do not require a formal initial estimate of the solution region and the nature of the search mechanism is stochastic, therefore GAs are capable of searching the entire solution space with more likelihood of finding the global optimum [28, 111]. They may have tendency to converge towards local optima if the fitness function is not handled properly. Depending on the amount of data and computational cost of the evaluation function, GA feature selection may be very time-consuming.

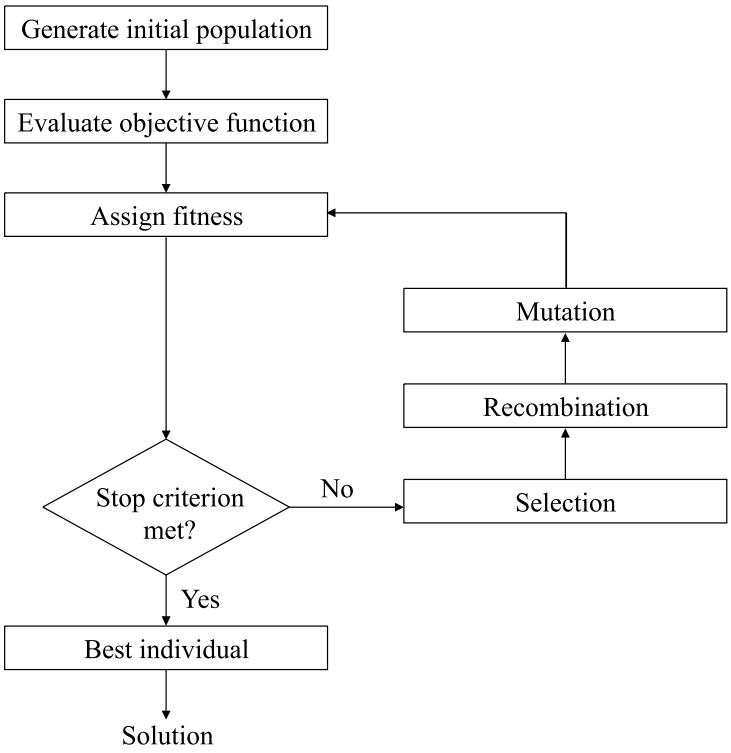

Figure 5.1: Schematic of Genetic Algorithms

5.5.2 Implementation

First, the data were divided into a training set and a test set. For each task, 240 windows were separated to use as test data, the remaining data were used to select the features and train the classifier.

Features were selected from the training data in two steps. According to their type, features were divided into different groups (α, β_1, β_2, λ, PLV(averages), Coh_{max}(averages), PLV(pairs), Coh_{max}(pairs)) and the best performing features were selected within each of these smaller groups. This approach has the advantage that the different types of features could be compared more directly.
In a second step, a final set of features was selected amongst these best features and its classification accuracy was evaluated on training data. The resulting 3-class discrimination was characterized by the classification rate (CR) and the error rate (ER), CR_{train} and ER_{train} respectively. This procedure was applied for each of the 3 pairs of tasks.
The generalization performance of the resulting set of features and classifier was then evaluated on the remaining, unseen, test data (CR_{test} and ER_{test}). The data were studied per day and per subject.

The details of the genetic algorithm were as follows. Initialization was done such that the amount of features selected (1's) was limited as compared to random initialization with 0's and 1's (about 50% of each). We generated a population of 20 individuals, represented by chromosomes. As we were to select amongst 290 features, the chromosomes had length 290. 10 chromosomes contained 5 random 1's and zeros otherwise. Another 2 times 5 chromosomes contained 10 and 20 random 1's respectively.
The training data were divided into 5 equal parts to evaluate the 5-fold CV classification accuracy. The objective function was the classification error resulting from this 5-fold CV of the feature subset corresponding to the chromosome. This was computationally very costly.
Stochastic Universal Sampling (SUS) was applied for selecting the individuals. The method provides a zero bias and minimum spread. The individuals are mapped to the segments of a line, in such a way that each individual's segment is equal in size to its fitness. A number of pointers, equalling the number of individuals to be selected (N_p), is placed over the line. The distance between the pointers is N_p and the position of the first pointer is determined randomly in the interval $1/N_p$.
Then, the recombination and mutation operators were applied. The probability of recombination occur-

ring between pairs of individuals was arbitrarily set to 0.7. Also the probability of a mutation to occur in the chromosome of an individual was 0.7. The "Generation Gap", the percentage of new individuals that are generated at each generation, was arbitrarily set to 0.9.
The stop criterion consisted of a maximum number of generations. The maximum number of generations was set to 15 for the first step of feature selection, and to 30 for the second step.

We apply this feature selection method to the 290 features described in Section 4.6.

5.5.3 Results using GAs

Table 5.1 presents, for the different types of features, the CRs obtained after the first step of feature selection (page 67) and the average number of features of this type selected for the final feature subset. The best results are obtained for PLV and α power features. These are also the types of features most often appearing in the final subset, except for the task pair L-R, where Coh_{max} averages are most often appearing in the final subset.

	L-R		L-W		R-W	
Type	CR (%)	♯	CR (%)	♯	CR (%)	♯
PLV pairs	67.3	2.5	71.7	3.9	70.8	4.1
Coh_{max} pairs	61.1	2.9	62.6	3.0	62.7	2.7
PLV avgs	64.5	3.0	71.0	3.1	70.8	4.4
Coh_{max} avgs	61.6	4.2	64.1	3.6	64.9	3.2
α	67.3	2.5	71.4	3.6	69.3	4.4
β_1	63.4	1.9	66.7	3.4	65.9	3.2
β_2	61.1	2.5	62.8	2.1	61.7	2.7
λ	66.5	2.4	69.9	3.6	69.1	3.3
Final	71.3	21.7	77.7	26.3	77.2	27.9
Synch left		2.7		2.3		2.9
Synch right		2.0		2.7		2.7
Power left		4.5		5.5		6.9
Power right		3.8		5.9		5.2

Table 5.1: CR obtained on training data for the different subgroups of features and the number of features of these subgroups selected for the final subset; CR obtained with final subset; Lower part of the table indicates the number of synchronization and power features selected in the left and right hemisphere. Average over all days, all subjects; GA

"Final" refers to the CR obtained after the second step of feature selection (page 67), which is, for the 3 task pairs, better than the CR obtained with any of the individual groups.
The lower part of Table 5.1 indicates the distribution of the selected EEG signals over the two hemispheres. For both synchronization and power features, EEG signals were more often selected in the left, right and left hemisphere for the respective task pairs L-R, L-W and R-W.

Table 5.2 shows the 3-class discrimination classification accuracies obtained on training and test data. In the average, a CR of 67.8% is obtained on training data. On test data, a CR of 50.5% was achieved.

Figure 5.2 shows, for the features computed for single EEG signals, averaged over the 3 days of recording and the 5 subjects, which electrode signals were selected. The PLV group averages and values for indi-

Subject	1	2	3	4	5	average
$\text{CR}_{\text{train}}(\%)$	75.3	59.0	77.0	61.3	66.3	67.8
$\text{ER}_{\text{train}}(\%)$	19.7	37.0	17.7	31.0	27.0	26.5
$\text{CR}_{\text{test}}(\%)$	61.0	36.7	62.0	42.3	50.3	50.5
$\text{ER}_{\text{test}}(\%)$	33.3	56.0	32.0	50.0	42.0	42.7

Table 5.2: Results for 3-class distinction obtained on training data and test data respectively. Features selected in 2 steps with GA. For each subject: average over 3 days of recording.

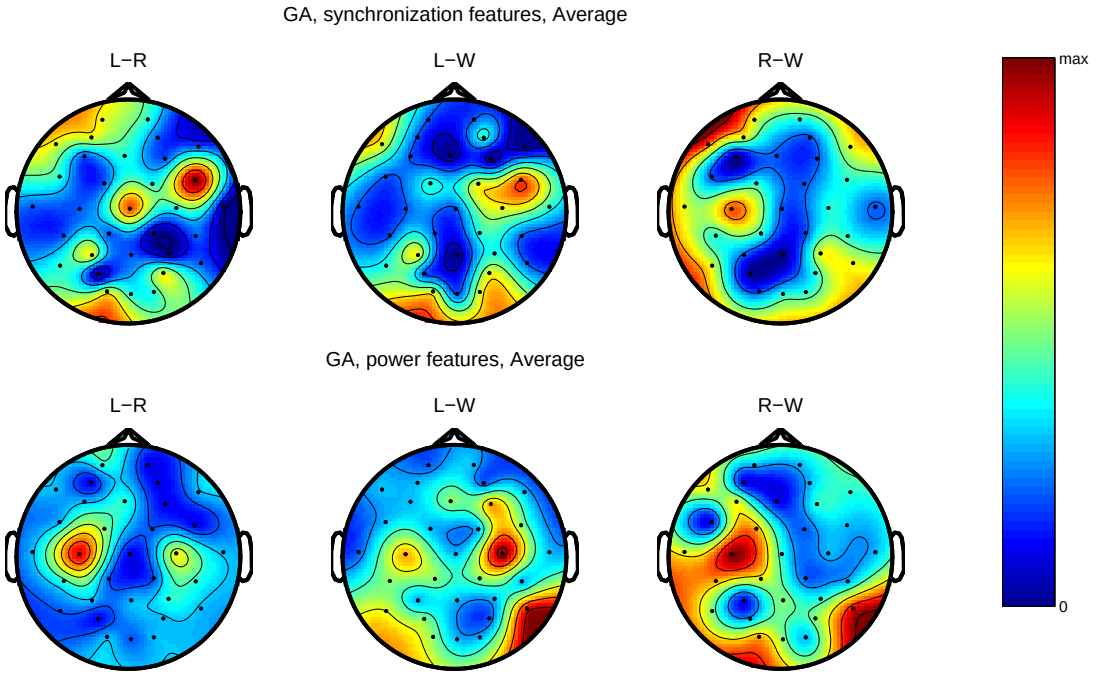

Figure 5.2: Left: Average over all days, all subjects: electrodes where features computed for individual electrodes were selected with GAs. Top: synchronization; Bottom: power; Right: Scale used in plots of electrode selection.

vidual electrode pairs were not considered in this visualization.
The colorbar for the head plots throughout this chapter is shown in Figure 5.2. 0 means no selections, "max" is the number of selections of the most often selected electrode in any of the head plots of a given figure.

Synchronization features are mostly selected in the right fronto-central region the task pairs L-R and L-W, and in the left central and frontal region for the task pair R-W.
In case of power features, we observe many selections for electrodes C3 and C4 for the task pair L-R. For L-W, mainly C4 and the occipital region seemed to be involved. To a lesser extent also C3. C3 is very often selected for the task pair R-W.
For imagined hand movements, the electrode signals from the contralateral hemisphere were most often selected. This is the hemisphere where the decrease of μ-rhythm accompanying the imagined hand movement is observed. In case of L-R, where both imagined left and right hand movement are involved, more features were selected in the left hemisphere than in the right one. From neurophysiological evidence, however, more selections in the nondominant hemisphere, the right hemisphere for right-handed subjects, would be expected as this hemisphere would make the greatest difference.

5.5.4 Conclusions

The proposed GA has been used successfully to select features in the framework of a BCI. We found that features were mostly selected in the central region, which corresponds to evidence from neurophysiology.

GAs have the disadvantage of being computationally very demanding. The above described procedures are very slow. The selection of features for one day of recording of one subject is of the order of days and therefore cannot be used practically in BCI systems.

To reduce the computation time, hybrid methods, including a preselection of features with a variant of the Relief algorithm [79], have been tested. The algorithm Relief gives each feature a relevance weighting indicating its level of relevance to the class label. To rank the features, Relief finds two nearest neighbors

for every sample of the training set: one from the same class (nearHit) and one from the opposite class (nearMiss). The weights of the features, initially zero, are updated by Relief according to the intuitive idea that a feature is more relevant if it has different values for the sample and its nearMiss, and similar values for a sample and its nearHit. Instead of finding just one nearHit and one NearMiss, we look for the 10 nearest hits and nearest misses. This allows us to cope with more noisy data. We do not search neighbors for all of the data samples. We limit ourselves to finding neighbors for a fixed number of samples (1000) only. In addition, we implemented the multi-class variant in order to deal with our 3 class problem. Furthermore, we evaluated preselection of features with the evaluation of redundancy.
However, due to the presence of GAs, computation time was still very high. Therefore, we investigated other FS techniques.

5.6 SVM-based Recursive Feature Elimination

Due to the high computation time, GAs proved not practically useable in this context. Computationally more efficient FS methods are necessary. We now propose SVM-based Recursive Feature Elimination (SVM-rfe). We first compare it to GA on the set of features presented in Section 4.6 and then use it to compare phase synchronization for all possible electrode pairs computed from EEG signals filtered in different, narrower and broader, frequency bands [66].

5.6.1 Feature Selection Algorithm

Guyon *et al* [58] introduced a new method of gene selection utilizing SVM methods based on Recursive Feature Elimination (rfe). Gene redundancy was automatically eliminated and this method yielded better and more compact gene subsets than other evaluated methods.
The weights w_i, i=1...l, of the decision function $D(\mathbf{x})$ are a function only of a small subset of the training examples, called 'support vectors'. Given the training examples $\{\mathbf{x}_1,\mathbf{x}_2,...\mathbf{x}_l\}$ and class labels $\{y_1,y_2,...y_l\}$, the following cost function is minimized over α_k:

$$J = \frac{1}{2}\sum_{ij} y_i y_j \alpha_i \alpha_j (\mathbf{x}_i^T \mathbf{x}_j + \lambda \delta_{ij}) - \sum_i \alpha_i \tag{5.1}$$

with $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i = 0$. λ and C are positive constants ensuring that the problem has a solution in case the data are not linearly separable and δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1\ \forall i = j$, $\delta_{ij} = 0\ \forall i \neq j$). The resulting decision function reads

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \tag{5.2}$$

where $\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$ and $b = \langle y_i - \mathbf{w}.\mathbf{x}_i \rangle_i$. The weight vector $\mathbf{w}$ is a linear combination of training patterns. Most weights α_i are zero. The training patterns with non-zero weights are support vectors.
SVM-rfe uses the weight magnitude as ranking criterion. At each iteration step, the feature with the smallest ranking criterion is eliminated. Guyon proposed $c_i = (w_i)^2$ as ranking criterium, as such removing the features that contribute least to margin maximization.
Lal *et al* adapted this method to do channel selection in BCIs [87].

5.6.2 Feature Selection Strategy

Training and test data were the same as in Section 5.5, as well as the groups of features from which features were selected in a 1st step. The 2-step FS procedure was similar to the one in Section 5.5, only have GAs been replaced with SVM-rfe. The number of best ranked features retained was 16 for the first step of feature selection and 30 for the second step. The n_f first ranked features were retained for the final feature subset, where n_f corresponds to the maximum 5-fold CV classification accuracy obtained from subsets with increasing size.

5.6.3 Application to the Set of 290 features Comprising PLV averages

Similarly to Table 5.1, Table 5.3 presents classification results for the different groups of features and their presence in the final feature subset. Best results were obtained with PLV, α and λ power features. The results obtained after the second step of feature selection were, again, better than those for any of the individual groups of features. Contrary to GAs, mainly PLV features were present in the final feature subset.
For the 3 respective task pairs, synchronization features were mostly selected in the right, right and left hemisphere, whereas power features were selected mostly in the left, right and left hemisphere.

	L-R		L-W		R-W	
Type	CR (%)	♯	CR (%)	♯	CR (%)	♯
PLV pairs	70.8	8.5	74.9	6.9	73.9	1.6
Cohmax pairs	50.5	0.5	56.7	0.4	52.9	0.7
PLV avgs	68.3	6.8	73.7	7.2	73.3	6.3
Cohmax avgs	48.5	2.6	58.1	3.1	51.1	3.1
α	69.1	0.9	72.5	0.9	70.5	0.5
β_1	66.1	0.9	68.6	1.1	67.9	0.5
β_2	62.5	2.3	65.5	2.7	62.9	2.5
λ	68.9	1.1	71.8	1.6	70.6	1.8
Final	71.3	23.6	76.8	23.9	75.4	22.8
Synch left		3.0		3.7		3.2
Synch right		3.5		3.9		3.0
Power left		2.6		2.3		2.3
Power right		1.6		2.9		1.9

Table 5.3: CR obtained on training data for the different subgroups of features and the number of features of these subgroups selected for the final subset; CR obtained with final subset; Lower part of the table indicates the number of synchronization and power features selected in the left and right hemisphere. Average over all days, all subjects; SVM-rfe

The results obtained with the final subset are presented in Table 5.4. They directly compare to the ones in Table 5.2.
The classification accuracies obtained for both methods are comparable. The subject-average CR for training data was slightly greater for SVM-rfe. The average CR on test data was was about 5% greater. In the average 63 features were selected.

The electrodes whose EEG signals were selected are presented in Figure 5.3. As compared to Figure 5.2, we see that C3 is less often selected for the task pairs R-W for synchronization features. In case of power features, less occipital EEG signals are selected for the task pairs L-W and R-W.

5.6.4 Application to Set of Features Comprising PLV for Individual Electrode Pairs

We applied SVM-rfe in a 2-step FS procedure to study PLV values for individual electrode pairs computed from broadband and narrowband EEG signals for the classification of spontaneous EEG recorded during 3 mental tasks.

Subject	1	2	3	4	5	average
$CR_{train}(\%)$	76.0	61.7	77.7	63.3	65.3	68.8
$ER_{train}(\%)$	18.7	32.3	16.7	29.7	28.0	25.1
$CR_{test}(\%)$	58.3	37.0	62.3	42.3	53.0	55.6
$ER_{test}(\%)$	36.0	55.7	29.3	50.3	41.0	42.5

Table 5.4: Results for 3-class distinction obtained on training data and test data respectively. Features selected from the set of 290 features in 2 steps with SVM-rfe. For each subject: average over 3 days of recording.

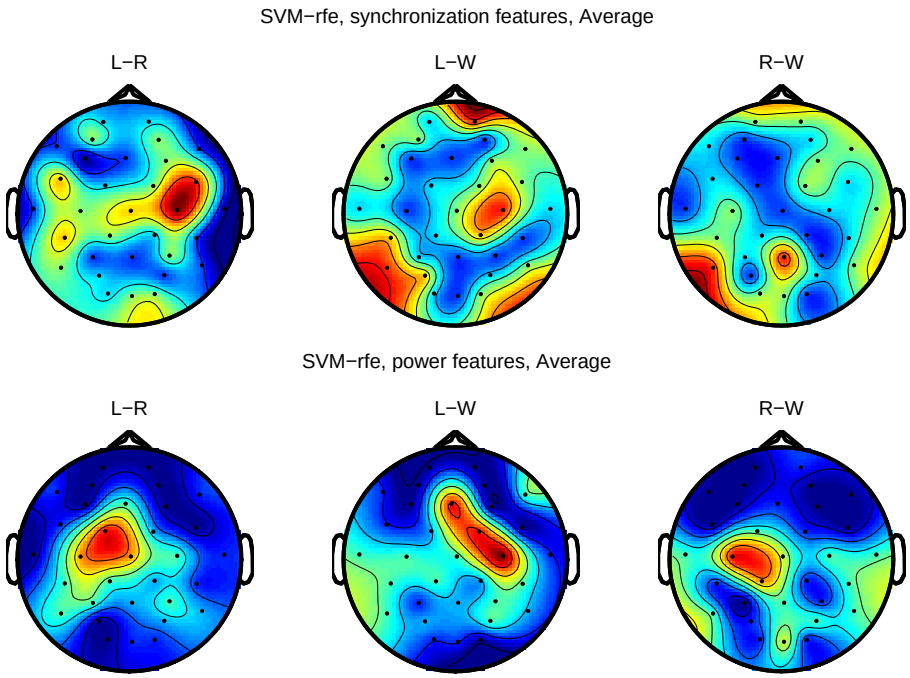

Figure 5.3: Average over all days, all subjects: electrodes where features computed for individual electrodes were selected with SVM-rfe. Top: synchronization; Bottom: power.

For the legibility, we do not consider the Coh_{max} and focus on PLV only. PLV-based features performed better than coherence features. In addition, they are much faster to compute and can be applied to shorter time windows, which makes them even more interesting for BCI applications.
For comparison, we will also consider 2 PSD based features: the power in some frequency band and the SMF.

Features were divided into different groups according to their type in order to compare the different types of features. In a first step, the best features within each of the 34 groups of features were selected by means of 5-fold cross validation (CV). For the first 32 groups, representing the PLV values between the signals recorded from the 32 respective electrodes and all other electrodes, feature selection finds for each electrode its most relevant partner electrodes. In what follows, each of the 32 PLV groups will be denoted with the name of the electrode with which all electrode pairs of this group were formed. Furthermore, the best electrode signals for Band Power (BP) and Statistical Mean Frequency (SMF) were selected.
For each feature group, from the 5 ranked lists of features, the 16 (out of 32) overall best ranked features were determined and retained for the second step of feature selection. The 5-fold CV classification performances obtained with the best features were compared for the different groups of features.

In a second step, again by means of 5-fold CV, the 50 best features amongst the best features of all groups were selected, after once removing the electrode pairs that were selected in two PLV groups (the pair CP6-C4, e.g., could be selected in both PLV groups CP6 and C4). The n_f first ranked features were retained for the final feature subset.

This procedure was applied for each of the 3 pairs of tasks and for the 4 FBs α (8-12Hz), β_1 (13-18Hz), β_2 (19-30Hz) and λ (8-30Hz). These frequency bands were chosen because they contain the μ and β rhythms, which have been proved to be involved in motor-tasks [116].

The following paragraphs will detail the preprocessing, feature extraction and results.

5.6.5 Data Preprocessing

The EEG recording and experimental paradigm are as described in Chapter 3. Contrary to the preprocessing of the first set of features, described in Section 4.6, artifacts were automatically detected and omitted in the training set. Further preprocessing is similar to what is described in Paragraph 4.6.2.

Artifacts

In an attempt to construct a better classifier, we detected eye blinking artifacts automatically with the method described in Paragraph 6.3.2 and rejected the windows where an artifact was detected for training the classifier and selecting the features. For testing, all windows of the test set were used, including the ones containing artifacts. Training and test set were as in Sections 5.5 and 5.6.
Section 6.4 details how artifact detection affects the CR.

Further Preprocessing

- Before the computation of the PLV, a linear phase FIR band-pass filter with different pass bands (α, β_1, β_2 and λ) was applied to the EEG signals, after subtracting the mean.

- For calculating the PSD, after subtracting the mean, a Laplacian filter was applied (see Paragraph 4.6.2). Then, a 4-40Hz linear phase FIR band-pass filter was applied. Finally, the signal was resampled at 1:4 of the original sampling rate to reduce the variance of the PSD coefficients.

5.6.6 Feature Extraction

For each of the FBs α, β_1, β_2 and λ, we extracted the following features from the preprocessed EEG signals:

- The PLV (see equation (4.1)) was computed for the 496 possible pairs of electrode signals, filtered in the frequency bands FB.

- We computed from the PSD the BP in the frequency band FB normalized with the total power, for all electrodes.

- In addition, we computed for all electrodes the normalized SMF using the PSD estimated at the frequencies between 8 and 30Hz:

$$SMF = \frac{1}{22}\left(\frac{\int_{8Hz}^{30Hz} P(f)f df}{\int_{8Hz}^{30Hz} P(f) df} - 8\right), \tag{5.3}$$

  where P is the estimation of the PSD at a given frequency f. The SMF is independent of the FB used to compute the other features.

Note that these features are normalized and thus a division by the average over all features of the same type at the same time instance as in Section 4.6 is not applied.

In summary, for each 1-second EEG window, we compute for each of the 32 electrode signals the PLV with the signals from all other electrodes, filtered in FB (32 groups); the BP in the frequency band FB for all electrode signals (1 group); and the normalized SMF for all electrode signals (1 group). Each of these 34 groups consists of 32 features (corresponding to the 32 electrodes).

5.6.7 Results using SVM-rfe

First, we describe the results of significance tests, testing for significant differences in classification performance between the different FBs for the different features and the different subjects. Second, the selected features are discussed. Then, the classification accuracies of different types of features are compared, presenting 5-fold cross validation results on training data for the 2-class discrimination problems. Finally, we evaluate the performance of the final subsets on unseen data for the 3-class discrimination problem.

Comparison of Different Frequency Bands

To assess the suitability of the different frequency bands for the task of discriminating mental tasks, we performed two series of significance tests, repeated for the 3 pairs of tasks. We used a 5% level of significance to reject the null hypothesis of equal performances for both FBs.

First, we tested for significant differences between the different FBs for the different groups of features. The FB dependent features were the PLV groups for each of the 32 electrodes and the BP. The CRs obtained with the best features of each of these 33 groups for the different days of recording from the different subjects were concatenated to perform a paired student's t-test.

Discrimination of L-R. In case of power features, α was significantly better than the other 3 FBs, β_1, β_2 and λ. Both β_2 and λ were significantly better than β_1.
λ was significantly better for almost all PLV groups. For very few PLV groups significant differences were noticed between the bands α/β_2 and β_1/β_2. In case of C3 and C4, λ was always significantly better, between the other frequency bands there were no significant differences.

Discrimination of L-W. For BP features, the power in the α band was significantly better than the β_1 and λ band. For the other pairs of FBs, power features did not show any significant differences. λ was significantly best for almost all PLV groups, α was significantly better than β_1 and β_2 for a lot of them.

Discrimination of R-W. In case of BP, α was significantly better than the other 3 FBs. For the other pairs of FBs, no significant differences were observed for BP.
For the PLV groups, λ was significantly best in almost all cases. 21 PLV groups showed significant differences only where the λ band was involved.

The second series of significance tests assessed significant differences in performance between the different frequency bands for the different subjects, concatenating the average results of all feature groups for that subject. The results for these tests are shown in Table 5.5. 1 and -1 mean that the CR was significantly different for the 3 task pairs, '1' indicating that the first FB was the best, and '-1' that the second FB was the best. 0 means that there were no significant differences in the same sense between both FBs for the 3 task pairs.

Subject	λ/α	λ/β_1	λ/β_2	α/β_1	α/β_2	β_1/β_2
1	1	1	1	1	1	-1
2	0	0	0	0	0	0
3	1	1	1	1	1	-1
4	1	0	0	0	0	-1
5	1	1	1	-1	0	1

Table 5.5: For the 5 subjects is shown whether there are significant differences between both FBs. 1 (-1): 1st (2nd) FB significantly better for the 3 pairs of tasks; 0: no significant differences.

For subjects 1 and 3, the classification results are significantly different for all FBs and all task pairs. For subject 2, none of the frequency bands were significantly different. CRs were always low, using another

FB did not increase the CR.
For the task pair L-R, the CRs are significantly different for all subjects between λ and α, and λ and β_2. In case of L-W, significant differences were noticed for λ/β_1 and λ/β_2, for all subjects. For R-W, λ is significantly better than the other 3 bands for the 5 subjects. Results for different frequency bands differ significantly for all pairs of bands, except for α/β_2.

Contrary to the synchronization features, yielding highest classification accuracy when computed from λ band filtered signals, band power features performed significantly better when computed in the narrower α than in the broader λ band.

Selected Features

- *Types:* 560 features, 496 PLV values (88.6%) and 64 PSD based features (11.4%), were computed from the 32 EEG signals. In the final feature subset for the task pair L-R, 12.9% of the features were PSD features and 87.1% PLV features. For L-W and R-W, 16.2% and 16.7% of the selected features were PSD features. These percentages result from averaging over all days of recording of all subjects.

- *Spatial location:* Considering the results combining BP and PLV features in their respective best FBs, the following EEG signals were selected.
  The left panel of Figure 5.4 shows in how many of the pairs selected for the final subset the different electrodes appear. Blue means no selections, yellow to red colors correspond to more selections. The bottom row shows the selections averaged over all days of recording of all subjects.
  The right panel of Figure 5.4 indicates where PSD features were selected. Where PSD features are rather selected on the motor cortex, the temporal and occipital regions seem to be more involved in coupling.

  The best performing PLV group was F4/C4 for L-R, and C4 and FC1 for L-W and R-W respectively. The most often selected electrodes in the final subset were O2, T8 and P3 for L-R, L-W and R-W respectively in case of PLV features. For PSD features, T7, C3, Oz, CP2 and C4 were the most often selected electrodes for L-R. For L-W and R-W, C4 and P8 were most often appearing in the final subset.
  For the 3 task pairs, more interhemispheric electrode pairs were selected than pairs located in the left or right hemisphere.

2-class Discrimination.

Table 5.6 presents the classification accuracy for the 2-class discrimination problem, CR(%), for the 3 pairs of tasks, obtained from 5-fold cross validation on training data. The training data were divided in 5 segments. The 5 respective segments were subsequently used for testing, while the other 4 were used for selecting the feature subset and training the classifier. For each subject, the performances are averaged over the 3 days of recording.
'El', 'PLV' and 'BP' refer to the first step of feature selection. 'El' is the best PLV group in the best frequency band, 'PLV' indicates the corresponding CR. The best FB for PLV was λ, for the 5 subjects. 'BP' indicates the CR obtained with BP features in the best FB. For subjects 1, 3, 4 and 5 the BP was best in the α band. For subject 2, β_2 was the best FB.

'Fin' indicates the CR resulting from the feature subset obtained after the second step of feature selection. This subset typically contains both PLV and PSD features.

