An HMM-PCA Approach for EEG-Based Brain Computer Interfaces (BCIs)

by

Ali Özgür Argunşah

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabanci University
March 2010

The thesis of Ali Özgür Argunşah is approved.

Assist. Prof. Dr. Hakan Erdoğan

Assoc. Prof. Dr. Berrin Yanıkoğlu

Dr. Devrim Ünay

Assist. Prof. Dr. Baran Çürüklü, Committee Co-Chair

Assist. Prof. Dr. Müjdat Çetin, Committee Chair

Sabanci University

2010

to my family

Acknowledgments

First of all, I wholeheartedly wish to thank my advisor, Assist. Prof. Dr. Müjdat Çetin, whose endless guidance and support made this thesis possible. It has been a wonderful experience to work with him, sharing his angle of scientific view and enthusiasm as well as his curiosity. He continuously has tried to make every step of my graduate life as simple and meaningful as possible.

I owe my deepest gratitude to Prof. Dr. Aytül Erçil who made me a part of a respectable research group VPALAB in Sabancı University. Without her encouragement and efforts, I would not have been able to come to the place i am at today.

I would also like to thank my co-advisor Assist. Prof. Dr. Baran Çürüklü. Starting from his very first day of joining VPALAB at Sabancı University, his ideas and personal experiences helped broaden our perception of science and nature.

I am grateful to Dr. Devrim Ünay, Assist. Prof. Dr. Hakan Erdoğan and Assoc. Prof. Dr. Berrin Yanıkoğlu for their participation in my thesis committee. Their constructive comments and feedback brought the thesis one step further.

My family has always believed in me and supported me though my whole life. For that, I am more than grateful to them. Back in Ankara, where i grew up, I have shared numerous deep discussions with my dearest friends, Mehmet Atay and Tarık Sivrikaya. These discussions and decisions that followed have been one of the important corner stones of my academic career. I also want to thank my friends at VPALAB who have not only supported me but also put up with me at times during my years in İstanbul. I will never forget our endless, though utopic at times, still amazingly fun conversations with Serhan Coşar, Batu Akan, Erkin Tekeli, Özge Batu, Harun Karabalkan, Saygın Topkaya, Berkay Topçu and Emrecan Çökelek.

I should also thank Gözde Şenocak and Pedro Garcia da Silva for reading the manuscript and Esen Sokullu for her endless support during my whole academic life.

I am also thankful to Turkish Scientific and Technological Research Council (TÜBİTAK) and Turkish Academy of Sciences (TÜBA) for their support during my graduate education.

Last but not least, I want to thank Prof. Dr. Emery N. Brown and Dr. Sydney Cash for giving me the opportunity of collaborating with their research groups in Massachusetts General Hospital, Charlestown, MA, US and Prof. Dr. Antonio Carlos Roque for accepting me to Latin American School on Computational Neuroscience (LASCON) in Ribeirao Preto, SP, Brazil.

An HMM-PCA Approach for EEG-Based Brain Computer Interfaces (BCIs)

Ali Özgür Argunşah

EECS, M.S. Thesis, 2010

Assist. Prof. Dr. Müjdat Çetin, Supervisor

Assist. Prof. Dr. Baran Çürüklü, Co-Supervisor

Keywords: electroencephalogram, brain computer interface, sensorimotor rhythms, hidden Markov models, principal component analysis

Abstract

Electroencephalography (EEG) based Brain-Computer Interface (BCI) systems are a new development in the field of applied neurophysiology. This new approach has been made possible thanks to progress in EEG analysis and in information technology which has led to a better understanding of psychophysical aspects of the EEG signals. BCI systems enable information flow from the brain directly to the outside world. For widespread use of brain signals for such objectives, effective signal analysis and pattern recognition techniques are needed.

In this thesis, we have developed a new technique based on hidden Markov models, and have demonstrated the effectiveness of our algorithms both on a standard dataset and on the data that we have collected in our laboratory.

We have used HMMs with AR features combined with PCA to classify two and four class single trial EEG data recorded during imagination of motor actions

type of BCI experiments. Results were compared with previous HMM based BCI classifiers and Mahalanobis distance classifier fed with two different state-of-the-art EEG features.

EEG Tabanlı Beyin Bilgisayar Arayüzleri için HMM-PCA Yaklaşımı

Ali Özgür Argunşah

EECS Yüksek Lisans Tezi, 2010

Asist. Prof. Dr. Müjdat Çetin, Tez Danışmanı

Asist. Prof. Dr. Baran Çürüklü, Tez Yardımcı Danışmanı

Anahtar Kelimeler: elektroensefalogram, beyin bilgisayar arayüzleri, saklı Markov modelleri, temel bileşen analizi

Özet

Elektroensefalografi (EEG) tabanlı Beyin-Bilgisayar arayüzü (BBA) sistemleri uygulamalı nörofizyoloji alanındaki yeni gelişmelerden biridir. Bu sistemler EEG analiz yöntemleri ve bilgi teknolojileri alanındaki gelişmeler ile EEG sinyalinin psikofiziksel özelliklerinin daha iyi anlaşılmasıyla mümkün hale gelmektedirler. BBA sistemleri beyinden dış dünyaya, çevresel sinir sistemini kullanmadan, doğrudan bir bilgi akışı sağlamayı amaçlar. Bu amaca ulaşılabilmesi için etkili sinyal işleme ve örüntü tanıma tekniklerine gerek vardır.

Bu tezde, saklı Markov modelleri (HMM) üzerine kurulu bir yaklaşım önerdik ve yaklaşımımızın etkinliğini genel kullanıma açık bir veri kümesi ve kendi laboratuvarımızda topladığımız veriler üzerindeki deneysel sonuçlar ile gösterdik.

Hareketin zihinde canlandırılması deneylerinden elde edilen iki ve dört sınıflı EEG verilerden kestirilen özbağlanımlı parametrelere (AR) dayalı öznitelikleri temel bileşen analizi (PCA) tabanlı boyut indirgeme ile birlikte kullandık ve elde ettiğimiz boyutu indirgenmiş öznitelikleri HMMler kullanarak sınıflandırdık. Elde ettiğimiz sonuçları daha önce yapılmış HMM tabanlı BBA sınıflandırıcıları ve Mahalanobis mesafesi sınıflandırıcısı ile karşılaştırdık.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

This introduction gives brief information on a general BCI system and explains the components of the EEG-based BCI system used in this thesis. We introduce and discuss the pattern recognition problem involved in such a system, followed by the main contributions and organization of the thesis.

1.1 Brain Computer Interfaces (BCIs)

Brain-Computer Interfaces (BCIs) are assistive systems for humans who have a healthy brain but a dysfunctional peripheral nervous system. Such individuals cannot convert their thoughts into actions. The purpose of a BCI system is to establish a communication channel between the brain and the device that will enable the user to perform an action. A general BCI system (Figure 1.1) is comprised by an interface to collect physiological data from the brain and a control system to process the collected data which converts thoughts into actions. The interface can either be a system such as Electroencephalograph (EEG), a Magnetoencephalograph (MEG) or a Magnetic Resonance Imaging (MRI) device to record electromagnetic neurophysiologic; or a near infrared spectroscope (NIRS) to record neurohematologic effects of working neuronal circuitry.

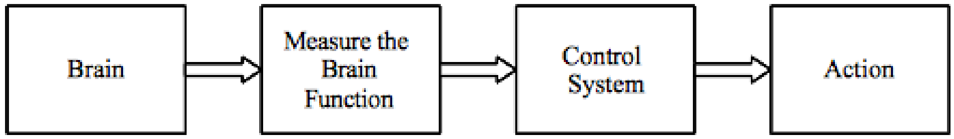
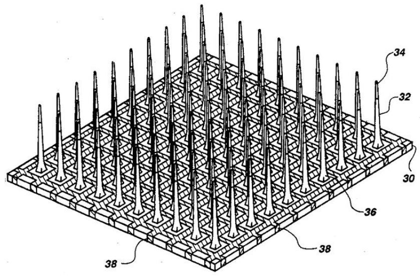


Figure 1.1: Brain Computer Interface System Scheme

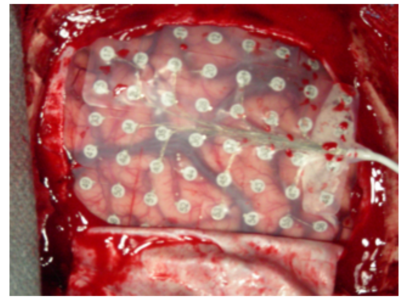
1.2 Electroencephalography (EEG) Based Brain Computer Interfaces (BCIs)

Electroencephalography (EEG) based Brain-Computer Interface (BCI) systems are a new development in the field of applied neurophysiology. These systems are being developed in order to enable people who cannot carry out normal motor functions, such as Amyotrophic Lateral Sclerosis (ALS) and Tetraplegic patients, to control computer based devices. Currently the most important application areas of BCI systems are: to facilitate the communication of ALS and paralyzed patients with their environments; and to enable people with spinal cord injuries to physically manipulate their surroundings by controlling neuroprostheses [5, 6].

Electrophysiology based BCI systems can be realized both in an invasive fashion through microelectrodes (Figure 1.2a) directly implanted in the brain or using an electrode matrix (Figure 1.2b) spread out on the brain; and in a non-invasive fashion through electrodes connected to the surface of the head (Figure 1.2c). In this thesis, we consider non-invasive measurement scenarios. Systems relying on signals measured from surface of the head are easier to use by patients in practice. However, in the process of signal propagation from inside the brain to the outside of the head, significant signal losses and superposition of various signals occur; furthermore noise creeps into the data. Consequently, automatic signal analysis becomes yet more challenging. An example, the one we have used for our study,



(a) Microelectrode



(b) ElectroCorticoGram

(c) ElectroEncephaloGram

Figure 1.2: Invasive and Non-invasive Recording Techniques

of an EEG-based BCI is given in figure 1.3.

1.3 EEG Pattern Recognition Problem

Imagine a room full of people talking about various topics. Lets assume their a television in this room and some of the people in the room talk louder or faster according to some images they saw on the television. Meanwhile let a couple of very good microphones record these people's voices from outside of the room and a scientist analyze these recordings to understand what a particular subset of this group, including those who speak louder, say. In this particular scenario we saw a combination of cocktail party and speech recognition problems.

EEG pattern recognition problem constitutes similar elements with the prob-

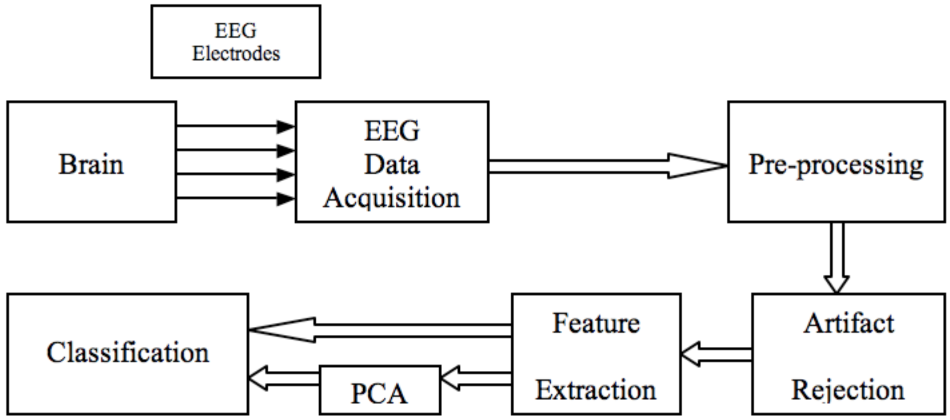


Figure 1.3: An EEG-Based Brain Computer Interface System Scheme

lem above. There are approximately 10^{11} neurons in the brain and 10^{14} connections (synapses) between them. Each neuronal circuitry has its own duty (Figure 1.4) and has interactions with many others constantly. Whenever we record electrical activity of the brain with electrodes connected to the scalp, we measure the electrical behaviors of a population of neurons (Figure 1.5). To deal with this highly convoluted signal, there is a need for extensive signal processing algorithms.

EEG is a time series signal. The purpose of an EEG-based BCI researcher is to develop signal processing techniques to find specific patterns, as in the case above, to carry specific actions in the brain using features extracted from this time series signal. Development of those techniques requires addressing various pieces of the problem including data acquisition, pre-processing, feature extraction, feature selection and classification [7]. Different groups in the BCI community attacked the problem from different angles depending on their area of expertise. If we divide the community into two major groups as Neurobiologists and Engineers, we will see that the former uses very simple pattern recognition methods such

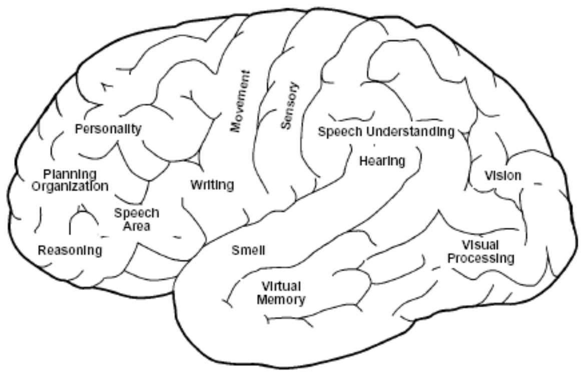


Figure 1.4: Different parts of the brain are responsible for different actions. (taken from epilepsyfoundation.org)

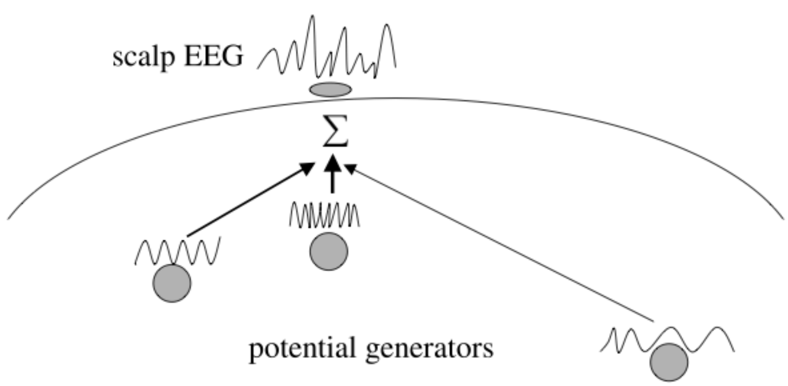


Figure 1.5: Volume conduction effect of brain tissues (Reprinted from [1])

as simple thresholding [8, 9] and focuses much more on subject training, as will be explained in section 2.2. More sophisticated methods, which will also be described in section 2.2 in detail, and which are mostly favored by the latter group of researchers, classify different EEG patterns using linear and nonlinear classifiers [10–12].