For L-R, the CRs for BP features range from 60.0% to 77.9% for the 5 subjects, those of PLV from 62.0% to 77.4%. The electrode corresponding to the best performing PLV group is in the frontal, central or occipital region.
For L-W, BP features yielded CRs between 68.6% and 82.1%, while the best PLV group, corresponding to electrodes in frontal or central region, yielded CRs ranging from 67.1% to 78.2%.
For R-W, CRs were between 68.7% and 79.9% and 66.7% and 80.4% for BP and PLV respectively. The

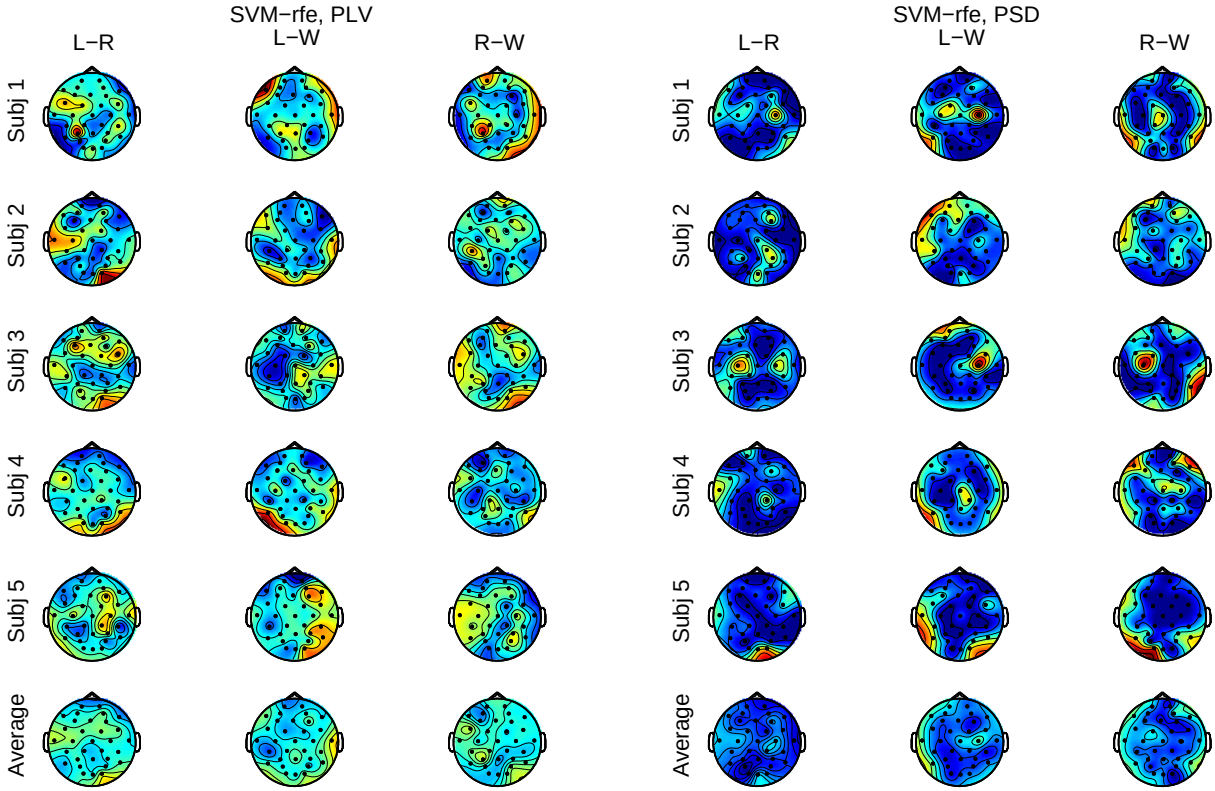

Figure 5.4: Left: Occurrence of the different electrodes in the PLV pairs selected for the final subset with SVM-rfe; Right: Electrodes where PSD features were selected for the final subset with SVM-rfe; Red corresponds to more selections.

best electrodes were mostly located in the frontal region.
For the 3 task pairs, the finally selected subset of features, containing as well PSD as PLV features, achieved about 9% better CRs, as compared to the CR of the best individual feature groups.
Comparing the 3 task pairs, L-R was less well discriminated than L-W and R-W.

3-class Discrimination

Figure 5.5 shows, for the 5 subjects and the 4 frequency bands, the generalization performance for discriminating the 3 classes. Subject 3 achieves the best generalization performance, for FB λ. Subject 2 and 4 hardly show differences in classification accuracy for the different FBs. This could be expected from Table 5.5, examining the differences between the FBs for the 3 task pairs. The CR and UR on training data, averaged over all subjects and all days were, for the frequency bands λ, α, β_1 and β_2 respectively, 78.6% and 5.7%, 74.9% and 6.2%, 73.5% and 6.6%, and 74.7% and 8.1%.

Table 5.7 presents the training and generalization results for 3-class discrimination, combining the PLV and BP features in their respective best frequency band. Comparison between Tables 5.5 and 5.7 indicates that the subjects not presenting significant differences in CRs for the different FBs, have the worse CR_{test}. This may suggest that they did not perform well the different mental tasks. The subjects presenting significant differences for all FBs and task pairs, achieve the highest generalization performance.
Comparing Tables 5.7 and 5.4 we see that for subjects 1, 3 and 5 the generalization error is greater when the set of features comprising PLV for all electrode pairs, BP and SMF is considered. For subjects 2 and 4, a smaller generalization error is achieved as compared to using the first set of 290 features.

Tasks	L-R				L-W				R-W			
	El	PLV	BP	Fin	El	PLV	BP	Fin	El	PLV	BP	Fin
Subj 1	F4	77.4	77.9	86.9	CP2	76.4	77.3	88.9	CP5	77.6	76.2	85.2
Subj 2	C3	67.4	66.1	76.0	FC2	69.9	73.1	82.6	FC2	66.7	68.7	79.3
Subj 3	F4	74.2	75.1	85.0	C4/F4	78.2	82.1	90.2	FC1	80.4	79.9	87.0
Subj 4	Oz	62.0	60.0	70.8	FC5	67.1	68.6	80.5	FC1/F4	66.9	68.8	79.7
Subj 5	C4/FC2	69.4	67.6	79.2	C4	71.6	69.5	79.8	Fz	72.5	69.1	86.6

Table 5.6: Results from 5-fold CV, obtained on training data for the 2-class discrimination problem, for the 3 pairs of tasks. El: PLV group for which the greatest CR was achieved; PLV: CR (%) for this PLV group; BP: CR (%) for BP features. BP and PLV features in their best FB. Fin: CR (%) for final subset (after the 2nd step of FS). Results were averaged over the 3 days of recording.

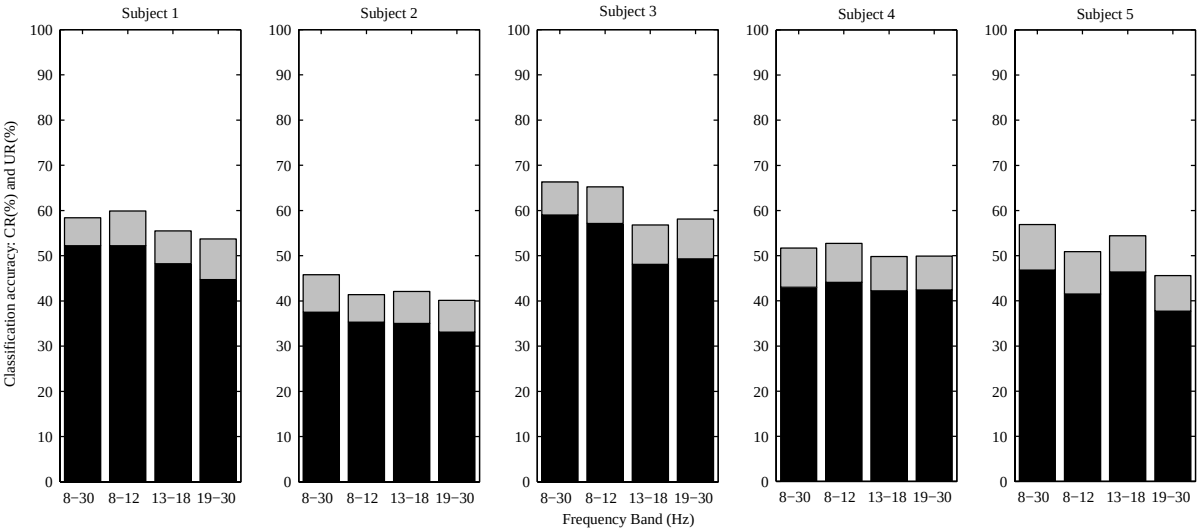

Figure 5.5: 3-class generalization performance using the finally selected subsets, averaged over the 3 days, for the 5 subjects and 4 FBs.

5.6.8 Discussion

We presented a feature selection procedure based on SVM-rfe and we applied it to study PLV computed from broadband and narrowband filtered EEG signals. Lal *et al* [87] used an adapted version of SVM-rfe to select EEG channels in a motor imagery paradigm. We used SVM-rfe in a 2-step feature selection procedure.

This FS strategy was originally validated by applying it to the 290 features presented in Section 4.6. PLV features were the most often occurring features in the final feature subsets. The most often selected electrodes were C4, C4 and P7 for the respective task pairs L-R, L-W, and R-W. We conclude that rather a decrease than an increase in μ rhythm is detected. Indeed, imagination of left hand movement causes the idling μ-rhythm to decrease in the activated right motor cortex.
Pfurtscheller and Neuper [116] showed that, independent of the required motor task -imagined versus real movement- the most prominent EEG changes were localized over the corresponding primary sensorimotor cortex. Our findings correspond well to neurophysiological evidence [116, 54] and validates the feature selection strategy. This FS strategy could thus be used in experiments where no prior knowledge about the mental tasks is available and to optimize the set of features for individual subjects.

We notice that the most often selected EEG signals when using the set of features comprising all individual electrode pairs, BP and SMF, correspond to the EEG signals selected from the 290 features presented in Section 4.6 in case of PSD features. In case of PLV features, however, the selected electrodes do not correspond. The averaged PLV values of neighboring electrode pairs used in Paragraph 4.6.5 could be more robust features than the PLV values for single electrode pairs used here. This could also explain

Subject	1	2	3	4	5	average
$CR_{train}(\%)$	85.1	75.5	85.9	72.1	77.3	79.18
$ER_{train}(\%)$	10.0	18.6	10.0	21.0	16.5	15.22
$CR_{test}(\%)$	55.1	39.6	59.4	43.0	47.9	49.00
$ER_{test}(\%)$	38.1	53.6	31.5	47.9	44.0	43.02

Table 5.7: Results for 3-class distinction obtained on training data and test data respectively. Features selected from the set of 34 groups in 2 steps with SVM-rfe. BP and PLV features were computed in their respective best frequency bands. For each subject: average over 3 days of recording.

the lower CR_{test} for subjects 1, 3 and 5.
With the features presented in Section 4.6, involving features obtained by averaging several PLV values, an average CR_{test} of 55.6% and $ER_{test} = 42.5\%$ were obtained. Here, using PLV values for single electrode pairs yields a CR_{test} of 49.00% and an $ER_{test} = 43.02\%$. In both cases, synchronization features were combined with PSD features in the feature subset selected by SVM-rfe. Using single PLV values between two EEG signals, reflecting more details of coupling than averaged values, does not result in a better classification of spontaneous EEG recorded during the 3 mental tasks 'left', 'right' and 'word' as compared to using averages of several PLV values.
Considering the 3 days of recording of the 5 subjects, we noticed that sometimes BP features were the best, sometimes PLV features. The finally selected feature subset contained both types features and achieved greater CRs. This may suggest some complementarity between single-channel PSD features and the bivariate PLV.

We observe that BP features perform significantly better in the narrow α frequency band. For PLV, significantly better results are achieved when computed from the broader λ band filtered signals, as compared to the narrower frequency bands α, β_1 and β_2. The interpretation of phase synchronization in case of broadband filtered signals, however, stays problematic.
Most people use narrowband signals to compute PLV or other synchronization measures. Chavez *et al* [27] found that in case of wide band analysis the spatio-temporal organizations obtained by the phase-synchrony index were sensitive to the clinical state. It could be explained by the complex interactions of different frequency contents during these periods. The spatio-temporal organizations were more specific in case of narrowband analysis.
This study, on the contrary, shows a better classification performance with PLV features derived from broadband than from narrowband signals. We have observed lower PLV values for the broader λ than for the narrower α band filtered signals. While the use of PLV for classification purposes has a different meaning than when used for studying brain function as in cognitive neuroscience, we think small PLV values may better discriminate mental tasks than larger ones. For stationary signals, the effect of filtering the signals under inspection into a narrow band forces the PLV towards higher values, due to two effects: 1) It reduces the noise on the signals' amplitudes and phases; 2) It increases the probability of having similar frequency content, which is a sufficient condition for having a high PLV value. By filtering signals in narrow bands, frequency components that could desynchronize the two signals are removed. By using broadband signals for computation of PLV, amplitude and phase noise increase, as well as the constraint on frequency matching. The local frequencies of the signals must be close for any synchronization to be possible. Synchronization is more difficult to obtain in case of broadband signals and, therefore, the PLV could be more discriminative in that case.

5.6.9 Conclusions

We presented a feature selection strategy that is validated by its correspondence to neurophysiological evidence. The strategy consists of applying SVM-rfe to different groups of features and applying the same algorithm a second time to select amongst the best features of all groups together. It allows to compare the classification performance that can be achieved with each type of feature.
In their respective best frequency bands, the PLV between signals recorded from different electrodes and the bandpower result in equally well performance for the classification of three pairs of tasks. Whereas the best FB for the bandpower was 8-12Hz, the best FB for PLV in this study was the broader frequency band 8-30Hz.

5.7 Fast Correlation-Based Filter

Computation time for SVM-rfe feature selection was of the order of hours, a significant improvement as compared to GAs. Still, we would like to have a good FS procedure at disposition that would be applicable between different recording sessions. Therefore, the Fast Correlation-Based Filter (FCBF) [159] was investigated. As it is a filter method, it is very fast.

We present here the application of a modified version of the FCBF to the comparison of phase synchronization and Power Spectral Density (PSD) features, computed from broadband and narrowband filtered EEG signals and their ability to discriminate 3 mental tasks [65]. To the best of our knowledge, this method had not yet been applied to BCIs.
We found that the features were selected from electrode signals corresponding to neurophysiological evidence, i.e. electrodes lying over the motor cortex. PSD and Phase Locking Value (PLV) features were more discriminative when computed from narrowband (8-12Hz) and broadband (8-30Hz) filtered signals respectively.

5.7.1 Feature Selection Algorithm

FCBF not only determines the relevant features, but it also includes a redundancy analysis.
First, the relevant features are determined from the symmetrical uncertainty between each feature and the class:

$$SU(X,Y) = 2\left(\frac{H(X) - H(X|Y)}{H(X) + H(Y)}\right), \tag{5.4}$$

where $H(X)$ is the entropy of variable X and $H(X|Y)$ is the entropy of X after observing values of Y.
Second, the concept of Markov blankets is used to remove redundant features. A feature M_i is said to be a Markov blanket for a feature F_i if M_i subsumes all information that F_i has about the class and about all of the other features. As the Markov blanket condition is difficult to verify, redundant features are detected by means of approximate Markov blankets. I.e., features F_i fulfilling $SU(M_i, F_i) > SU(F_i, C)$ are removed.

We used a modified version of the algorithm presented by Yu and Liu. Application of their method to features computed from noisy EEG data resulted in almost all features being removed. This is because the correlation between the individual features and the class is rather low. Therefore, we did not remove all features for which a considered predominant feature was an approximate Markov blanket, i.e., we did not remove all features for which the correlation with the predominant feature was greater than the correlation of the feature with the class. We removed only the 25% of features for which the difference between the correlation with the predominant feature and the correlation with the class, $SU(M_i, F_i) - SU(F_i, C)$, was greatest. For simplicity of writing, we will refer to this modified version with FBCF in what follows.

This algorithm was applied in a two-step feature selection procedure, as in Section 5.6. In a first step, features were selected within each of the 34 groups. In a second step, a final feature subset was selected amongst the features selected in the different groups. This was done for each of the 3 pairs of tasks.
Training and test set were also as described in Section 5.6.

5.7.2 Data Preprocessing - Feature Extraction

Data preprocessing was exactly as in Paragraph 5.6. The extracted features were the same as the ones presented in Paragraph 5.6, but only the frequency bands α and λ were considered.

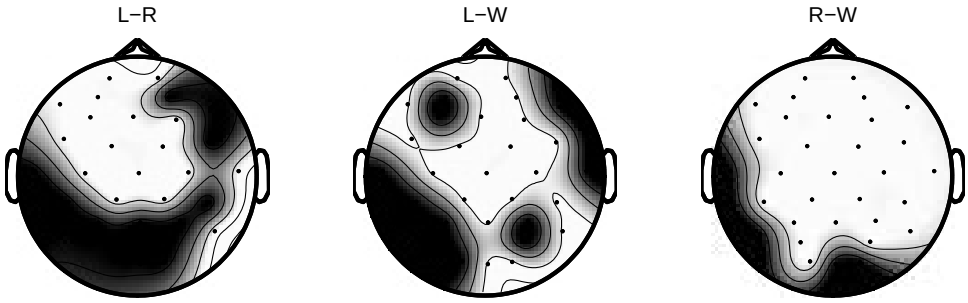

Figure 5.6: White: electrodes for which the PLV presented significant differences between λ and α frequency band.

5.7.3 Results using FCBF

Comparison of Different Frequency Bands

We performed a paired Student's T-test to look for significant differences between the 2 frequency bands for the different groups of features. We used a 5% level of significance to reject the null hypothesis of equal performances for both FBs. The results obtained for the respective feature groups for the 3 days of recording and the 5 subjects were concatenated to perform the significance test. The significance test was repeated for the 3 pairs of tasks. We found that if there was a significant difference in performance between the α- and the λ-frequency band, the λ-band was better for the PLV features and the α-band was better for PSD features. Figure 5.6 shows for which electrodes PLV presented significant differences.

Selected Features

Generally (averaged over all subjects all days), C4 was the most often selected electrode for the distinction of the task pairs L-R and L-W and FC1 for R-W. Contrary to [66], for L-R and L-W most of the electrode pairs whose PLV was selected were in the right hemisphere. For R-W pairs were mostly selected in the left hemisphere.
Considering PLV and PSD features separately, learned that in case of PLV the most often selected electrodes are C4, CP2/C4 and FC1 for the task pairs L-R, L-W and R-W respectively. For PSD, most often selected electrodes were C4, C4 and F7/FC1/C3 for the 3 respective task pairs. Both PSD and PLV features point towards the motor cortex, contrary to Paragraph 5.6.7 where selected PLV features corresponded less well to neurophysiological evidence for this set of features. The difference between FCBF and SVM-rfe may be due to the high dimensional and noisy feature space. Also Weston *et al* [154] point out that SVMs can suffer in high dimensional spaces where many features are irrelevant.

The incidence of selection of the 32 electrodes in the final feature subset for the 3 task pairs is reflected in the left and right panel of Figure 5.7, for PLV features and PSD-based features (BP and SMF) respectively. The most often selected electrodes appear in red.

3-class Discrimination

Table 5.8 shows the 3-class classification accuracies for the 5 subjects. As compared to results in Table 5.7, the classification accuracy on training data is much lower for FCBF than for SVM-rfe. The average classification accuracy on unseen test data was 49.00% in case of SVM-rfe feature selection. This was obtained with in the average 117 features. FCBF yields an average classification accuracy of 48.56% on test data, but obtains it with in the average 19 features only. FCBF is less prone to overfit the data. In addition, FCBF is very fast and the selected features correspond better to neurophysiological evidence.

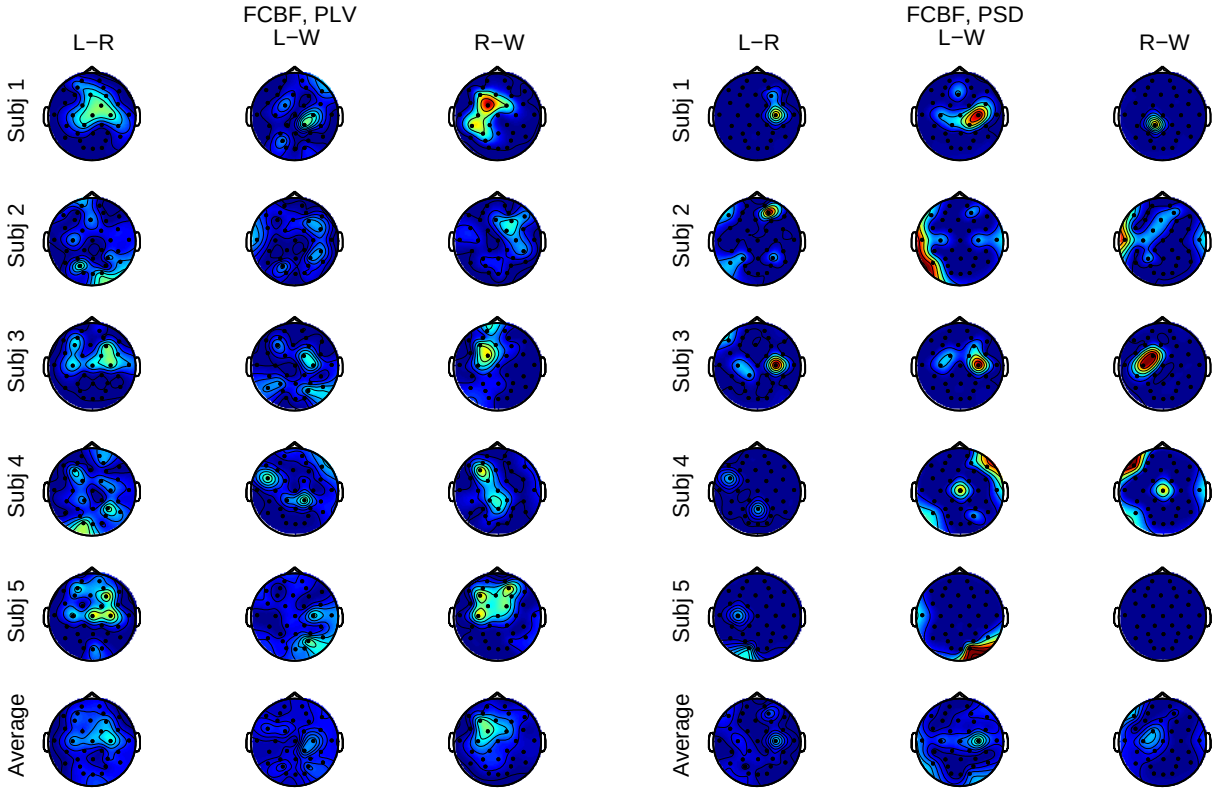

Figure 5.7: Left: Occurrence of the different electrodes in the PLV pairs selected for the final subset with FCBF; Right: Electrodes where PSD features were selected for the final subset with FCBF; Red color corresponds to more selections.

Subject	1	2	3	4	5	Average
$CR_{train}(\%)$	66.51	52.61	70.26	51.05	57.88	59.66
$ER_{train}(\%)$	27.95	41.46	23.61	41.12	35.47	33.92
$CR_{test}(\%)$	56.37	36.82	61.93	40.07	47.63	48.56
$ER_{test}(\%)$	35.93	57.63	30.67	51.11	44.44	43.96

Table 5.8: Classification accuracies and error rates for 3-class distinction, averaged over the 3 days of recording of each subject combining the respective best frequency bands for BP and PLV.

5.7.4 Discussion

The modified FCBF-algorithm used in a 2-step feature selection procedure allows to select neurophysiologically relevant features and compare the classification performance achieved with PSD and PLV features computed from broadband and narrowband filtered signals.

FCBF is a very fast feature selection method. It is a so called filter method, i.e., it does not use a classifier to select the features. FCBF uses the symmetrical uncertainty to determine the relevant features and remove the redundant ones.
Fast feature selection methods may allow for improvement of BCIs. Between two sessions new, actual, best features could be selected and used for classification in the next session.
In addition, FCBF selects significantly less features, allowing for an additional acceleration of the classification algorithm. A possible disadvantage of filter methods could be that the selected features are not well suited to the classifier, but in our experiments no significant generalization performance decrease has been detected, compared to the previously used SVM-rfe method.

Both PSD and PLV features were mostly selected from the central electrode signals. Both types of features point to the same brain region, the motor cortex. In case of PLV, the involved brain region seems to be wider, including also more frontally located areas.

The correspondence between the selected electrode signals and the brain regions predicted by neurophysiology to be involved in the considered mental tasks makes FCBF a valuable analysis tool. It could be used to investigate new protocols where we do not know a priori which brain regions are involved and which type of features carry the most discriminant information.

For the two frequency bands considered, 8-12Hz and 8-30Hz, PSD features yielded better results in the narrower 8-12Hz frequency band. This frequency band corresponds to the μ-rhythm, associated with executed or imagined motor tasks.
On the contrary, in case of PLV features better results were obtained using broadband, 8-30Hz, filtered EEG signals.

5.7.5 Conclusions

In the framework of a Brain-Computer Interface, we have applied a modified FCBF-algorithm in a 2-step feature selection procedure to data recorded during the mental tasks left and right hand movement imagination and word generation. The algorithm selected features in the brain regions predicted by neurophysiological evidence to be involved in these tasks. Better classification results were obtained using PSD and PLV computed from signals filtered in the narrower frequency band 8-12Hz and the broader 8-30Hz frequency band respectively. The generalization performance is as good as the one obtained with SVM-rfe, but this algorithm is faster and selects fewer features. These properties may make FCBF a valuable tool for further improvement of BCIs.

5.8 Application to BCI Competition Data

We participated in the BCI competition III[2]. We classified the data of two of the proposed datasets, dataset I and dataset IVa. The set of features we used contained power features in several frequency bands, PLV for all electrode pairs, the Cluster Participation Index (CPI) and features derived from the Empirical Mode Decomposition (EMD) of the EEG signals. Feature selection was done with the FCBF and for classification both SVM and DA classifiers were considered.
A description of both datasets and the procedure followed to classify the blind data is given in Paragraphs 5.8.3 and 5.8.8. We first introduce CPI and EMD.

5.8.1 Cluster Participation Index

Allefeld and Kurths [5] introduce the concept of a statistical phase synchronization cluster, in which oscillators participate in different degrees, ranging from no to perfect agreement with the cluster dynamics. The cluster participation index (CPI) measures the degree of participation of the oscillator in the cluster and ranges from 0 to 1. It quantifies how close an oscillator follows the common rhythm and how important its contribution to the cluster is. The objective of multivariate phase synchronization analysis is to derive from the matrix of bivariate indices some information about the synchronization state of all N oscillators.

5.8.2 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) [122, 48] seemed worth our attention in the framework of BCIs. EMD is a nonlinear technique for adaptively representing nonstationary signals as sums of zero-mean AM-FM components. EMD considers oscillations in signals at a very local level. A high-frequency part, or local detail, is defined and separated from the identified low-frequency, local trend. This is done for all the oscillations composing the entire signal, and the procedure can subsequently be applied on the residual consisting of all local trends. Constitutive components of a signal can therefore be iteratively extracted.

[2] http://ida.first.fraunhofer.de/projects/bci/competition_iii/

A sifting process makes the detail signal zero-mean according to some stopping criterion. The detail is then referred to as Intrinsic Mode Function (IMF). Such a component is locally, at the scale of one oscillation, in the highest frequency band. This method can be regarded as an automatic and adaptive, signal dependent, time-variant filtering.

5.8.3 Dataset I

The first dataset [86] consisted of ECoG recorded during a cued motor imagery paradigm (left pinky, tongue) and aimed to test session-to-session transfer. Training and (blind) test data are ECoG recordings from two different sessions with about one week in between.

5.8.4 Dataset I: The Thrill

The design of a classifier for a BCI system is very challenging when a classifier that was trained on the first day shall classify data recorded during following days (if possible, without retraining): the patient might be in a different state concerning motivation, fatigue etc. so that his brain will show different electrical activity. In addition, the recording system might have undergone slight changes concerning electrode positions and impedances. This data set reflects this situation: training data and test data were recorded from the same subject and with the same task, but on two different days with about 1 week in between. The goal was to learn on the data of the first session and achieve a classification accuracy as high as possible on the second session.

5.8.5 Dataset I: Experiment

During the BCI experiment, a subject had to perform imagined movements of either the left small finger or the tongue. The time series of the electrical brain activity was picked up during these trials using a 8x8 ECoG platinum electrode grid which was placed on the contralateral (right) motor cortex (64 ECoG channels, 0.016-300Hz). The grid was assumed to cover the right motor cortex completely, but due to its size (approx. 8x8cm) it partly covered also surrounding cortex areas. All recordings were performed with a sampling rate of 1000Hz. After amplification the recorded potentials were stored as microvolt values. Every trial consisted of either an imagined tongue or an imagined finger movement and was recorded for 3 seconds duration. To avoid visually evoked potentials being reflected by the data, the recording intervals started 0.5 seconds after the visual cue had ended. 278 trials were available for training; the test set consisted of 100 blind trials.

5.8.6 Dataset I: Methods

After visual inspection of the power spectra of the signals, it was decided to compute the band power in the frequency bands 7-9Hz, 6-10Hz, 9-17Hz, 17-19Hz, 16-20Hz, 25-29Hz and 7-30Hz, for all electrode signals. They were normalized with the band power in 4-40Hz. We also computed the SMF, as in equation (5.3), but using the frequency band 7-30Hz this time. The PLV was computed for all possible pairs of electrode signals, filtered in 8-30Hz. From the matrix of PLV values, the CPI was calculated.
Furthermore, the first 3 IMFs resulting from the EMD were retained for computation of additional features. For each component, the SMF and the ratio of the band power of this component and the band power of the original filtered signal was computed. For the second component, PLV was computed for all possible electrode pairs.
For every trial, features were computed from 5 overlapping 1s windows.

FCBF feature selection was applied in two steps. First, the best features within each feature group were selected. Then a final set was selected amongst all best features.
Our modified version of the FCBF algorithm has a parameter (x) to set which is the percentage of features retained in each iteration of the algorithm. This parameter has to be set for both the first and second step of feature selection.
We selected subsets of features using a few different values for this parameter. Subsequently, classification

accuracy achieved with the selected subset was evaluated on training data with 5-fold CV. Every trial was classified once and was training data 4 times. To classify a trial, majority voting was applied to the class obtained for the 5 overlapping 1s windows. Both SVMs and robust DA classifiers were used for comparison.

We studied the 5-fold CV results obtained on the training trials to determine for which of the two classes most errors occurred using the different feature subsets and classifiers. We compared these findings to the number of times each class was obtained for the test trials with the same features and classifier.
Particular interest was given to those cases where the mistakes made were different on the test set. E.g., when for the training data, many more trials are classified in class '1' than in class '-1', and on the test data there is an equilibrium of both classes or or a lot more trials are classified '-1', it could mean the the classifier is really "sure" this class is the correct one.

For the SVM-classifier, trials were classified much more often in 1 class than in the other, for both training data and test set. As the most appearing class was the same for training and test set, probably the same errors were made on the test set as on the training set.
For the DA-classifier, the appearance of the 2 classes was more balanced, and sometimes the inversion of disequilibrium between the two classes obtained for training and test set occurred.
We also studied the generalization error and compared the labels given by the best feature sets.

Finally, the parameter combination $x_1 = 0.50$ and $x_2 = 0.90$ was chosen, resulting in 19 selected features:

- The PLV for the following electrode pairs: 29-55, 31-32, 31-55, 37-45, 38-17, 38-39, 39-55
- The CPI for electrode 7
- The PLV between 2nd IMFs for the following electrode pair: 14-45
- Power 6-10Hz for electrode 22
- Power 9-17Hz for electrode 38
- Power 17-19Hz for electrode 61
- Power 7-30Hz for electrode 12
- Power 1st IMF for electrode 38, 48, 54
- Power 2nd IMF for electrode 31, 38
- Power 3rd IMF for electrode 18

With these features and and DA classification, a generalization error of 79.857% was obtained on the training data by means of 5-fold CV.

5.8.7 Dataset I: Results on blind test data

A classification accuracy of 79% was obtained for the blind test data. This accuracy was nearly the same as the classification accuracy obtained by 5-fold cross-validation on the training data. This result yielded a 13th place in the ranking (27 ranked). The group that won the competition for this data set obtained a classification accuracy of 91%. Average classification accuracy of all solutions submitted was 70.44%.

5.8.8 Dataset IVa

This dataset related to cued motor imagery (right hand, foot) and aimed at finding solutions for small training sets. Data were recorded from 5 subjects. From 2 subjects most trials were labelled (resp. 80% and 60%), while from the other 3 less and less training data were given (resp. 30%, 20% and 10%), see Table 5.9. The challenge is to make a good classification even from little training data, thereby maybe using information from other subjects with many labeled trials.

Subject	'al'	'aa'	'av'	'aw'	'ay'
Class 1	112	80	42	30	18
Class 2	112	88	42	26	10

Table 5.9: Available number of labeled trials for the respective subjects, other trials unlabeled; Total: 280 trials per subject

5.8.9 Dataset IVa: The Thrill

When taking a machine learning approach to Brain-Computer Interfacing, one has to have labeled training data to teach the classifier. To this end, the user usually performs a boring calibration measurement before starting with BCI feedback applications. One important objective in BCI research is to reduce the time needed for the initial measurement. This data set poses the challenge of getting along with only a little amount of training data. One approach to the problem is to use information from other subjects' measurements to reduce the amount of training data needed for a new subject. Of course, competitors may also try algorithms that work on small training sets without using the information from other subjects.

5.8.10 Dataset IVa: Experiment

This data set was recorded from five healthy subjects. Subjects sat in a comfortable chair with arms resting on armrests. This data set contains only data from the 4 initial sessions without feedback. Visual cues indicated for 3.5 s which of the following 3 motor imageries the subject should perform: (L) left hand, (R) right hand, (F) right foot. The presentation of target cues were intermitted by periods of random length, 1.75 to 2.25s, in which the subject could relax.
There were two types of visual stimulation: (1) where targets were indicated by letters appearing behind a fixation cross (which might nevertheless induce little target-correlated eye movements), and (2) where a randomly moving object indicated targets (inducing target-uncorrelated eye movements). From subjects 'al' and 'aw' 2 sessions of both types were recorded, while from the other subjects 3 sessions of type (2) and 1 session of type (1) were recorded. Data were recorded with 118 EEG channels at a sampling rate of 1000Hz and filtered in 0.05-200Hz. 280 trials were recorded for each subject.

5.8.11 Dataset IVa: Methods

The procedure followed was similar to the one presented in Paragraph 5.8.6. The computed features were the same. Again, for characterizing each trial 5 overlapping 1s windows were used. Data from the 5 subjects were used to select features with the FCBF. Also the evaluation of the feature sets obtained for the different parameter combinations x_1 and x_2 and the final choice of classifier based on an evaluation of the mistakes made by the classifier was as in Paragraph 5.8.6.
The methods for evaluation of dataset I and dataset IVa differ in the selection of the features and the training of the classifiers. For some subjects very few training data were at disposition, therefore, we decided for the following strategy.

For each subject, 10 labeled trials were randomly selected for each of the two classes. These data were used for feature selection. The number of trials selected corresponds to the minimum number of trials available for each class for the different subjects. This way every subject was given an equal importance for feature selection and training of the common classifier. Data of the different subjects were combined in an attempt to select a more robust set of features (yielding a better generalization performance) for both the subjects of whom a lot of labeled trials were available by not getting too specific for individual subjects and those for whom not many labeled trials were at disposition by using more information available and.

Then a common classifier was trained with the selected features and the randomly selected trials of the different subjects. An "individual" classifier was trained with the features selected on all subjects but using labeled data from the considered subject only. Both SVM and DA were used for classification purposes.
We also experimented with repeating (20 times) the FS and training of the classifier for different sets of randomly selected trials for the subjects. We then retained the features that were selected more than

Subject	'al'	'aa'	'av'	'aw'	'ay'
FS repeated	no	yes	no	yes	no
x_1	50	85	90	95	85
x_2	95	85	90	95	85
Classifier	ind	ind	ind	ind	comm
Type	DA	SVM	SVM	SVM	DA
$CR_{gen}(\%)$	91.5	73.8	70.2	82.1	92.9
$CR_{comp}(\%)$	96.4	69.6	64.3	69.6	61.9

Table 5.10: Selected classifier for the different subjects

once or twice. The common classifier retained was the one trained on the set of random trials that yielded the best classification accuracy.
Using all labeled trials of all subjects for feature selection and training of the common classifier instead of just 10 trials per class and per subject, did not yield better results on training data.

Again, the mistakes made by the classifiers on the training set and the distribution of the unlabeled data over the two classes was studied to decide for the feature set and classifier. Table 5.10 shows for each subject whether FS was repeated for different sets of randomly selected trials, the finally chosen feature set (determined by x_1 and x_2), the type of classifier and whether the common or individual one was used. CR_{gen} is the 5-fold CV classification accuracy obtained from training data. The classification accuracy obtained on the blind data is indicated as CR_{comp}.

For all 5 subjects, the finally chosen set of features comprised mostly PLV values for pairs of filtered EEG signals and for pairs of 2nd IMFs. Furthermore, CPI, the SMF for the first IMF of some electrode signals and the power computed from some electrode signals sometimes occurred in the selected set of features.

5.8.12 Dataset IVa: Results on blind test data

In Table 5.10 the subjects are ranked according to the number of labeled trials available for training. For Subject 'al', for whom most labeled trials were available, the CR obtained on blind data is greater than the 5-fold cross-validation CR obtained on training data. For the other subjects, however, the obtained CRs are smaller than the 5-fold CV CR on training data and the difference increases with decreasing number of labeled trials available for training. For the first 4 subjects, the individual classifier was preferred to the common one.
The first ranked group, in the competition, obtained an accumulated classification accuracy of 94.17%. They used CSP, AR and Temporal Waves of the readiness potential as features and a LDA classifier. Their approach, however, is not applicable in real-world problems, as they used adaptation by former classified test samples as extended training samples for the datasets with small amount of training data. With the method here described, an accumulated classification accuracy of 67.86% was obtained, good for a 7th place, out of 14 ranked. Average classification accuracy of all submitted solutions was 67.71%.

5.8.13 Conclusions

The results obtained for the BCI competition III demonstrate the relevance of the presented features and feature selection algorithms for the considered problems. From the evaluation of the selected features, we conclude that especially PLV features and PSD and SMF computed from the EMD components are interesting in this context. Further investigation into how to better use these features and how to combine them with other features is necessary to further exploit the potential of the presented methods.

5.9 Comparison of Different Sets of Features with FCBF

5.9.1 Introduction

In this section, we apply FCBF feature selection to study different sets of features. FCBF feature selection is independent of the classification algorithm. Different classifiers can thus be applied to the selected features. For these experiments, we used linear SVMs and linear DA classifiers. Descriptions of these classifiers can be found in Section 3.5.

5.9.2 Feature sets used

We study 3 different sets of features. The first set is the one presented in Section 4.6 and comprises averaged PLV and Coh_{max} values, PLV and Coh_{max} values for some individual electrode pairs and PSD features. The second one is very similar to the first one, but features are normalized differently and some features were added to the first set of 290 features. The third set comprises PLV for all individual electrode pairs, power features and features derived from the IMFs. No artifact detection or noise reduction algorithms were applied.
To exclude the influence of the number of features on the CR, we select, with FCBF, the same number of features from each of the 3 feature sets. We then compare the classification accuracy and study the selected features.
FCBF was chosen because it proved to be a suitable tool for feature selection in previous experiments and it is very fast. The parameters x_1 and x_2 were chosen such that 9 features were selected, for each pair of tasks.

1st set: Power and Synchronization Averages

The first set of features is the one used in Section 4.6 [59] and consists of power, PLV and Coh_{max} features amounting for a total of 290 features. Power features were computed in the α, β_1, β_2 and λ band. PLV and Coh_{max} features were mostly averages over several individual PLV and Coh_{max} values. In addition, some PLV and Coh_{max} values for individual electrode pairs in the fronto-centro-parietal region were included in the set features. At every time instance, the features were normalized with the average over all features of the same type at that time instance.

2nd set: Power and Synchronization Averages, different normalization

The second set of features is similar to the first one, but with a different normalization. Instead of averaging over several features at the same time instance, power features were normalized by the total power in the band 4-40Hz. Synchronization features are between 0 and 1 and were used as such. A few features have been added as compared to the first set: the cluster participation index, the spatial center of gravity, and the SMF.

Spatial Center of Gravity. By performing (imagined) left and right hand movements, the μ-rhythm will attenuate in the right and left motor cortex respectively. Therefore, we designed a feature, which we call the Spatial Center of Gravity (SCG), to reflect how the power in a certain frequency band moved over the scalp, from left to right (x-component) and from occipital to frontal region (y-component). We defined a global and local SCG of which the x- and y-coordinates were computed as follows:

$$\begin{aligned} SCGglobal_x \;=\; & (-4(BP(F7)+BP(T7)+BP(P7)) \\ & -3(BP(Fp1)+BP(FC5)+BP(CP5)+BP(01)) \\ & -2(BP(AF3)+BP(F3)+BP(C3)+BP(P3)+BP(PO3)) \\ & -1(BP(FC1)+BP(CP1))+1(BP(FC2)+BP(CP2)) \\ & +2(BP(AF4)+BP(F4)+BP(C4)+BP(P4)+BP(PO4)) \\ & +3(BP(Fp2)+BP(FC6)+BP(CP6)+BP(O2)) \\ & +4(BP(F8)+BP(T8)+BP(P8)))/BP(El) \end{aligned}$$

(5.5)

$$\begin{aligned} SCGglobal_y &= (-4(BP(O1)+BP(O2))-3(BP(PO3)+BP(PO4)) \\ &\quad -2(BP(P7)+BP(P3)+BP(Pz)+BP(P4)+BP(P8)) \\ &\quad -1(BP(CP5)+BP(CP1)+BP(CP2)+BP(CP6)) \\ &\quad +1(BP(FC5)+BP(FC1)+BP(FC2)+BP(FC6)) \\ &\quad +2(BP(F7)+BP(F3)+BP(Fz)+BP(F4)+BP(F8)) \\ &\quad +3(BP(AF3)+BP(AF4))+4(BP(Fp1)+BP(Fp2)))/BP(El) \end{aligned}$$

(5.6)

$$\begin{aligned} SCGglobal_x &= (-4(BP(T7)+BP(P7))-3(BP(FC5)+BP(CP5)) \\ &\quad -2(BP(C3)+BP(P3))-1(BP(FC1)+BP(CP1)) \\ &\quad +1(BP(FC2)+BP(CP2))+2(BP(C4)+BP(P4)) \\ &\quad +3(BP(FC6)+BP(CP6))+4(BP(T8)+BP(P8)))/BP(El) \end{aligned}$$

(5.7)

$$\begin{aligned} SCGglobal_y &= (-2(BP(P7)+BP(P3)+BP(P4)+BP(P8)) \\ &\quad -1(BP(CP5)+BP(CP1)+BP(CP2)+BP(CP6)) \\ &\quad +1(BP(T7)+BP(C3)+BP(C4)+BP(T8)) \\ &\quad +2(BP(FC5)+BP(FC1)+BP(FC2)+BP(FC6)))/BP(El) \end{aligned}$$

(5.8)

BP(El) is the sum of the BP of all electrodes used for computation of the specific coordinate (x or y) of the considered SCG (global or local). The BPs considered were the α and β band respectively.

3rd set: Power, PLV for Individual Electrode Pairs, Features Computed from IMFs

At every time instance, a 79x32 matrix of features is computed:

- The first 32 rows correspond to the PLV between the 32 respective electrodes and all other electrodes. PLV is computed from EEG signals filtered in 8-30Hz.
- From the PLV values for all electrode pairs, the CPI is computed for all 32 electrode signals
- The SMF was computed as in Equation (5.3) using the PSD estimated at the frequencies between 8 and 30Hz.
- Further feature groups are the power in α frequency band, the β_1 band, the β_2 band, and the λ band, for all electrodes. These band powers were computed from EEG signals filtered in 4-40Hz and normalized with the power in the frequency band 4-40Hz.
- Then, for each of the first 3 IMFs, the SMF and the ratio between the power in the IMF and the total power (4-40Hz) was computed.
- For the second IMF, approaching μ-rhythm frequency wise, also the PLV features were computed.
- The CPI based on PLV values for the second IMF is computed.
- At last, the PLV interaction with neighbors and group averages were added to the feature set. There were only 11 group averages, so 21 columns in the last row of the feature matrix were empty.

5.9.3 Comparison of Feature Sets

FCBF was applied to select features from each of the 3 sets of features. This was done with 5-fold cross validation. For each day of recording of each subject, 4 sessions were used for selection of features and for training the classifier, the 5th session was used for testing. This was repeated 5 times, such that every session was test set once. The obtained 5-fold CV CRs are reported in the tables below. Selected features are visualized on head plots.

1st set of Features

Classification Accuracies. Features were selected from the set of 290 features and from this set without Coh_{max} features and without PLV features respectively. Furthermore, we evaluated this set without Coh_{max} features and PLV values for individual electrode pairs, and for the two sets composing the latter: only power features and only PLV averages.
The classification accuracies obtained for these 6 cases are presented in Table 5.11 for SVMs and Table 5.12 for DA classification. Average CRs are smaller for classification with robust Discriminant Analysis than with SVMs.
In case of SVM classification, we observe for Subjects 1, 3, 4 and 5 that CR is smaller when PLV features are left out than when Coh_{max} features are left out. For Subjects 3, power features combined with Coh_{max} features yield better results. For Subjects 2 and 4, the CR obtained with features selected from the entire set improves slightly when Coh features and PLV values for individual electrode pairs are left out. For Subjects 3 and 5, this causes a slight decrease in CR. Using only power features yields better results as compared to using only PLV averages. The combination of both, however, results in a CR better than the ones obtained for the individual sets, except for Subject 4.
The average CR obtained with features selected from the set of features without Coh features and PLV values for single electrode pairs is 51.11% (3-class discrimination).

Subject		1	2	3	4	5	Average
1st set	CR	61.56	39.96	**63.79**	36.94	**52.86**	51.02
290 features	UR	6.08	4.72	6.87	9.66	5.93	6.65
1 FS step	ER	32.36	55.33	29.34	53.40	41.21	42.33
1st set, no Coh_{max}	CR	**62.30**	39.00	61.73	39.80	51.53	50.87
209 features	UR	6.36	4.40	7.08	7.78	5.90	6.30
1 FS step	ER	31.34	39.48	31.19	52.42	42.58	39.40
1st set, no PLV	CR	59.30	39.72	61.65	38.23	47.77	49.33
209 features	UR	7.13	5.08	7.91	11.25	7.36	7.75
1 FS step	ER	33.57	55.20	30.44	50.53	44.87	42.92
1st set, no Coh_{max}, no PLV pairs	CR	61.56	**40.27**	63.50	39.49	50.71	**51.11**
169 features	UR	6.19	4.06	7.20	8.02	6.95	6.49
1 FS step	ER	32.25	55.67	29.30	52.49	42.33	42.41
1st set, only power	CR	59.11	40.24	59.97	39.96	47.77	49.41
128 features	UR	7.52	5.11	7.51	8.37	7.73	7.25
1 FS step	ER	33.37	54.65	32.52	51.67	44.59	43.36
1st set, only PLV avgs	CR	54.27	38.23	50.71	34.65	44.25	44.42
41 features	UR	6.04	4.14	4.52	5.43	7.04	5.43
1 FS step	ER	39.69	57.63	44.77	59.92	48.71	50.15

Table 5.11: Results for 1st set of features, SVM classifier

Selected Features. The different panels in Figure 5.8 show, for the 5 respective subjects and averaged over all subjects, the power features selected in the frequency bands α, β_1, β_2 and λ from the set of 290 features for the 3 pairs of tasks. The amount of times an electrode is selected for the different types of features is reflected in the figures according to the scale presented in Figure 5.2, where 0 means no selections and max corresponds to the number of selections of the most selected electrode. For each figure (comprising several head plots) the maximum value corresponding to the red color was determined from the maximum value appearing in the head plots contained in that figure.

Subject		1	2	3	4	5	Average
1st set	CR	59.62	35.45	60.46	35.26	51.47	48.45
290 features	UR	6.69	7.09	5.94	9.87	6.46	7.21
1 FS step	ER	33.69	57.46	33.60	54.86	42.08	44.34
1st set, no Cohmax	CR	60.06	36.12	59.01	39.62	49.85	48.93
209 features	UR	6.48	5.97	5.94	7.87	8.41	6.93
1 FS step	ER	33.46	57.91	35.05	52.52	41.74	44.14
1st set, no PLV	CR	57.78	37.69	59.12	36.12	46.56	47.45
209 features	UR	6.95	7.00	7.61	10.88	7.79	8.05
1 FS step	ER	35.27	55.31	33.27	53.01	45.65	44.50
1st set, no Cohmax, no PLV pairs	CR	59.62	36.97	60.19	39.51	49.47	49.15
169 features	UR	5.06	5.39	6.28	8.22	6.89	6.37
1 FS step	ER	35.33	57.64	33.53	52.27	43.64	44.48
1st set, only power	CR	58.28	36.76	57.78	**41.28**	45.72	47.96
128 features	UR	7.00	5.61	6.84	7.72	8.44	7.12
1 FS step	ER	34.71	57.64	35.38	51.01	45.83	44.91
1st set, only PLV avgs	CR	49.49	33.27	53.49	35.95	45.31	43.50
41 features	UR	6.17	4.69	5.02	6.02	6.26	5.63
1 FS step	ER	44.33	62.04	41.49	58.03	48.43	50.86

Table 5.12: Results for 1st set of features, DA classifier

For Subject 1, α band, we see the greatest number of selections on the motor cortex. Only few features are selected in the other frequency bands. In case of Subject 2, we can see selection of electrode C4 for the task pairs L-W and R-W in the frequency bands α and β_1. However, other electrodes are more often selected, especially in the frontal and temporal region.
The figures resulting from feature selection for Subject 3 show a strong detection of μ activity on the motor cortex. For the task pair L-R, α power is often selected for both electrode signals C3 and C4. C4, which is in the nondominant hemisphere, is selected more often than C3. For the task pairs L-W and R-W, α power is selected from electrodes C4 and C3 respectively. This indicates a detection of attenuation of the μ-rhythm for the motor imagery tasks. Very few features were selected for the other frequency bands.
For Subject 4, features are selected in several frequency bands and from several electrodes. The highest number of selections was from the Cz electrode in the β_2 frequency band. For Subject 5, most often selected was electrode C3 in the frequency band β_1 for the task pair R-W. For the other two task pairs we can observe a few selections of the electrodes C3 and C4 in this frequency band. In the α band, we can observe selection of electrode C3 for the task pair L-R and of the occipital region for task pairs L-W and R-W. This may suggest a visual imagery of the words generated during the word task.

For the subjects achieving CRs above 60%, we see clear involvement of the motor cortex, which corresponds well to neurophysiological evidence. Averaged, over all subjects, C3 and C4 are most often selected, mostly in the alpha band and less often in the β_1 frequency band. Furthermore, we observe some involvement of the occipital region in the alpha band where "Word" is one of the tasks and of the temporal region in the β_2 frequency band for the task pair L-W.

The selected synchronization features are presented in Figure 5.9, for the different subjects individually and averaged over all subjects. The scale [0 4] is used for all plots in this figure.
For Subject 1, 3 and 5, we can observe an involvement C4 for the task pairs L-W. For these subjects there is also some occipital activation for L-W and R-W.

The types of features selected are reflected in Table 5.13. Averaged over all subjects and all days, PLV and α power features are most often appearing in the final subset. Synchronization features were mostly selected in the right, right and left hemisphere for the 3 respective task pairs. Power features were mostly selected in the left, right and left hemisphere.

		L-R					
Subject		1	2	3	4	5	Av
PLV group	9	2.67	2.00	0	0	0.67	1.07
Cohmax group	9	0	0	0	1.67	1.33	0.60
PLV pairs	40	12.00	9.33	7.00	1.00	8.67	7.60
Cohmax pairs	40	0.33	0	1.67	24.00	0	5.20
PLV electr	32	3.67	6.67	3.67	0.33	2.33	3.33
Cohmax electr	32	3.67	0	5.33	13.67	1.67	4.87
α	32	15.33	2.33	18.00	2.00	10.00	9.53
β_1	32	2.33	5.67	1.67	0.67	6.67	3.40
β_2	32	1.00	10.33	1.00	1.33	3.67	3.47
λ	32	4.00	8.67	6.67	0.33	10.00	5.93
Synch left		1.67	2.67	1.67	7.00	1.67	2.93
Synch right		5.67	2.67	7.00	2.00	2.00	3.87
Power left		13.33	12.00	14.33	2.00	15.33	11.40
Power right		9.33	14.00	12.00	2.00	12.00	9.87
		L-W					
Subject		1	2	3	4	5	Av
PLV group	9	3.00	1.33	4.33	0.67	4.00	2.67
Cohmax group	9	0	0	0.33	0.33	0.33	0.20
PLV pairs	40	14.33	8.67	17.33	4.00	6.00	10.07
Cohmax pairs	40	0	0.33	1.00	24.67	0	5.20
PLV electr	32	11.00	5.67	10.00	0.00	8.33	7.00
Cohmax electr	32	0	0.67	1.33	4.67	0.33	1.40
α	32	11.33	3.00	7.67	4.00	12.67	7.73
β_1	32	0.67	6.00	0	1.33	5.67	2.73
β_2	32	0	14.33	0	4.33	3.67	4.47
λ	32	4.67	5.00	3.00	1.00	4.00	3.53
Synch left		4.00	2.00	6.00	1.67	0	2.73
Synch right		6.33	3.33	5.00	2.00	8.00	4.93
Power left		6.00	9.67	0.67	4.00	12.33	6.53
Power right		8.67	17.67	8.33	4.33	12.00	10.20
		R-W					
Subject		1	2	3	4	5	Av
PLV group	9	5.33	2.00	4.67	0	2.00	2.80
Cohmax group	9	0	0	1.33	1.33	2.67	1.07
PLV pairs	40	9.00	7.33	12.00	4.00	13.33	9.13
Cohmax pairs	40	0	0.33	0.67	12.00	0	2.60
PLV electr	32	11.33	8.67	11.00	1.33	6.67	7.80
Cohmax electr	32	2.67	2.00	5.33	12.67	4.33	5.40
α	32	10.00	5.33	8.33	7.33	5.33	7.27
β_1	32	0.33	6.67	0	2.33	5.67	3.00
β_2	32	2.00	3.33	0	3.00	1.33	1.93
λ	32	4.33	9.33	1.67	1.00	3.67	4.00
Synch left		5.00	6.33	9.67	6.67	3.33	6.20
Synch right		6.67	4.33	5.00	3.33	7.67	5.40
Power left		12.33	11.67	9.67	5.00	10.67	9.87
Power right		0.67	11.33	0.33	5.67	4.33	4.47

Table 5.13: Types of features selected for 1st set of features and left-right distribution

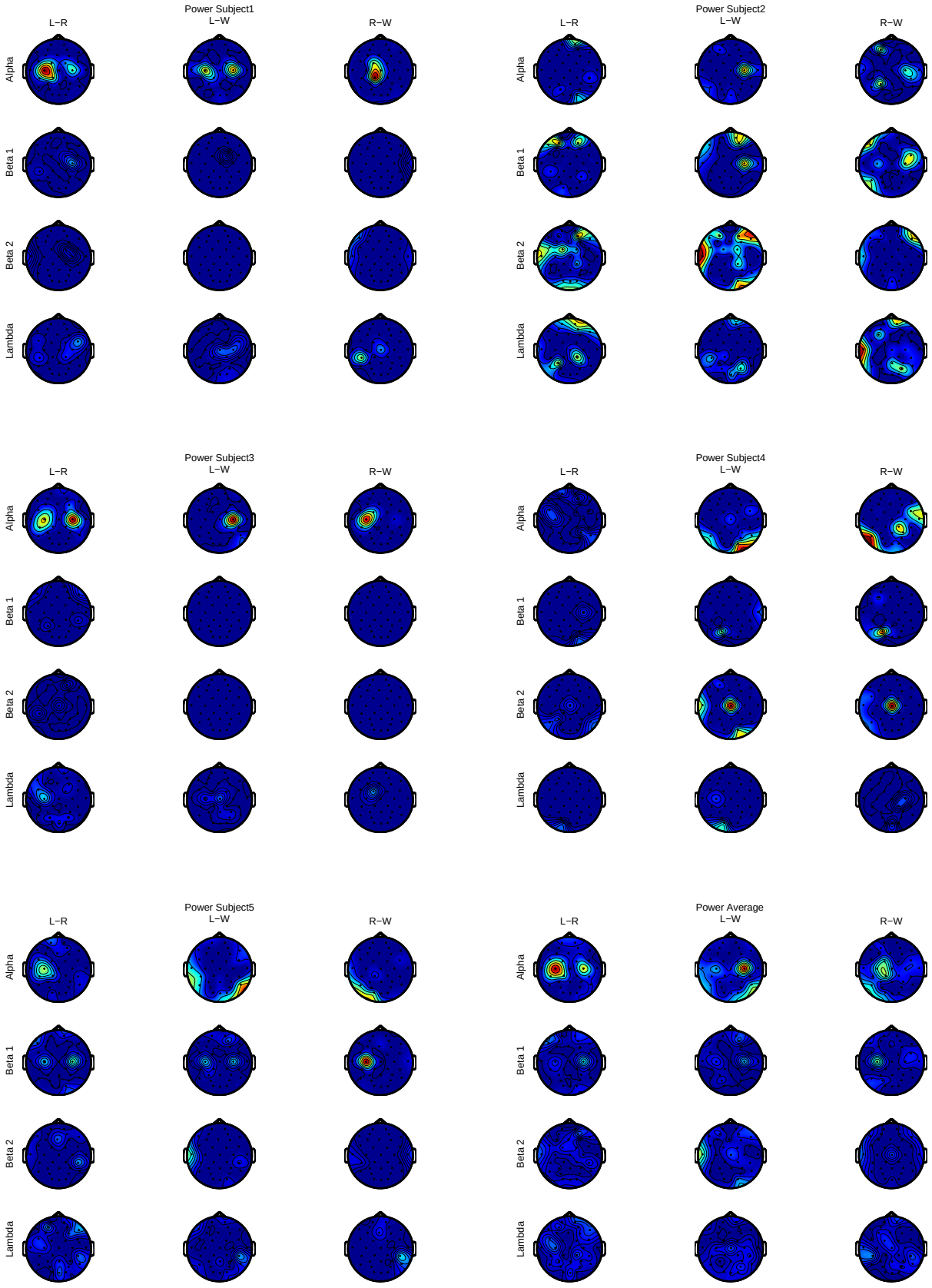

Figure 5.8: Selected features in α, β_1, β_2 and λ band, 1st set of features, for subjects 1, 2, 3, 4, 5 and averaged over all subjects

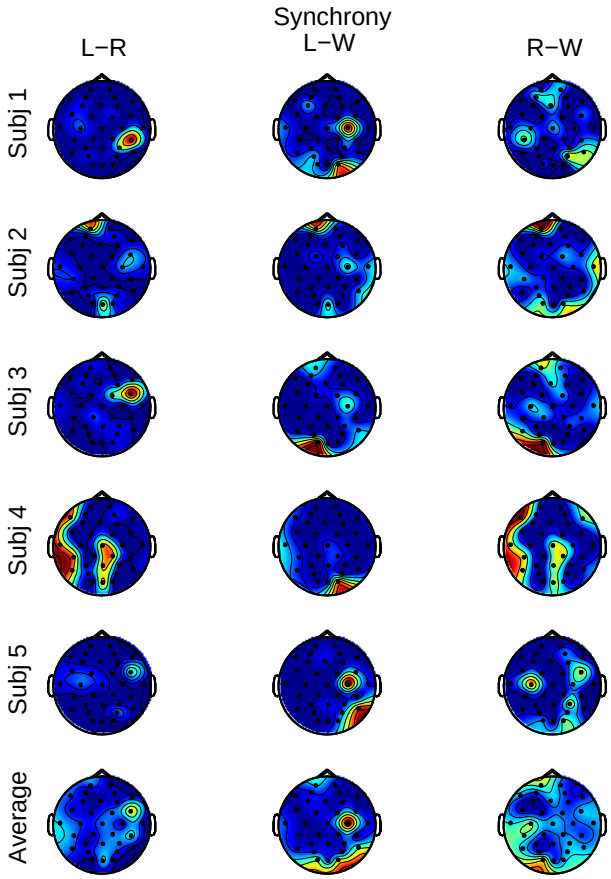

Figure 5.9: Selected synchronization features, 1st set of features

2nd set of Features

Classification Accuracies. Results obtained with SVMs (Table 5.14) outperform the ones obtained with DA (Table 5.15). Selecting features from all features of the 2nd set, an average CR of 50.25% was obtained. Reducing this set of features to a set of 290 features similar to the 1st set yields a CR of 50.77%. The only difference between this reduced set and the 1st set is that power features were normalized with the 8-30Hz band power and that λ features were replaced with SMF. There was no division by the mean of all features of the same type at the same time instance. As compared to the results in Table 5.11, these differences do not significantly affect the classification accuracy. Leaving out all Coh and all PLV features respectively yielded decreased CRs. The decrease was greater when PLV features were left out.

Subject		1	2	3	4	5	Average
2nd set	CR	58.81	**39.51**	62.62	39.96	50.35	50.25
362 features	UR	5.95	6.49	6.90	6.99	7.63	6.79
1 FS step	ER	35.24	54.01	30.48	53.05	42.02	42.96
2nd set	CR	**60.52**	38.93	62.56	41.91	49.95	**50.77**
290 features	UR	6.29	7.17	7.00	6.71	7.69	6.97
1 FS step	ER	33.19	53.90	30.44	51.38	42.36	42.26
2nd set, no Coh_{max}	CR	59.15	37.59	**63.01**	40.30	**51.62**	50.33
249 features	UR	5.79	6.18	6.16	6.12	7.36	6.32
1 FS step	ER	35.05	56.23	30.84	53.59	41.03	43.35
2nd set, no PLV	CR	56.73	39.06	60.23	41.93	45.72	48.73
249 features	UR	7.09	6.20	7.33	7.44	7.73	7.16
1 FS step	ER	36.19	54.74	32.44	50.63	46.55	44.11

Table 5.14: Results for 2nd set of features, SVM classifier

Subject		1	2	3	4	5	Average
2nd set	CR	60.16	35.17	61.40	38.63	47.36	48.55
362 features	UR	6.81	4.59	5.66	5.70	6.46	5.85
1 FS step	ER	33.03	60.24	32.93	55.67	46.18	45.61
2nd set	CR	56.67	34.41	59.65	**42.24**	48.92	48.38
290 features	UR	6.45	6.04	5.57	6.71	8.22	6.60
1 FS step	ER	36.88	59.55	34.78	51.05	42.86	45.02
2nd set, no Coh_{max}	CR	59.82	34.27	60.36	39.83	47.83	48.42
249 features	UR	5.06	5.14	5.48	7.21	6.58	5.89
1 FS step	ER	35.12	60.59	34.15	52.96	45.59	45.68
2nd set, no PLV	CR	55.28	33.21	56.24	40.56	45.29	46.12
249 features	UR	6.16	7.00	6.35	7.85	7.76	7.02
1 FS step	ER	38.55	59.78	37.41	51.60	46.95	46.86

Table 5.15: Results for 2nd set of features, DA classifier

Selected Features. Because of the division by the mean of all features of the same type at the same time instance for the 1st set of features, all electrode signals were involved in the computation of a feature at a certain electrode. This is not the case for this 2nd set of features.