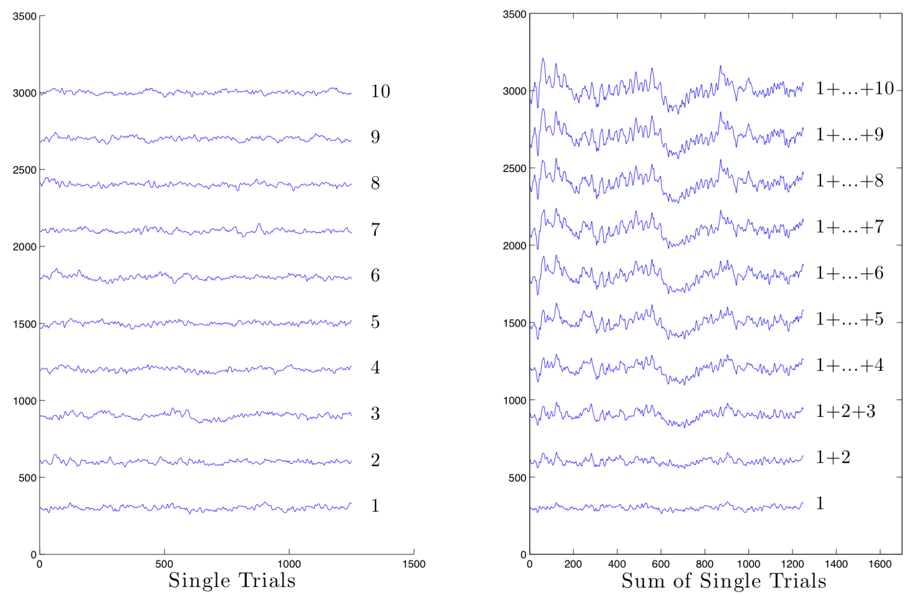


Figure 1.6: Ten Individual Single Trials recorded during a Finger Movement Experiment

When a subject performs a simple motor action such as moving a finger, a corresponding signal can be recorded using EEG electrodes from the related part of the brain, the motor cortex. In Figure 1.6, various EEG signal epochs during a finger movement task recorded over this region of the brain has been shown. Finger movement action was performed 10 times each lasting approximately 5 seconds. In theory, similar EEG patterns for each individual movement can be expected. However, when we inspect each trial, one can see that these signals significantly differ from each other. Our purpose is to find a way to extract the similar patterns in the epochs for classification.This requires a series of steps such as increasing the signal-to-noise (SNR) ratio, characterizing the signal structure and designing a robust classifier. It is possible to increase the SNR of EEG signal by averaging multiple single trials (Figure 1.6). However, averaging needs repetition of trial and leads to loss of time. It is very important to work on single trial epochs for that reason. Of course there is a trade off between SNR and classification speed. Depending on the application, single trial classification could be more important. For that reason, many studies have been performed on single trial classification [13–18].

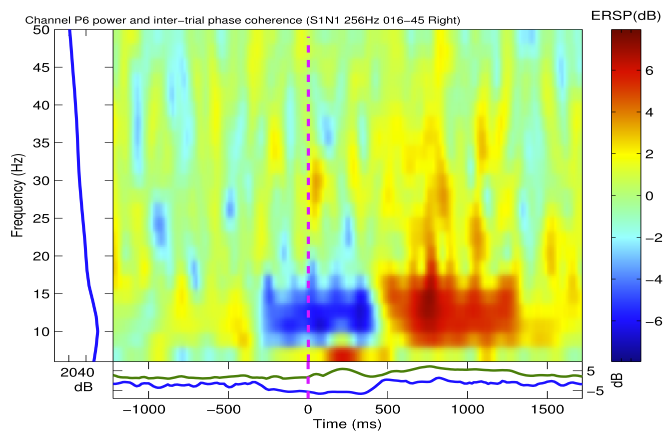


Figure 1.7: Time-Frequency Analysis (TFA) of 10 averaged single trials recorded during a finger movement task. One can easily see the TF pattern.

In this thesis, we have focused on single trial classification of EEG epochs recorded during 2 different motor imagery tasks. During the imagining of motor actions, frequency structure of the EEG signal changes through time. An example of a time-frequency analysis (TFA) of such a signal is shown in Figure 1.7. To increase the SNR, ten individual single trials have been averaged and the TFA has been performed. Averaging of multiple trials makes it easier to highlight the evolution of the frequency structure through time. It is possible to interpret the TFA figure as there are certain changes in the frequency structure of the signal which might be thought as a change in the state of the brain. Throughout the thesis, we have developed techniques that performed well despite the low SNR and characterize the change of the frequency structure of the signal as well as different states of the brain.

1.4 Contribution of this Thesis

The first contribution of this thesis is creating a publicly available BCI database. This database is the first national BCI database which will be publicly available in Turkey. There are a few BCI databases which were made publicly available via international BCI competitions, none of which has 12 subjects and different recording scenarios.

The second contribution involves a novel application of the hidden Markov model (HMM) framework to the EEG based BCI problem. We have used HMMs for a 4-class BCI problem for the first time. HMMs have been used for 2-class BCI problems before [19] and worked pretty well because of the dynamic structure of the classifier. However, previous studies used Hjorth features to train HMMs whereas we have used auro-regressive (AR) features.

The third contribution of this thesis is a novel application of principal component analysis (PCA) to extract and reduce EEG features. PCA has also been used many times in the BCI context. Our contribution involves using PCA for feature dimensionality reduction, and then using HMMs on the reduced dimensional features.

1.5 Organization

In chapter 2, we review the current state-of-the-art BCI systems and classification methods. In chapter 3, EEG feature extraction methods and statistical tools that used in this thesis are summarized. In Chapter 4, the SU-BCI dataset that we have recorded in our laboratory is described in detail and another data set we use in our work is introduced. Chapter 5 contains our feature extraction approach, PCA-based dimensionality reduction procedure, and HMM-based classifier, together with experimental results. Chapter 6 finalizes the thesis with conclusions and potential future perspectives.

CHAPTER 2

Background

In this chapter, we provide some information on the state-of-the-art in EEG based BCI research and summarize current classification methods used in this field.

2.1 EEG based BCIs

EEG-based BCI research can be classified under two types of phenomena [20]:

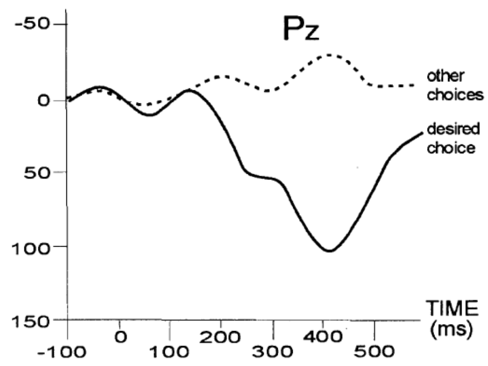
- **Evoked Potentials (EPs):** EPs are uncontrolled responses that are generated by the subject to an external stimulus. EPs can be thought of as a reflex of the brain to a given stimulus. Steady-State evoked potentials (SSEP) and P300 responses are two types of EPs.

- **Spontaneous EEG:** These signals are consciously generated by the subject [7] before, during or after performing an action. Sensory-Motor rhythms (SMR), Bereitschaftspotential (BP or Readiness Potential) and Slow Cortical Potentials (SCPs) are the members of this type BCIs.

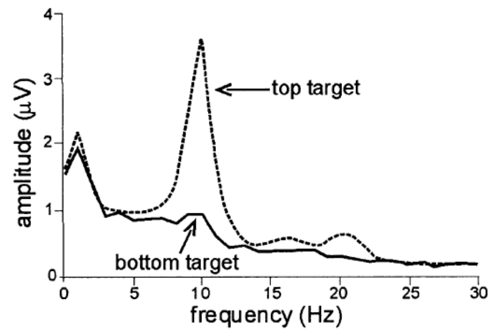
We describe each of these two types of BCIs in the following subsections.

2.1.1 Evoked Potential-based BCIs

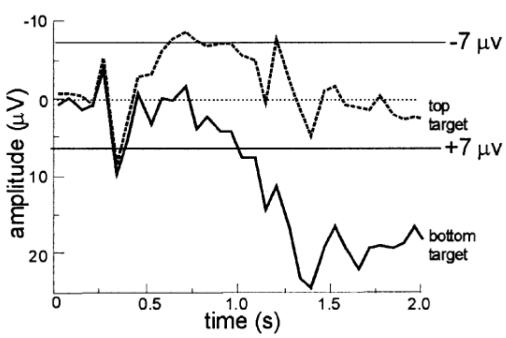
Evoked potentials (EP) are electrical brain responses that are triggered by occurrence of particular events or stimuli. It is hard to detect this activity due to



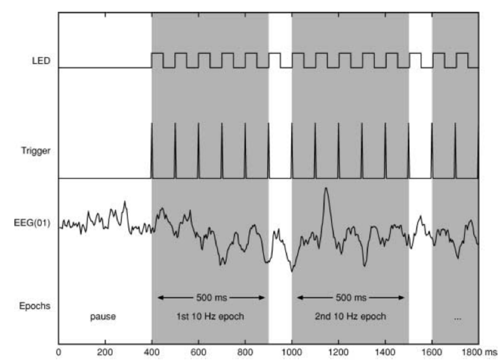
(a) P300



(b) Sensory Motor



(c) Slow-Cortical Potentials



(d) SSVEP

Figure 2.1: Examples of the types of phenomena used in EEG-based BCI techniques. (Taken from [2] and [3])

larger ongoing or background activity. Dawson [21] tried to solve this problem by superposing a number of time-locking activities. This method enhances the event related activity by suppressing the background activity. The main advantage of EP-based BCIs is that they do not need subject training. There are two major types of this kind of BCIs: Steady-state evoked potentials and P300 evoked potentials.

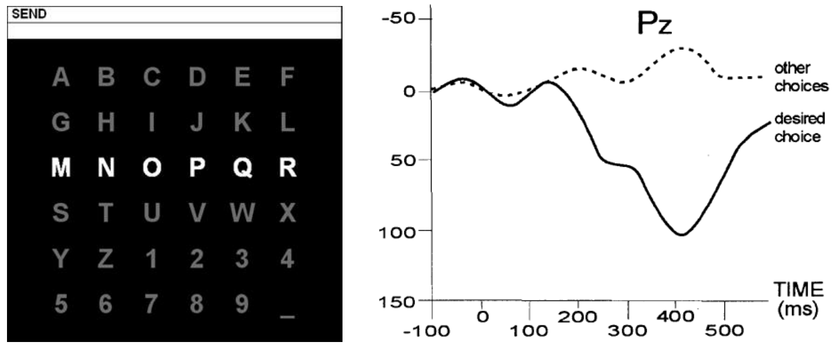
2.1.1.1 Steady-State Evoked Potentials (SSEPs)

SSEPs appear during an exposure to auditory, visual or tactile periodic stimulus such as flashing lights or repeating beeps. Amplitude or frequency of SSEPs can be used as features.

Steady-state VEPs (SSVEPs) were used to enable subjects to use virtual keyboards and perform other types of tasks. One particular setup to exploit SSVEPs involves a number of flashing lights appearing on the screen as virtual buttons. Each button flashes at different frequencies. When the subject focuses on one of these buttons, EEG activity at that frequency increases (Figure 2.1d) [22,23]. That is to say, when stimulated by a signal at a particular frequency, the brain generates an oscillatory response at the same frequency and its harmonics [24,25].

2.1.1.2 P300 Evoked Potentials

P300 is an increase in the EEG amplitude (Figure 2.1a) approximately 300 miliseconds after the presentation of a task-relevant stimulus [25]. Subjects are given two types of stimuli one of which is less presented than the other (e.g. with probability ratio of 1 to 10) which is called the "odd-ball" paradigm. P300 phenomenon has been introduced to the BCI community by Farwell and Donchin



(a) Given stimulus (b) P300 Response

Figure 2.2: P300 spelling device

in 1988 [26], through the so-called P300-spelling device (Figure 2.2). In a P300 spelling device, rows and columns of a letter matrix flash randomly. Each time the row or column of the chosen letter flashes, a P300 signal occurs in the brain. After a couple of repetitions of this process and averaging over trials and matching the row/column flashing pairs, the goal is to automatically detect the intended letter based on the EEG signal and in this way enable the subject to write letters on the screen.

2.1.2 Spontaneous EEG-based BCIs

2.1.2.1 Sensory Motor Rhythms (SMR)

There are four main rhythms in EEG known as alpha (α or μ, 8-12 Hz), beta(β, 13-higher), theta (θ, 4-8 Hz) and delta (δ, 0.1-4 Hz).

When someone does not process a sensory input or produce a motor output, primary sensory or motor cortical areas display a 8-12 Hz activity knows as the μ rhythm [2] (Figure 2.1b). The μ rhythm decreases with movement, preparation of movement or imagination of that movement [27], particularly in contra-lateral areas of the brain to the limb movement. This decrease is called event-related desynchronization (ERD). Similarly increased rhythm is called event related synchronization (ERS) and often appears after movement [27].

Two of the prominent research groups (Wadsworth Center BCI Group, Albany, NY and Graz BCI Group, Graz, Austria) in the field of EEG-based BCI research use SMRs. In Wolpaw et al. [6, 28] patients with motor impairments were able to learn to control their μ and β rhythm amplitudes and in this way performed 2 dimensional movements on computer screen. Similarly, Graz-BCI system enabled patients to control a virtual keyboard [29, 30] and an arm prosthesis [31, 32] .

2.1.2.2 Bereitschaftspotential (BP)

The Bereitschaftspotential or BP (from German, "readiness potential"), also called the pre-motor potential or readiness potential (RP), is slow negative potentials in the premotor and motor cortex of the brain leading up to voluntary muscle movement. The BP is a manifestation of cortical contribution to the premotor planning of volitional movement. RPs are independent of the perception or processing of stimuli [25]. It was first recorded and reported in 1964 by Hans Helmut Kornhuber and Lder Deecke at the University of Freiburg in Germany. In 1965 the full publication appeared after many control experiments. Since the sensorimotor cortex has a somatotopic organization, the body part that will be moved, or for which a movement is imagined, can be inferred from the location of greatest amplitude of the RP. This phenomenon has been used in combination with sensorimotor rhythms in a BCI based on motor imagery [25, 33, 34].

2.1.2.3 Slow Cortical Potentials

SCPs are slow drifts in EEG from 300ms up to a couple of seconds (Figure 2.1c). Positive SCPs means an increase in the cortical excitability and negative SCPs the opposite. As in the SMR case, subjects can also learn to control their SCPs with feedback training which takes longer than SMR training but apparently SCPs are more robust than SMRs [7, 35]. Birbaumer et al. [8] used slow cortical potentials (SCPs) to make patients select letters, words or pictograms in a computerized language support program called the thought translation device.

2.2 BCI Pattern Classification Methods

The purpose of a BCI is to translate neuronal activity into a command for a computer. To achieve this goal, various machine learning algorithms have been used. These algorithms are used to identify patterns of brain activity. Below we provide a brief review of some of the major machine learning algorithms used for EEG-based BCIs so far.

Most of the signal processing and machine learning tools such as linear classifiers, neural networks, nonlinear Bayesian classifiers, nearest neighbor classifiers and the combinations of any of the above were used for EEG-based BCI systems. Table 2.1 summarizes classification methods used for EEG-based BCI systems so far.This table is taken from [4].

Linear classifiers are one of the most popular algorithms for BCI applications due to their simplicity. Linear Discriminant Analysis (LDA) is the most popular linear classifier and it is based on discriminant algorithms that use linear functions to distinguish classes. LDA has assumption of multivariate normal distribution and all groups have the same covariance matrix. LDA has been used in many BCI systems such as motor imagery based BCI [27], P300 speller [36], multiclass [37] or asynchronous [38] BCI. The biggest shortage of LDA is its linearity that can lead to poor results on complex nonlinear EEG data [?, 4].

Support Vector Machines (SVMs) also use a discriminant hyperplane to identify classes. In BCI research the SVM has been applied to P300 data [39, 40] and motor imagery data [41]. Linear SVM enables classification using linear decision boundaries and has been applied to a relatively large number of synchronous BCI problems [10, 37]. Although they have good classification accuracy, several problems exist that hinder the application of SVMs in practical BCI systems. One major issue involves the optimization problems that have to be solved when training SVMs. Solving these problems can be very time consuming, especially when a large number of training examples is used.

Neural networks (NN) are, together with linear classifiers, the most used classifiers in BCI research (see, e.g., [42]). Multi Layer Perceptron (MLP), one of the most favorite NN used in classification, has been applied to almost all BCI problems such as synchronous [43], asynchronous [44], binary [45] or multiclass [42], BCI. When the network complexity is high, MLP classifiers are sensitive to overtraining, especially with such noisy and non-stationary data as EEG [46]. This is why careful architecture selection and regularization is needed.

Bayesian techniques have been rarely used in BCI research. Bayesian graphical network (BGN) has been employed for BCI but currently it is not fast enough for real-time BCI applications [47, 48]. Bayesian classification aims at assigning to a feature vector the class it belongs to with the highest probability [49]. The Bayes rule is used to compute a posteriori probability that a feature vector has of belonging to a given class. Roberts and Penny [50] used autoregressive (AR) model to extract EEG features recorded while the subject performed either men-

Table 2.1: Classifiers used in BCI Research and their properties (taken from [4])

Classifier	Linear	Nonlinear	Generative	Discriminant	Dynamic	Static	Regularized	Stable	Unstable	High Dim. Robust
FLDA	X			X		X		X		
RFLDA	X			X		X	X	X		
Linear-SVM	X			X		X	X	X		X
RBF-SVM		X		X		X	X	X		X
MLP		X		X		X			X	
BLR NN		X		X		X			X	
ALN NN		X		X		X			X	
TDNN		X		X	X				X	
FIRNN		X		X	X				X	
GDNN		X		X	X				X	
Gaussian NN		X		X		X			X	
LVQ NN		X		X		X			X	
Perceptron	X			X		X		X		
RBF-NN		X		X		X			X	
PeGNC		X		X		X	X		X	
Fuzzy		X		X		X			X	
ARTMAP										
NN	X			X		X		X		
HMM		X	X		X				X	
IOHMM		X		X	X				X	
Bayes-Quad		X	X			X			X	
Bayes-GraphNet		X	X			X			X	
kNN		X		X		X			X	
Mahalonobis		X		X		X			X	

tal arithmetic or imagined hand movements. Using the MAP (maximum a posteriori) rule the class of this feature vector can be estimated. Bayesian classifiers are also applied with success to motor imagery [51, 52] and mental task classification [53, 54]. Other interesting examples for the use of Bayesian methodology in BCI systems can be found in [55].

Hidden Markov Models (HMM) are popular dynamic classifiers used in a variety of fields, perhaps most widely in the field of speech recognition [56]. An HMM is a kind of probabilistic automaton that can provide the probability of observing a given sequence of feature vectors. Each state of the automaton can modelize the probability of observing a given feature vector. For BCI, these probabilities usually are Gaussian mixture models (GMM), e.g., [30]. HMM are perfectly suitable algorithms for the classification of time series. As EEG components used to drive BCI have specific time courses, HMM have been applied to the classification of temporal sequences of BCI features [19, 30, 57] and even to the classification of raw EEG [58]. HMM are not very widespread within the BCI community but these studies revealed that they were promising classifiers for BCI systems. Another kind of HMM which has been used to design BCI is the inputoutput HMM (IOHMM) [44]. IOHMM is not a generative classifier but a discriminative one. The main advantage of this classifier is that one IOHMM can discriminate several classes, whereas one HMM per class is needed to achieve the same operation.

Generative algorithms have been used less frequently in BCI research than discriminative methods. Generative algorithms based on Gaussian distributions have been applied with success to the classification of motor imagery data [51] and the classification of other cognitive tasks [53, 59]. A potential advantage of using the generative approach in a BCI system is that a priori knowledge about neurophysiologic signals can be modeled relatively easily [44]. Further advantages are that generative approaches can readily be used for multi-class problems, that generative approaches can easily deal with missing data, and that a probabilistic output is given. A potential disadvantage is that in generative approaches often too many parameters have to be learned.

In this thesis we have used HMM to classify 2 class and 4 class sensory motor data. To overcome the "too many parameters" problem he have combined HMM with PCA.

CHAPTER 3

Preliminaries

In this chapter, we give background information on EEG feature extraction methods and dimension reduction as well as two classification algorithms of interest in this thesis.

3.1 Feature Extraction

Two of the common techniques to extract EEG features in spontaneous EEG-based BCI research, which we will focus on in this thesis, are: Autoregressive (AR) Features and Hjorth Features. Autoregressive features are good at representing frequency related features and Hjorth features are more related with complexity of the data and its change through time [60].

3.1.1 Autoregressive (AR) Features

Parametric modeling of a random process has three steps; selecting the model and the model order, estimating the model parameters, and obtaining the spectral estimates by substituting the estimated model parameters into theoretical power spectral density (PSD) functions [61]. AR parameter estimation is one of the most commonly used techniques to extract frequency related EEG features. The p^{th} order autoregressive (AR) model describes an EEG signal $y_k(t)$ at channel (electrode) k as:

$$y_k(t) = a_{1,k}y(t-1) + a_{2,k}y(t-2) + ... + a_{p,k}y(t-p) + E(t)$$

Here, $a_{i,k}$ denotes the i^{th} order AR parameter modeling the EEG signal at channel k and $E(t)$ is white noise with zero mean and finite variance. There is a direct correspondence between the AR parameters and the autocorrelation function of the process, and this correspondence can be inverted to determine the parameters from the autocorrelation function using the Yule-Walker equations. We have estimated AR parameters for each EEG channel that we have used for this study using least-squares (LS) estimation.

Let K be the number of EEG channels to be used, p be the order of the AR model and n is the number of time points; raw EEG data Y can be represented as,

$$Y = \begin{bmatrix} y_1(t_1) & y_1(t_2) & \cdots & y_1(t_n) \\ y_2(t_1) & y_2(t_2) & \cdots & y_2(t_n) \\ \vdots & \vdots & \cdots & \vdots \\ y_K(t_1) & y_K(t_2) & \cdots & y_K(t_n) \end{bmatrix}_{K \times n} \tag{3.1}$$

while estimated AR feature space of Y is,

$$\begin{aligned} v_{AR}(t) = \{&(a_{1,t}, a_{2,t}, ..., a_{p,t})_1, \\ &(a_{1,t}, a_{2,t}, ..., a_{p,t})_2, \\ &\vdots \\ &(a_{1,t}, a_{2,t}, ..., a_{p,t})_K\} \end{aligned}$$

3.1.2 Hjorth Features

The Hjorth parameters [62] namely activity, mobility and complexity describing the properties of the EEG signal y(t) are calculated as follows:

$$\begin{aligned}
Activity(y(t)) &= Var(y(t)), \\
Mobility(y(t)) &= \sqrt{\frac{Activity(\frac{dy(t)}{dt})}{Activity(y(t))}}, \\
Complexity(y(t)) &= \frac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t)}
\end{aligned}$$

$$\begin{aligned}
v_{Hjorth}(t) = \{&(Activity, Mobility, Complexity)_1, \\
&(Activity, Mobility, Complexity)_2, \\
&\vdots \\
&(Activity, Mobility, Complexity)_K\}
\end{aligned}$$

Hjorth features are mostly used for the classification of motor imagery [12].

3.2 Principal Component Analysis

Principal component analysis (PCA), also known as discrete Karhunen-Loeve transform (KLT) or Hotelling transform, is an orthogonal linear transformation that maps the data into a new space, so called eigenspace, such that the elements of the transformed data are uncorrelated with each other.

Let f be the r dimensional feature vector with zero mean and covariance matrix Σ.

$$\Sigma = Cov(X) = \begin{bmatrix} var(f_1) & cov(f_1, f_2) & \cdots & cov(f_1, f_r) \\ cov(f_2, f_1) & var(f_2) & \cdots & cov(f_2, f_r) \\ \vdots & \vdots & \ddots & \vdots \\ cov(f_r, f_1) & cov(f_r, f_2) & \cdots & var(f_r) \end{bmatrix}_{r \times r} \tag{3.2}$$

Eigenvalues and eigenvectors can be calculated using eigendecomposition:

$$\Lambda = W^T \Sigma W$$

where W is the eigenvector matrix of the covariance matrix Σ, and Λ is the corresponding diagonal matrix of eigenvalues. These eigenvectors in this case are known as the principal components.

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix}_{r \times r}$$

Consequently, projection to the eigenspace is achieved by

$$z = Wf$$

One can reduce the dimension of the feature vector by ordering eigenvalues and selecting the corresponding first s columns of W where $s < r$. PCA can be seen as a linear projection $\mathcal{R}^r \to \mathcal{R}^s$ onto the lower-dimensional subspace corresponding to the maximal eigenvalues [63].

3.3 Classification

Effective applications of hidden Markov models (HMMs) to multiclass EEG data is one of the main purposes of this thesis. Since HMM is a dynamic classifier there is a need for a mean of comparison to a static classifier. In this section the general idea of classification and details of used classifiers will be discussed.

Classification is a task of grouping things. If we come home from grocery with some apples and oranges we can clearly claim that we bought two different things since apples and oranges have very discriminative features like their color,shape,taste etc.

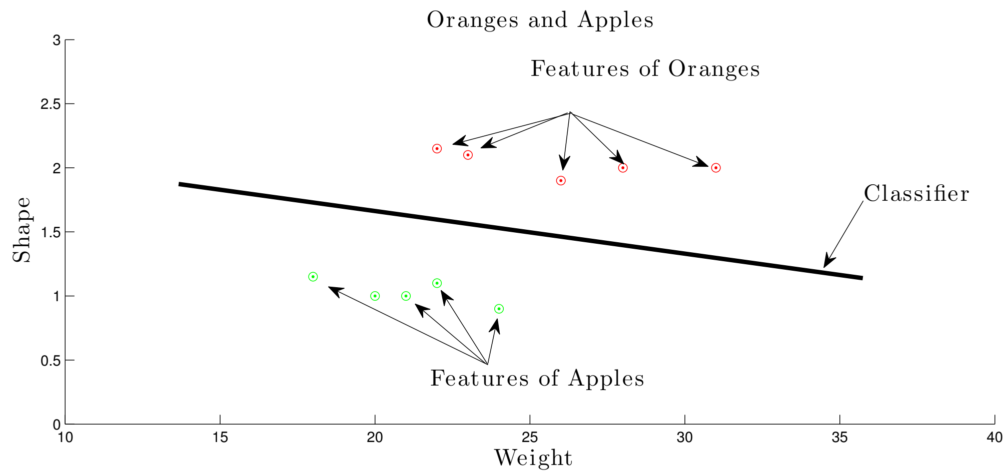


Figure 3.1: Classification of 5 Oranges and 5 Apples according to their Weight and Color

Classification starts with determining the most discriminative features of the classes that we are trying to separate. Discriminative features of the data can either be learnt from the data or one can chose those features according to some beliefs he or she has about corresponding classes depending on their experience. There may be different features of oranges and apples chosen to make a classification. Depending on the chosen features and the type of classification method, one can come up with different results of separating oranges and apples. Once the class-based feature densities are learnt or predetermined, a new data depending on its discriminative features can be assigned to one of the possible classes by calculating some distance between the new data and possible candidate classes.

The classifier in figure 3.1 has a linear and static structure. Mahalanobis distance classifier can be shown as an example of a linear-static classifier. Mahalanobis distance classifier is a simple yet robust classifier that was proven to be effective in classifying multi-class and asynchronous BCI problems [7].

3.3.1 Mahalanobis Distance

The Mahalanobis distance takes into account the covariance among the variables in calculating distances. With this measure, the problems of scale and correlation inherent in the Euclidean distance are no longer an issue. To understand how this works consider that, when using Euclidean distance, the set of points equidistant from a given location is a sphere. The Mahalanobis distance stretches this sphere to correct for the respective scales of the different variables, and to account for correlation among variables.

Mahalanobis distance between vectors X and Y is calculated as follows:

$$d_{XY}^2 = (X - Y)\Sigma^{-1}(X - Y)^T \tag{3.3}$$

where Σ is sample covariance matrix.

The classification of a new feature point Z (Green spot in Figure 3.2) has been done by measuring the Mahalanobis distance d from Z to each of the means (There are 4 means in case of 4 class problem and 2 means for 2 class problem), and assigning Z to the class for which the Mahalanobis distance is minimum.

The Mahalanobis distance can be applied directly to modeling problems as a replacement for the Euclidean distance, as in radial basis function neural networks.