Synchronization features (see Figure 5.10) were selected in the central and centroparietal region for the 3 respective pairs of tasks for Subject 1. For Subject 2, synchronization features were rather selected in frontal and temporal region. Subject 3 shows involvement of the fronto-central, central and centro-parietal region respectively for the different pairs of tasks. For subject 4, frontal and temporal synchronization features were selected; for Subject 5, fronto-central, central and frontal EEG signals were selected for the respective task pairs.

EEG signals selected for power features are depicted in Figure 5.11. For Subject 1, features are primarily selected from the central electrodes C3 and C4, in the frequency bands α, β_2 and for the SMF. Subject 2 rather presents an involvement of temporal and frontal areas. For the α frequency band, an involvement

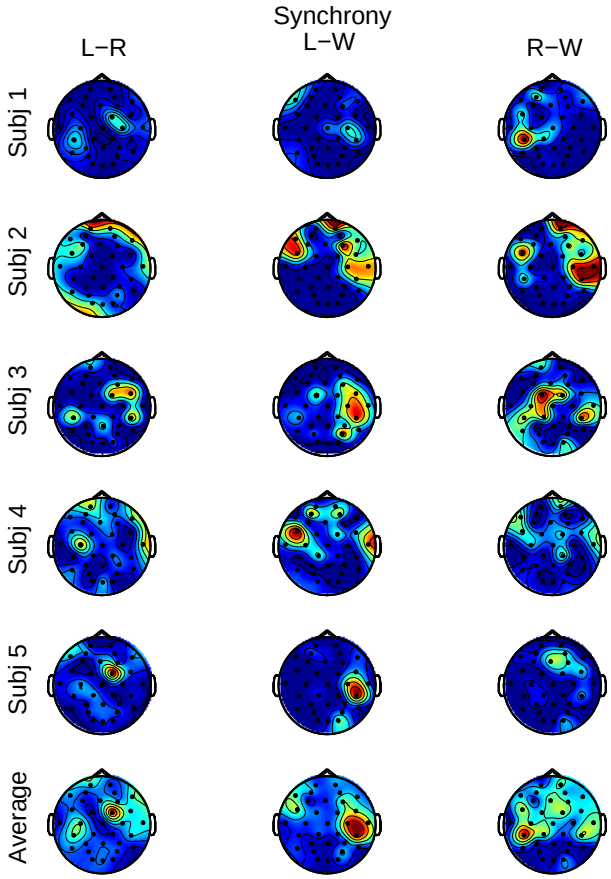

Figure 5.10: Selected synchronization features, 2nd set of features

of electrode C4 can be seen for L-W and R-W.
For Subject 3, electrodes C3 and C4 are selected in the α band and for SMF for the task pair L-R. For the task pairs L-W and R-W, α and SMF features computed from electrode signals C4 and C3 were selected respectively. This agrees well with neurophysiology.
In case of Subject 4, features are selected from temporal, occipital and frontal electrodes. We can also observe a selection of the Cz electrode for α, β_2 and SMF features. For Subject 5, rather β_1 than α power features are selected, from C3 for L-R, C4 for L-W and C3 for R-W. Furthermore, some SMF features were selected, from central, centro-parietal and occipital electrodes.
Averaged over all subjects, mainly α and SMF features were selected, from C3 and C4 for L-R, from C4 for L-W and from C3 for R-W. Again, selected EEG signals correspond well to predictions from neurophysiology.

Table 5.16 presents the types of features selected for the final subset. The most often selected types of features were the CPI and the SMF. For both synchronization and power features, the most often selected hemisphere was the right one for the task pairs L-R and L-W. For the task pair R-W, synchronization features were equally selected in both hemispheres and power features were most often selected in the left hemisphere.

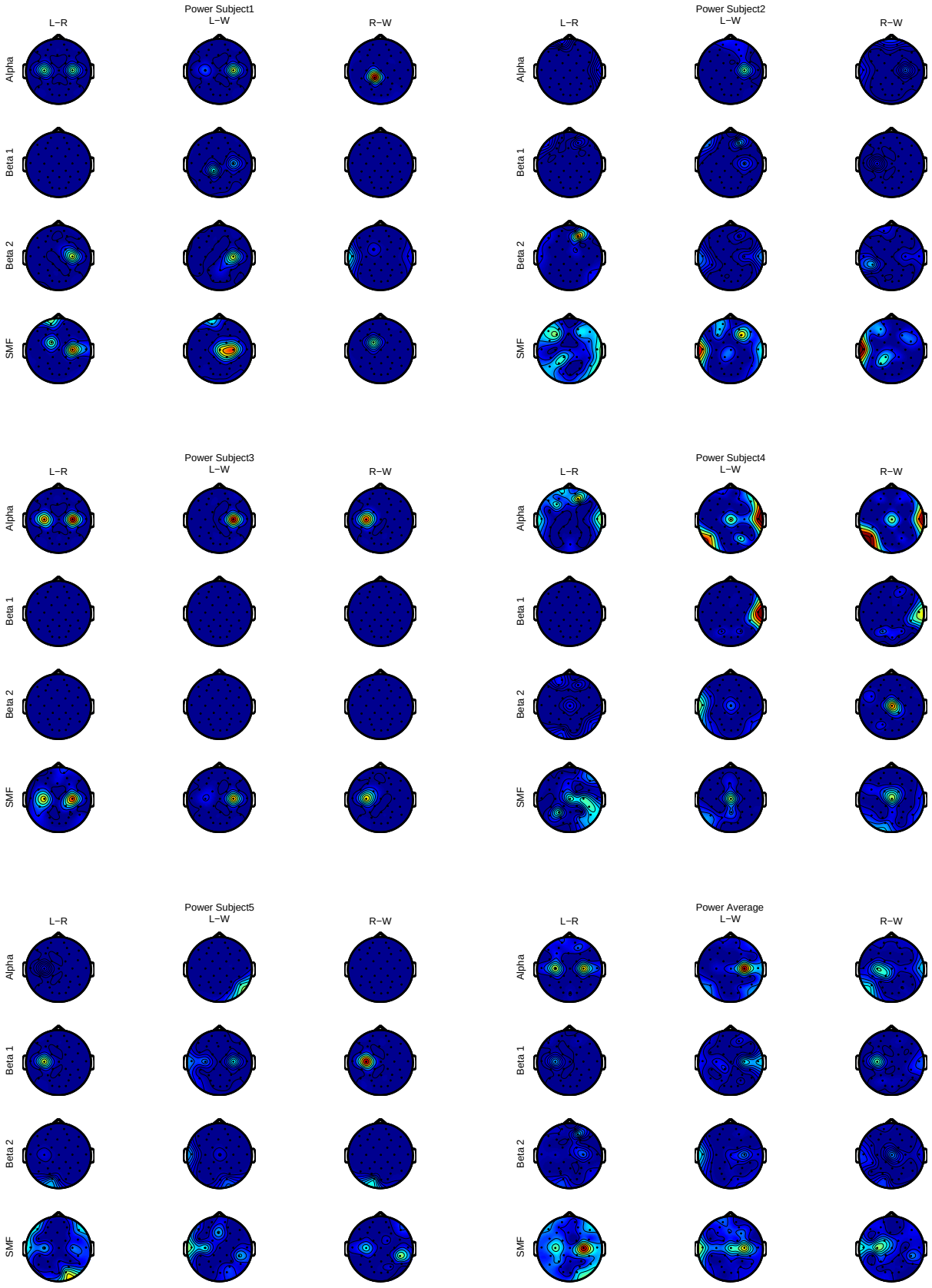

Figure 5.11: Selected features in α, β_1 and β_2 band and for SMF, 2nd set of features, for subjects 1, 2, 3, 4, 5 and averaged over all subjects

		L-R					
Subject		1	2	3	4	5	Av
PLV group	9	1.67	0.33	0.33	1.33	0.67	0.87
Cohmax group	9	0.67	0	0	0	0.67	0.27
PLV pairs	40	6.00	5.67	3.33	3.67	9.00	5.53
Cohmax pairs	40	1.67	0	1.67	0.67	0.67	0.93
PLV electr	32	3.33	7.00	3.00	2.67	2.33	3.67
Cohmax electr	32	4.33	0.33	3.33	1.67	6.33	3.20
PLV CPI	32	11.00	7.33	6.67	12.33	5.67	8.60
Cohmax CPI	32	0	0	2.33	1.00	2.00	1.07
α	32	3.33	0.67	9.00	9.00	0.33	4.47
β_1	32	0	0.67	0	0	2.67	0.67
β_2	32	2.67	4.33	0	2.00	1.67	2.13
SMF	32	7.00	18.00	13.67	10.33	13.00	12.40
SCG	8	3.33	0.67	1.67	0.33	0	1.20
Synch left		8.00	7.33	4.33	8.00	6.67	6.87
Synch right		9.67	6.67	9.67	7.33	7.33	8.13
Power left		4.67	11.67	10.00	7.67	8.67	8.53
Power right		8.33	11.67	12.33	11.00	8.00	10.27
		L-W					
Subject		1	2	3	4	5	Av
PLV group	9	3.00	0.67	1.33	0.33	3.67	1.80
Cohmax group	9	0	1.33	0	0	0	0.27
PLV pairs	40	9.00	5.67	11.00	2.00	8.67	7.27
Cohmax pairs	40	0	0	0	1.67	1.00	0.53
PLV electr	32	4.67	5.33	9.67	0	5.00	4.93
Cohmax electr	32	0	2.67	0	7.33	2.33	2.47
PLV CPI	32	11.67	9.00	13.00	10.33	5.67	9.93
Cohmax CPI	32	0	0	0	1.00	0	0.20
α	32	2.67	3.00	5.00	10.67	2.00	4.67
β_1	32	2.33	2.33	0	4.67	4.00	2.67
β_2	32	3.00	2.33	0	3.00	2.33	2.13
SMF	32	8.67	12.67	4.67	3.67	10.00	7.93
SCG	8	0	0	0.33	0.33	0.33	0.20
Synch left		6.00	5.00	4.67	10.33	1.00	5.40
Synch right		8.67	11.67	18.00	7.67	11.33	11.47
Power left		3.00	8.33	1.00	5.67	8.00	5.20
Power right		11.33	11.33	8.67	11.33	8.00	10.13
		R-W					
Subject		1	2	3	4	5	Av
PLV group	9	1.33	1.33	2.33	0.67	3.67	1.87
Cohmax group	9	0	1.00	0	0.67	1.33	0.60
PLV pairs	40	10.33	5.00	8.33	4.67	15.00	8.67
Cohmax pairs	40	0	0.67	0	2.00	0.33	0.60
PLV electr	32	5.00	7.33	7.00	1.33	2.33	4.60
Cohmax electr	32	4.33	3.00	2.00	4.00	5.33	3.73
PLV CPI	32	17.33	9.33	15.67	8.00	3.33	10.73
Cohmax CPI	32	0	0	0	0.33	0.33	0.13
α	32	3.33	1.67	5.33	9.33	0	3.93
β_1	32	0	0.33	0	4.33	4.33	1.80
β_2	32	1.67	2.33	0	3.67	1.33	1.80
SMF	32	1.67	11.00	4.33	5.67	6.00	5.73
SCG	8	0	2.00	0	0.33	1.67	0.80
Synch left		18.67	4.33	13.33	5.00	2.00	8.67
Synch right		4.33	15.00	10.00	7.00	7.33	8.73
Power left		6.33	11.33	9.67	7.33	7.67	8.47
Power right		0.33	3.67	0	9.00	4.00	3.40

Table 5.16: Types of features selected for 2nd set of features and left-right distribution

3rd set of Features

Classification Accuracies. Results are presented in Tables 5.17 and 5.19 for SVM and DA classification respectively. FS was done both in 2 steps and 1 step. Features were used as described above or renormalized by dividing all features of one row with the average of that row ("Renorm"). This was done at each time instance. Finally, we also chose 9x32 features in order to have a number of features of the same order of magnitude as the 1st set. This was done to approach the conditions of the features of the first set.
For both the SVM and DA classifier there is no important difference in classification accuracy for the feature sets with and without renormalization. Selecting features in two steps resulted in slightly improved results as compared to selection of features in only one step. This could be because removing irrelevant features may be less difficult in smaller feature sets.

Some further results are presented in Tables 5.18 and 5.20 for SVM and DA classification respectively. Here, only the 5 groups of PSD features (α, β_1, β_2, λ and SMF), the two groups of synchronization averages and the power contained in the 3 IMFs are considered.
We can see that the power of contained in the 3 IMFs deteriorated the classification performance obtained by the 5 PSD feature groups and 2 groups of PLV averages. Furthermore, we notice that combining PLV and PSD features improves the classification accuracy obtained by the separate groups.

Subject		1	2	3	4	5	Average
3rd set, no renorm	CR	57.94	**41.33**	64.22	42.32	50.01	51.16
79X32 features	UR	7.71	6.80	7.27	8.21	7.14	7.43
2 FS steps	ER	34.35	51.87	28.51	49.46	42.85	41.41
3rd set, no renorm	CR	56.30	41.12	**64.68**	41.62	49.76	50.69
79X32 features	UR	7.08	6.98	6.69	8.18	7.33	7.25
1 FS steps	ER	36.62	51.90	28.64	50.20	42.91	42.05
3rd set, renorm	CR	**60.19**	40.44	62.54	41.44	**52.24**	**51.37**
79X32 features	UR	7.19	7.88	8.04	7.79	6.67	7.51
2 FS steps	ER	32.62	51.68	29.42	50.77	41.09	41.12
3rd set, renorm	CR	60.07	39.24	62.72	41.58	50.94	50.91
79X32 features	UR	8.09	8.22	7.06	8.71	7.61	7.94
1 FS step	ER	31.85	52.54	30.22	49.70	41.46	41.15
3rd set, no renorm	CR	59.24	40.93	64.37	42.07	49.35	51.19
9X32 features	UR	7.10	7.01	7.15	8.15	6.80	7.24
1 FS step	ER	33.66	52.06	28.48	49.79	43.85	41.57
3rd set, renorm	CR	58.94	40.05	62.16	41.36	49.80	49.56
9X32 features	UR	7.92	6.98	7.80	8.31	7.39	7.46
1 FS step	ER	33.14	52.98	30.05	50.33	42.81	41.57

Table 5.17: Results for 3rd set of features, SVM classifier

Subject		1	2	3	4	5	Average
3rd set, renorm	CR	58.03	38.76	60.84	41.74	50.25	49.92
5 PSD, 2 PLV, 3 IMF	UR	7.19	7.29	7.85	8.50	7.82	7.73
1 FS step	ER	34.78	53.94	31.31	49.76	41.93	42.34
3rd set, renorm	CR	59.37	39.69	61.26	40.64	52.03	50.60
5 PSD, 2 PLV	UR	6.91	6.43	8.17	8.21	6.67	7.28
1 FS step	ER	33.72	53.89	30.57	51.14	41.30	42.13
3rd set, renorm	CR	57.37	37.04	57.59	42.12	48.36	48.50
5 PSD	UR	7.65	6.83	8.26	8.37	8.60	7.94
1 FS step	ER	34.99	56.13	34.15	49.51	43.04	43.56
3rd set, renorm	CR	51.29	40.27	53.28	36.42	47.90	45.83
2 PLV	UR	7.06	5.39	7.42	9.00	6.73	7.12
1 FS step	ER	41.65	54.34	39.30	54.58	45.37	47.05

Table 5.18: Further results for 3rd set of features using only PSD features, PLV averages and the power of the IMFs, SVM classifier

Subject		1	2	3	4	5	Average
3rd set, no renorm	CR	57.72	39.68	63.44	42.14	49.14	50.42
79X32 features	UR	7.09	7.32	7.18	8.59	7.63	7.56
2 FS steps	ER	35.19	53.00	29.38	49.27	43.23	42.01
3rd set, no renorm	CR	56.56	39.28	62.69	42.04	49.88	50.09
79X32 features	UR	6.63	8.12	7.33	8.69	8.56	7.87
1 FS steps	ER	36.82	52.60	29.98	49.27	41.56	42.04
3rd set, renorm	CR	58.88	36.89	61.59	41.03	51.13	49.90
79X32 features	UR	7.30	9.21	8.26	8.35	7.39	8.10
2 FS steps	ER	33.82	53.91	30.16	50.62	41.49	42.00
3rd set, renorm	CR	58.60	36.27	61.58	41.09	50.91	49.69
79X32 features	UR	7.21	8.72	7.70	9.12	8.32	8.21
1 FS step	ER	34.19	55.01	30.71	49.79	40.77	42.10
3rd set, no renorm	CR	57.94	36.85	62.32	**42.48**	48.20	49.56
9X32 features	UR	7.22	7.69	7.73	7.27	7.39	7.46
1 FS step	ER	34.84	55.45	29.95	50.24	44.42	42.98
3rd set, renorm	CR	57.36	36.64	62.25	40.54	50.20	49.40
9X32 features	UR	6.94	8.28	7.74	8.31	7.23	7.70
1 FS step	ER	35.70	55.08	30.01	51.14	42.57	42.90

Table 5.19: Results for 3rd set of features, DA classifier

Subject		1	2	3	4	5	Average
3rd set, renorm	CR	55.42	35.47	61.72	42.11	48.85	48.71
5 PSD, 2 PLV, 3 IMF	UR	7.25	8.10	6.49	7.59	7.27	7.34
1 FS step	ER	37.33	56.43	31.80	50.30	43.88	43.95
3rd set, renorm	CR	59.11	36.33	61.52	41.46	51.13	49.91
5 PSD, 2 PLV	UR	6.97	7.38	7.83	8.15	7.54	7.57
1 FS step	ER	33.92	56.29	30.65	50.39	41.33	42.52
3rd set, renorm	CR	56.76	34.27	57.84	41.50	47.56	47.58
5 PSD	UR	7.21	8.29	7.89	8.97	8.50	8.17
1 FS step	ER	36.02	57.44	34.28	49.53	43.94	44.24
3rd set, renorm	CR	52.18	36.29	51.73	35.17	47.78	44.63
2 PLV	UR	7.09	8.40	8.63	9.31	7.32	8.15
1 FS step	ER	40.73	55.31	39.65	55.52	44.90	47.22

Table 5.20: Further results for 3rd set of features using only PSD features, PLV averages and the power of the IMFs, DA classifier

Selected Features. As depicted in Figure 5.12, for subjects 1 and 3 and 5, PLV features were selected in the fronto-central region for the task pairs L-R and R-W. For Subject 2, PLV features were selected from CP6 for L-W and R-W. For Subject 4, only few PLV features were selected, in the centro-parietal region.

The different panels in Figures 5.13, 5.14 and 5.15 show, for the 5 respective subjects and for the all subject average, where power features computed from the EEG signals and its first 3 IMFs have been selected. For Subject 1, α features are selected from electrodes C4, C4 and CP1 for the task pairs L-R, L-W and R-W respectively. Also SMF features are selected in the central region. IMF features were mostly selected in the central region.
For Subject 2, α features are selected from C4 electrode, λ features in the frontal area and SMF features in the temporal area, especially when the "word" task was involved.
For Subject 3, mostly α features are selected, from C3 and C4 electrodes for L-R and from C4 and C3 electrodes respectively for L-W and R-W. IMF features were mainly selected from the second IMF, from the EEG signals C3 and C4, C4, and C3 for the 3 respective task pairs.
For Subject 4, L-W and R-W, α features are selected from the central, temporal and occipital region. For the same task pairs, SMF features are selected centrally and frontally. IMF features were mostly selected in parietal and frontal regions.
For Subject 5, the β_1 band seems to be the most relevant frequency band. β_1 features are selected from C3 electrode for L-R and R-W and from C4 electrode for L-W. IMF features were often selected from

electrode CP6 for the task pair L-W.

In the average, α power features are selected from C3 and C4 for L-R, from C4 for L-W and from C3 for R-W. Furthermore, for R-W, β_1 power features computed from C3 and frontal SMF features are selected. For power features computed from the IMFs, central EEG signals were selected for the 2nd IMF and centro-parietal EEG signals for the 3rd IMF.

The types of features selected are reflected in Tables 5.21, 5.22 and 5.23 for the 3 respective pairs of tasks. Most features in the final subset were PLV features. The most often appearing EEG power and IMF power features were α and IMF3, for the 3 task pairs. For the task pairs L-R and L-W, most of the selected PLV pairs were in the left hemisphere, as well as most of the selected PLV features computed at individual electrodes and power features. In case of R-W distinction slightly more PLV electrodes were selected in the right hemisphere. PLV pairs and power features were mostly in the left hemisphere.

		L-R					
Subject		1	2	3	4	5	Av
EEG PLV	496	24.67	6.67	23.67	10.67	27.00	18.53
EEG CPI	32	0	1.00	0.67	0	2.33	0.80
PLV electr	32	3.00	1.00	2.33	1.00	7.67	3.00
PLV group	8	2.33	0	0.33	0	0	0.53
IMF2 PLV	496	1.67	6.67	0	8.67	0.33	3.47
IMF2 CPI	32	0	2.67	0	0.67	0.33	0.73
SMF	32	1.67	1.67	0	3.67	0	1.40
α	32	3.33	2.33	10.00	2.67	0	3.67
β_1	32	0	0.67	0	0.67	2.67	0.80
β_2	32	0	2.67	0	1.67	0	0.87
λ	32	0	0.67	0	0.67	0.67	0.40
IMF1 power	64	2.67	8.00	1.00	2.00	3.33	3.40
IMF2 power	64	3.00	5.00	5.67	3.33	0	3.40
IMF3 power	64	2.67	6.00	1.33	9.33	0.67	4.00
Synch electr left		1.00	2.33	0.67	1.33	3.00	1.67
Synch electr right		2.00	1.33	2.00	0.33	4.00	1.93
Synch pairs left		10.67	5.33	5.33	4.67	6.67	6.53
Synch pairs right		12.00	6.00	16.00	10.67	13.67	11.67
Synch interhem pairs		3.67	2.00	2.33	4.00	7.00	3.80
Power left		3.00	12.00	7.33	8.00	4.33	6.93
Power right		10.00	13.67	10.67	13.33	3.00	10.13

Table 5.21: Types of features selected for 3rd set of features and left-right distribution, task pairs L-R

5.9.4 Conclusions

The following types of features were selected from the 3 respective sets of features. From the first set of features, mostly PLV features and PSD in the α band were selected. Coh_{max} features were selected to a lesser extent.
The second feature set is similar to the first one but contains the following features in addition: CPI for PLV and Coh_{max}, the SCG and the SMF instead of λ power. We observe that most of the selected features were PLV features, CPIs computed from PLV, and SMFs. As compared to the first set, Coh_{max} and power features computed in the α, β_1 and β_2 frequency bands were less often selected.
For the third set, Coh_{max} features were omitted and PLV values for all possible electrode pairs and power and PLV features computed from the EMD EEG signals were added. Most of the selected features were PLV features. To a lesser extent the power contained in the third IMF and α power features were selected. As compared to the second subset, SMF and CPI appeared less often in the selected feature subset.

Generally, features were selected from both motor cortices for the task pair L-R, from the right motor cortex for L-W and from the left motor cortex for R-W. When the "Word" task was involved, more features were selected in the temporal region. These findings correspond well to neurophysiological evidence.

		L-W					
Subject		1	2	3	4	5	Av
EEG PLV	496	12.33	11.67	18.67	15.33	14.33	14.47
EEG CPI	32	0	0.33	0	0.33	0	0.13
PLV electr	32	0.67	4.00	7.00	2.33	2.67	3.33
PLV group	8	0	1.00	1.00	2.00	0	0.80
IMF2 PLV	496	1.33	0	0	0	0.33	0.33
IMF2 CPI	32	0.33	0.33	0	0	0.67	0.27
SMF	32	5.67	6.67	1.67	9.00	1.00	4.80
α	32	5.00	3.33	8.33	6.67	3.00	5.27
β_1	32	0.33	1.00	0	2.67	4.00	1.60
β_2	32	0	1.67	0.67	0	2.00	0.87
λ	32	0.33	1.33	0	0	4.00	1.13
IMF1 power	64	4.67	2.67	0	0.33	1.00	1.73
IMF2 power	64	2.67	3.33	5.00	0.67	2.00	2.73
IMF3 power	64	11.67	7.67	2.67	5.67	10.00	7.53
Synch electr left		0.67	1.00	1.33	2.67	2.00	1.53
Synch electr right		0.33	3.67	4.00	0	0.67	1.73
Synch pairs left		0	0	7.67	5.33	2.33	3.07
Synch pairs right		10.67	11.00	10.00	2.00	10.00	8.73
Synch interhem pairs		3.00	0.67	1.00	8.00	2.33	3.00
Power left		11.00	11.00	1.00	5.33	6.67	7.00
Power right		17.33	16.67	16.00	13.33	19.33	16.53

Table 5.22: Types of features selected for 3rd set of features and left-right distribution, task pairs L-W

		R-W					
Subject		1	2	3	4	5	Av
EEG PLV	496	12.33	14.67	19.67	20.67	20.00	17.47
EEG CPI	32	0	2.00	0	067	0.67	0.67
PLV electr	32	7.67	5.67	3.67	0.33	10.33	5.53
PLV group	8	2.00	0	0	1.33	0.33	0.73
IMF2 PLV	496	4.00	0.33	0	0.33	0	0.93
IMF2 CPI	32	0	0	0	0	0	0
SMF	32	2.00	5.33	1.67	9.33	0.33	3.73
α	32	5.67	1.00	11.00	6.00	0	4.73
β_1	32	0.67	1.00	0	1.00	5.00	1.53
β_2	32	1.00	1.67	0	1.00	0.33	0.80
λ	32	0.33	2.33	0	0	0.33	0.60
IMF1 power	64	2.00	1.67	0.67	0.33	0	0.93
IMF2 power	64	2.00	2.33	3.67	0.67	1.67	2.07
IMF3 power	64	5.33	7.00	4.67	3.33	6.00	5.27
Synch electr left		6.00	1.67	2.33	0.33	2.00	2.47
Synch electr right		1.00	6.00	0.33	0.33	5.33	2.60
Synch pairs left		9.67	2.00	13.67	9.33	6.33	8.20
Synch pairs right		1.67	13.00	3.00	4.00	8.33	6.00
Synch interhem pairs		5.00	0	3.00	7.67	5.33	4.20
Power left		18.33	13.00	18.67	9.67	6.00	13.13
Power right		0.67	9.33	1.33	5.33	7.67	4.87

Table 5.23: Types of features selected for 3rd set of features and left-right distribution, task pairs R-W

SVMs usually outperform DA. In case of SVM classification, the best result obtained with features selected from the 3 respective feature sets was 51.11%, 50.77% and 51.37%. For DA classification, we obtained classification rates of 49.15%, 48.55% and 50.42%. The best performance was achieved applying FCBF twice to the 3rd set of features.
Subjects 2, 3 and 4 achieved best results with the third set of features. For the Subjects 1 and 5, the first set proved best. The 3rd set, however, comprises 2528 features, while the 1st set yielding an average CR of 51.11%, only comprised 169 features. Considering the details of coupling by retaining the PLV values for all individual electrode pairs improved the CR very little while heavily increasing the computational cost and memory load.

For some subjects, the classification accuracy improved by dividing a given feature at a given electrode by the all-electrode average value for that feature at that time instance.

5.10 Conclusions

In this chapter, several feature selection techniques have been investigated. GAs proved not practically usable for BCI applications because of the high computational cost, but selected otherwise physiologically sound features. Selection of features with SVM-rfe was significantly faster, but still needed computation times on the order of hours. A better approach seemed to be the Fast Correlation Based Filter. The algorithm selected very few features, with which good classification accuracies could be achieved. These features were selected from EEG signals recorded from brain regions involved in the considered mental tasks.
When the size of the feature set increases, however, FCBF may perform less well. Therefore, improved results may be obtained by omitting some types of features from the original set of features.

Application of feature selection algorithms to narrowband and broadband BP and PLV features showed us that BP is preferably computed in the 8-12Hz frequency band and that PLV features are best computed from 8-30Hz filtered EEG signals.

FCBF feature selection was also applied to blind data from the third BCI competition (sets I and IVa). The two data sets from the BCI competition considered as well as our data dealt with motor imagery tasks. The first data set consisted of intracranially recorded EEG, while data set IVa and our data were recorded from the scalp.
The BCI competition data were classified using features selected by FCBF. On the one hand, the set of features comprised BP features, the SMF, PLV features for individual electrode pairs and the CPI, computed from the filtered EEG signals. On the other hand, power and PLV features features computed from the Empirical Mode Decomposed EEG signals were included. A similar set of features, the third set of features described in Paragraph 5.9.2, was computed from our data and submitted to the FCBF feature selection. Significant classification rates were obtained, also in case of blind data.
For the first data set, a set of features comprising PLV features, BP features and the power content of the IMFs was selected. In case of data set IVa, mostly PLV features were selected, complemented by some CPI, SMF and and IMF power features. For our own data, again, PLV features were most often selected. Also IMF power features and α power features appeared to a lesser extent in the selected subset.

Finally, we evaluated 3 different sets of features with FCBF. The best set of features proved different for the different subjects. For the considered paradigm, usefull features are α power and SMF features, the CPI and PLV averages, and the power contained in the IMF components. Considering PLV for all individual electrode pairs yielded a slightly improved CR, but at the same time substantially increased the size of the feature set. The EEG signals from which features were selected corresponded to the brain regions involved in the considered mental tasks.