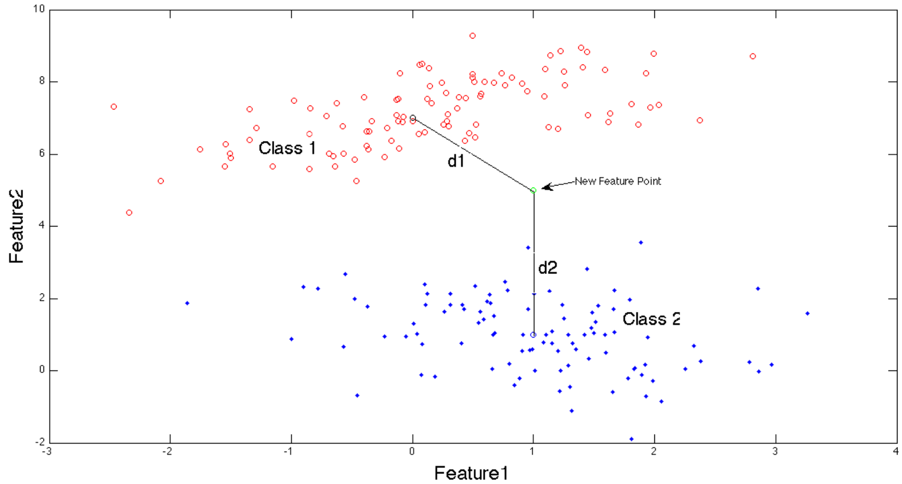


Figure 3.2: Mahalanobis Distance Classifier

3.3.2 Hidden Markov Models (HMMs)

Classifying things that has static, unchanging, features may easily be done using a simple classifier as in above case. However, if one tries to classify features that have a dynamical structure, simple static classifiers may not work well as in the case of time series signals. If a signal exhibits different patterns thought-out time, taking account of this evolution of the pattern may introduce advantages for classification. Hidden Markov Models (HMMs) can model this change as different states of the process and by assigning probabilities of possible transitions among different states. Then, based on some given test data, they can be used to dynamically determine the most probable state sequences. HMMs are one of the most popular dynamic classifiers in the speech processing community [56].

EEG is a time series signal. What we want to do is to model the temporal frequency variations of the EEG epochs thought-out time. In the previous section we have described two different feature extraction methods, AR and Hjorth.

During performing a sensory motor task, when estimated in overlapping windows, AR parameters and Hjorth features exhibit structural changes through-out time. These changes may be interpreted as different states of the mind. Most likely state sequences can be calculated using estimated class specific densities. Gaussian mixtures is one (widely used) choice, but other densities could also be used within HMMs. In this section, the theory of Hidden Markov models (HMM) will be introduced in the context of continuous multiple observation sequences recognition based on Rabiner's tutorial on HMM's [56, 64]

3.3.2.1 Hidden Markov Models for Continuous Observation Densities

An HMM can be considered as the simplest dynamic Bayesian network with a one dimensional chain structure. At each time step t it changes its state from state i to state j with a transition probability of a_{ij} and at each time t that a state j is entered, an observation is generated by jth state from the probability distribution $b_j(\boldsymbol{x}_t)$ which is estimated by Gaussian mixtures in our study. The Mathematical representation of the Gaussian mixture density that we have used is

$$b_j(\boldsymbol{x}_t) = \sum_{m=2}^{M} c_{jm} \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_{jm}|}} \exp(-\frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_{jm})\boldsymbol{\Sigma}_{jm}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_{jm})), \qquad (3.4)$$

where M is the number of mixtures, n is the dimension of the observation vector and $\boldsymbol{\mu}_{jm}$, $\boldsymbol{\Sigma}_{jm}$, c_{jm} are the mean, the covariance and the weight of mixture m of state j.

An HMM has following parameters [56]:

- Number of states ($\boldsymbol{N}$),
- Set of state transition probabilities ($\boldsymbol{A} = \{a_{ij}\}$) from hidden state i to hidden state j ,

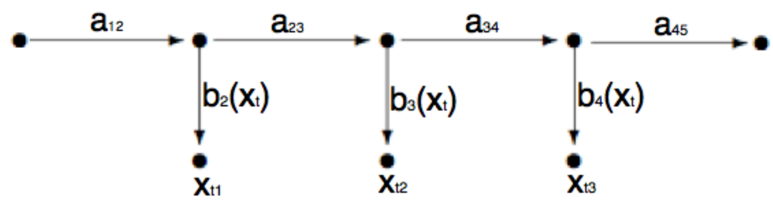


Figure 3.3: Graphical structure of a simple left-to-right HMM

- Set of observation probability distributions in state j ($\boldsymbol{B} = \{b_j(\boldsymbol{x}_t)\}$),
- Initial state distribution, i.e., the set of probabilities of state i being the initial state. ($\boldsymbol{\Pi} = \{\pi_i\}$)

There are three central issues to focus on HMMs:

1. **The Evaluation Problem:** Given the parameters of the model, compute the probability of a particular output sequence. This requires summation over all possible state sequences, but can be done efficiently using the forward algorithm,

2. **The Decoding Problem:** Given the parameters of the model and a particular output sequence, find the state sequence that is most likely to have generated that output sequence. This requires finding a maximum over all possible state sequences, but can similarly be solved efficiently by the Viterbi algorithm.,

3. **The Learning Problem:** Given an output sequence or a set of such sequences, find the most likely set of state transition and output probabilities. In other words, derive the maximum likelihood estimate of the parameters of the HMM given a dataset of output sequences. No tractable algorithm is known for solving this problem exactly, but a local maximum likelihood can be derived efficiently using the Baum-Welch algorithm. The Baum-Welch

algorithm is also known as the forward-backward algorithm, and is a special case of the Expectation-Maximization (EM) algorithm.

In contexts like speech recognition and EEG, a special type of HMM named left-to-right HMM is favored. In a left-to-right HMM, only transitions from a state to itself or to the following state is possible and the entry and exit states are both do not generate any observations.

A three-state, left-to-right HMM topology is given in Figure 3.3.

The parameter N has to be determined a priori. There is a tradeoff between too few states and too many states. Too few states may be inadequate to model the structure of the data and too many states may model the noise too [64].

3.3.2.2 Learning Parameters of the Model

Learning the parameters of an HMM is determining the parameter set $\boldsymbol{\lambda} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi}\}$. The parameter $\boldsymbol{B}$ for Gaussian mixture distribution is equivalent to the parameters $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, c\}$ which are the means, covariances and weights of the mixtures. For left-to-right HMMs, the parameter $\boldsymbol{\Pi}$ is not relevant since the initial state is known to be the non-emitting entry state. The parameters $\boldsymbol{A}$ and $\boldsymbol{B}$ are estimated recursively by the Baum-Welch Algorithm, also known as the Forward-Backward Algorithm.

The forward probability $\alpha_j(t)$, defined as the probability of observing first t observation vectors and being in state j, can be recursively computed by

$$\alpha_j(t) = \sum_{i=2}^{N-1} [\alpha_i(t-1)a_{ij}]b_j(\boldsymbol{x}_t) \tag{3.5}$$

with initial conditions

$$\alpha_1(1) = 1, \tag{3.6}$$

$$\alpha_j(1) = a_{1j} b_j(\boldsymbol{x}_1), \tag{3.7}$$

for $1 < j < N$ and the final condition

$$\alpha_N(T) = \sum_{i=2}^{N-1} [\alpha_i(T) a_{iN}]. \tag{3.8}$$

Notice here that from the definition of $\alpha_j(t)$

$$P(\boldsymbol{X}|M) = \alpha_N(T). \tag{3.9}$$

Similarly, the backward probability $\beta_i(t)$, defined as the probability of observing the observation vectors from $t+1$ to T and being in state i, can be recursively computed by

$$\beta_i(t) = \sum_{j=2}^{N-1} a_{ij} b_j(\boldsymbol{x}_{t+1}) \beta_j(t+1) \tag{3.10}$$

with initial conditions

$$\beta_i(T) = a_{iN}, \tag{3.11}$$

for $1 < i < N$ and the final condition

$$\beta_1(1) = \sum_{j=2}^{N-1} a_{1j} b_j(\boldsymbol{x}_1) \beta_j(1). \tag{3.12}$$

Multiplying the forward and backward probabilities, the probability of being in state i at time t and in state j at time $t+1$, $\xi_t(i,j)$, is derived as

$$\xi_t(i,j) = \frac{\alpha_i(t) a_{ij} b_j(\boldsymbol{x}_{t+1}) \beta_j(t+1)}{\sum_{i=1}^{N} \sum_{i=1}^{N} \alpha_i(t) a_{ij} b_j(\boldsymbol{x}_{t+1}) \beta_j(t+1)}, \tag{3.13}$$

and the probability of being in state i at time t, $\gamma_i(t)$, is derived as

$$\gamma_i(t) = \sum_{j=1}^{N} \xi_t(i,j). \tag{3.14}$$

Given the above definitions, the re-estimation formulae can be expressed as

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_i(t)}, \tag{3.15}$$

$$\boldsymbol{\mu}_{jm} = \frac{\sum_{t=1}^{T} \gamma_{jm}(t)\boldsymbol{x}_t}{\sum_{t=1}^{T} \gamma_{jm}(t)}, \tag{3.16}$$

$$\boldsymbol{\Sigma}_{jm} = \frac{\sum_{t=1}^{T} \gamma_{jm}(t)(\boldsymbol{x}_t - \boldsymbol{\mu}_{jm})(\boldsymbol{o}_t - \boldsymbol{\mu}_{jm})'}{\sum_{t=1}^{T} \gamma_{jm}(t)}, \tag{3.17}$$

$$c_{jm} = \frac{\sum_{t=1}^{T} \gamma_{jm}(t)}{\sum_{t=1}^{T} \gamma_j(t)}. \tag{3.18}$$

Needless to say, the parameters to be estimated have to be initialized before the re-estimation procedure. The initial estimates can be chosen such that

$$a_{ij} = 0.5 \qquad 1 < i < N-1, \quad 2 < j < N, \tag{3.19}$$

$$\boldsymbol{\mu}_{jm} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_t, \tag{3.20}$$

$$\boldsymbol{\Sigma}_{jm} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\mu}_{jm})(\boldsymbol{x}_t - \boldsymbol{\mu}_{jm})', \tag{3.21}$$

$$c_{jm} = \frac{1}{M}. \tag{3.22}$$

3.3.2.3 Computing Likelihood of a New Trial

After estimating the $\boldsymbol{\lambda} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\Pi}\}$ for each task, recognition can be performed based on the model likelihoods. The likelihoods $P(\boldsymbol{X}|\lambda)$ are calculated for each model over the most likely state sequence. The most likely state sequence can be identified using the Viterbi Algorithm as discussed before.

In Viterbi Algorithm, for a given model λ, $\phi_j(t)$ representing the maximum likelihood of observing first t observation vectors and being in state j is recursively computed by

$$\phi_j(t) = \max_i \{\phi_i(t-1) a_{ij} b_j(\boldsymbol{x}_t)\}, \tag{3.23}$$

where

$$\phi_1(1) = 1, \tag{3.24}$$

$$\phi_j(1) = a_{1j} b_j(\boldsymbol{x}_1), \tag{3.25}$$

for $1 < j < N$ which gives the best state sequence. Eventually, the likelihood $P(\boldsymbol{X}|\lambda)$ can be evaluated by

$$P(\boldsymbol{X}|\lambda) = \phi_N(T) = \max_i \{\phi_i(T) a_{iN}\} \tag{3.26}$$

and the model with the highest likelihood is decided to be the imagined movement.

CHAPTER 4

Data

We have focused on spontaneous EEG-based BCI applications for this thesis. Two different sensory-motor datasets have been used. First dataset, SU-BCI-SM-2009, recorded in the EEG Room built in Computer Vision and Pattern Analysis Laboratory in Sabancı University from 12 subjects and has 2 classes. This dataset will be publicly available to researchers in Turkey. Second dataset is a publicly available dataset, BCI Competition IV-2a, recorded in Graz University of Technology from 9 subjects and has 4 classes. Details of the datasets are given below.

4.1 SU-BCI Dataset

4.1.1 EEG Room

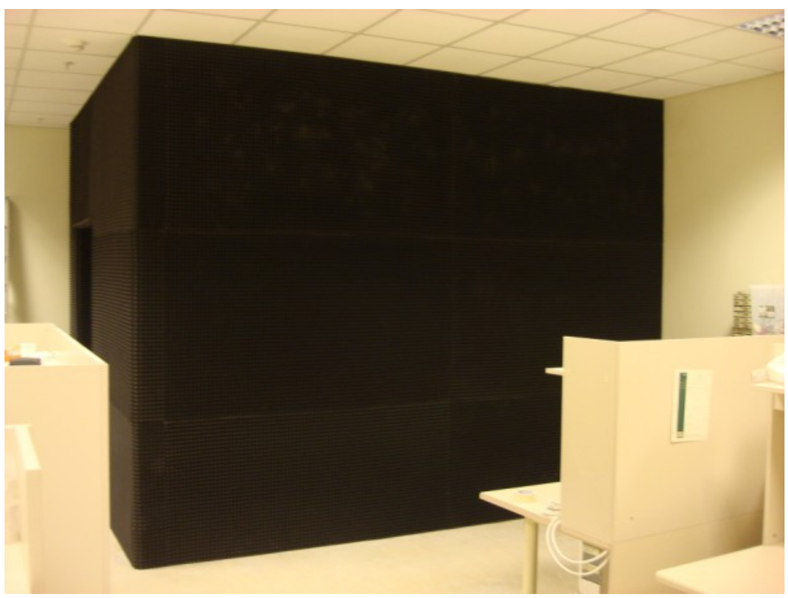
The most important step in EEG research is data collection. Traditional way to record EEG signals from subjects is to use an instrumentation amplifier with high Common Mode Rejection Ratio (CMRR) in an electro-magnetically (EM) shielded room to prevent 50 Hz line noise. Since we wanted to record our own EEG data, we had to built our own EM shielded EEG room. Steps of building the VPALAB EEG Room are as follows:

(a) EEG Room Section in VPALAB (L020)

(b) EEG Room Floor Coating

(c) Faraday Cage

(d) Plaster Coating



(e) Soundproofing

Figure 4.1: EEG Recording Room (Outside)

(a) Door from Outside

(b) Door from Inside

(c) Interface Box

(d) DC Lighting

Figure 4.2: EEG Recording Room (Inside)

- Step 1: 1×3 meter framed $3mm$ galvanized steel plates were prepared.
- Step 2: First of all, floor plates were laid (figure 4.1(b)).
- Step 3: A $4 \times 3 \times 2.6$ meter room (Faraday cage) had been built using galvanized steel plates (figure 4.1(c)).
- Step 4: This Faraday cage was coated with plaster for protection and holding the filling material between the galvanized steel and plaster (figure 4.1(d)).
- Step 5: The outside of the room was coated with black pyramidal sponges to prevent outside noise (figure 4.1(e))

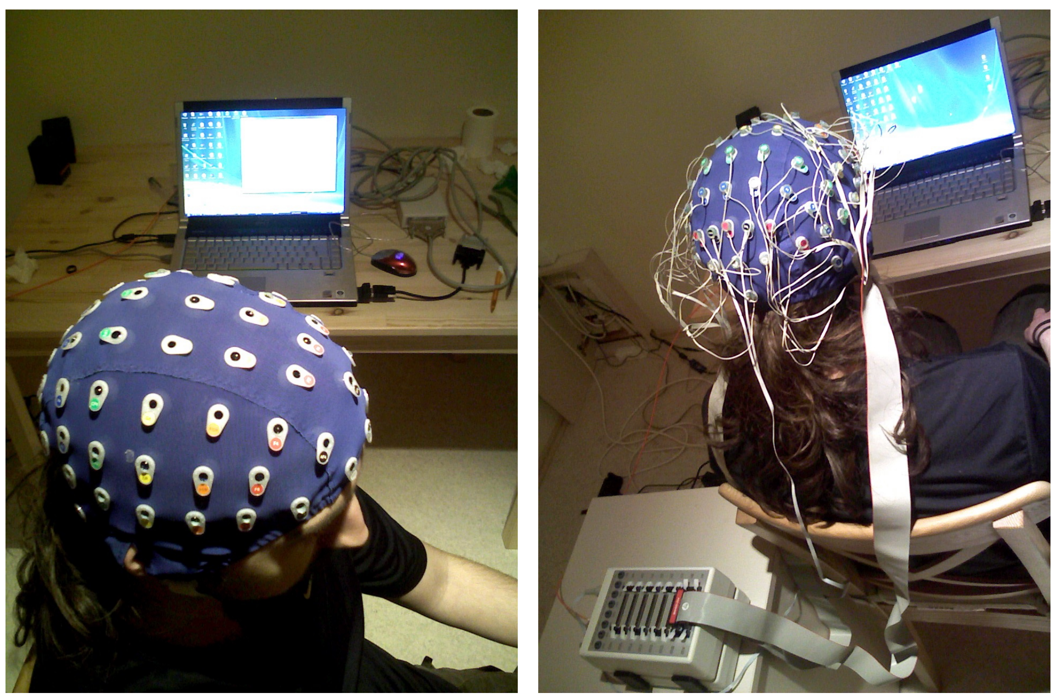
The EEG Room has a special RF shielding which has copper whisks around to unify the Faraday cage when the door is closed (figure 4.2(a),4.2(b)). There is an interface box (figure 4.2(c)) on the corner of the room which runs 20 Ampere filters (100dB, 1 GHz). Cables that run from the outside in are filtered through this filter, which prevents 50 Hz line noise. Lighting is done using three 12V DC bulbs.

4.1.2 Experimental Procedure

SU-BCI-SM-2009 dataset is recorded using Biosemi Active 2 system. Biosemi Active 2 system has up to 256 active electrodes (figure 4.3) depending on the configuration. We used a 64 channel system. The data were recorded at a sampling frequency of 2048Hz and resampled into 256Hz and filtered (Figure 4.6) with a 4^{th} order FIR filter with a passband between 0.16 and 45 Hz for further analysis (An example of a time-frequency analysis of the data can be seen in figure 4.9). Recording was made unreferenced. During the processing average-reference

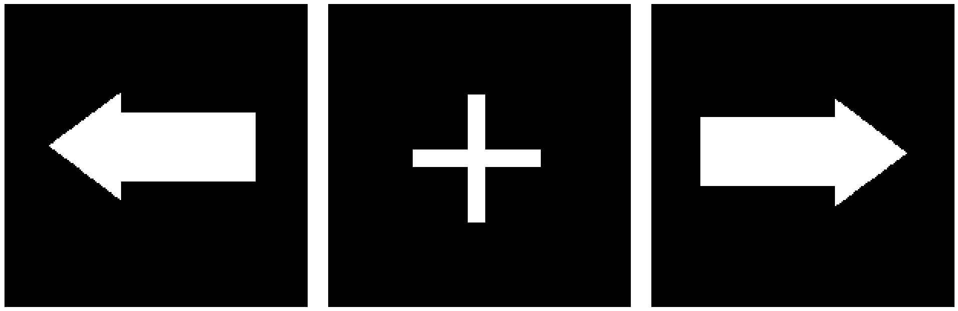
(a) Active Electrodes (b) Electrode Bundle with 32 Electrodes (c) Biosemi Active 2 Amplifier and Battery

Figure 4.3: Biosemi Active 2 System and Active Electrodes



(a) Before Sticking Electrodes (b) 64 Electrode Connected to Head

Figure 4.4: Preparation for EEG Recording



(a) Left Arrow (b) Fixation Cross (c) Right Arrow

Figure 4.5: Figures shown during Experiments

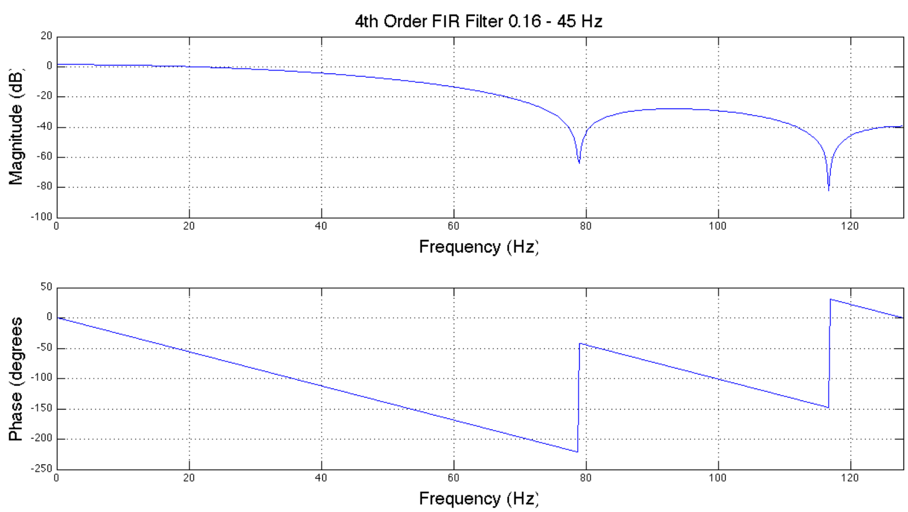


Figure 4.6: 4th Order Bandpass FIR Filter Response

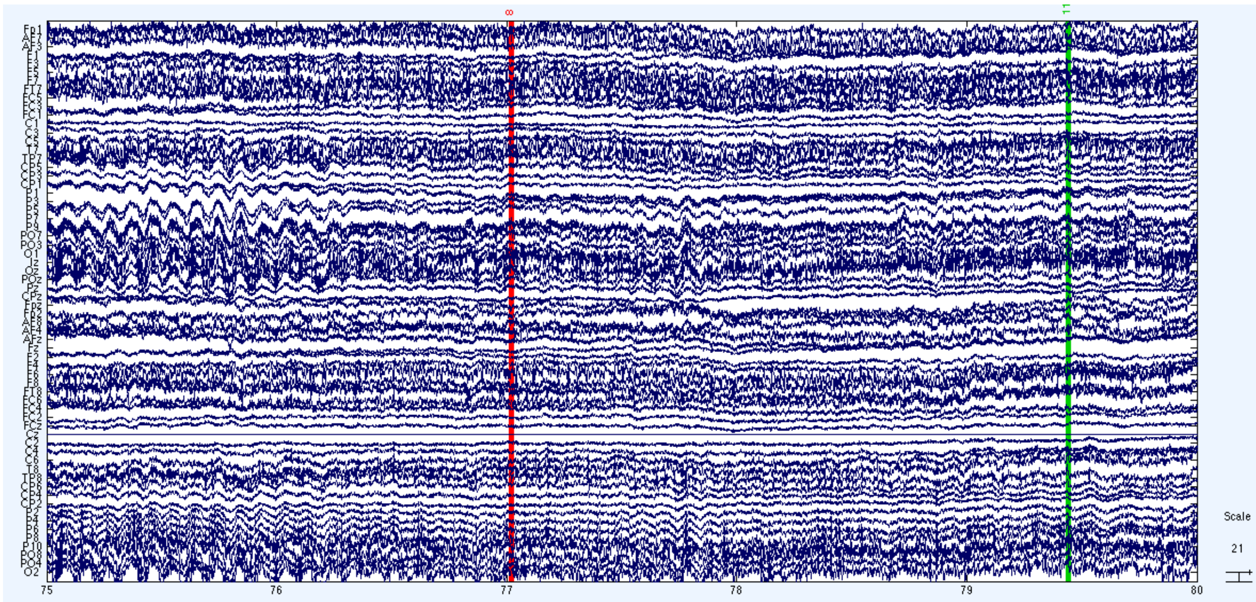
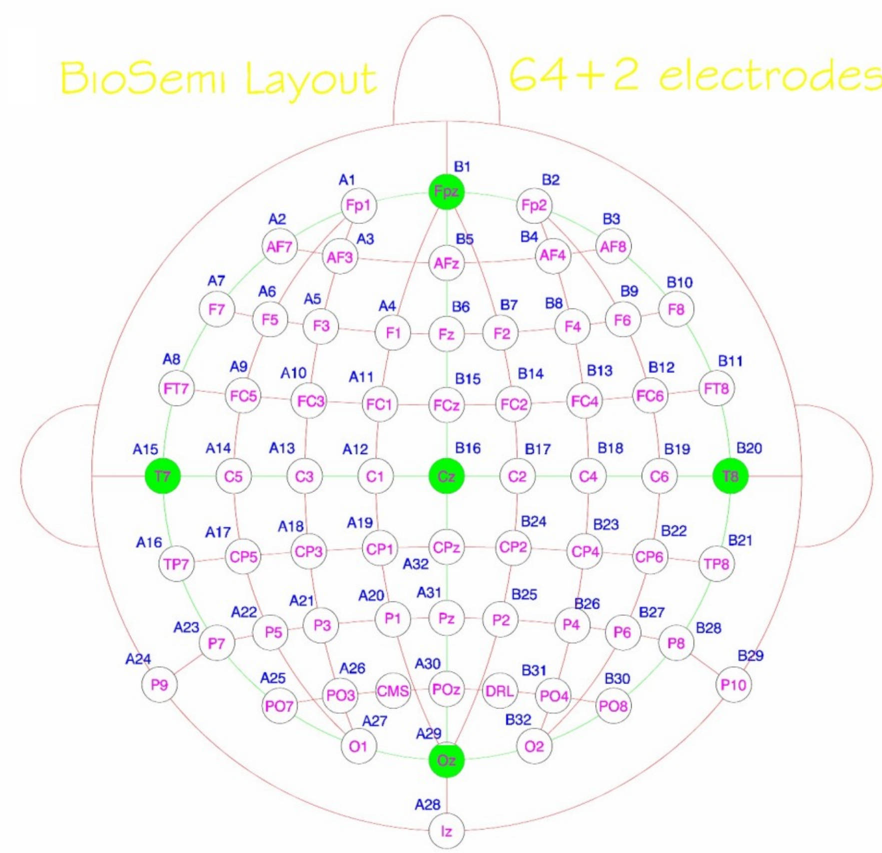


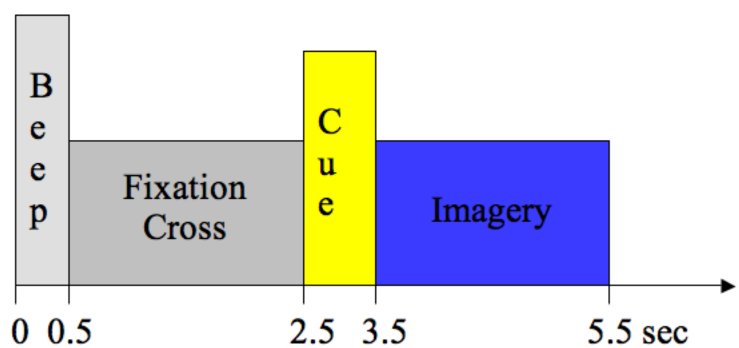
Figure 4.7: Sample SU-BCI data

data have been used. This data set consists of EEG data from 12 subjects.

The cue-based BCI paradigm consisted of two different motor tasks, namely the movement or imagination of movement of the left index finger (class 1 - left) and movement or imagination of movement of the right index finger (class 2 - right). Two sessions on the same day were recorded for each subject. Each session is comprised of 3 different experiments. In the first experiment (SU-BCI-SM-2009N1), subjects performed actual movement of right or left index fingers according to randomly appearing cues on the screen. In the second experiment (SU-BCI-SM-2009Im1), instead of actually moving fingers, subjects were asked to imagine moving their corresponding fingers depending on the cue. In the last experiment (SU-BCI-SM-2009V1) subjects moved their right or left index finger depending on an uninformative cue (No arrow, only fixation cross in (Figure 4.5b)). One run consists of 40 trials (20 for each of the classes for the first



(a) Electrode Configuration



(b) Timing scheme of the paradigm

Figure 4.8: Electrodes and Timing for Dataset 1

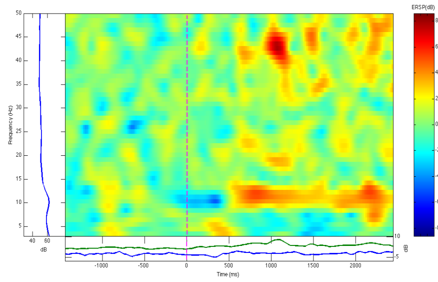
two experiments and subjects were asked to try to perform approximately 20 movement for each classes in the third experiment). Subjects were sitting on a wooden armchair in front of a computer screen. At the beginning of a trial (t = 0 s), a fixation cross appeared on the black screen (Figure 4.5b). In addition, a short acoustic warning tone was presented. After two seconds (t = 2 s), a cue in the form of an arrow pointing either to the left or right(corresponding to one of the two classes left hand or right hand) appeared and stayed on the screen for 1.25 s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at t = 6 s. A short break followed where the screen was black again. The paradigm is illustrated in Figure 4.8.

In this thesis, SU-BCI-SM-2009Im1 dataset was used.