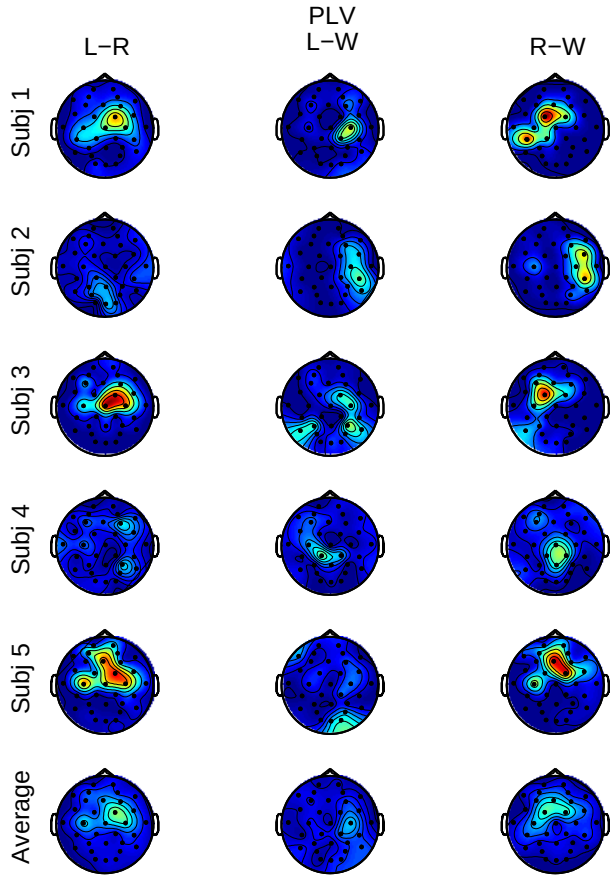

Figure 5.12: Selected PLV features, 3rd set of features

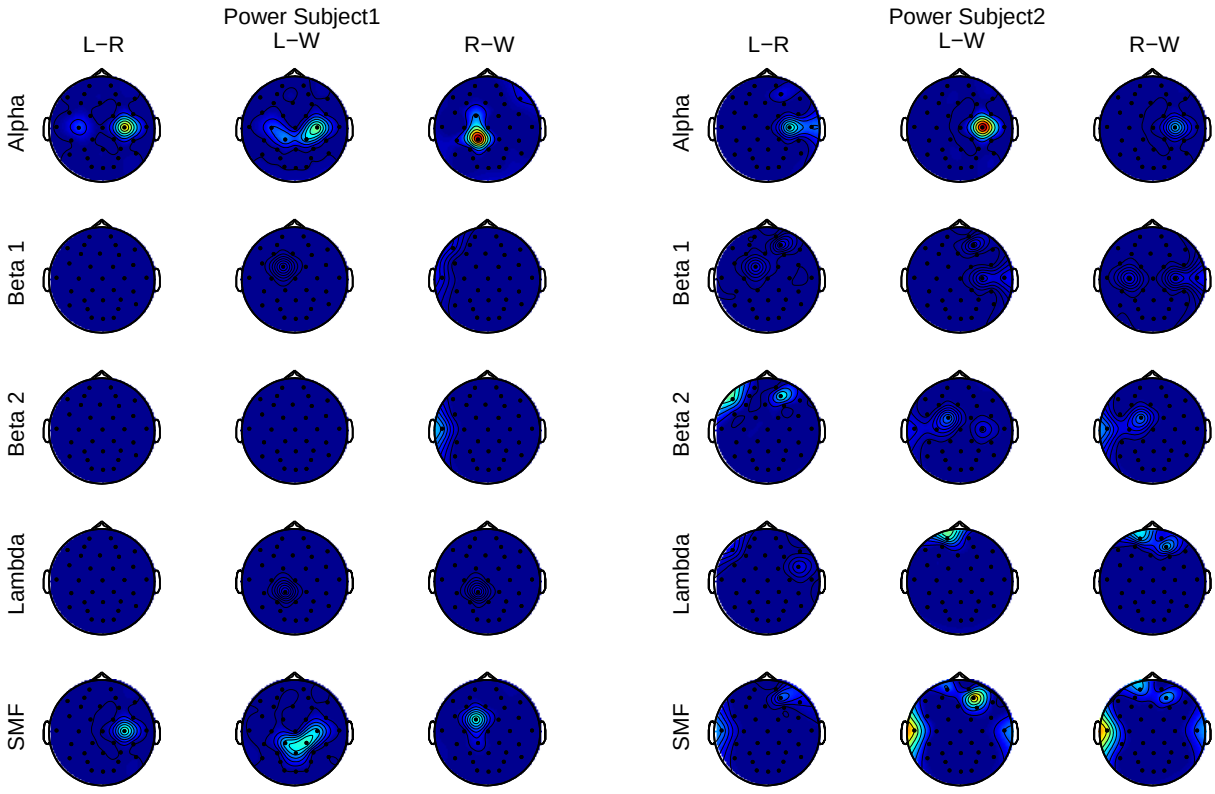

Figure 5.13: Selected features in α, β_1, β_2 and λ band and for SMF, 3rd set of features, for subjects 1 and 2.

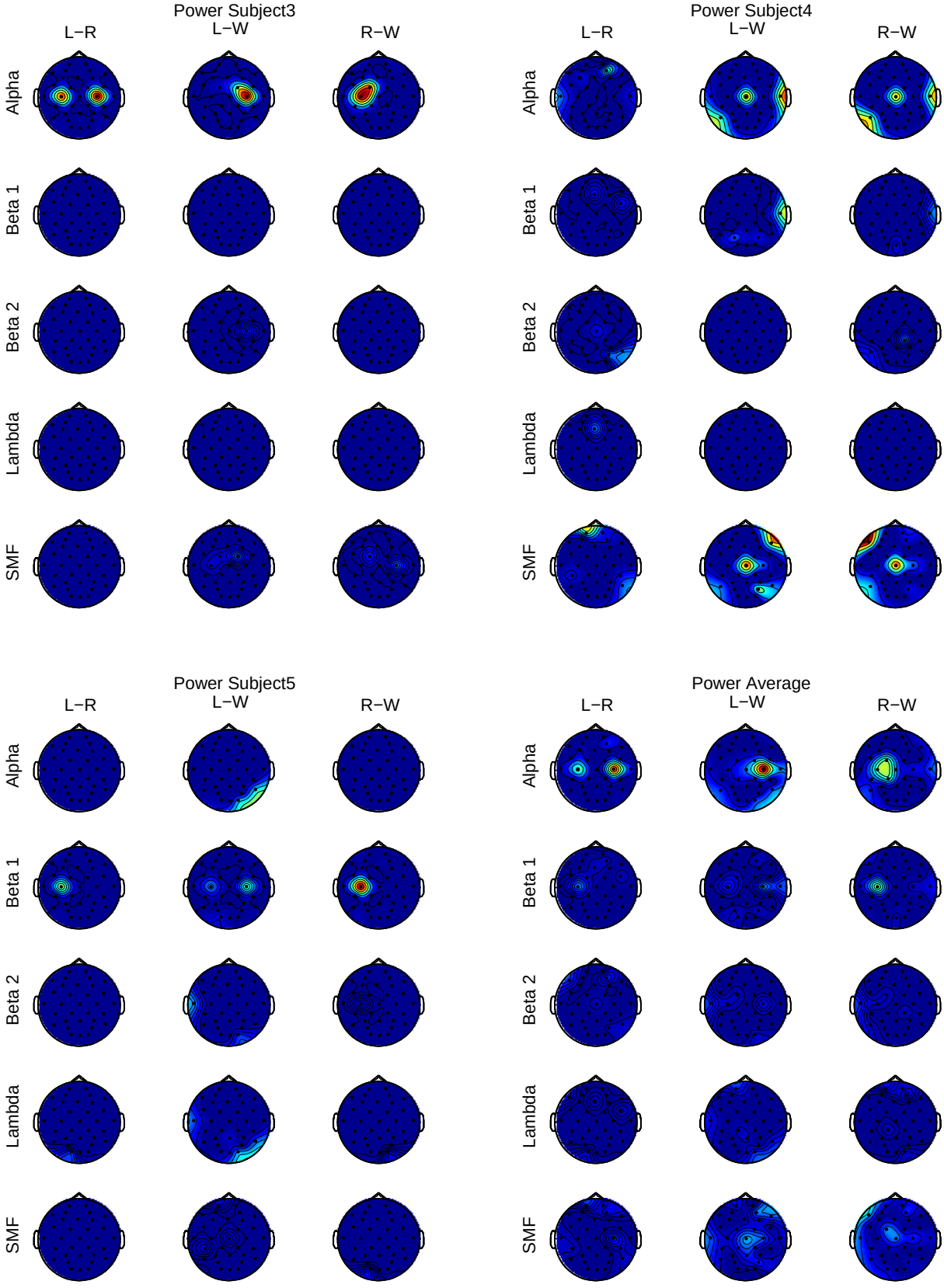

Figure 5.14: Selected features in α, β_1, β_2 and λ band and for SMF, 3rd set of features, for subjects 3, 4, 5 and averaged over all subjects

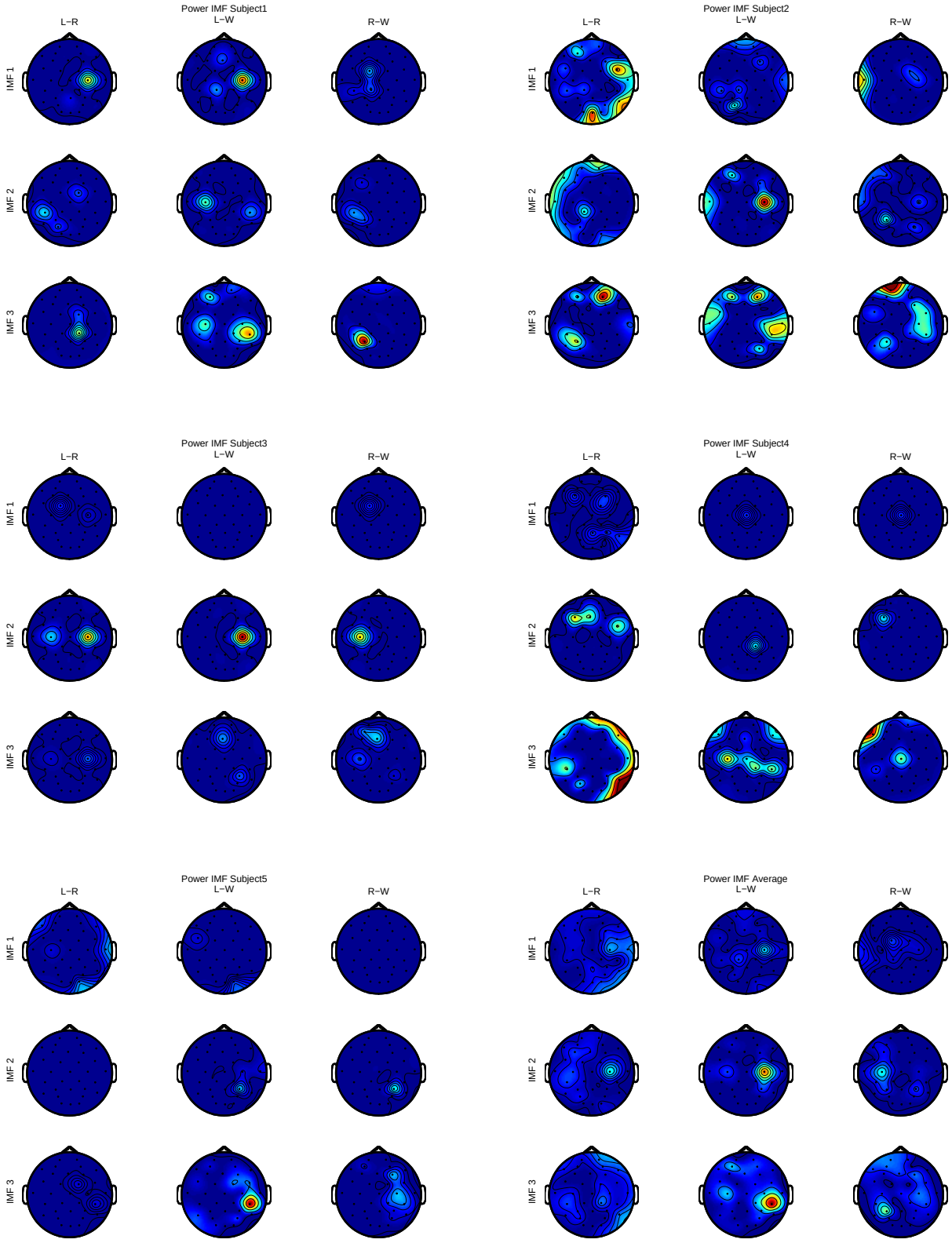

Figure 5.15: Selected power features for IMF 1, IMF 2 and IMF 3, 3rd set of features, for subjects 1, 2, 3, 4, 5 and averaged over all subjects

Chapter 6

Artifact Detection and Denoising

6.1 Introduction

In this chapter, we investigate how the application of artifact detection and denoising techniques prior to computation of features affects the classification accuracies.

For artifact detection, we used the algorithm proposed by Celka *et al* [26] with an additional preprocessing step to make the algorithm better suited for short time windows.

We implemented and applied different denoising techniques to test whether they are able to remove irrelevant information from the EEG signals while keeping the relevant information the classifier is using to distinguish the different mental tasks. It is difficult to define which part of the EEG signals actually corresponds to noise and which part contains the relevant information. Therefore, denoising algorithms were evaluated by their improvement or deterioration of the CR.

Both artifact detection and noise reduction algorithms have been applied to the short windows used for computation of features, and not to the entire signal. This was to be closer to an online paradigm, i.e. using no information of future and able to deal with short time windows.

6.2 Artifacts and Noise in EEG: their Origin

The purpose of BCIs is to command a device by altering brain activity. It is important to control the recorded brain signals for artifacts to ensure that the results are accounted for by brain activity and not by artifacts created by some kind of (voluntary) movement, as we want to develop a system to be operated by motor-disabled people. The most important artifacts are EOG (electro-oculogram) and muscle artifacts, EMG (electromyogram) [107, 26]. Furthermore, an EEG can also contain ECG (electrocardiogram) artifacts.

The EOG signal reflects eye movements and eye-blinking. Eye movements produce electrical activity on the scalp because they correspond a moving (rotating) dipole source. This interference with the EEG can mask relevant features in the EEG. EMG artifacts result from the electrical activity accompanying muscle contractions. Muscle artifacts regularly observed in the EEG are the ones resulting from opening and closing the mouth, lip movements, swallowing and neck movements. ECG artifacts result from the fact that electrodes used to measure EEG signals are sensitive to the electrical activity of the heart, present throughout the body.

Recorded EEG signals also contain measurement noise. More noise can result from displacements of the recording electrodes and bad contacts. In addition, every EEG electrode will register some background EEG activity, not specifically related to the brain region the electrode is supposed to register activity from.

6.3 Artifact Detection

6.3.1 Artifacts in our Data

To study how many artifacts occurred and to which extent they were still present in the frequency band used in this study, the raw and the 8-30Hz filtered EEG were visually inspected for artifacts.

- Ocular Artifacts occurred in signals recorded from Fp1, Fp2, AF3 and AF4 in 16%, 12% and 10% of the windows for day 1, 2 and 3 of recording from subject 1. More artifacts occurred during the task 'word'. Subject 2 presented artifacts in 5%, 14% and 2% of the windows. For subject 3, artifacts occurred in 5%, 3% and 2% of the windows on day 1, 2 and 3 respectively. The raw EEG of subjects 4 and 5 presented ocular artifacts in 10%, 4% and 4% and in 4%, 2% and 12% of the windows on days 1, 2 and 3 respectively. The percentage of windows containing artifacts did not differ more than 2.5% for the 3 different tasks, except for subject 3, day 2, and subject 5, day 3, where more artifacts occurred during the task 'left' and 'word' respectively.
  Inspection of the EEG band-pass filtered in 8-30Hz, the frequency band considered in this study, showed that these artifacts were mostly filtered out. Indeed, ocular artifacts generally appear under 4Hz [55].

- Muscle Artifacts were present in the EEG of subject 3 for 4 sessions of day 1 and 3 sessions of day 2. They appear on channels F7, T7, F8 and T8 during long periods, covering the 3 different tasks. For one session of day 3 of recording from subject 5, 3 consecutive windows out of 69 windows coinciding with the task 'word' contained muscle artifact.
  Generally, the muscle artifacts occurred during the 3 tasks. In addition, most of their energy is not contained in the 8-30Hz frequency band considered here [55].

Automatic EEG artifact detection and rejection/removal techniques (see [55, 73, 133, 30] and references therein) have been developed to process long-term EEG recordings, to improve evoked potential studies where the corruption of a few seconds of EEG turns out to be critical, or for systems where automated EEG preprocessing is necessary.

Considering spontaneous EEG signals, three situations can occur. First, if the artifact occurrence rate (AOR) is low, and artifacts thus have no or few impact on the analysis, it is not worth the effort to detect and reject or remove them. Second, if the AOR is high, artifact rejection/removal may also remove most of the EEG information content. Third, if the AOR is in between these two extremes, artifact rejection/removal could be of interest. Still, removing artifacts is a very delicate task because the information content of the spontaneous EEG is widely spread in the space and frequency domain. Hence, artifact removal could distort or destroy the brain activity present in the EEG. The best solution is thus a simple rejection of the window containing the artifact.
In most sessions considered in our study, the AOR was quite low (few percents). For 7 sessions of subject 3, the muscle AOR was quite high. As it was not necessary to exclude artifacts to be sure of classifying brain activity, and because AOR in different sessions was either very low, either very high, we decided not to treat them any further.

In an attempt to construct a more robust classifier, however, we detected eye blinking artifacts automatically with the method described in [26] and rejected the windows where an artifact was detected for training the classifier and selecting the features. For testing, all windows of the test set were used, including the ones containing artifacts.

6.3.2 Algorithm

The algorithm [26] is an adaptive one that does not use a reference signal. Its structure is based on a fixed-weight leakage normalized stochastic least mean fourth (NLMF) algorithm in a predictor mode. The LMF algorithm has been proven to be efficient in non gaussian environment. The input signal

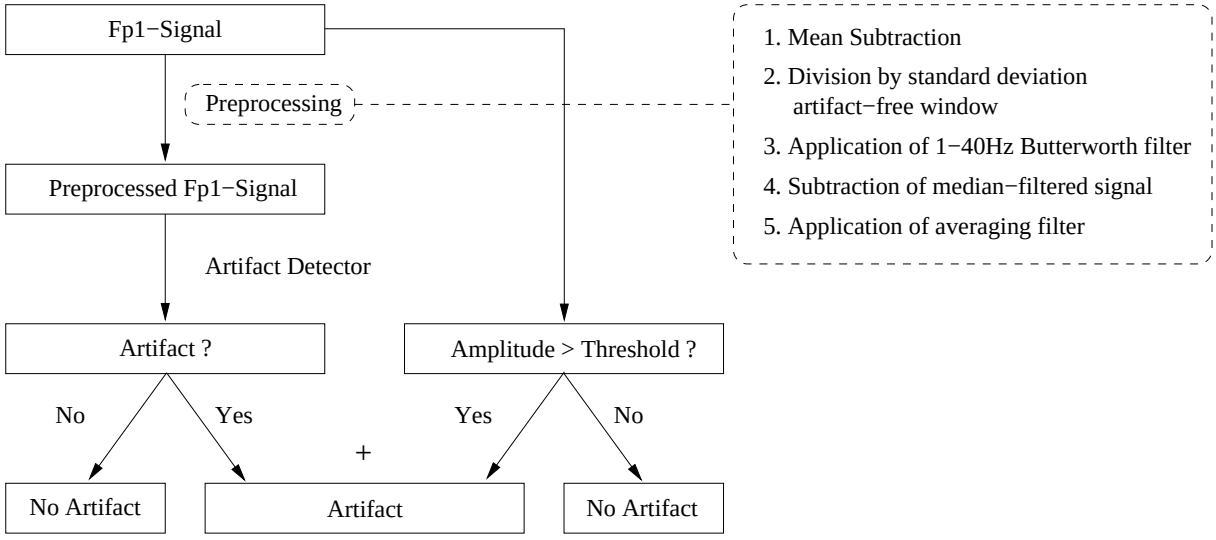

Figure 6.1: Method for detecting artifacts from 2-second windows using the artifact detection algorithm presented in [26].

$r(n) = s(n) + \hat{a}(n)$ is the zero-mean noisy EEG. The filter $W^T = [w_1(n), ..., w_N(n)]$ detects abrupt changes with a poor prediction ability. The recursive equation for W is

$$W(n+1) = (1-\gamma)W(n) + \frac{\mu e^3(n) r(n-1)}{\|r(n-1)\|^2} \tag{6.1}$$

$$e(n) = r(n) - W^T(n) r(n-1) \tag{6.2}$$

A GA was used to find the set of parameters yielding a minimal number of false positive and maximal number of true positive detections. The optimal set of parameters was found to be $\mu = 0.03$, $\gamma = 0.1$ and $N = 80$. After applying a threshold function $\Theta_\epsilon(x) = 1$ if $|x| > \epsilon$ and $\Theta_\epsilon(x) = 0$ if $|x| < \epsilon$, we apply a morphological operation that is a dilation that results in a signal composed of connected ones and zeros. This segments the signal into artifacts and non-artifacts. Depending on the application, artifacts can then be omitted or replaced some predicted EEG.
The algorithm reliably (nearly 100%) detected artifacts from the entire session. Detecting artifacts from 2-second windows, needed for online protocols, however, was problematic. A lot of false positives occurred. Therefore, the preprocessing was refined as follows (see Figure 6.1).
Only channel Fp1 was used for automatic eye blinking detection. First, the mean of the signal was subtracted. Then, the signal was divided by the standard deviation of an artifact-free window. This was once manually determined from the first artifact-free windows of the first session of the first day of recording and was set to 24, 16, 9, 19 and 19 for subjects 1, 2, 3, 4 and 5 respectively. Subsequently, a butterworth filter with pass band from 1 to 40 Hz was applied. To reduce false positive detections, we then subtracted the median filtered signal (window length 0.5s) from the signal and applied an averaging filter (window length 30 samples). Finally, a threshold was applied to the amplitude of the Fp1-signal at the time instances an artifact was detected. The threshold was set to 1 for all subjects.
The performance of this automated artifact detection method, depicted in Figure 6.1, varied amongst different subjects. The false positives ranged from 0 to 6.6038%, false negatives from 0 to 13.27%. The average FP and FN rates were 2.40% and 2.69% respectively.

6.4 Influence of Artifacts on Classification Accuracy: Results

6.4.1 Introduction

We repeated the computations presented in Section 5.9, this time automatically rejecting the artifacts for selecting features and training the classifier. The idea was to make the classifier more robust by constructing it without the data resulting from artifacts.

For testing, all data were used, whether or not contaminated with artifacts.

6.4.2 1st set of Features

The results for the first set of features are presented in Tables 6.1 and 6.2 and can be compared to the results in Tables 5.11 and 5.12 respectively. When the average CR obtained after removal of artifacts is better than the one obtained without rejecting artifacts in the training set, it is in italic. For DA classification, average results are slightly better when artifacts are excluded for training, but generally not more than 1%. For each subject, the best CR is indicated in bold font. The best all subject average CR obtained for the 1st set of features is 51.11%, when leaving out Coh features and using SVM classification.

Subject		1	2	3	4	5	Average
1st set	CR	61.39	38.86	62.24	37.15	51.28	50.18
290 features	UR	6.18	4.04	6.53	9.19	6.61	6.51
1 FS step	ER	32.44	57.10	31.23	53.66	42.11	43.31
1st set, no Coh_{max}	CR	**61.93**	**39.81**	61.59	40.27	**51.96**	***51.11***
209 features	UR	5.44	3.97	7.03	7.59	6.48	6.10
1 FS step	ER	32.63	56.21	31.38	52.15	41.55	42.79
1st set, no PLV	CR	58.97	38.70	61.18	37.09	48.30	48.85
209 features	UR	6.47	5.67	7.48	10.44	8.16	7.64
1 FS step	ER	34.57	55.64	31.35	52.47	43.54	43.51
1st set, no Coh_{max}, no PLV pairs	CR	60.82	39.31	**62.78**	40.84	50.09	50.77
169 features	UR	5.70	4.04	6.99	7.34	7.17	6.25
1 FS step	ER	33.47	56.65	30.23	51.81	42.74	42.98
1st set, only power	CR	57.88	38.45	61.10	39.46	48.92	49.16
128 features	UR	6.94	5.21	7.35	7.87	7.04	6.88
1 FS step	ER	35.18	56.34	31.55	52.68	44.04	43.96
1st set, only PLV avgs	CR	54.04	37.21	50.86	34.63	44.31	44.21
41 features	UR	5.15	6.58	3.96	5.93	6.23	5.57
1 FS step	ER	40.81	56.21	45.18	59.44	49.45	50.22

Table 6.1: Results for 1st set of features, artifacts automatically rejected for training, SVM classifier

Subject		1	2	3	4	5	Average
1st set	CR	59.66	35.18	59.49	34.29	50.63	47.85
290 features	UR	5.80	4.50	6.46	10.34	5.77	6.57
1 FS step	ER	34.54	60.32	34.04	55.37	43.26	45.51
1st set, no Coh_{max}	CR	59.71	36.90	58.88	**41.00**	49.45	*49.19*
209 features	UR	5.46	6.73	5.79	8.30	7.45	6.75
1 FS step	ER	34.83	56.37	35.33	50.70	43.10	44.07
1st set, no PLV	CR	58.19	37.87	57.99	35.80	47.55	*47.48*
209 features	UR	7.68	6.31	7.64	10.81	7.76	8.04
1 FS step	ER	34.13	55.82	34.37	53.39	44.69	44.48
1st set, no Coh_{max}, no PLV pairs	CR	60.17	36.08	59.53	40.84	49.66	*49.26*
169 features	UR	5.34	3.88	6.60	8.53	8.51	6.57
1 FS step	ER	34.49	60.04	33.87	50.63	41.83	44.17
1st set, only power	CR	58.00	35.60	57.04	40.86	45.75	47.45
128 features	UR	7.23	6.60	7.80	7.55	8.47	7.53
1 FS step	ER	34.77	57.80	35.16	51.59	45.78	45.02
1st set, only PLV avgs	CR	49.38	35.14	54.08	36.13	44.13	*43.77*
41 features	UR	5.24	4.69	7.74	5.84	5.86	5.27
1 FS step	ER	45.37	60.17	41.18	58.03	50.01	50.95

Table 6.2: Results for 1st set of features, artifacts automatically rejected for training, DA classifier

6.4.3 2nd set of Features

As compared to no artifact detection and rejection, averaged CRs did not improve for SVM classification (see Table 6.3). For DA classification (see Table 6.4), small improvements were observed. The best CR was obtained with SVM classification and amounted to 50.53%, averaged over all days of recording and all subjects.

Subject		1	2	3	4	5	Average
2nd set	CR	58.52	39.13	**62.57**	40.31	49.92	50.09
362 features	UR	6.24	5.98	6.46	7.91	7.76	6.87
1 FS step	ER	35.24	54.89	30.97	51.79	42.33	43.04
2nd set	CR	59.54	**39.41**	62.28	40.67	**50.75**	**50.53**
290 features	UR	6.60	6.52	6.87	7.21	7.16	6.87
1 FS step	ER	33.86	54.08	30.85	52.12	42.08	42.60
2nd set, no Coh_{max}	CR	57.97	37.66	62.35	40.57	50.00	49.71
249 features	UR	5.71	5.99	7.34	7.28	6.82	6.63
1 FS step	ER	36.32	56.36	30.31	52.15	43.18	43.66
2nd set, no PLV	CR	56.85	38.57	60.08	**41.84**	46.22	48.71
249 features	UR	6.78	5.52	6.96	7.47	7.48	6.84
1 FS step	ER	36.38	55.91	32.96	50.69	46.30	44.45

Table 6.3: Results for 2nd set of features, artifacts automatically rejected for training, SVM classifier

Subject		1	2	3	4	5	Average
2nd set	CR	**60.30**	36.96	59.07	40.17	49.86	*49.27*
362 features	UR	6.10	4.86	5.72	6.84	6.86	6.08
1 FS step	ER	33.60	58.17	35.21	52.99	43.29	44.65
2nd set	CR	59.87	33.91	59.59	41.19	49.67	*48.84*
290 features	UR	5.55	6.95	5.58	7.43	6.14	6.33
1 FS step	ER	34.58	59.14	34.83	51.38	44.19	44.83
2nd set, no Cohmax	CR	59.36	35.63	59.65	40.90	47.17	*48.54*
249 features	UR	6.01	5.57	5.56	6.21	6.61	5.99
1 FS step	ER	34.63	58.79	34.78	52.89	46.21	45.46
2nd set, no PLV	CR	56.04	35.77	53.95	41.16	46.12	*46.61*
249 features	UR	6.26	5.84	5.92	8.43	7.36	6.76
1 FS step	ER	37.70	58.39	40.14	50.41	46.52	46.63

Table 6.4: Results for 2nd set of features, artifacts automatically rejected for training, DA classifier

6.4.4 3rd set of Features

Results obtained on the third set of features are presented in Tables 6.5 and 6.6 for SVM and DA classification respectively. As compared to Tables 5.17 and 5.19 average CRs are sometimes better, sometimes worse. The best CR obtained was 51.65%.

6.4.5 Conclusions

In case of SVM classification, the best result obtained with features selected from the 3 respective feature sets is 51.11%, 50.40% and 51.37%, without artifact removal. Automatically removing eye movement artifacts for selecting the features and training the classifier, resulted in CRs of 51.11% 50.60% and 51.65% for the 3 respective feature sets. Given the additional computational cost and the minor increase of CR, automatic artifact detection is not recommended.