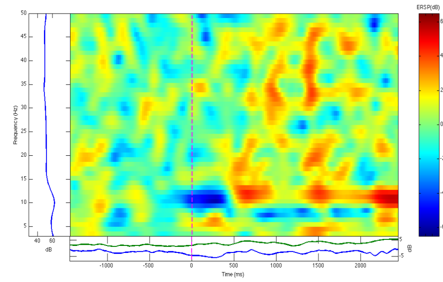
4.2 BCI Competition Dataset

This data set consists of EEG data from 9 subjects. The cue-based BCI paradigm consisted of four different motor imagery tasks, namely the imagination of movement of the left hand (class 1 - left), right hand (class 2 - right), both feet (class 3 - down), and tongue (class 4 - up). Two sessions on different days were recorded for each subject. Each session is comprised of 6 runs separated by short breaks. One run consists of 48 trials (12 for each of the four possible classes), yielding a total of 288 trials per session.

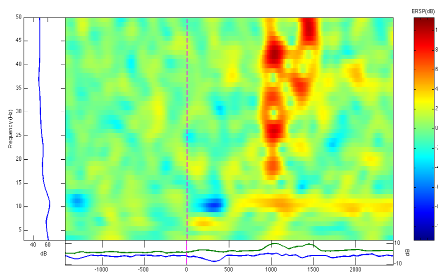
At the beginning of each session, a recording of approximately 5 minutes was performed to estimate the EOG in sequence. The recording was divided into 3 blocks: (1) two minutes with eyes open (looking at a fixation cross on the screen),



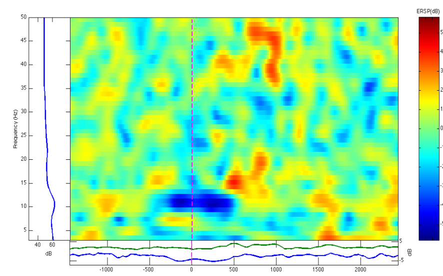
(a) C3-Left



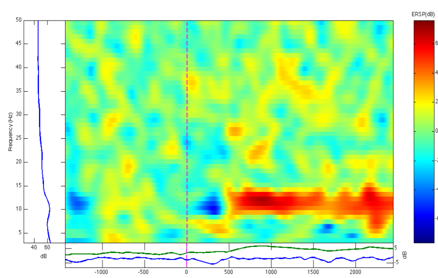
(b) C3-Right



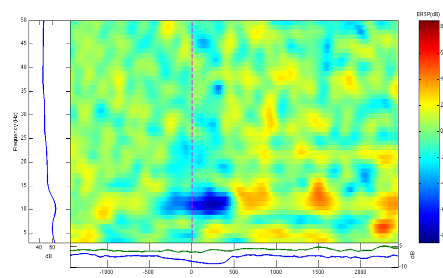
(c) C4-Left



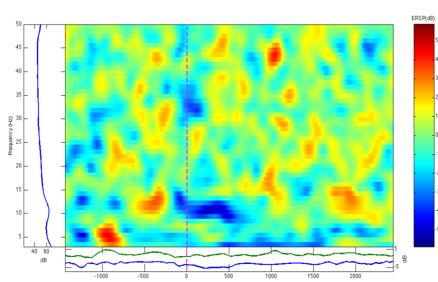
(d) C4-Right



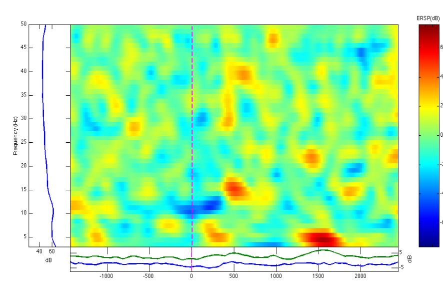
(e) Cz-Left



(f) Cz-Right



(g) Pz-Left



(h) Pz-Right

Figure 4.9: Time-Frequency Maps of 4-Electrodes (C3,C4,Cz,Pz) for Left (a,c,e,g) and Right (b,d,f,h) Finger Movement.

(2) one minute with eyes closed, and (3) one minute with eye movements.

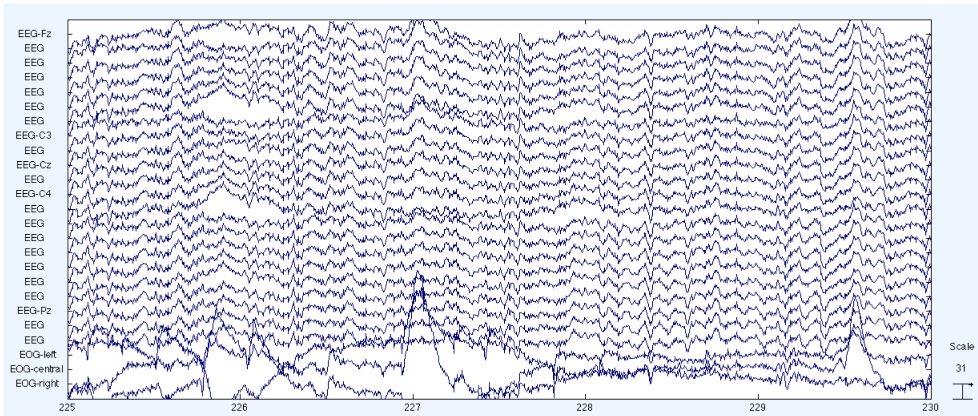
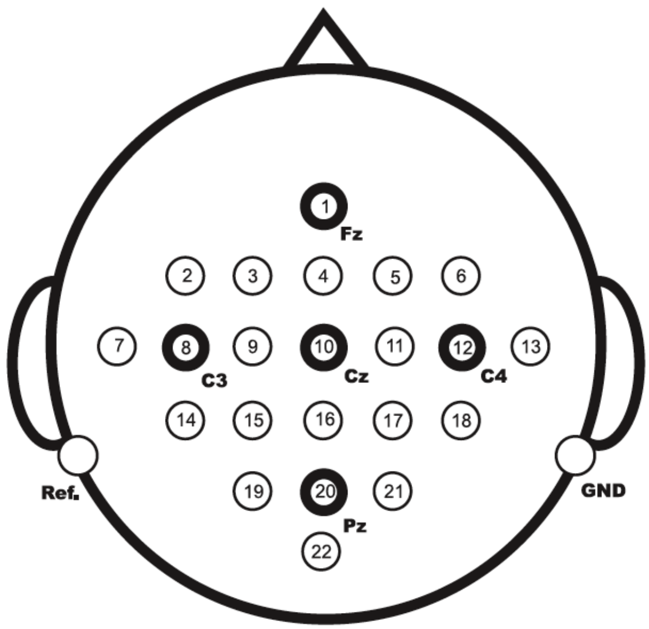


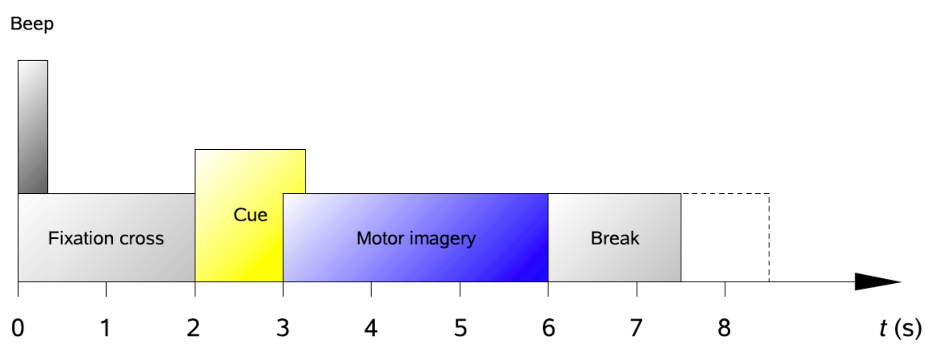
Figure 4.10: Sample BCI Comp. data

The subjects were sitting in a comfortable armchair in front of a computer screen. At the beginning of a trial (t = 0 s), a fixation cross appeared on the black screen. In addition, a short acoustic warning tone was presented. After two seconds (t = 2 s), a cue in the form of an arrow pointing either to the left, right, down or up (corresponding to one of the four classes left hand, right hand, foot or tongue) appeared and stayed on the screen for 1.25 s. This prompted the subjects to perform the desired motor imagery task. No feedback was provided. The subjects were asked to carry out the motor imagery task until the fixation cross disappeared from the screen at t = 6 s. A short break followed where the screen was black again. The paradigm is illustrated in Figure 4.11.

Twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm) were used to record the EEG; the montage is shown in Figure 4.11a up. All signals were recorded monopolarly with the left mastoid serving as reference and the right mastoid as ground. The signals were sampled at 250 Hz and bandpass filtered between 0.5 Hz and 100 Hz. An additional 50 Hz notch filter was enabled to suppress line noise.



(a) Electrode Configuration corresponding to international 10-20 system

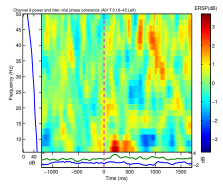


(b) Timing scheme of the paradigm

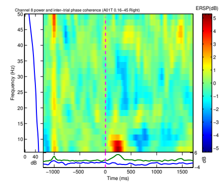
Figure 4.11: Electrodes and Timing for Dataset 2

In addition to the 22 EEG channels, 3 monopolar EOG channels were recorded and also sampled with 250 Hz (see Figure 4.10). They were bandpass filtered between 0.5 Hz and 100 Hz (with the 50 Hz notch filter enabled), and the sensitivity of the amplifier was set to 1 mV. The EOG channels are provided for the subsequent application of artifact processing methods [65] and must not be used for classification. A visual inspection of all data sets was carried out by an expert and trials containing artifacts were marked. Eight out of the total of nine data-sets were analyzed in [66, 67].

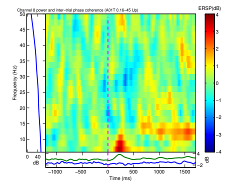
All data sets are stored in the General Data Format for biomedical signals (GDF), one file per subject and session.



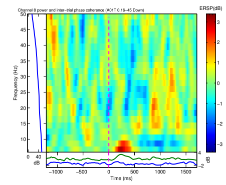
(a) C3-Left



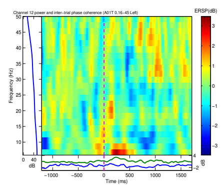
(b) C3-Right



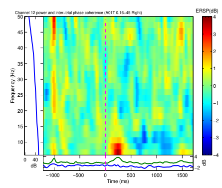
(c) C3-Tongue



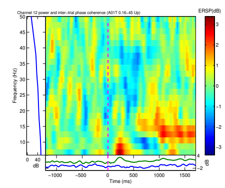
(d) C3-Feet



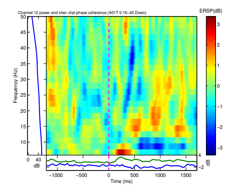
(e) C4-Left



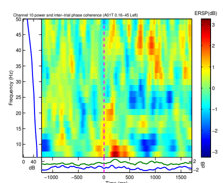
(f) C4-Right



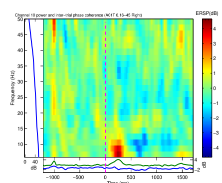
(g) C4-Tongue



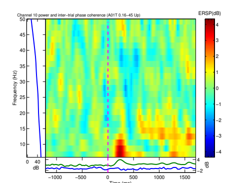
(h) C4-Feet



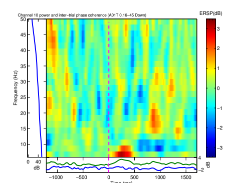
(i) Cz-Left



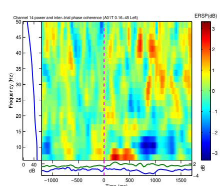
(j) Cz-Right



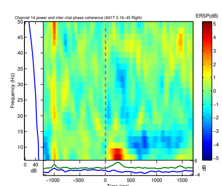
(k) Cz-Tongue



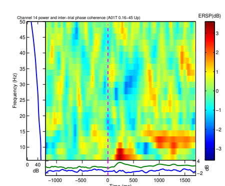
(l) Cz-Feet



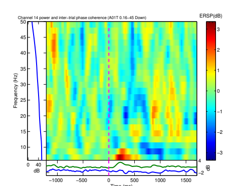
(m) Pz-Left



(n) Pz-Right



(o) Pz-Tongue



(p) Pz-Feet

Figure 4.12: Time-Frequency Maps of 4-Electrodes (C3,C4,Cz,Pz) for Left (a,e,i,m) and Right (b,f,j,n) Finger Movement and Tongue (c,g,k,o) and Feet (d,h,l,p) Movements

CHAPTER 5

AR-PCA-HMM Approach for BCI

Obermeier et al. [30] used HMMs combined with Hjorth features (Hjorth-HMM) for two class offline [19] and [30] online BCI scenarios and compared their results with those obtained through a linear discriminant classifier (LDC). They used AR features for LDC (AR-LDC) while they preferred Hjorth features for the HMM classifier due to their lower dimensionality.

Since AR features give more time-frequency related information about the EEG data compared to Hjorth features, and what we actually try to model is the temporal change in the frequency characteristic of the EEG signal, we have decided to explore the question of whether HMMs with AR features would provide better classification performance than HMMs with Hjorth features. However, as rightfully pointed out by Obermeier et al., high dimensionality of the feature space is an issue that one needs to address. We propose using PCA to reduce the dimensionality of AR features and using these features to train an HMM for classification.

Another issue is intersubject variability which is one of the biggest problems [40] for the EEG-based BCI systems on the way of creating a unified classifier. As already shown by many studies [2,30,39,40,44], a classifier aiming to discriminate different EEG patterns should be trained with data from the same subject. This is why every feature-classifier pair used in this thesis is person specific. Following procedures were applied to each subject one-by-one.

5.1 Feature Extraction

We have used four EEG channels ($C3, C4, Cz$ and Pz) to extract features because of their proven significance in movement and imagination of movement experiments [12].

$$Y = \begin{bmatrix} y_{C3}(t_1) & y_{C3}(t_2) & \cdots & y_{C3}(t_n) \\ y_{C4}(t_1) & y_{C4}(t_2) & \cdots & y_{C4}(t_n) \\ y_{Cz}(t_1) & y_{Cz}(t_2) & \cdots & y_{Cz}(t_n) \\ y_{Pz}(t_1) & y_{Pz}(t_2) & \cdots & y_{Pz}(t_n) \end{bmatrix}_{4\times n} \tag{5.1}$$

where n is the number of time points in a trial.

As discussed in Blanco et al. [68] EEG signal is stationary ranging from several seconds to several minutes depending on the state of the brain. Assuming that the 5 second EEG epoch that we work on would be non-stationary, we decided to extract features in $500ms$ windows and perform further processing in these windows. There is a $200ms$ overlap between windows.

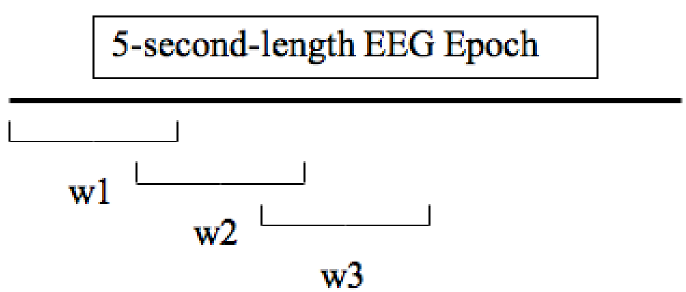


Figure 5.1: Features were extracted in overlapping sliding windows

We have extracted AR features using the procedures that we have discussed in section 3.1.1:

$$F = \begin{bmatrix} a_{1,C3}(1), & \cdots & , a_{1,C3}(M) \\ \vdots & \vdots & \vdots \\ a_{p,C3}(1), & \cdots & , a_{p,C3}(M) \\ a_{1,C4}(1), & \cdots & , a_{1,C4}(M) \\ \vdots & \vdots & \vdots \\ a_{p,C4}(1), & \cdots & , a_{p,C4}(M) \\ a_{1,Cz}(1), & \cdots & , a_{1,Cz}(M) \\ \vdots & \vdots & \vdots \\ a_{p,Cz}(1), & \cdots & , a_{p,Cz}(M) \\ a_{1,Pz}(1), & \cdots & , a_{1,Pz}(M) \\ \vdots & \vdots & \vdots \\ a_{p,Pz}(1), & \cdots & , a_{p,Pz}(M) \end{bmatrix}_{4p \times M}$$

Here $a_{i,k}(m)$ is the AR parameter at the m^{th} window and M is the number of overlapping windows. We have concatenated the features of different channels in each window. Each window is assumed to have stationary EEG features.

Similarly, Hjorth features were also extracted using the method explained in section 3.1.2:

$$F_{Hjorth} = \left[\begin{array}{cccc} Activity_{C3}(1) & Activity_{C3}(2) & \cdots & Activity_{C3}(M) \\ Mobility_{C3}(1) & Mobility_{C3}(2) & \cdots & Mobility_{C3}(M) \\ Complexity_{C3}(1) & Complexity_{C3}(2) & \cdots & Complexity_{C3}(M) \\ \hline Activity_{C4}(1) & Activity_{C4}(2) & \cdots & Activity_{C4}(M) \\ Mobility_{C4}(1) & Mobility_{C4}(2) & \cdots & Mobility_{C4}(M) \\ Complexity_{C4}(1) & Complexity_{C4}(2) & \cdots & Complexity_{C4}(M) \\ \hline Activity_{Cz}(1) & Activity_{Cz}(2) & \cdots & Activity_{Cz}(M) \\ Mobility_{Cz}(1) & Mobility_{Cz}(2) & \cdots & Mobility_{Cz}(M) \\ Complexity_{Cz}(1) & Complexity_{Cz}(2) & \cdots & Complexity_{Cz}(M) \\ \hline Activity_{Pz}(1) & Activity_{Pz}(2) & \cdots & Activity_{Pz}(M) \\ Mobility_{Pz}(1) & Mobility_{Pz}(2) & \cdots & Mobility_{Pz}(M) \\ Complexity_{Pz}(1) & Complexity_{Pz}(2) & \cdots & Complexity_{Pz}(M) \end{array}\right]_{12 \times M}$$

From now on, each column of F either is an AR or Hjorth feature, will be represented as f_m^c if data are labeled and as f_m if the data are not labeled, where $m \in [1, \ldots, M]$ and $c \in [1, \ldots, 4]$, with c denoting the class label.

5.2 Dimensionality Reduction using PCA

We compute one eigenvector matrix for each overlapping window separately. To do that, first we estimate features for each trial using four electrodes as explained in Section 5.1. Then, we create the following matrix G_m^n for each overlapping window by concatenating the data from same window of different classes:

$$G_m^n = \left[\; f_m^1, \ldots, f_m^4 \; \right]_{4.p \times 4}$$

where $n \in [1, ..., N]$ denotes the trial number and N corresponds to the total number of trials. For the sake of notational simplicity we ignore the dependence of f_m^c on the trial index n. Then, corresponding features from the corresponding windows of each trial are concatenated:

$$H_m = \begin{bmatrix} G_m^1, & \cdots & , G_m^N \end{bmatrix}_{4.p \times 4.N}$$

For each overlapping window m, we estimate the covariance matrix of f_m as $\hat{\Sigma}_m = H_m^T H_m$. Then W_m is found as the matrix of eigenvectors of $\hat{\Sigma}_m$.

First s rows ($s < 4p$) of each overlapping window specific W_m matrix is used and represented as W_m^s to calculate each reduced dimensional feature vector, where s is the number of principal components that we want to reduce the dimension to.

$$j_m^c = W_m^s f_m^c$$

Finally, by concatenating the reduced dimensional feature vectors we get the following reduced dimensional feature matrix J_{train}^c for each class:

$$J_{train}^c = [j_1^c, j_2^c, \ldots, j_M^c]_{s \times M}$$

We apply the learned matrix W_m^s for dimensionality reduction of unlabeled test data.

5.3 HMM-Based Classifier

We learn a different HMM for each class in each particular classification experiment. We model the conditional probability densities of the reduced dimensional feature vectors given the underlying states with Gaussian mixtures. For each of these models, we consider two sets of parameters to be learned. The first one, which we denote model-order parameters includes the number of states (NoS) of the HMM, the number of Gaussian mixtures (NoGM), the AR model order p, and the reduced dimension s. We denote the second set of parameters as *model*

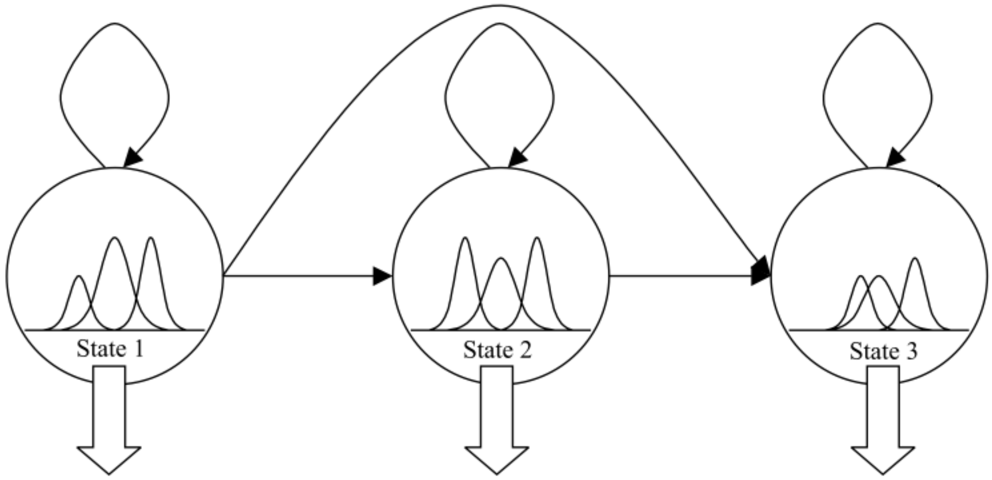


Figure 5.2: Left-to-Right HMM model for 3 states and 3 mixture of gaussians (Reprint from Obermaier et al. 2001)

parameters $\boldsymbol{\lambda_c} = (\boldsymbol{A},\boldsymbol{B},\boldsymbol{\Pi})$. Here $\boldsymbol{A}$ denotes the state transition probabilities, $\boldsymbol{B}$ is the means and variances of the observation probability distributions and $\boldsymbol{\Pi}$ is initial state distributions (see [56]). We split the data into three namely $\boldsymbol{F^c_{train}}$ (40%), $\boldsymbol{F^c_{validation}}$ (30%), $\boldsymbol{F_{test}}$ (30%). For a fixed set of model-order parameters, we learn the model parameters based on training data $\boldsymbol{F^c_{train}}$ using the expectation-maximization (EM) algorithm so that the likelihood of the observed training sequences is locally maximized. We learn the model-order parameters by maximizing the classification performance on validation data $\boldsymbol{F^c_{validation}}$. We then test our classification approach on test data $\boldsymbol{F_{test}}$ based on these learned parameters.

5.4 Experimental Analysis and Results

Probability of correct classification (PCC) of different feature extraction, feature reduction and classification algorithm settings are presented in this section. We investigated whether our proposed $\boldsymbol{AR \rightarrow PCA \rightarrow HMM}$ method is superior to $\boldsymbol{AR \rightarrow HMM}$, $\boldsymbol{AR \rightarrow Mahalanobis}$ and $\boldsymbol{Hjorth \rightarrow HMM}$ methods for two-class and four-class BCI systems. We also present a comparison of AR-PCA-HMM with the top techniques in BCI Competition IV on BCI Competition IV-2a dataset.

For BCI Competition IV-2a dataset, we report three different results: validation and two different test results. Result Test-1 is obtained after training with the first $\mathbf{40\%}(\boldsymbol{F^c_{train}})$ of the data and testing with last $\mathbf{30\%}(\boldsymbol{F_{test}})$ of the data. Result Test-2 is obtained after training with the first $\mathbf{70\%}(\boldsymbol{F^c_{train}}\mathbf{(40\%)} + \boldsymbol{F^c_{validation}}\mathbf{(30\%)})$ of the data and testing with last $\mathbf{30\%}(\boldsymbol{F_{test}})$ of the data.

For SU-BCI-SM-2009Im1 dataset, we report test results which are obtained after training with the first $\mathbf{60\%}(\boldsymbol{F^c_{train}})$ of the data and testing with last $\mathbf{40\%}(\boldsymbol{F_{test}})$ of the data.

5.4.1 Four-Class Results

BCI Competition IV-2a dataset consists of EEG data from 9 subjects. We have used the first sessions of first 8 of the subjects. These same 8 subjects were also used in [67] and [66].

We interpret the probability of correct classification (PCC) performance of the validation set as the degree of learning of the classifier. Figure 5.3 reports the best PCC performances of different classifiers over the validation dataset. Parameters that lead to the reported PCC performances are listed in Table 5.1.

Figure 5.4 and 5.5 report PCC performances over the unseen data $\boldsymbol{F_{test}}$. For 7 of 8 subjects, our approach achieves the highest PCC. In Table 5.2, we present a comparison of AR-PCA-HMM with the top techniques in BCI Competition IV on this dataset in terms of the $\boldsymbol{\kappa}$ coefficient. We observe that AR-PCA-HMM

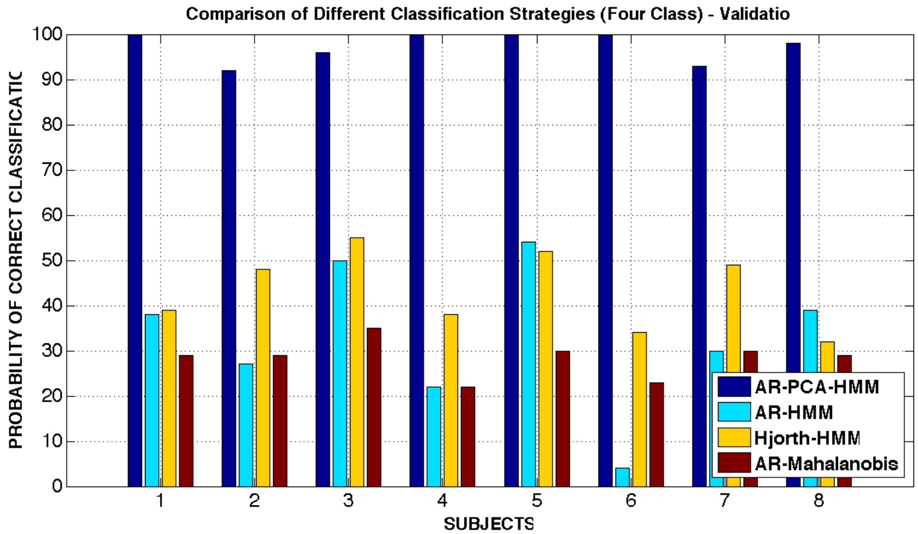


Figure 5.3: Validation Performance of the proposed AR-PCA-HMM approach as compared to other techniques.

achieves the best performance.

Table 5.1: Model-order parameters that lead to the best probability of correct classification results in the validation dataset and used in the test dataset.

Method	Parameters	S1	S2	S3	S4	S5	S6	S7	S8
AR-PCA-HMM	p	14	15	14	14	15	14	15	15
	s	10	8	8	10	10	7	9	8
	$NoGM$	3	3	3	3	3	2	3	3
	NoS	5	5	5	3	5	5	5	5
AR-HMM	p	9	5	5	8	6	15	13	6
	$NoGM$	2	1	1	1	2	1	1	3
	NoS	4	2	3	4	1	2	3	3
Hjorth-HMM	$NoGM$	3	1	3	2	1	1	1	3
	NoS	5	3	1	5	3	1	4	1
AR-Mahal	p	10	8	5	7	9	5	6	9

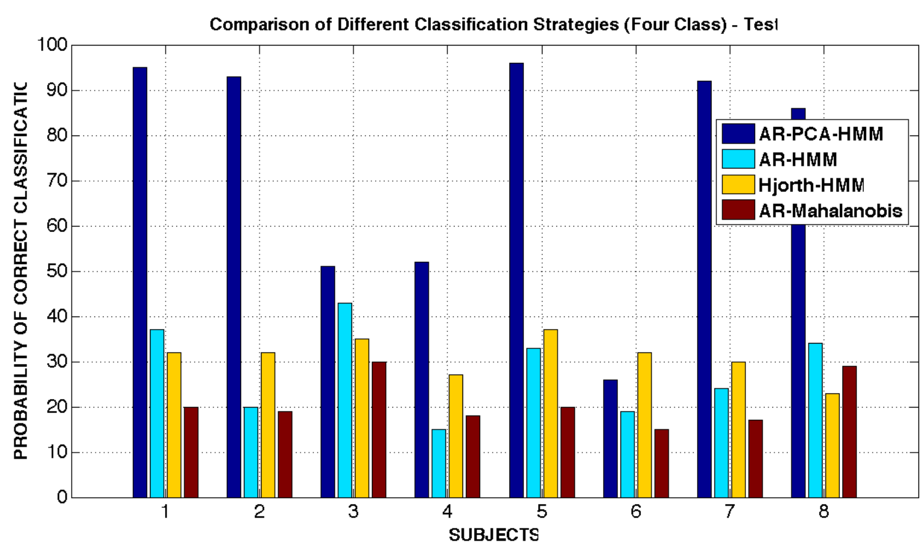


Figure 5.4: Performance of the proposed AR-PCA-HMM approach as compared to other techniques (Test1).