Subject		1	2	3	4	5	Average
3rd set, no renorm	CR	57.16	**41.33**	63.93	41.42	50.75	50.92
79X32 features	UR	7.16	6.86	6.84	7.90	7.33	7.22
2 FS steps	ER	35.68	51.81	29.24	50.69	41.92	41.87
3rd set, no renorm	CR	56.81	40.27	63.34	41.44	**53.08**	***51.65***
79X32 features	UR	6.97	6.16	7.89	7.77	6.66	7.20
1 FS steps	ER	36.22	53.56	28.76	50.78	40.26	41.15
3rd set, renorm	CR	59.14	39.64	62.07	40.99	51.06	50.58
79X32 features	UR	6.44	7.18	7.21	8.56	7.36	7.35
2 FS steps	ER	34.41	53.18	30.72	50.45	41.58	42.07
3rd set, renorm	CR	59.67	39.21	60.61	40.59	51.65	50.34
79X32 features	UR	5.77	7.57	7.27	8.46	7.26	7.17
1 FS step	ER	34.56	53.22	32.12	50.95	41.09	42.49
3rd set, no renorm	CR	57.54	41.21	**64.46**	42.08	48.64	50.79
9X32 features	UR	7.83	20.61	7.12	8.15	6.08	9.96
1 FS step	ER	34.63	53.38	28.42	49.77	45.28	42.29
3rd set, renorm	CR	59.01	40.26	63.31	40.45	50.76	*50.76*
9X32 features	UR	7.27	6.74	6.99	7.18	6.49	6.93
1 FS step	ER	33.72	53.00	29.70	52.37	42.76	42.31

Table 6.5: small Results for 3rd set of features, artifacts automatically rejected for training, SVM classifier

Subject		1	2	3	4	5	Average
3rd set, no renorm	CR	56.68	39.22	62.66	41.55	50.28	50.08
79X32 features	UR	7.61	7.48	7.03	8.02	6.95	7.42
2 FS steps	ER	35.71	53.80	30.31	50.42	42.76	42.50
3rd set, no renorm	CR	56.31	37.85	63.44	41.99	51.53	*51.07*
79X32 features	UR	7.18	7.69	6.59	7.81	6.52	7.17
1 FS steps	ER	36.51	54.46	299.97	50.20	41.94	41.76
3rd set, renorm	CR	**59.72**	37.20	61.80	41.58	50.16	*50.09*
79X32 features	UR	6.02	8.22	6.80	8.53	8.26	7.57
2 FS steps	ER	34.26	54.59	31.40	49.89	41.58	42.34
3rd set, renorm	CR	57.85	37.26	59.91	41.87	51.28	49.40
79X32 features	UR	7.16	7.70	7.80	8.34	7.82	7.76
1 FS step	ER	34.99	55.05	32.29	49.79	40.90	42.60
3rd set, no renorm	CR	57.99	37.86	62.93	**42.77**	48.45	*50.00*
9X32 features	UR	7.59	7.18	7.30	7.88	7.42	7.47
1 FS step	ER	34.42	54.96	29.77	49.35	44.13	42.52
3rd set, renorm	CR	57.64	36.34	63.56	39.79	49.89	*49.44*
9X32 features	UR	7.24	8.25	6.34	7.72	7.82	7.47
1 FS step	ER	35.11	55.42	30.10	42.49	42.29	43.08

Table 6.6: Results for 3rd set of features, artifacts automatically rejected for training, DA classifier

6.5 Denoising Techniques

6.5.1 Introduction

We investigated 3 different denoising techniques. One is based on Independent Component Analysis (ICA) [70] and two others on Principal Component Analysis (PCA). We applied the different denoising methods to the data and then computed the 1st set of features without Coh features and without PLV values for individual electrode pairs. Only one feature set was considered as testing several algorithms with several settings of the parameters for all feature sets considered in this study would be very time-consuming and would not improve the legibility.
Although results achieved with the 3rd set of features were slightly better, this set was chosen because it contained only 169 features. In addition, it does not include Coh_{max} features, which require long computation time.

Both ICA and PCA are statistical and computational techniques for revealing hidden factors that underlie sets of random variables, measurements, or signals. They search a transformation of the data such that the

transformed variables give information that is otherwise hidden in the large data set. Every component y_i is expressed as a linear combination of observed variables x_i.

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \mathbf{W} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} \tag{6.3}$$

where the coefficients w_{ij} define the representation and $t = 0 \ldots N-1$, N defining the length of the variables. ICA and PCA can be regarded as the problem of determining the coefficients w_{ij}. The matrix $\mathbf{W}$ could be determined by the statistical properties of the transformed components y_i.

6.5.2 ICA-based Method

ICA results in components that are both statistically independent and non-gaussian. The value of any one of the statistically independent components gives no information on the values of the other components. For gaussian data, uncorrelated components are always independent. ICA, however, tries to find statistically independent components in the general case where the data are non-gaussian. In practical situations, in general it is not possible to find a representation where the components are really independent, one can find components that are as independent as possible.

We applied FastICA (`http://www.cis.hut.fi/projects/ica/fastica/`) to decompose our 32-channel EEG data and reconstructed a denoised EEG signal from the components that had their SMF in the interval 8-30Hz and 4-40Hz respectively.

From Tables 6.7 and 6.8 we observe that for both SVM and DA classifiers the CR decreased, suggesting that more information relevant to the discrimination of the mental tasks than noise has been removed from the original signals.

ICA denoising	Subj	1	2	3	4	5	Average
None	CR	**61.56**	40.27	**63.50**	39.49	50.71	**51.11**
	UR	6.19	4.06	7.20	8.02	6.95	6.49
	ER	32.25	55.67	29.30	52.49	42.33	42.41
SMF 8-30Hz	CR	51.00	38.84	41.98	33.09	36.41	40.26
	UR	6.67	5.08	6.12	6.49	5.99	6.07
	ER	42.34	56.08	51.89	60.43	57.60	53.67
SMF 4-40Hz	CR	56.22	40.04	54.29	36.38	45.57	46.50
	UR	6.91	5.37	6.41	6.83	6.98	6.50
	ER	36.88	55.59	39.31	56.79	47.44	47.20

Table 6.7: Results for ICA denoising, applied to the first feature set without Coh_{max} features and without PLV features for individual electrode pairs, SVM classifier

ICA denoising	Subj	1	2	3	4	5	Average
None	CR	59.62	36.97	60.19	39.51	49.47	49.15
	UR	5.06	5.39	6.28	8.22	6.89	6.37
	ER	35.33	57.64	33.53	52.27	43.64	44.48
SMF 8-30Hz	CR	50.38	33.84	41.19	35.39	36.47	39.46
	UR	7.16	6.25	6.48	6.64	6.80	6.66
	ER	42.46	59.91	52.33	57.97	56.74	53.88
SMF 4-40Hz	CR	54.38	33.05	53.89	37.33	44.24	44.58
	UR	6.01	7.01	6.35	5.92	7.54	6.57
	ER	39.61	59.94	39.76	56.76	48.23	48.86

Table 6.8: Results for ICA denoising, applied to the first feature set without Coh_{max} features and without PLV features for individual electrode pairs, DA classifier

6.5.3 Global PCA-based Method

In case of Principal Component Analysis or factor analysis [72, 148], the statistical principle for choosing the matrix **W** in Equation (6.3) is to limit the number of components y_i to be small and to determine **W** so that the y_i contain as much information on the data as possible. **W** is determined by finding a single linear combination such that it explained the maximum amount of the variation in the data.

The signal $x(t)$ is reconstructed in a D-dimensional space of delayed coordinates (D = embedding dimension). When it is assumed that the noise is random and white while the signal is predictable up to some extent, the noise is present uniformly in this space, while the trajectory of the signal is confined to a lower dimensional subspace of dimension $d < D$ (d = projection dimension).
Consider the observed signals $x(t)$, consisting of a signal and noise part:

$$x(t) = s(t) + n(t), t = 0, \ldots, N-1, \tag{6.4}$$

where N is the length of the signal. The space of delayed coordinates is defined by:

$$\mathbf{x}_n(t) = [x(t), x(t-\delta), \ldots x(t-(D-1)\delta)]^T, n = 1 \ldots N-(D-1)\delta, \tag{6.5}$$

where δ is the embedding lag and D the dimension of the reconstructed state space. The number of points in the global state space, $\mathcal{G}$, is $N_g = N-(D-1)\delta$.

Noise reduction is obtained by omitting the noise subspace for signal reconstruction. To this end, we use PCA of the zero mean data vector $\mathbf{x}_{n0} = \mathbf{x}_n - \bar{\mathbf{x}}$, $\forall \mathbf{x}_n \in \mathcal{G}$ defined by the orthogonal transformation:

$$\mathbf{x}_{n0}(t) = \sum_{j=1}^{D} a_{nj}(t)\mathbf{\Phi}_j \tag{6.6}$$

with $\mathbf{\Phi}_j, j = 1, \ldots D$ are the eigenvectors of the local covariance matrix **C**:

$$\mathbf{C} = \sum_{\mathbf{x}_n(t) \in \mathcal{G}} \mathbf{x}_{n0}(t)\mathbf{x}_{n0}(t)^T. \tag{6.7}$$

As **C** is a symmetric non-negative matrix of dimension DxD, it determines a complete set of orthogonal eigenvectors $\mathbf{\Phi}_j, j = 1, \ldots D$. The principal components $a_{nj}(t)$ are obtained by projection of the data onto the eigenvectors:

$$a_{nj}(t) = \mathbf{x}_{n0}^T(t)\mathbf{\Phi}_j \qquad j = 1, \ldots, D \tag{6.8}$$

The delayed vector with reduced noise is obtained as follows

$$\hat{s}_n(t) = \sum_{j=1}^{d} a_{nj}(t)\mathbf{\Phi}_j + \bar{\mathbf{x}} \qquad d < D \tag{6.9}$$

Algorithms for denoising in the global state space have been provided by Celka and allow for the parameter options mentioned in Table 6.9. "Mode_lag" determines whether a lag = 1 or a lag computed from the correlation function will be used for state space embedding. D determines the embedding dimension and corresponds to the maximum number of eigenvalues.
The dimension d of the projection space can be determined manually or automatically by MDL (Minimum Description Length) [123]. MDL computes d from the eigenvalues of the autocovariance matrix of the reconstructed signal, C. When the projection dimension is determined with MDL, "mode_eig" allows for choosing the L_1 or L_2 norm, i.e. whether the singular values or the eigenvalues of C are used for MDL.

The purpose of statistical modeling is to discover regularities in observed data. The success in finding such regularities can be measured by the length with which the data can be described. This is the rationale behind the Minimum Description Length (MDL) Principle, introduced by Rissanen in 1978. Any regularity in a given set of data can be used to compress the data.
MDL is a consistent model order estimator, especially for short time series. It selects the model that produces the minimum code length for the given data. The shorter the stochastic complexity, the better the model class. The amount of useful and learnable 'information' in the data as well as the amount that

SNR	Basic noise level	10
mode_lag	Lag = 1 if 'n'; computed from correlation if 'y'	'y'-'n'
D	Dimension of the embedding space (10 - 60)	20-10
mode_MDL	use of MDL for determining the dimension of the projection space if 'y'	'y'-'n'
mode_eig	use L_2 norm if 'y'; use L_1 norm if 'n'	'y'-'n'
d	Number of eigenvalues used for projection if mode_MDL='n'	4-5-10

Table 6.9: Parameter options for Global PCA denoising and values used in our experiments

cannot be explained as incompressible 'noise' can formally be measured.

Tables 6.10 and 6.11 present results obtained for SVM and DA classification respectively with features computed after global PCA denoising. For SVMs, none of the denoising methods achieved a better average CR than the CR obtained without application of denoising. For subjects 2 and 4, however, sometimes an improvement of CR was detected. Originally, 51.11% classification accuracy was obtained, the best denoising method yielded and average CR of 51.06%. In case of DA classification, the average CR increased from 49.15% to 49.20% in the best case. For subjects 3 and 5 no results better than the ones without denoising could be obtained.

Global PCA	Subj	1	2	3	4	5	Average
None	CR	**61.56**	40.27	**63.50**	39.49	50.71	**51.11**
	UR	6.19	4.06	7.20	8.02	6.95	6.49
	ER	32.25	55.67	29.30	52.49	42.33	42.41
20-10-4-n-n-y	CR	60.59	40.76	62.81	39.85	50.38	50.88
	UR	6.42	4.67	6.87	8.28	6.92	6.63
	ER	33.00	54.57	30.32	51.88	42.70	42.49
20-10-4-n-n-n	CR	60.20	39.78	62.82	40.33	50.53	50.73
	UR	6.54	5.17	6.47	8.34	7.14	6.73
	ER	33.26	55.05	30.71	51.34	42.32	42.54
20-10-4-y-n-n	CR	53.74	38.72	43.72	40.13	43.43	43.95
	UR	4.75	4.22	6.10	5.46	5.28	5.16
	ER	41.84	57.06	50.18	54.41	51.29	50.96
20-10-4-n-y-n	CR	60.20	39.78	62.82	40.33	50.53	50.73
	UR	6.51	5.17	6.47	8.34	7.14	6.72
	ER	33.26	55.05	60.71	51.34	42.32	42.54
20-10-4-y-y-y	CR	58.46	39.05	59.97	40.67	48.08	49.25
	UR	6.72	5.76	7.50	7.40	7.48	6.97
	ER	34.82	55.19	32.54	51.92	44.44	43.78
10-10-5-n-n-y	CR	60.59	41.25	62.81	39.98	50.54	51.03
	UR	6.39	4.27	6.65	8.87	6.83	6.60
	ER	33.03	54.48	30.54	51.15	42.63	42.37
20-10-10-n-n-n	CR	60.62	41.16	63.00	40.19	50.35	51.06
	UR	6.39	4.67	6.78	8.34	6.89	6.61
	ER	33.00	54.17	30.22	51.47	42.76	42.32
10-10-5-n-n-n	CR	60.65	41.31	62.60	39.76	50.57	50.98
	UR	6.45	4.33	6.44	8.74	7.08	6.61
	ER	32.90	54.35	30.96	51.50	42.35	42.41

Table 6.10: Results for Global PCA denoising (D, SNR, d, mode_lag, mode_eig, mode_MDL), applied to the first feature set without Coh_{max} features and without PLV features for individual electrode pairs, SVM classifier

The maximum number of eigenvalues did not significantly change the obtained CRs when the dimension of the projection space was determined by MDL or defined as half of the dimension of the reconstruction space. For the same reconstruction dimension, the dimension of the projection space did not significantly affect the CR. The choice of L_1 or L_2 norm does not have a great influence on the CR either. Only the choice of "mode_lag" seemed to significantly change the CR. It seems better to use a lag = 1, rather than

Global PCA	Subj	1	2	3	4	5	Average
None	CR	59.62	36.97	60.19	39.51	49.47	49.15
	UR	5.06	5.39	6.28	8.22	6.89	6.37
	ER	35.33	57.64	33.53	52.27	43.64	44.48
20-10-4-n-n-y	CR	59.59	36.61	59.54	40.07	48.74	48.91
	UR	6.58	6.71	6.78	8.90	7.63	7.32
	ER	33.82	56.67	33.68	51.03	43.63	43.77
20-10-4-n-n-n	CR	59.03	37.61	59.60	41.20	48.56	49.20
	UR	6.91	7.27	7.37	8.77	7.70	7.60
	ER	34.06	55.13	33.03	50.02	43.74	43.20
20-10-4-y-n-n	CR	49.98	36.58	43.40	36.13	42.12	41.64
	UR	5.80	5.51	5.52	5.39	5.25	5.49
	ER	44.22	57.91	51.08	58.48	52.63	52.86
20-10-4-n-y-n	CR	59.03	37.61	59.60	41.20	48.56	47.01
	UR	6.91	7.27	7.37	8.77	7.70	7.60
	ER	34.06	55.12	33.03	50.02	43.74	43.20
20-10-4-y-y-y	CR	58.90	35.97	59.14	40.20	45.00	45.23
	UR	6.28	6.10	6.87	7.40	6.67	6.66
	ER	34.82	57.93	33.99	52.40	48.33	45.49
10-10-5-n-n-y	CR	59.63	36.87	59.69	40.59	48.84	49.12
	UR	6.83	7.21	6.81	8.67	7.82	7.47
	ER	33.55	55.93	33.49	50.74	43.35	43.41
20-10-10-n-n-n	CR	59.14	37.08	59.51	40.58	48.81	49.02
	UR	6.89	6.80	6.91	8.71	7.54	7.37
	ER	33.98	56.12	33.59	50.72	43.66	43.61
10-10-5-n-n-n	CR	58.77	37.27	59.50	40.67	48.37	48.92
	UR	7.44	6.84	7.16	9.05	8.00	7.70
	ER	33.79	55.90	33.34	50.27	43.63	43.38

Table 6.11: Results for Global PCA denoising (D, SNR, d, mode_lag, mode_eig, mode_MDL), applied to the first feature set without $\text{Coh}_{\max}$ features and without PLV features for individual electrode pairs, DA classifier

using the lag computed from the correlation function.

6.5.4 Local PCA-based Method

Generally, the linear approximation of PCA is not adequate, therefore some PCA-based local linear projective noise reduction algorithms have been proposed. PCA is applied to local clusters of nearest-neighbor vector points on the attractor which are close enough together for the local region to be considered linear.

Consider again a signal $x(t)$ as in equation (6.4) and reconstruct the signal in a global state space, as in equation (6.5). In this global state space, local neighborhoods $\mathcal{N}^k$ can be defined around the center of mass defined by:

$$\bar{\mathbf{x}}^{(k)} = \frac{1}{|\mathcal{N}^k|} \sum_{\mathbf{x}_n(t) \in \mathcal{N}^k} \mathbf{x}_n(t) \tag{6.10}$$

$k = 1 \dots K$ and K is the number of neighborhoods. Noise reduction is obtained by omitting the noise subspace for signal reconstruction. To this end, we use PCA of the local zero mean data vector $\mathbf{x}_{n0} = \mathbf{x}_n - \bar{\mathbf{x}}^{(k)}$, $\forall \mathbf{x}_n \in \mathcal{N}^k$ defined by the orthogonal transformation:

$$\mathbf{x}_{n0}(t) = \sum_{j=1}^{D} a_{nj}(t) \mathbf{\Phi}_j^{(k)} \tag{6.11}$$

with $\mathbf{\Phi}_j^{(k)}, j = 1, \dots D$ are the eigenvectors of the local covariance matrix $\mathbf{C}^{(k)}$:

$$\mathbf{C}^{(k)} = \frac{1}{|\mathcal{N}^k|} \sum_{\mathbf{x}_n(t) \in \mathcal{N}^k} \mathbf{x}_{n0}(t) \mathbf{x}_{n0}(t)^T. \tag{6.12}$$

As $\mathbf{C}^{(k)}$ is a symmetric non-negative matrix of dimension DxD, it determines a complete set of orthogonal eigenvectors $\mathbf{\Phi}_j^{(k)}, j = 1, \ldots D$. The principal components $a_{nj}(t)$ are obtained by projection of the data onto the eigenvectors:

$$a_{nj}(t) = \mathbf{x}_{n0}^T(t)\mathbf{\Phi}_j^{(k)} \qquad j = 1, \ldots, D \tag{6.13}$$

The delayed vector with reduced noise is obtained as follows

$$\hat{s}_n(t) = \sum_{j=1}^{d^{(k)}} a_{nj}(t)\mathbf{\Phi}_j^{(k)} + \bar{\mathbf{x}}^{(k)} \qquad d^{(k)} < D \tag{6.14}$$

The denoised signal is obtained by merging all the local approximations. The approaches discussed below differ in the choice of local neighborhoods and parameters.

Cawley and Hsu [24] present a local-geometric-projection method for noise reduction. Noise reduction near $\bar{p}$ is achieved by averaging these points with their projections. Instead of projection of the whole state space onto one subspace of lower dimension, several boxes, all of them together covering the entire trajectory, are constructed. Projection to a lower dimensional subspace is performed in each of these boxes. Their implementation is such that every point $\mathbf{x}_n$ of the global state space appears in one neighborhood $\mathcal{N}^k$ only and thus is projected only once. Orbits are generated by delay coordinate construction. Parameters D, d, ν (number of neighbors), δ (delay), A_n (Noise Amplitude), N (trajectory length) were to be defined manually.
They developed a systematic procedure for covering the embedded object. The elements of the cover are ordered according to the magnitudes of r_k^{-1} for various neighborhoods $\mathcal{N}^k$ where r_k denotes the radius of the ball enclosing a fixed number of points nearest neighbor to a given corresponding reference point $\bar{\mathbf{x}}^{(k)}$. This produces a rough separation of densely and sparsely populated regions of the noisy attractor. The resulting measure-ordered cover allows for designing a natural scheme to prevent multiple projections. First, points in cover elements having small radius r_k are projected since the manifolds should be better approximated in these elements of the cover. Once a point has been projected, it is prohibited from being projected again. Their algorithm consisted of 1) a covering procedure, 2) a projection procedure and 3) an iteration protocol.

1. Select a random set of K candidate reference points $\mathbf{x}_c$ amongst the N_g points in the state space. Find for each of these reference points its nn nearest neighbors and form the neighborhood. Every point of the state space can be part of one neighborhood only. K should be the smallest number such that the neighborhoods cover the entire trajectory.
   For each neighborhood, compute and evaluate the sample average $r_k = \max_{0<j<\nu}|\bar{\mathbf{x}}^{(k)} - \mathbf{x}_j|, \forall \mathbf{x}_j \in \mathcal{N}^k$. Order the neighborhoods such that $r_1^{-1} \geq r_2^{-1} \geq \ldots \geq r_K^{-1}$

2. Starting from the first neighborhood , estimate the sample average $\bar{\mathbf{x}}^{(k)}$
   Denote by Q the projection operator, replace the points of the neighborhood by a new collection

$$\mathbf{x}_j^{(k)} \rightarrow \hat{\mathbf{x}}_j^{(k)} = \bar{\mathbf{x}}^{(k)} + f(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}^{(k)}) + (1-f)Q(\mathbf{x}_j^{(k)} - \bar{\mathbf{x}}^{(k)}),$$
$$0 \leq f \leq 1, j = 0, 1, ..., \nu. \tag{6.15}$$

   The original projection points are averaged with the original data state points by means of the factor f to take the local curvatures of the manifold in a rough way into account. Repeat this for each point of the neighborhood without projecting points previously projected.
   The scalar time series corresponding to the state vector time series resulting from previous steps are computed.

3. After replacing the original time series $x(t)$ by $\hat{s}(t)$, the signal is re-embedded and the previous steps are repeated for some choice of D, d, ν, δ and τ_s (the algorithm sampling time). The algorithm is iterated until the SNR improvement, measured by a method proposed for estimating the noise present in the data, achieved a stable maximum.

Vetter *et al* [149] developed a method for automatic nonlinear noise reduction using local PCA where the parameters are chosen automatically using MDL. Orbits where generated by delay coordinate construction with the delay δ set to 1. Each point in the state space appeared in one box only. Every point $\mathbf{x}_n$

can appear in one neighborhood $\mathcal{N}^k$ only.
They use the MDL criterium to determine $d^{(k)}$ in each local domain.

The choice of the global parameters embedding dimension D and neighborhood size $|\mathcal{N}^k|$ or, equivalently, the number of local regions K, is based on the MDL criterion. To this end, a family of competing models and density functions is considered. The parameters D and K are chosen such that the residual errors of an approximation of the data set are most likely to be Gaussian white noise. The likelihood is proposed to be based on the singular values of the covariance matrix of the delayed embedding of the residual noise. This selection of D and K is computationally demanding. If prior knowledge is at disposition, the range of D and K can be restricted accordingly to diminish computational load.

We implemented an approach to noise reduction based on local PCA allowing for delays greater than 1,and where the points in the global state space can appear in more than one neighborhood. The denoised signal is then reconstructed from multiple projections of a point in multiple neighborhoods.

The optimal embedding delay is computed from the autocorrelation function. If the delay δ is very big, there will be only few points in the state space. We put a maximum value on the delay, equal to $\delta < \frac{N-6\nu}{D-1}$ so that we have at least 6 times more points in the local space than the number of nearest neighbors.
The algorithm is as follows:
The signal is embedded in state space of dimension D. Choose K ($= N_g/10$ in our experiments) random points in this space and find the ν (50 in our experiments) nearest neighbors to construct the neighborhood. Then, d is chosen manually or determined with MDL and the data of the neighborhood are projected onto the d-dimensional subspace. A denoised signal can then be obtained from the projections along the different projection dimensions. Finally, the denoised signal can be reconstructed from the segments resulting from the different neighborhoods.

Table 6.12 shows that for SVM classifiers, the average CR obtained for local PCA denoising never outperforms the one obtained without denoising. The best average was CR = 50.96% as compared to 51.11% originally. For subject 2, 4 and 5, local PCA denoising could yield improved results.
As compared to global PCA denoising, the best average CRs were about the same: 51.06% and 50.96% for global and local PCA denoising respectively.
The different settings of parameters considered did not affect the CR significantly. Using MDL to determine the dimension of the subspace yielded slightly better results than the ones obtained with a manually chosen d.
For DA classification, a slightly better average CR of 49.50% (compared to 49.15%) could be obtained. For subjects 1, 2 and 4 local PCA denoising could yield an improved classification accuracy as compared to no denoising.

The denoising results for the different subjects are summarized in the Figures 6.2, 6.3, 6.4, 6.5 and 6.6. The horizontal dashed line shows the CR achieved without denoising. The vertical dotted lines group the denoising methods: no denoising, ICA denoising, global PCA denoising, local PCA denoising.

6.6 Conclusions

In this chapter, we investigated whether automatically detecting and removing windows containing artifacts from the training data improves the classification results obtained on test data. Also whether, similarly, denoising yields increased classification accuracies.
Artifact detection only slightly improved the best average CR.
The ICA-based method we applied for denoising significantly deteriorated the classification accuracies.
The PCA based approaches, especially local PCA-based denoising, yielded improved results for 3 out of 5 subjects. The best CR averaged over the 5 subjects, however, did not outperform the result obtained without denoising.
The mixed performance of the considered PCA-based denoising methods may be due to the fact that they enhance the deterministic part of the signal. By suppressing random contributions, the "dynamics" of the features used for classification may be reduced, which may cause a decreased performance for certain

Local PCA	Subj	1	2	3	4	5	Average
None	CR	**61.56**	40.27	**63.50**	39.49	50.71	**51.11**
	UR	6.19	4.06	7.20	8.02	6.95	6.49
	ER	32.25	55.67	29.30	52.49	42.33	42.41
50-10-5-10-n	CR	60.34	41.66	62.53	39.94	49.45	50.79
	UR	6.70	5.30	7.40	8.06	8.29	7.15
	ER	32.96	53.04	30.07	52.00	42.27	42.07
50-10-5-10-y	CR	59.83	41.59	62.69	39.93	50.57	50.92
	UR	7.16	4.27	7.88	8.82	7.26	7.08
	ER	33.02	54.14	29.43	51.25	42.17	42.00
50-10-5-10-n,	CR	60.86	40.60	62.75	39.86	50.38	50.89
delay=1	UR	6.52	4.30	7.15	8.87	7.29	6.83
	ER	32.62	55.10	30.10	51.27	42.32	42.28
50-10-5-10-y,	CR	60.28	41.25	62.97	39.54	**50.79**	50.96
delay=1	UR	6.60	4.18	6.99	8.65	6.70	6.63
	ER	33.12	54.57	30.04	51.81	42.51	42.41
50-20-4-10-n	CR	60.11	**41.71**	59.37	40.09	50.29	50.31
	UR	6.45	5.29	7.86	7.93	7.36	6.98
	ER	33.45	53.00	32.77	51.98	42.36	42.71
50-20-4-10-y	CR	60.46	40.67	62.87	40.04	50.35	50.88
	UR	6.51	4.12	7.23	7.90	7.26	6.61
	ER	33.03	55.21	29.89	52.06	42.38	42.52

Table 6.12: Results for Local PCA denoising (ν, D, d, SNR, mode_MDL), applied to the first feature set without Coh_{max} features and without PLV features for individual electrode pairs, SVM classifier

Local PCA	Subj	1	2	3	4	5	Average
None	CR	59.62	36.97	60.19	39.51	49.47	49.15
	UR	5.06	5.39	6.28	8.22	6.89	6.37
	ER	35.33	57.64	33.53	52.27	43.64	44.48
50-10-5-10-n	CR	59.86	38.43	59.57	41.11	48.52	49.50
	UR	7.58	6.78	6.81	8.50	7.45	7.42
	ER	32.56	54.79	33.62	50.40	44.03	43.08
50-10-5-10-y	CR	58.52	36.70	59.88	40.73	48.69	48.90
	UR	6.76	7.02	7.89	9.06	7.44	7.64
	ER	34.72	56.27	32.23	50.21	43.87	43.46
50-10-5-10-n,	CR	59.50	37.04	59.17	41.02	48.31	49.01
delay=1	UR	6.68	6.90	7.16	8.71	8.58	7.61
	ER	33.82	56.06	33.68	50.27	43.01	43.37
50-10-5-10-y,	CR	59.44	36.90	59.82	**41.33**	48.84	49.26
delay=1	UR	27.07	7.24	6.54	8.65	8.50	11.60
	ER	33.49	55.87	33.64	50.02	42.67	43.14
50-20-4-10-n	CR	59.95	38.96	56.29	41.29	48.70	49.04
	UR	7.37	6.44	8.26	7.78	7.35	7.44
	ER	32.68	54.66	35.45	50.94	43.94	43.53
50-20-4-10-y	CR	59.66	36.80	59.63	40.99	48.59	49.13
	UR	6.64	6.84	6.78	8.62	7.91	7.36
	ER	33.70	56.36	33.59	50.39	43.50	43.51

Table 6.13: Results for Local PCA denoising (ν, D, d, SNR, mode_MDL), applied to the first feature set without Coh_{max} features and without PLV features for individual electrode pairs, DA classifier

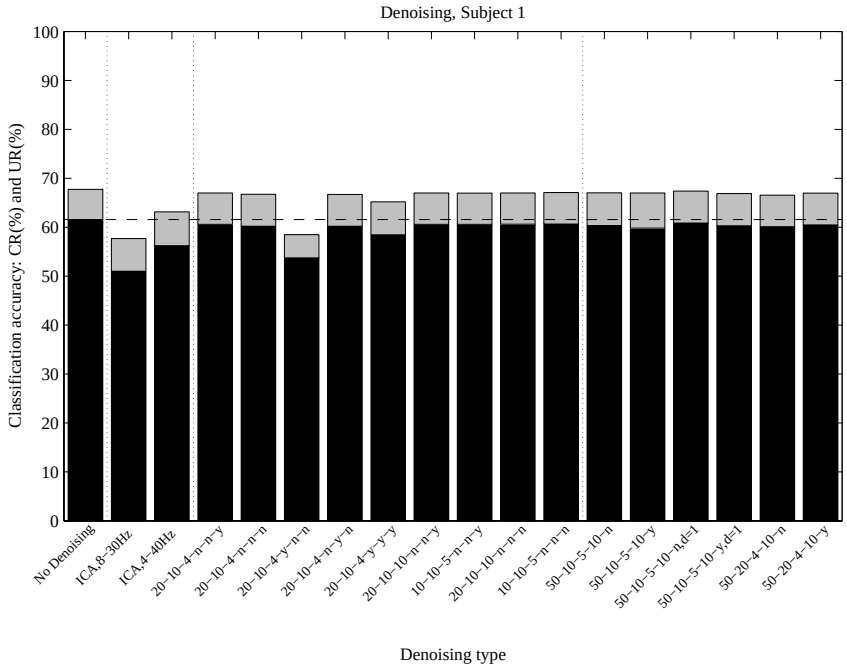

Figure 6.2: CR and UR obtained with SVM for different denoising methods (no denoising, ICA denoising, global PCA denoising, local PCA denoising), Subject 1

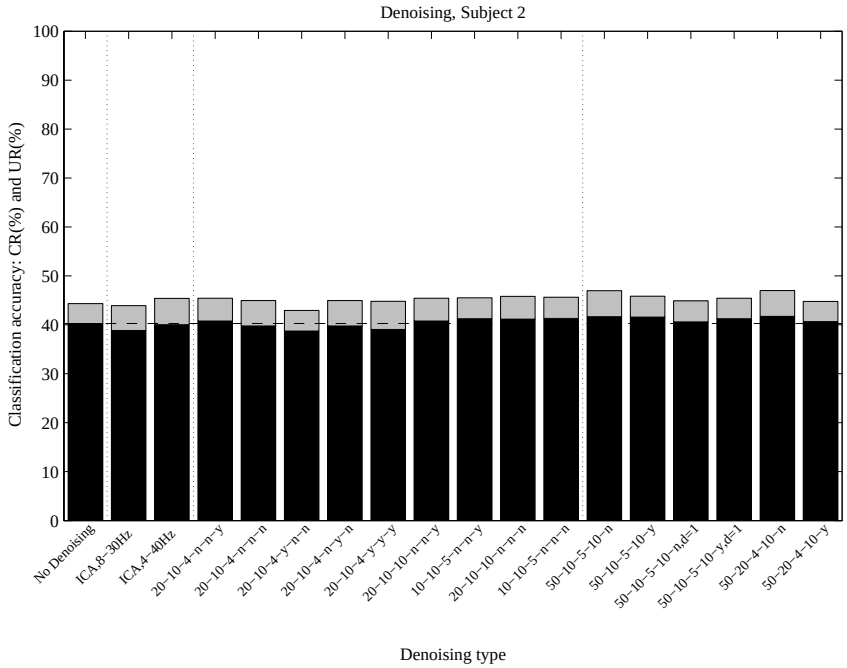

Figure 6.3: CR and UR obtained with SVM for different denoising methods (no denoising, ICA denoising, global PCA denoising, local PCA denoising), Subject 2

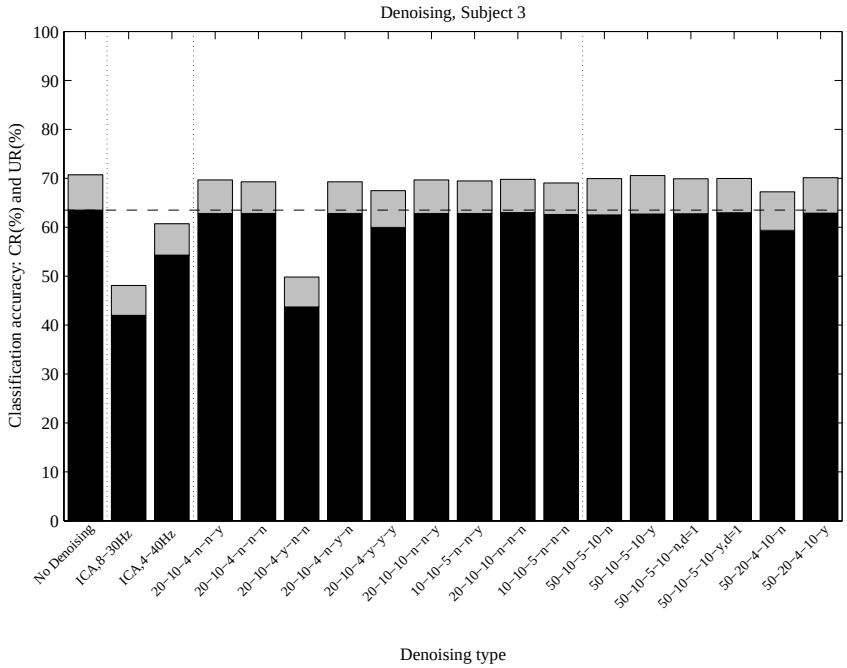

Figure 6.4: CR and UR obtained with SVM for different denoising methods (no denoising, ICA denoising, global PCA denoising, local PCA denoising), Subject 3

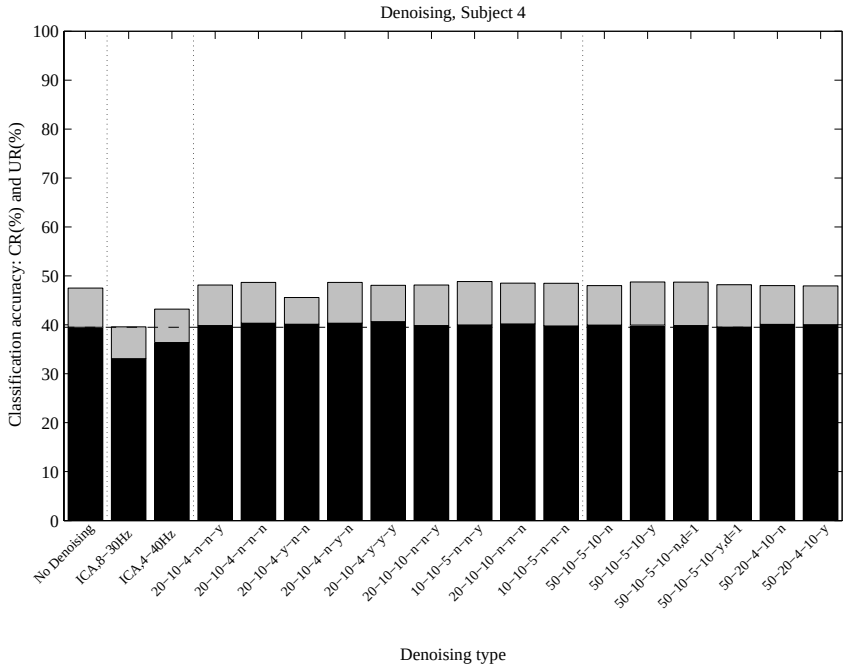

Figure 6.5: CR and UR obtained with SVM for different denoising methods (no denoising, ICA denoising, global PCA denoising, local PCA denoising), Subject 4

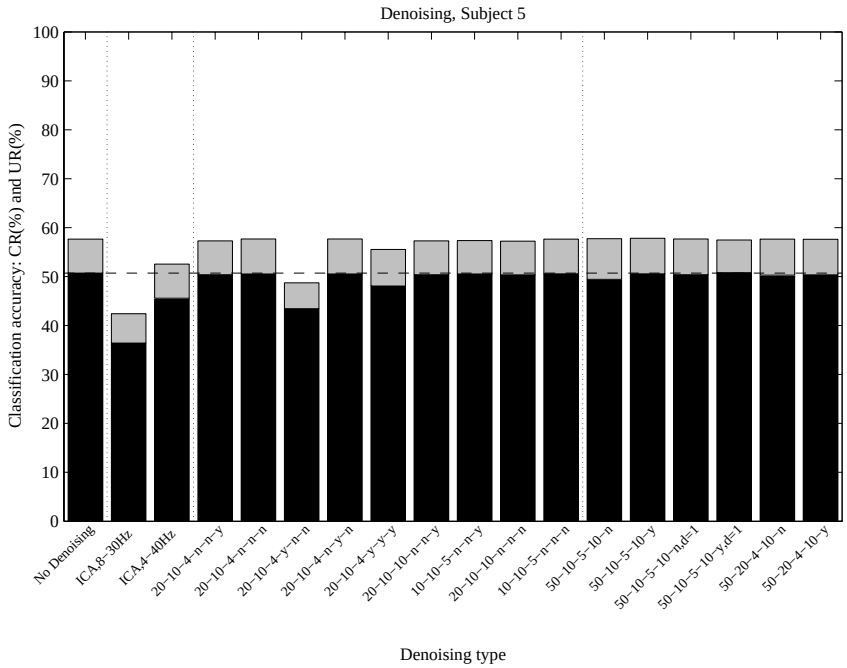

Figure 6.6: CR and UR obtained with SVM for different denoising methods (no denoising, ICA denoising, global PCA denoising, local PCA denoising), Subject 5

types of features. This is not necessarily true for all types of features.

Chapter 7

Conclusions

7.1 Introduction

This chapter discusses the achievements reported in this thesis. Section 7.2 summarizes the previous chapters. The method proposed for analysis of surface EEG in the framework of BCIs is described in Section 7.3. Section 7.4 proposes issues to explore and Section 7.5 suggests other possible applications of the work presented here. We conclude this chapter with some ethical remarks related to BCI (Section 7.6).

7.2 Summary of results

In Chapter 2 we reviewed BCIs and the state-of-the-art features employed in BCIs. As we have seen, most BCIs use spectral estimates and AR coefficients. Some exploit spatial information by using MVAR coefficients, CSPs or TFSC. After describing the BCI paradigm considered in this study, in Chapter 3, we proposed, in Chapter 4, synchronization features, novel to BCI, to extract task-relevant information from the EEG signals. After introducing PLV and coherence, their properties with regard to resolution and influence of noise were discussed and illustrated on sine waves and uncoupled and coupled Rössler oscillators.
For very short time windows, only PLV distinguished the different regimes of coupling. PLV and Coh_{max} were quite robust to noise added to the signal and the amplitude. They were less robust to noise added to the phase and not at all robust to noise added to the frequency.
Furthermore, we suggest that PLV is more meaningful when computed from broadband filtered signals, because narrowband filtering artificially increases the the synchronization between signals and reduces the PLV to a detection of presence of certain frequencies in the signals. For studying frequency-specific coupling, the spectral coherence should be used instead.

We demonstrated that synchronization measures are relevant for discrimination of spontaneous EEG recorded during different mental tasks. Generally, PLV yielded better classification accuracies than coherence-based features. Synchronization features and PSD features generally did not outperform each other, but a combination of both usually led to (significantly) improved results and never degraded performances as compared to the separate subsets. This suggests some complementarity between the bivariate synchronization measures and the single-channel PSD and shows new possibilities for future BCI research. The PLV may be especially suited, because of its fast computation, necessary in online feedback setups. In addition, it deals better with short time windows than coherence.

In Chapter 5, several feature selection algorithms were discussed. GAs as well as SVM-rfe showed to select features from EEG signals recorded from brain regions involved in the considered mental tasks, but were very slow. FCBF is a much faster algorithm, has a good generalization performance and shows an even stronger correspondence with neurophysiological evidence.

Application of feature selection algorithms to α, β_1, β_2 and λ power features as well as PLV features computed from EEG signals filtered in the α, β_1, β_2 and λ frequency bands, showed us that BP and

PLV features are preferably computed from the α and λ frequency band respectively. For broadband signals, the phase does not have a clear physical meaning. We think PLV could be more discriminative in case of broadband signals, because in this case synchronization is more difficult to achieve. Filtering the EEG signals in narrow frequency bands removes the components that could desynchronize the two signals.

Results obtained by applying of FCBF to PSD, PLV, and IMF-based features computed from BCI competition data were presented. Significant classification results have been obtained on blind data.
Furthermore, FCBF was applied to different sets of features in order to evaluate them. Studying the results presented in Tables 5.11 and 5.12, we concluded that including Coh_{max} features and PLV values for individual electrode pairs in the fronto-centro-parietal region in the feature set did not yield improved results. Combining PLV averages and power features yielded better results than using these features separately. Tables 5.14 and 5.15 also suggest Coh_{max} can be left out of the feature set without significantly decreasing the CR. Tables 5.17 and 5.19 indicate that considering PLV values for all possible electrode pairs slightly improves the classification accuracy for some subjects. The minor increase of classification accuracy, however, has to be weighed against the additional computational load and memory consumption. In some cases, better results were obtained by dividing the features computed for a single EEG signal by the average over all EEG signals.

Whether artifact detection and rejection and noise reduction algorithms improved the classification accuracies was verified in Chapter 6. SVMs yielded again better results than DA and were considered for drawing the following conclusions. The CR often increased slightly when automatically detecting artifacts and rejecting the corresponding data window for feature selection and training the classifier. However, if any at all, the increase in classification accuracy was generally not greater than 1%.
In all cases, the ICA-based noise reduction method we proposed decreased the classification accuracies significantly. Noise reduction based on global or local PCA never yielded an improved result for two of the subjects. The denoising algorithms may reduce the performance of some features considered by reducing their "dynamics". For the other 3 subjects, the CR increased with up to 1.69% and up to 1.84% for global and local PCA respectively.

Studies with motor-disabled individuals are necessary to confirm that for disabled subjects results similar to those for healthy subjects can be achieved. Magnetic Resonance Imaging research indicates that underlying motor maps are maintained, even after years of paralysis [142].

7.3 Proposed Analysis Strategy

With the methods developed during this PhD, significant classification accuracies have been obtained on blind data. This validates the potential of the proposed techniques. If refined, they could contribute to further BCI development.

Based on the results presented in the preceding chapters, the following recommendations are proposed for analyzing EEG signals in the framework of BCIs:

If automatic artifact detection and rejection and noise reduction algorithms improved the discrimination of EEG recorded during different mental tasks, this improvement was only minor. Therefore, we do not generally recommend the application of these algorithms.

From the results presented in Chapter 5, we suggest that the set of features to select features from should comprise:

- The SMF for all electrodes.
  The center of gravity of the power spectrum seems to be related to the mental tasks (see Figure 5.11). This feature is selected for electrode signals recorded from brain regions involved in the mental tasks.

- The ratio between the power in the frequency bands usually used in EEG studies (δ, θ, α, β and γ) and the total power of the (filtered) signal.

When prior knowledge about the mental tasks and/or the subject's EEG is available, these frequency bands could be restricted, or narrower frequency bands may be chosen. For motor imagery paradigms, e.g., considering the power in the α and possibly the β_1 frequency band will generally be sufficient.

- The SMF of the first 3 Intrinsic Mode Functions of the empirical mode decomposed EEG signals.
- The ratio of the power of the first 3 IMFs respectively and the power of the original signal.
  In this thesis, features derived from the EMD of the EEG signals were selected especially for intracranially recorded data. They resulted in a good long-term (1 week) generalization and yielded high classification accuracies in combination with PLV and BP features.
- PLV features computed from broadband filtered EEG signals.
  Considering PLV values for all possible electrode pairs may result in increased CRs, but we recommend to first consider averages of PLV values for all pairs of electrodes within different groups of electrodes and for averages of the PLV values of the pairs formed by an electrode and its neighbors. When using averaged PLV values, it may be of interest to also consider the CPI.

It may be worthwile to renormalize the features, i.e. to divide a feature for a given EEG signal by the average of that feature over all EEG signals, and to try a selection from a subset of the above-mentioned features.

For feature selection, the modified FCBF algorithm is preferred to Genetic Algorithms and SVM-based rfe because:

- For the FCBF selected features, the ratio between performance on test data and training data is better than for SVM-rfe and GAs. FCBF has less tendency to overtrain.
- It is very fast, an important asset in the framework of BCIs. It would allow for selecting new features between different sessions of BCI communication.
- FCBF achieves good classification accuracies with very few features. This allows for an acceleration of the classification step.
- The electrode signals selected by FCBF correspond well to the brain regions indicated by neurophysiological studies to be involved in the considered mental tasks.

For classification purposes, generally, linear SVMs outperformed robust linear DA, but in some cases DA may achieve better results.

Practically, online application should be possible. PLV and PSD features are fast to compute. Only the EMD may pose some computational problems. In that case, the proposed set of features could be reduced to retain only PLV and PSD based features. After recording some data without feedback, FCBF could be applied to find the types of features and electrode signals relevant for a specific experimental paradigm and subject. From the computational point of view, FCBF can easily be applied between two recording sessions. With a standard PC with a 3GHz processor, the time required to select features with FCBF and train an SVM classifier is on the order of minutes. This allows for updating the subset of features in between recording sessions.

We believe that the study of synchronization features and the exploration of features with the Fast-Correlation Based Filter can improve the development of BCIs. Further studies are needed to fully exploit the capacities of these methods.

7.4 Issues to Explore

This thesis was limited to 3 years and hence did not allow to explore the following ideas, maybe worthwile to study in the framework of BCIs:

- Use of interhemispheric PLV/coherence as features (PLV and coherence between signals recorded from corresponding electrodes in the left and right hemisphere).

- Link rates (number of transitions from unsynchronized to synchronized regime), proposed by van Putten [145], could be explored in the framework of BCI. Song [138] has explored binarization of PLV, but transitions from unsynchronized to synchronized regimes could be more robust because they take into account both significant synchronization and significant desynchronization.

- Investigation of extraction of other features from PLV values. Song, e.g., presents a desynchronization network analysis for the recognition of imagined movement [137].

- Interhemispheric differences in, e.g., spectral power.

- Interhemispheric asymmetries:

$$A_F(P_{e_1e_2}) = \frac{F(e_1) - F(e_2)}{F(e_1) + F(e_2)},$$

  Where $A_F(P_{e_1e_2})$ is the asymmetry of feature F for the interhemispheric electrode pair $e_1 - e_2$, e_2 being the corresponding electrode in the right hemisphere of e_1 in the left hemisphere. F(e_1) represents the value of a feature computed from the electrode signal e_1 and F(e_2) the value of the same feature computed from e_2.

- Transformations of the original features that result in better discriminable features could be explored. The disadvantage is that in this case we do not know what the selected features correspond to. But this is less important for practical use of BCI than a possible improvement of classification accuracy. Torkkola and Campbell [143], e.g., propose feature transformations based on Mutual Information.

- The optimal number of features should be determined in a more automatic and refined way. To this end, MDL [123] and Akaike's Information Criterion (AIC) [4] could be investigated. These information criteria introduce a performance measure and a penalty in function of the number of features to determine the optimal number of features.

- It is maybe worthwhile to investigate whether selecting pairs of features (the same feature for two corresponding electrodes in the different hemispheres) yields better results than selecting single features.

- Synchronization could be used to take the effects of learning into account [136]. To exploit this, synchronization features should be combined with adaptive classifiers.

- Data selection [50] could be applied to obtain better classifiers. Blum and Langley [18] give 3 reasons for applying data selection: 1) increase computational efficiency, 2) when cost of labeling is high, 3) increase the rate of learning by focusing attention on informative examples only.

7.5 Other Applications

Techniques developed during this PhD could be applied to other problems related to automatic discrimination of EEG.

One such application is the anticipation of epileptic seizures. In epilepsy research, this subject received a lot of attention in the last years. The quality of life of epilepsy patients could be improved significantly if their seizures could be anticipated. Synchronization features have already been studied in this framework [101, 100, 27, 145, 144]. Automated discrimination of EEG signals based on synchronization features has been proposed in this work. In order to train a reliable classifier that distinguishes preictal from interictal EEG, a sufficient amount of training data (interictal, preictal and ictal) is required. However, one may need a lot of time to acquire enough data.

Studies on intracranially recorded EEG [101, 91], REM sleep recordings [95] and long surface EEG recordings [64] of epilepsy patients indicate that synchronization is asymmetric between the two hemispheres and that this asymmetry is correlated to the hemisphere in which the epileptic focus is located. The features and feature selection algorithms presented in this study may be utilized to determine the epileptic

hemisphere and/or the type of seizure.

Differences between EEG of subjects with no reported psychiatric or neurological disorders, subjects with maniac symptoms and epilepsy patients have been reported [10]. It could be imagined to study the use of the techniques presented in this thesis to automatically diagnose epilepsy, mania, depression and schizophrenia.

Other applications could include the detection of sleep stages, mental effort, and the level of concentration.

In all cases, enough data should be acquired to reliably select features and train reliable classifiers. The data acquisition should be done carefully, in order to exclude deterioration due to, e.g., medication.

7.6 Ethical Issues

Individuals can learn how to change and control their own brainwave activity if they are given immediate feedback in an adequate form. This is how EEG biofeedback training works. EEG biofeedback, also called neurofeedback, uses a protocol similar to the one deployed in BCIs to treat ADHD (Attention Deficit Hyperactivity Disorder) patients. The aim of the neurofeedback is to alter the brainwave activity using correcting signals to obtain an increased attention span, increased learning, and decreased temper and impulsivity. Neurofeedback is also investigated as a technique to suppress seizures.
One can wonder whether such protocols, altering brain activity in ADHD patients and epilepsy patients, induces altered brain activity in healthy subjects participating in BCI experiments.
Wolpaw says that during the years of BCI research in their group, they have never noticed changed behavior of the subjects, or induced seizures. He argues that healthy brain activity is much more difficult to alter than pathological brain activity.
Nevertheless, it is remarkable that while people applying neurofeedback are required to be certified, almost no researchers in the BCI community are certified to carry out neurofeedback experiments.

Foster *et al* [49] discuss some ethical aspects related to the development of brain imaging techniques and BCIs and raise the following interesting questions. When, if ever, is it acceptable to manipulate someone's brain to make him perform actions that he would not ordinarily take voluntary? How far should we go in using technology to enhance our abilities and who should have access to and control over the technologies we use? A central problem in neuroethics is to establish appropriate limits of human intervention in our cognitive and affective functioning.
As with all new technologies, the consequences of new technologies derived from neuroscience are hard to predict. Even if we can never fully anticipate the impact of employing these technologies, it is important to attempt to do so. BCI devices have the potential to improve the lives of people, but these technologies will almost certainly bring negative consequences that we cannot clearly foresee. While many researchers investigating BCIs are motivated to develop a communication channel for disabled people, U.S. military researches the same technique (`http://www.sciencedaily.com/releases/2002/08/020820071329.htm`) for war making. A DARPA (Defense Advanced Research Projects Agency) spokesperson stated that the agency is interested in technology that might, for example, enable soldiers to push buttons with their brains, giving them speedier control of submarines and aircraft and enabling them to more adeptly manipulate robotic arms that move munitions [155].

We can only hope that BCIs will be used only to improve the quality of life of those (disabled) people who really need them. And in this respect I hope that this dissertation can be a contribution to the development of better BCIs.

Appendix A

EEG

The Society for Neuroscience publishes a 52-page primer on the brain and nervous system, "Brain Facts", which can be found on `http://apu.sfn.org/content/Publications/BrainFacts/brainfacts.pdf`.

This appendix uses material from the Wikipedia articles "Electroencephalography" (`http://en.wikipedia.org/wiki/Electroencephalography`) and "Human Brain" (`http://en.wikipedia.org/wiki/Human_brain`), licensed under the GNU Free Documentation License (`http://www.gnu.org/copyleft/fdl.html`).

Electroencephalography is the method and science of recording and interpreting traces of brain electrical activity as recorded from the scalp. The resulting traces are known as an electroencephalogram (EEG) and represent so-called brainwaves. The first EEG, shown in Figure A.1 (this image has been released into the public domain by the copyright holder, its copyright has expired, or it is ineligible for copyright. This applies worldwide), was obtained by Hans Berger.
Since the thirties, numerous studies have been done with the scalp EEG in efforts to correlate brain activity with sensory and behavioral events. Neuroscientists and biological psychiatrists use EEGs to study the function of the brain by recording brainwaves during controlled behavior of human volunteers and animals in lab experiments. Theories to explain sleep often rely on EEG patterns recorded during sleep sessions. Furthermore, EEG is used to assess brain damage, epilepsy and other problems, such as brain death. EEG can also be used in conjunction with other types of brain imaging.

A.1 EEG recording

The recording is obtained by placing electrodes on the scalp, usually after preparing the scalp area by light abrasion and application of a conductive gel to reduce impedance. The way to place electrodes on the scalp has been standardized and is referred to as the 10-20 electrode placement system [71]. Each electrode is connected to an input of a differential amplifier (one amplifier per pair of electrodes), which amplifies the voltage between them (typically 1,000-100,000 times, or 60-100 dB of voltage gain), and then inputs it to a computer and displays it on a screen. The amplitude of the EEG is about 10-100 μV when measured on the scalp, and about 1-2 mV when measured on the surface of the brain.

The electrode-amplifier relationships are typically arranged in one of three ways:

- Common reference derivation: One terminal of each amplifier is connected to the same electrode,

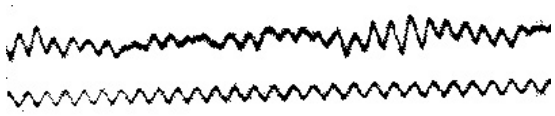

Figure A.1: The first EEG recording, obtained by Hans Berger in 1929.

and all other electrodes are measured relative to this single point. It is typical to use a reference electrode placed somewhere along the scalp midline, or a reference that links both earlobe electrodes.

- Average reference derivation: The outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each amplifier.

- Bipolar derivation: The electrodes are connected in series to an equal number of amplifiers. For example, amplifier 1 measures the difference between electrodes A and B, amplifier 2 measures the difference between B and C, and so on.

EEG has several limitations. Scalp EEG cannot pick out individual action potentials, the electric unit of signaling in the brain. Instead, the EEG picks up synchronization of neurons, which produces a greater voltage than the firing of an individual neuron. Secondly, EEG has limited anatomical specificity when compared with other functional brain imaging techniques such as functional magnetic resonance imaging (fMRI). Some anatomical specificity can be gained by using a large number of electrodes to triangulate the source of the electrical activity.

A.2 EEG rhythms

Historically five major types of continuous rhythmic sinusoidal EEG waves are recognized (gamma, beta, alpha, delta and theta). There is no precise agreement on the frequency range for each type.

- Gamma is approximately the frequency range 30-80 Hz. Gamma rhythms appear to be involved in higher mental activity, including perception and consciousness. E.g., it disappears with general anesthesia.

- Beta is the frequency range 12 Hz and 30 Hz approximately. Disorganized, low amplitude beta is often associated with active, busy or anxious thinking and active concentration. Rhythmic beta is associated with various pathologies and drug effects.

- Alpha (Berger's wave) is the frequency range from 8.5 Hz to 12 Hz. It is characteristic of a relaxed, alert state of consciousness and is present by the age of two years. Alpha rhythms are best detected with the eyes closed. Alpha attenuates with drowsiness and open eyes, and is best seen over the occipital (visual) cortex. An alpha-like normal variant called mu is sometimes seen over the motor cortex (central scalp) and attenuates with movement, imagined movement, or with the intention to move.

- Theta is the frequency range from 4.5 Hz to 8 Hz and is associated with drowsiness, childhood, adolescence and young adulthood. This EEG frequency can sometimes be produced by hyperventilation. Theta waves can be seen during hypnagogic states such as trances, hypnosis, deep day dreams, lucid dreaming and light sleep and the preconscious state just upon waking, and just before falling asleep. Controlled meditation or yogic meditation and/or breathing also produces theta waves.

- Delta is the frequency range up to 4 Hz and is often associated with the very young and certain encephalopathies and underlying lesions. It is seen in deep sleep.

In addition to the above types of rhythmic activity, individual transient waveforms such as sharp waves, spikes, spike-and-wave complexes occur in epilepsy, and other types of transients occur during sleep.

A.3 Entraining brain rhythms

There is some debate over whether one can entrain one's own brain to achieve a specific brainwave frequency by using various techniques, including visual and auditory stimuli, and neurofeedback.

A.3.1 Rhythmic stimuli

One method proposed for attempting to entrain EEG rhythms is through sensory prompting. Medical EEG testing commonly includes intense light stimulation at various frequencies (to test for seizures). Some research indicates that software that generates various frequencies is able to stimulate the brain and alter the brain's frequency via a subtle audio or visual stimulus. Other research counters these claims. Probably the best summary at this time is that entrainment, or "following" does occur in many but not all individuals, and that the result may have a transitory influence on the dominant subjective state reported by that person.

A.3.2 Neurofeedback

A particularly controversial method proposed for "training the brain", or training for improved function, is through neurofeedback, a process in which the subject is given real-time biofeedback of their EEG rhythms. This technique has been proposed for the treatment of epilepsy, Attention Deficit Hyperactive Disorder(ADHD), and other diseases and disorders. Neurofeedback during clinical trials has attempted to remediate some mental problems, by conducting an assessment and then attempting to train specific changes in order to "normalize" the EEG. This sometimes results in improved function, although any benefit has often been attributed to the placebo effect. Some manufacturers and researchers have proposed that the operant conditioning of EEG may produce lasting positive functional changes (Current bibliography: `http://www.isnr.org/nfbarch/nbiblio.htm`). However, many ADHD researchers, most notably Barkley, state that it is not true. Barkley claims that there is no evidence that neurofeedback works in any way. He, and some in the neurology research community, maintain that the supporting research is sloppy and, to date, has not managed to successfully prove anything.

Measuring the brain's activity is a standard procedure in medicine, and can give interesting indicators of how the brain works. Whether EEGs can be used to balance the cerebral hemispheres, induce relaxation, improve mental functioning, or change or benefit the brain in any way is still inconclusive and can at best be considered controversial. Although brain training companies are likely to quote some studies that show some benefits, the overall view in neuroscience through meta analyzes and thorough reviews of research is that electronic brain training is unreliable. A great deal of further research is needed before any kind of brain training can be regarded as reliable treatments. In the meantime, far more reliable methods of treatment have been successful in helping patients.

Bibliography

[1] `http://www.robots.ox.ac.uk/~parg/projects/bci/rev4.html`.

[2] `http://www.boosting.org`.

[3] December 2004. Milwaukee Journal Sentinel.

[4] H. Akaike. A new look at statistical model identification. *IEEE Trans Automat Contr*, 19:716–723, 1974.

[5] C. Allefeld and J. Kurths. An approach to multivariate phase synchronization analysis and its application to event-related potentials. *Int J Bif Chaos*, 14(2):417–426, 2004.

[6] C.W. Anderson. Effects of variations in neural network topology and output averaging on the discrimination of mental tasks from spontaneous electroencephalogram. *Journal of Intelligent Systems*, 7(1-2):165–190, 1997.

[7] C.W. Anderson and M.J. Kirby. EEG subspace representations and feature selection for Brain-Computer Interfaces. In *Proc 1st IEEE Workshop on Computer Vision and Pattern Recognition for Human Computer Interaction, Madison, Wisconsin*, June 2003.

[8] C.W. Anderson, E.A. Stolz, and S. Shamsunder. Discriminating mental tasks using EEG represented by AR models. In *Proc of the 1995 IEEE Eng Med Biol Annual Conf*, pages on CD–ROM, September 1995.

[9] R. Beisteiner, P. Höllinger, G. Lindinger, W. Lang, and A. Berthoz. Mental representations of movements. brain potentials associated with imagination of hand movements. *Electroenceph. Clin. Neurophysiol.*, 96:183–193, 1995.

[10] J. Bhattacharya. Reduced degree of long-range phase synchrony in pathological human brain. *Acta Neurobiol Exp*, 61:309–318, 2001.

[11] J. Bhattacharya and H. Petsche. Shadows of artistry: cortical synchrony during perception and imagery of visual art. *Cognitive Brain Research*, 13:179–186, 2002.

[12] J. Bhattacharya, H. Petsche, U. Feldmann, and B. Rescher. EEG gamma-band phase synchronization between posterior and frontal cortex during mental rotation in humans. *Neuroscience Letters*, 311:29–32, 2001.

[13] N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor. The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans Rehab Eng*, 8(2):190–193, June 2000.

[14] G.E. Birch and S.G. Mason. Brain-computer interface research at the Neil Squire Foundation. *IEEE Trans Rehab Eng*, 8(2):193–195, June 2000.

[15] B. Blankertz, G. Curio, and K.R. Müller. Classifying single trial EEG: Towards Brain Computer Interfacing. In T.G. Diettrich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems (NIPS 01)*, volume 14, pages 157–164, 2002.

[16] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.R. Müller, V. Kunzmann, F. Losch, and G. Curio. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Trans Neur Syst Rehab Eng*, 11(2), 2003.

[17] B. Blankertz, K.-R. Müller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer. The BCI Competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Trans Biomed Eng*, 51(6):1044–1051, June 2004.

[18] A.L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artif. Intell. (Special Issue on Relevance)*, 97:245–271, 1997.

[19] B. Boashash. Estimating and interpreting the instantaneous frequency of a signal - part 1: Fundamentals. *Proceedings of the IEEE*, 80(4):520–538, April 1992.

[20] R. Boostani and M.H. Moradi. A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier. *J. Neural Eng.*, 1:212–217, 2004.

[21] J.F. Borisoff, S.G. Mason, A. Bashashati, and G. Birch. Brain-Computer Interface design for asynchronous control applications: Improvements to the LF-ASD asynchronous brain switch. *IEEE Trans Biomed Eng*, 51(6):985–992, June 2004.

[22] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[23] C. Carmeli, M.G. Knyazeva, G. Innocenti, and O. De Feo. Assessment of EEG synchronization based on state-space analysis. *NeuroImage*, 25:339–354, 2005.

[24] R. Cawley and G.-H. Hsu. Local-geometric-projection method for noise reduction in chaotic maps and flows. *Physical Review A*, 46(6), September 1992.

[25] P. Celka. Statistical analysis of the Phase Locking Value. Private Communication.

[26] P. Celka, B. Boashash, and P. Colditz. Preprocessing and time-frequency analysis of newborn eeg seizures. *IEEE EMBS Magazine*, 20:30–39, 2001.

[27] M. Chávez, M. Le Van Quyen, V. Navarro, M. Baulac, and J. Martinerie. Spatio-temporal dynamics prior to neocortical seizures: Amplitude versus phase couplings. *IEEE Trans Biomed Eng*, 50(5):571–583, May 2003.

[28] A.J. Chipperfield, P.J. Fleming, and C.M. Fonseca. Genetic algorithm tools for control systems engineering. In *Adaptive Computing in Engineering Design and Control*, 1994.

[29] L. Citi, R. Poli, C. Cinel, and F. Sepulveda. Feature selection and classification in Brain Computer Interfaces by a Genetic Algorithm. In *Genetic and Evolutionary Computation Conference (GECCO)*, June 2004.

[30] W. De Clercq. *Advanced preprocessing techniques and nonlinear signal analysis applied to scalp electroencephalograms for the prediction of epileptic seizures.* PhD thesis, Faculty of Engineering, K.U.Leuven, Leuven, Belgium, 2005.

[31] S. Coyle, T. Ward, and C. Markham. An Optical Brain Computer Interface. In *Proc of the 2nd Int Brain-Computer Interface Workshop and Training Course 2004*, pages 45–46, September 2004. Published as supplementary volume of "Biomedizinische Technik".

[32] C. Croux and K. Joossens. Empirical comparison of the classification performance of robust linear and quadratic discriminant analysis. *Birkhäuser Verlag Basel, Switzerland*, pages 131–140, 2005.

[33] E. Curran, P. Sykacek, S. Roberts, W. Penny, M. Stokes, I. Johnsrude, and A. Owen. Cognitive tasks for driving a brain computer interfacing system: A pilot study. *IEEE Trans Neur Syst Rehab Eng*, 12(1):48–54, 2004.

[34] E.A. Curran and M. Stokes. Learning to control brain activity: A review of the production and control of eeg components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, 51:326–336, 2003.

[35] O. David, D. Cosmelli, J.-P. Lachaux, S. Baillet, L. Garnero, and J. Martinerie. A theoretical and experimental introduction to the non-invasive study of large-scale neural phase synchronization in human beings. *Int. J. Computational Cognition*, 1(4):53–77, December 2003.

[36] J. del R. Millán, M. Franzé, J. Mouriño, F. Cincotti, and F. Babiloni. Relevant EEG features for the classification of spontaneous motor-related tasks. *Biol. Cybern.*, pages 89–95, 2002.

[37] J. del R. Millán, J. Mouriño, F. Babiloni, F. Cincotti, M. Varsta, and J. Heikkonen. Local neural classifier for EEG-based recognition of mental tasks. volume 3, pages 3632–3636. IEEE-INNS-ENNS Int Joint Conf on Neural Networks, July 2000.

[38] J. del R. Millán, F. Renkens, J. Mouriño, and W. Gerstner. Non-invasive brain-actuated control of a mobile robot by human EEG. *IEEE Trans Biomed Eng*, 51(6):1026–1033, June 2004.

[39] A. Delorme and S. Makeig. EEG changes accompanying learning regulation of the 12-hz EEG activity. *IEEE Trans Rehab Eng*, 11(2):133–136, 2003.

[40] E.M. Dewan. Occipital alpha rhythm eye position and lens accomodation. *Nature*, 214:975–977, June 1967.

[41] J.P. Donoghue. Connecting cortex to machines: recent advances in brain interfaces. *Nat neurosci. supplement*, 5:1085–1088, November 2002.

[42] G. Dornhege, B. Blankertz, and G. Curio. Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials. In *Proc 1st Int IEEE EMBS Conf on Neural Engineering*, pages 591–594, 2003.

[43] G. Dornhege, B. Blankertz, G. Curio, and K.R. Müller. Combining features for BCI. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS 02)*, volume 15. MIT Press: Cambridge, 2003.

[44] G. Dornhege, B. Blankertz, G. Curio, and K.R. Müller. Increase information transfer rates in bci by csp extension to multi-class. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems (NIPS 03)*, volume 16. MIT Press: Cambridge, 2004.

[45] T. Ebrahimi, J.M. Vesin, and G. Garcia. Brain-Computer Interface in multimedia communication. *IEEE Signal Proc Mag*, 20(1):14–24, January 2003.

[46] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroenceph. Clin. Neurophysiol.*, 70:510–523, 1988.

[47] J. Fell, P. Klaver, K. Lehnertz, T. Grunwald, C. Schaller, C.E. Elger, and G. Fernandez. Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nat Neurosci.*, 2001.

[48] P. Flandrin, G. Rilling, and P. Gonçalvès. Empirical mode decomposition as a filter bank. *IEEE Sig. Proc. Letters*, 11(2):112–114, 2004.

[49] K.R. Foster, P.R. Wolpe, and A.L. Caplan. Bioethics & the brain. *IEEE Spectrum*, pages 34–39, June 2003.

[50] G. Fung and O.L. Mangasarian. Data selection for support vector machine classifiers. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 64–70. ACM Press, New York, USA, 2000.

[51] G.N. Garcia, T. Ebrahimi, and J.M. Vesin. Classification of EEG signals in the ambiguity domain for Brain Computer Interface applications. In *Proc of the 14th Int Conf in Digital Signal Processing*, 2002.

[52] G.N. Garcia, T. Ebrahimi, and J.M. Vesin. Support vector EEG classification in the Fourier and time-frequency correlation domains. In *Proc of the IEEE-EMBS first Int Conf on Neural Engineering*, 2003.

[53] G.N. Garcia, U. Hoffmann, T. Ebrahimi, and J.-M. Vesin. Direct brain-computer communication through EEG signals. IEEE EMBS Book Series on Neural Engineering, 2004.

[54] C. Gerloff, J. Richard, J. Hadley, A.E. Schulman, M. Honda, and M. Hallett. Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements. *Brain*, 121:1513–1531, 1998.

[55] A.S. Gevins, C.L. Yeager, G.M. Zeitlin, S. Ancoli, and M.F. Dedon. On-line computer rejection of EEG artifact. *Electroenceph. Clin. Neurophysiol.*, 42:267–274, 1977.

[56] C. Guger, H. Ramoser, and G. Pfurtscheller. Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface. *IEEE Trans Rehab Eng*, 8(4):447–56, December 2000.

[57] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Machine Learning Research*, (3):1157–1182, 2003.

[58] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *J. Machine Learning Research*, (3):1439–1461, 2003.

[59] E. Gysels and P. Celka. Phase synchronization for the recognition of mental tasks in a Brain Computer Interface. *IEEE Trans Neur Syst Rehab Eng*, 12(4):406–415, December 2004.

[60] E. Gysels and P. Celka. Selection of synchronization and power features in Brain Machine Interfaces. In *Proc of the 2nd Int Brain-Computer Interface Workshop and Training Course 2004*, pages 51–52, September 2004. Published as supplementary volume of "Biomedizinische Technik".

[61] E. Gysels, J. del R. Millán, S. Chiappa, and P. Celka. Studying phase synchrony for classification of mental tasks in brain machine interfaces. In *Proc of the 14th World Congress of the Int Society for Brain Electromagnetic Topography*, November 2003. Abstract, appeared in Brain Topography.

[62] E. Gysels, P. Desomere, K. Vonck, R. Van de Walle, P. Boon, and I. Lemahieu. Measuring synchrony and similarity for anticipating epileptic seizures: a case study. *Acta Neurologica Belgica*, 102:194–195, 2002. Abstract.

[63] E. Gysels and M. Knyazeva. A strategy for EEG synchronization analysis. In *Proc of the 16th World Congress of the Int Society for Brain Electromagnetic Topography*, volume 18, 2005. Abstract, appeared in Brain Topography.

[64] E. Gysels, M. Le Van Quyen, J. Martinerie, P. Boon, K. Vonck, I. Lemahieu, and R. Van de Walle. Long-term evaluation of synchronization between scalp EEG signals in partial epilepsy. volume 3, pages 1495–1498. 9th Int Conf on Neural Information Processing, November 2002.

[65] E. Gysels, P. Renevey, and P. Celka. Fast feature selection to compare broadband with narrowband phase synchronization in Brain-Computer Interfaces. In *Proc of the 5th International Workshop on Biosignal Interpretation BSI 2005*, September 2005.

[66] E. Gysels, P. Renevey, and P. Celka. SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces. *Signal Processing, Special Issue on Neuronal coordination in the brain: A Signal Processing Perspective*, 85:2178–2189, 2005.

[67] B. Hjorth. EEG analysis based on time domain properties. *Electroenceph. Clin. Neurophysiol.*, 29:306–310, 1970.

[68] U. Hoffmann, G. Garcia, J.-M. Vesin, and T. Ebrahimi. Application of the evidence framework to Brain-Computer Interfaces. In *Proc of the Int Conf IEEE Eng. in Med. and Biol. Society (EMBS) 2004*, September.

[69] M. Hoke, K. Lehnertz, C. Pantev, and B. Lütkenhï¿½er. *Brain Dynamics, Progress and Perspectives*, chapter Spatiotemporal Aspects of Synergetic Processes in the Auditory Cortex as Revealed by the Magnetoencephalogram. Springer Series in Brain Dynamics. Springer-Verlag, 1989.

[70] A. Hyvärinen and J. Karhunen E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[71] H.H. Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr Clin neurophysiol*, 10:371–375, 1958.

[72] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.

[73] T.-P. Jung, S. Makeig, C. Humphries, T.W. Lee, M.J. McKeown, V. Iragui, and T.J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000.

[74] M.J. Katz. Fractals and the analysis of waveforms. *Comput Biol Med.*, 18(3):145–156, 1988.

[75] Z.A. Keirn and J.I. Aunon. A new mode of communication between man and his surroundings. *IEEE Trans Biomed Eng*, 37(12):1209–1214, December 1990.

[76] P. Kennedy, D. Andreasen, P. Ehirim, B. King, T. Kirby, H. Mao, and M. Moore. Using human extra-cortical local field potentials to control a switch. *J. Neural Eng.*, 1:72–77, 2004.

[77] M.G. Knyazeva, E. Fornari, R. Meuli, G. Innocenti, and P. Maeder. Imaging of a synchronous neuronal assembly in the human visual brain. *Neuroimage*, September 2005.

[78] D. Koller and M. Sahami. Toward optimal feature selection. In *Proc. of the 13th Int Conf on Machine Learning*, pages 284–292, 1996.

[79] I. Kononenko, M. Robnik-Sikonja, and U. Pompe. ReliefF for estimation and discretization of attributes in classification, regression and ILP problems. In A. Ramsay, editor, *Artificial Intelligence: Methodology, Systems, Applications: Proceedings of AIMSA'96*, pages 31–40. IOS Press, 1996.

[80] A. Kostov and M. Polak. Prospects of computer access using voluntary modulated EEG signal. In Lj. Rakić, G. Kostopoulos, D. Raković, and Dj. Koruga, editors, *Brain and Consciousness*, pages 233–236. ECPD Symposium, September 1997.

[81] A. Kostov and M. Polak. Parallel man-machine training in development of EEG-based cursor control. *IEEE Trans Rehab Eng*, 8(2):203–205, June 2000.

[82] R. Krepki, B. Blankertz, G. Curio, and K.R. Müller. The Berlin Brain-Computer Interface (BBCI): Towards a new communication channel for online control of multimedia applications and computer games. 9th Int Conf on Distributed Multimedia Systems (DMS'03), 2003.

[83] W.N. Kuhlman. Functional topography of the human mu rhythm. *Electroenceph. Clin. Neurophysiol.*, 44:83–93, 1978.

[84] J.P. Lachaux, E. Rodriguez, M. Le Van Quyen, A. Lutz, J. Martinerie, and F.J. Varela. Studying single-trials of phase synchronous activity in the brain. *Int J Bif Chaos*, 10(10):2429–2439, 2000.

[85] J.P. Lachaux, E. Rodriguez, J. Martinerie, and F.J. Varela. Measuring phase synchrony in brain signals. *Hum Br Map*, 8:194–208, 1999.

[86] T.N. Lal, T. Hinterberger, G. Widman, M. Schröder, N.J. Hill, W. Rosenstiel, C.E. Elger, B. Schölkopf, and N. Birbaumer. Methods towards invasive human brain computer interfaces. In L.K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems (NIPS 04)*, volume 17. MIT Press: Cambridge, 2005.

[87] T.N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf. Support vector channel selection in BCI. *IEEE Trans Biomed Eng*, 51(6):1003–1010, June 2004.

[88] M. Laubach, J. Wessberg, and M.A.L. Nicolelis. Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. *Nature*, 405:567–571, June 2000.

[89] M. Le Van Quyen, J. Foucher, J.P. Lachaux, E. Rodriguez, A. Lutz, J. Martinerie, and F.J. Varela. Comparison of hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J Neurosci Methods*, 111:83–98, 2001.

[90] M. Le Van Quyen, J. Martinerie, V. Navarro, M. Baulac, and F.J. Varela. Characterizing neurodynamic changes before seizures. *J. Clin. Neurophysiol.*, 18(3):191–208, 2001.

[91] K. Lehnertz and C.E. Elger. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy. *Electroenceph. clin. Neurophysiol.*, 95:108–117, 1995.

[92] E.C. Leuthardt, G. Schalk, J.R. Wolpaw, J.G. Ojemann, and D.W. Moran. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.*, 1:63–71, 2004.

[93] P.J. Loughlin. Do bounded signals have bounded amplitudes? *Multidimensional Systems and Signal Processing*, 9:419–424, 1998.

[94] S. Makeig, S. Enghoff, T.-P. Jung, and T.S. Sejnowski. A natural basis for efficient brain-actuated control. *IEEE Trans Rehab Eng*, 8(2):208–211, June 2000.

[95] B.A. Malow and M.S. Aldrich. Localizing value of rapid eye movement sleep in temporal lobe epilepsy. *Sleep Medicine*, 1:57–60, 2000.

[96] K. Mardia. *Probability and Mathematical Statistics: Statistics of Directional Data.* London: Academic, 1972.

[97] D.J. McFarland, L.M. McCane, S.V. David, and J.R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroenceph. Clin. Neurophysiol.*, 103:386–394, 1997.

[98] P. Meinicke, M. Kaper, F. Hoppe, M. Heumann, and H. Ritter. Improving transfer rates in Brain Computer Interfacing: a case study. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems (NIPS 02)*, volume 15. MIT Press: Cambridge, 2003.

[99] W.H.R. Miltner, C. Braun, M. Arnold, H. Witte, and E. Taub. Coherence of gamma-band EEG activity as a basis for associative learning. *Nature*, 397:434–436, February 1999.

[100] F. Mormann, T. Kreuz, R.G. Andrzejak, P. David, K. Lehnertz, and C.E. Elger. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Research*, 53:173–185, 2003.

[101] F. Mormann, K. Lehnertz, P. David, and C.E. Elger. Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D*, 144:358–369, 2000.

[102] K.-R. Müller, C.W. Anderson, and G.E. Birch. Linear and nonlinear methods for brain-computer interfaces. *IEEE Trans Neur Syst Rehab Eng*, 11(2):165–169, June 2003.

[103] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clin Neurophysiol.*, 110(5):787–798, May 1999.

[104] S. Musallam, B.D. Corneil, B. Greger ad H. Scherberger, and R.A. Andersen. Cognitive control signals for neural prosthetics. *Science*, 305, July 2004.

[105] M.A.L. Nicolelis. Actions from thoughts. *Nature*, 409:403–407, January 2001.

[106] M.A.L. Nicolelis and J.K. Chapin. Controlling robots with the mind. *Scientific American*, 2002.

[107] E. Niedermayer and F. Lopes Da Silva, editors. *Electroencephalography: Basic Principles, Clinical Applications and Related Fields.* Williams & Wilkins, 1999.

[108] P.L. Nunez, R. Srinivasan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, and P.J. Cadusch R.B. Silberstein. EEG coherency I: Statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales. *Electroenceph. Clin. Neurophysiol.*, 103(5):499–515, 1997.

[109] S. Ohara, T. Mima, K. Baba, A. Ikeda, T. Kunieda, R. Matsumoto, J. Yamamoto, M. Matsuhashi, T. Nagamine, K. Hirasawa, T. Hori, T. Mihara, N. Hashimoto, S. Salenius, and H. Shibasaki. Increased synchronization of cortical oscillatory activities between human supplementary motor and primary sensorimotor areas during voluntary movements. *J Neuroscience*, 21(23), December 2001.

[110] A.V. Oppenheim and R.W. Schafer. *Discrete-Time Signal Processing.* Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1999. Chapter 11.

[111] C.A. Peña-Reyes and M. Sipper. Evolutionary computation in medicine: an overview. *Artificial Intelligence in Medicine*, 19:1–23, 2000.

[112] W.D. Penny and S.J. Roberts. EEG-based communication via dynamic neural network models. In *Proc Int Joint Conf on Neural Networks*, 1999.

[113] W.D. Penny, S.J. Roberts, and M.J. Stokes. Imagined hand movements identified from the EEG mu-rhythm. Technical report, Department of Electrical Engineering, Imperial College, 1998.

[114] F. Perrin, J. Pernier, O. Bertrand, and J.F. Echallier. Spherical splines for scalp potential and current density mapping. *Electroenceph. Clin. Neurophysiol.*, 72:184–187, 1989.

[115] F. Perrin, J. Pernier, O. Bertrand, and J.F. Echallier. Corrigendum EEG 02274. *Electroenceph. Clin. Neurophysiol.*, 76:565, 1990.

[116] G. Pfurtscheller and C. Neuper. Motor imagery and direct Brain-Computer communication. *Proc of IEEE*, 89(7):1123–1134, July 2001.

[117] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schï¿½l, B. Obermaier, and M. Pregenzer. Current trands in Graz brain-computer interface (BCI) research. *IEEE Trans Rehab Eng*, 8(2), June 2000.

[118] J.R. Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers Inc., 1993.

[119] R.Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Physical Review E*, 65(041903), 2002.

[120] R. Rastogi and K. Shim. PUBLIC: A decision tree classifier that integrates building and pruning. *Data Mining and Knowledge Discovery*, 2000.

[121] F. Renkens and J. del R. Millán. Brain-actuated control of a mobile platform. In *Proc of the 7th Int Conf on Simulation of Adaptive Behavior*, 2002. Workshop Motor Control in Humans and Robots, Edingburgh.

[122] G. Rilling, P. Flandrin, and P. Gonçalvès. On empirical mode decomposition and its algorithms. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, 2003.

[123] J. Rissanen. *Stochastic complexity in statistical engineering*. World scientific, 1989.

[124] S.J. Roberts, W. Penny, and I.Rezek. Temporal and spatial complexity measures for EEG-based brain-computer interfacing. *Medical and Biological Engineering and Computing*, 37(1):93–99, 1998.

[125] S.J. Roberts and W.D. Penny. Real-time Brain-Computer Interfacing: a preliminary study using bayesian learning. *Medical and Biological Engineering and Computing*, 38(1):56–61, 2000.

[126] E. Rodriguez, N. George, J.P. Lachaux, J. Martinerie, B. Renault, and F.J. Varela. Perception's shadow: Long-distance synchronization of human brain activity. *Nature*, 397:430–433, February 1999.

[127] M. Rosenblum, A. Pikovsky, J. Kurths, C. Schï¿½er, and P. Tass. *Handbook of Biological Physics*, volume 4, chapter Phase synchronization: from theory to data analysis, pages 279–321. Elsevier Science, 2001.

[128] S. Ruggieri. Efficient C4.5. *IEEE Transactions on Knowledge and Data Engineering*, 14(2):438–444, March 2002.

[129] J. Sarnthein, H. Petsche, P. Rappelsberger, G.L. Shaw, and A. von Stein. Synchronization between prefrontal and posterior association cortex during human working memory. *Proc Natl Acad Sci USA*, 95:7092–7096, 1998.

[130] B. Schack, P. Rappelsberger, C. Anders, S. Weiss, and E. Möller. Quantification of synchronization processes by coherence and phase and its application in analysis of electrophysiological signals. *Int J Bif Chaos*, 10(11):2565–2586, 2000.

[131] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. Bci2000: a general-purpose Brain-Computer Interface. *IEEE Trans Biomed Eng*, 51(6):1034–1043, June 2004.

[132] R. Scherer, G.R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller. An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate. *IEEE Trans Biomed Eng*, 51(6):979–984, June 2004.

[133] A. Schlögl, P. Anderer, M.-J. Barbanoj, G. Klösch, G. Gruber, J.L. Lorenzo, O. Filz, M. Koivuluoma, I. Rezek, S.J. Roberts, A. Värri, P. Rappelsberger, G. Pfurtscheller, and G. Dorffner. Artefact processing of the sleep EEG in the 'SIESTA' project. *Proceedings of the EMBEC'99*, 1:1644–1645, 1999.

[134] A. Schlögl, G. Pfurtscheller, and B. Schack. Single-trial EEG analysis using an adaptive autoregressive model. In *Proc of 4th Int Symp on Central Nervous Monitoring*, September 1996.

[135] M. Schröder, M. Bogdan, W. Rosenstiel, T. Hinterberger, and N. Birbaumer. Automated EEG feature selection for Brain Computer Interfaces. In *Proc of the 1st IntIEEE EMBS Conf on Neural Engineering*, pages 626–629, 2003.

[136] E. Sommerfeld, A. Hensel, and A. Hildebrandt. Cooperation of frontal and parietal brain areas as a function of cognitive training.

[137] L. Song. Desynchronization network analysis for the recognition of imagined movement. In *Proc. 27th IEEE EMBS Annual International Conference*, September 2005.

[138] L. Song, E. Gysels, and E. Gordon. Phase synchrony rates for the recognition of motor imageries in BCIs. In *Advances in Neural Information Processing Systems (NIPS 05)*, volume 18. MIT Press: Cambridge, 2006.

[139] C.J. Stam, D.L.J. Tavy, and R.W.M. Keunen. Quantification of alpha rhythm desynchronization using the acceleration spectrum entropy of the EEG. *Clinical Electroencephalography*, 24(3):104–109, 1993.

[140] E.E. Sutter. The brain response interface: communication through visually induced electrical brain responses. *J. of Microcomputer Applications*, 15:31–45, 1992.

[141] P. Sykacek, S. Roberts, M. Stokes, E. Curran, M. Gibbs, and L. Pickup. Probabilistic methods in BCI research. *IEEE Trans Neur Syst Rehab Eng*, 11(2):192–195, June 2003.

[142] D.M. Taylor, S.I.H. Tillery, and A.B Schwartz. Direct cortical control of 3D neuroprosthetic devices. *Science*, 296:1829–1832, June 2002.

[143] K. Torkkola and W.M. Campbell. Mutual information in learning feature transformations. In *Proc. 17th International Conf. on Machine Learning*, pages 1015–1022. Morgan Kaufmann, San Francisco, CA, 2000.

[144] M.J.A.M van Putten. Nearest neighbor phase synchronization as a measure to detect seizure activity from scalp EEG recordings. *J. Clin. Neurophysiol.*, 20(5):320–325, 2003.

[145] M.J.A.M. van Putten. Proposed link rates in the human brain. *J Neurosci Methods*, 127:1–10, 2003.

[146] F.J. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2:229–239, 2001.

[147] T.M. Vaughan, W.J. Heetderks, L.J. Trejo, W.Z. Rymer, M. Weinrich, M.M. Moore, A. Kübler, B.H. Dobkin, N. Birbaumer, E. Donchin, E.W. Wolpaw, and J.R. Wolpaw. Brain-computer interface technology: a review of the second international meeting. *IEEE Trans Neur Syst Rehab Eng*, 11(2):94–109, June 2003.

[148] R. Vautard, P. Yiou, and M. Ghil. Singular-spectrum analysis: a toolkit for short, noisy chaotic signals. *Physica D*, 58:95–126, 1992.

[149] R. Vetter, J.-M. Vesin, P. Celka, P. Renevey, and J. Krauss. Automatic nonlinear noise reduction using local principal component analysis and MDL parameter selection. In *IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, 2002.

[150] J.J. Vidal. Real-time detection of brain events in EEG. *Proc. IEEE*, 65:633–664, May 1977.

[151] T. Wang and B. He. An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a Brain-Computer Interface. *J. Neural Eng.*, 1:1–7, 2004.

[152] N. Weiskopf, K. Mathiak, S.W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N. Birbaumer. Principles of a Brain-Computer Interface (BCI) based on real-time functional Magnnetic Resonance Imaging (fMRI). *IEEE Trans Biomed Eng*, 51(6), June 2004.

[153] J. Wessberg, C.R. Stambaugh, J.D. Kralik, P.D. Beck, M. Laubach, J.K. Chapin, J. Kim, S.J. Biggs, M. A. Srinivasan, and M.A.L. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365, November 2000.

[154] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik. Feature selection for SVMs. In S. Solla, T. Leen, and K.-R.Müller, editors, *Advances in Neural Information Processing Systems (NIPS 00)*, volume 13. MIT Press: Cambridge, 2001.

[155] I. Wickelgren. Tapping the mind. *Science*, 299:496–499, 2003.

[156] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan. Brain-computer interface technology: A review of the first international meeting. *IEEE Trans Rehab Eng*, 8(2):164–173, 2000.

[157] J.R. Wolpaw, D.J. McFarland, and T.M. Vaughan. Brain-computer interface research at the Wadsworth Center. *IEEE Trans Rehab Eng*, 8(2):222–226, June 2000.

[158] E. Yom-Tov and G.F. Inbar. Feature selection for the classification of movements from single movement-related potentials. *IEEE Trans Neur Syst Rehab Eng*, 10(3):170–177, September 2002.

[159] L. Yu and H. Liu. Efficient feature selection via analysis of relevance and redundancy. *J. Machine Learning Research*, 5:1205–1224, 2004.

Curriculum Vitae

Personal Data

Name: Elly Gysels
Date of Birth: 23.04.1978
Place of Birth: Aalst, Belgium

Education

- 2001: Graduation Physics Engineering with great distinction,
- 2000-2001: 5th year Physics Engineering studies, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland,
- 1996-2000: First 4 years Physics Engineering studies, Ghent University, Ghent, Belgium,
- 1990-1996: Secondary Education, Denderleeuw, Belgium. Specialization: Science-Mathematics.

Experience

- 2002-2005: PhD student at CSEM (Swiss Center for Electronics and Microtechnology),
- 2001-2002: Scientific collaborator at the Department of Electronics and Information Systems (ELIS), Ghent University, Belgium,
- 2000-2001: Diploma work on Scanning Optical Coherence Tomography (OCT) at the Biomedical Photonics Group (BPG), Institute of Applied Optics (IOA), Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland,
- 2000: Traineeship at the institute for Materials Science, Welding and Forming, Graz, Austria,
- 1999-2000: President of VVN, de Vereniging Voor Natuurkunde (association for physics students) at Ghent University, Belgium,
- 1998-2000: Representative of the physics engineering students in the educational commission for physics engineering at Ghent University, Belgium,
- Computer:
  - Good working experience with both Linux and Windows operating systems,
  - Good knowledge of Pascal, MATLAB, LabVIEW, LaTeX,
  - Basic knowledge of Mathcad and P-Spice.

Language skills:

- Dutch: Mother tongue,
- English: Fluent, both expression and comprehension,
- French: Fluent, both expression and comprehension,
- German: Fair expression, good comprehension,
- Swiss German: Fair expression, good comprehension.

List of Publications

International Journals

1. E. Gysels, P. Renevey and P. Celka. SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces. In *Signal Processing, Special Issue on Neuronal coordination in the brain: A Signal Processing Perspective,* 85: 2178-2189, 2005.

2. E. Gysels and P. Celka. Phase Synchronization for the Recognition of Mental Tasks in a Brain Computer Interface. In *IEEE Trans Neur Syst Rehab Eng,* 12(4):406-415, 2004.

International Conferences and Workshops

1. L. Song, E. Gysels and E. Gordon. Phase Synchrony Rates for the Recognition of Motor Imageries in BCIs. In *Advances in Neural Information Processing Systems (NIPS 05),* 18, 2006.

2. E. Gysels and M. Knyazeva. A Strategy for EEG synchronization analysis. In *Proc of the 16th World Congress of the Int Society for Brain Electromagnetic Topography,* 2005. Oral presentation.

3. E. Gysels, P. Renevey and P. Celka. Fast Feature Selection to Compare Broadband with Narrowband Phase Synchronization in Brain-Computer Interfaces. In *Proc of the 5th International Workshop on Biosignal Interpretation BSI 2005,* 2005.

4. E. Gysels and P. Celka. Selection of synchronization and power features in Brain Machine Interfaces. In *Proc of the 2nd Int Brain-Computer Interface Workshop and Training Course 2004,* 51-52, 2004. Published as supplementary volume of "Biomedizinische Technik".

5. E. Gysels, J. del R. Millán, S. Chiappa, P. Celka. Studying Phase Synchrony for Classification of Mental Tasks in Brain Machine Interfaces. In *Proc of the 14th World Congress of the Int Society for Brain Electromagnetic Topography,* 2003. Abstract, appeared in Brain Topography.

6. E. Gysels, M. Le Van Quyen, J. Martinerie, P. Boon, K. Vonck, I. Lemahieu, R. Van de Walle. Long-Term Evaluation of Synchronization between Scalp EEG Signals in Partial Epilepsy. In *Proc 9th Int Conf on Neural Information Processing,* 1495-1498, 2002.

Local Conferences and Journals

1. E. Gysels, P. Desomere, K. Vonck, R. Van De Walle, P. Boon and I. Lemahieu. Measuring synchrony and similarity for anticipating epileptic seizures: a case study. In *Acta Neurologica Belgica,* 102:194-195, 2002. Abstract.