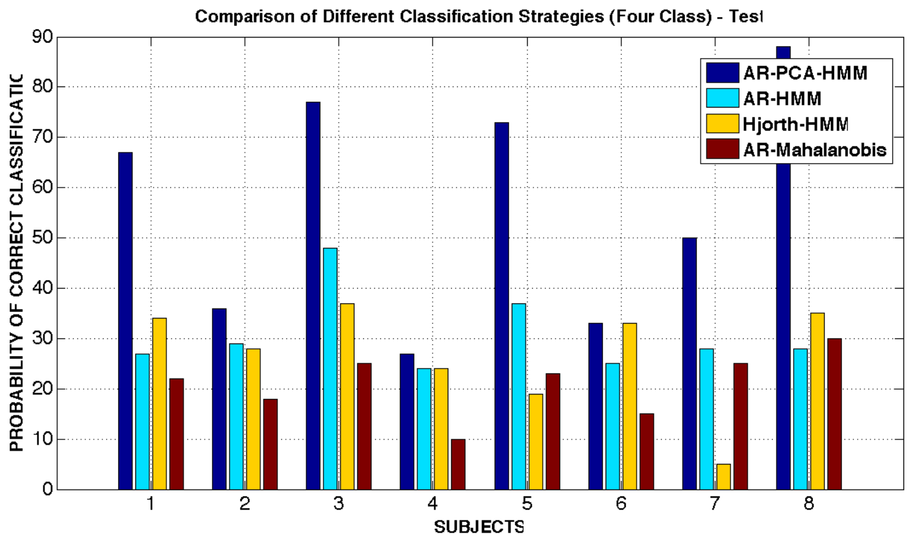


Figure 5.5: Performance of the proposed AR-PCA-HMM approach as compared to other techniques (Test2).

Table 5.2: AR-PCI-HMM vs. State-of-the-art Techniques. $\boldsymbol{\kappa = (C \times PCC - 1)/(1 - C)}$. $\boldsymbol{\kappa}$ approaches zero as PCC approaches $\boldsymbol{1/C}$ where $\boldsymbol{C}$ is the number of classes.

Feature / Classifier	κ_{mean}	S1	S2	S3	S4	S5	S6	S7	S8
AR-PCA / HMM	.65	.93	.91	.34	.36	.95	.01	.89	.81
FBCSP / Naive Bayes	.57	.68	.42	.75	.48	.40	.27	.77	.75
CSP / LDA-Bayes	.51	.69	.34	.71	.44	.16	.21	.66	.73
CSP / SVM-Voting	.30	.38	.18	.48	.33	.07	.14	.29	.49
CSP / LDA-SVM	.29	.46	.25	.65	.31	.12	.07	.00	.46
CSP / SVM	.28	.41	.17	.39	.25	.06	.16	.34	.45

5.4.2 Two-Class Results

Two-Class problem is relatively easier to solve compared to four-class one. However, due to insufficient amount of data (20 trials for each class), training a classifier and testing with validated model order parameters was not a robust option. In Figure 5.6 we present optimized PCC performances. In particular, in this experiment we present results based on model-order parameters that maximize PCC on the test data for which we report classification results. Hence our experiment here leaves out the issue of effective learning the model-order parameters.

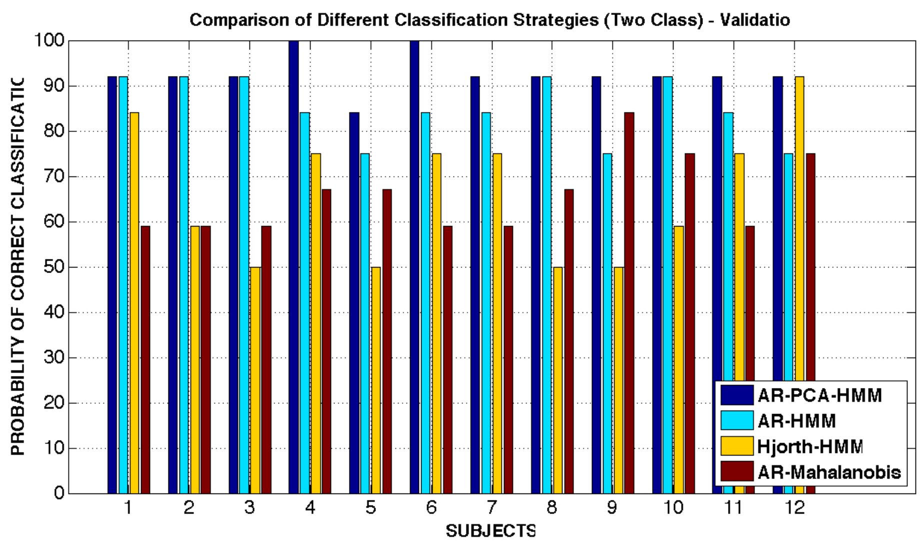


Figure 5.6: Performance of the proposed AR-PCA-HMM approach as compared to other techniques with optimized model-order parameters.

Although, AR-PCA-HMM again reached the highest PCC performance as in the four-class problem, insufficient amount of two-class data prevents us from performing more intensive investigations on classification performance of the proposed and existing classification methods.

Table 5.3: Model-order parameters that gave best probability of correct classification results on the test dataset.

Method	Parameters	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
AR-PCA-HMM	$\boldsymbol{p}$	6	12	13	14	10	13	6	12	9	5	6	5
	$\boldsymbol{s}$	4	3	5	10	2	6	4	3	2	3	2	5
	$\boldsymbol{NoGM}$	1	2	3	3	3	1	1	2	2	2	3	3
	$\boldsymbol{NoS}$	4	3	2	5	5	3	5	3	4	1	4	3
AR-HMM	$\boldsymbol{p}$	15	7	10	8	7	5	5	10	7	5	10	6
	$\boldsymbol{NoGM}$	3	1	3	1	3	2	1	1	1	2	2	1
	$\boldsymbol{NoS}$	5	3	3	3	3	5	5	2	4	5	2	4
Hjorth-HMM	$\boldsymbol{NoGM}$	3	1	3	3	3	3	3	2	3	1	3	3
	$\boldsymbol{NoS}$	5	4	5	5	5	4	5	5	5	4	4	4
AR-Mahal	$\boldsymbol{p}$	9	10	8	5	8	12	7	11	10	5	9	8

CHAPTER 6

Conclusions and Future Work

6.1 Summary and Conclusions

In this thesis, we have proposed an AR-PCA based approach combined with hidden Markov Models (HMMs) to solve the EEG-based BCI classification problem. The main idea of the thesis was supporting the use of AR features with HMM classifiers for EEG-based BCI problem. AR features are good frequency estimators however their relatively high dimensions are problematic for the convergence of HMM parameters during training. To overcome this problem, we have proposed decreasing the dimensions of AR features using PCA to make them more suitable for HMM classifiers.

Four-class classification results suggest that AR features are better features for HMM based EEG-BCI classifiers and dimension reduction is crucial for EEG classification. Comparison with the state-of-the-art classification methods shows that dynamic structure of the HMMs combined with a good frequency estimator results in better performance than all static classifiers and their combinations. Depending on the subject, it is possible to reach 80% classification accuracies using AR-PCA-HMM in the four-class problem which is a quite reasonable result compared to AR-Mahalanobis which could only achieve up to 35% which is not very far from classification by chance.

Two-class classification results suggest that HMM classifiers coupled with AR features lead to better PCC performances. Due to limitations in the amount of data, our experimental analysis on the two-class dataset was not as thorough as that on the the four-class dataset. In particular, we have avoided the model-order parameter learning problem and presented (optimistic) results for each method based on model-order parameters optimized on the test data. We aim to collect more data, and perform a validation-based analysis as in the four-class experiments in the future. As the number of trials increases, subjects may perform the imagination of movement tasks better trial by trial which will lead to a decrease in the inter-trial variability.

It is clear that the four-class problem is harder to solve and needs better modeling of the data. Results suggest that modeling the evolution of the frequency structure of the EEG data with HMM outcompetes the static approach more for the four-class problem compared to the two-class one. As the complexity of the problem increases, the need for the use of HMMs increases.

Dimension reduction is very important for EEG classification. For the four-class problem PCA increases the classification performance. Using these results, we can conclude that PCA should be applied to EEG data after the feature extraction independent of the feature extraction method.

6.2 Future Work

HMMs are generative models. They are typically trained to maximize the joint likelihood of observation and label pairs. To do this, enumeration of all possible observations is needed. Because of that reason, it is possible that the inference problem may be intractable. Conditional random fields (CRFs) can overcome this problem as proposed by Lafferty et al. [69].

We have modeled the temporal, time-frequency, structure of the EEG data by using a one dimensional graphical model, HMMs. One important future direction in this research may be modeling not only the temporal but also the spatial structure of the EEG data. In this thesis, we have used different EEG channels for classification but we have not modeled the spatial relation between those channels. Modeling the information flow between different EEG channels combined with temporal evolution of the frequency change may give much more better results. However, localization of the EEG activity is crucial at this point since the volume conduction effects of the head tissues, the EEG signal at different electrodes seem to carry very similar information.

We have proposed one particular way of using PCA to reduce the dimension of AR features. We have calculated PCA mapping matrices for each 500 ms length overlapping windows. One can investigate the use of different window lengths and different number of principal components as well as their coevolution.

References

[1] M. Honal, "Determining user state and mental task demand from electroencephalographic data," *Master's Thesis, Universitat Karlsruhe (TH), Karlsruhe, Germany*, 2005.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control." *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, June 2002. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/12048038

[3] C. S. Herrmann, "Human EEG responses to 1100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experimental Brain Research*, vol. 137, no. 3, pp. 346–353, April 2001. [Online]. Available: http://dx.doi.org/10.1007/s002210100682

[4] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "Journal of neural engineering; a review of classification algorithms for EEG-based brain-computer interfaces." vol. 4, no. 2, 2007.

[5] G. Pfurtscheller, R. Leeb, C. Keinrath, D. Friedman, C. Neuper, C. Guger, and M. Slater, "Walking from thought." *Brain Res*, vol. 1071, no. 1, pp. 145–152, February 2006. [Online]. Available: http://dx.doi.org/10.1016/j.brainres.2005.11.083

[6] J. R. Wolpaw, "Brain-computer interfaces (BCIs) for communication and control." in *ASSETS*, E. Pontelli and S. Trewin, Eds. ACM, 2007, pp. 1–2.

[7] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces." *Journal of neural engineering*, vol. 4, no. 2, June 2007. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/17409472

[8] N. Birbaumer, A. Kubler, 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," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 190–193, 2000. [Online]. Available: http://dx.doi.org/10.1109/86.847812

[9] S. Levine, J. Huggins, S. BeMent, R. Kushwaha, L. Schuh, M. Rohde, E. Passaro, D. Ross, K. Elisevich, and B. Smith, "A direct brain interface based on event-related potentials," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 180–185, Jun 2000.

[10] B. Blankertz, G. Curio, and K. R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," in *Advances in Neural Inf. Proc. Systems (NIPS 01)*, T. G. Diettrich, S. Becker, and Z. Ghahramani, Eds., vol. 14, January 2002, pp. 157–164.

[11] J. Del R Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni, "A local neural classifier for the recognition of EEG patterns associated to mental tasks." *IEEE Trans Neural Netw*, vol. 13, no. 3, pp. 678–686, 2002. [Online]. Available: http://dx.doi.org/10.1109/TNN.2002.1000132

[12] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul 2001.

[13] G. Birch, P. Lawrence, and R. Hare, "Single-trial processing of event-related potentials using outlier information," *Biomedical Engineering, IEEE Transactions on*, vol. 40, no. 1, pp. 59–73, Jan. 1993.

[14] D. Lange, H. Pratt, and G. Inbar, "Segmented matched filtering of single event related evoked potentials," *Biomedical Engineering, IEEE Transactions on*, vol. 42, no. 3, pp. 317–321, March 1995.

[15] D. Popivanov, J. Dushanova, A. Mineva, and I. Krekule, "Detection of successive changes in dynamics of EEG time series: linear and nonlinear approach," in *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, vol. 4, Oct-3 Nov 1996, pp. 1590–1591 vol.4.

[16] A. Schlögl, D. Flotzinger, and G. Pfurtscheller, "Biomedizinische technik. biomedical engineering; adaptive autoregressive modeling used for single-trial EEG classification." vol. 42, no. 6, pp. 162–167, 1997.

[17] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms: a conceptual tutorial." *Brain Lang*, vol. 66, no. 1, pp. 7–60, January 1999. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/10080864

[18] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 10, pp. 1297–1307, 2000. [Online]. Available: http://dx.doi.org/10.1109/10.871402

[19] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, "Hidden Markov models for online classification of single trial EEG data," *Pattern Recogn. Lett.*, vol. 22, no. 12, pp. 1299–1309, 2001.

[20] G. Pfurthscheller and C. Neuper, "EEG-based brain computer interfaces," pp. 1265–1271, 2005.

[21] G. Dawson, "A summation technique for the detection of small evoked potentials," *'Electroencephalogr. Clin Neurophysiol.*, vol. 1, no. 6, pp. 65–84, 1954.

[22] E. C. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly, and G. McDarby, "Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 3156–3164, 2005.

[23] M. Middendorf, G. Mcmillan, G. Calhoun, and K. S. Jones, "Brain-computer interfaces based on the steady-state visual-evoked response," *Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Neural Systems and Rehabilitation]*, vol. 8, no. 2, pp. 211–214, 2000. [Online]. Available: http://dx.doi.org/10.1109/86.847819

[24] C. S. Herrmann and R. T. Knight, "Mechanisms of human attention: event-related potentials and oscillations." *Neuroscience and biobehavioral reviews*, vol. 25, no. 6, pp. 465–476, August 2001. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/11595268

[25] U. Hoffmann, "Bayesian machine learning applied in a brain-computer interface for disabled users," *EPFL, Ph.D. Thesis*, 2007.

[26] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials." *Electroencephalogr Clin Neurophysiol*, vol. 70, no. 6, pp. 510–523, December 1988. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/2461285

[27] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles." *Clin Neurophysiol*, vol. 110, no. 11, pp. 1842–1857, November 1999. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/10576479

[28] 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 Rehabil*

Eng, vol. 8, no. 2, pp. 164–173, June 2000. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/10896178

[29] C. Neuper, G. Muller, A. Kubler, N. Birbaumer, and G. Pfurtscheller, "Clinical application of an EEG-based brain computer interface: a case study in a patient with severe motor impairment," *Clinical Neurophysiology*, vol. 114, no. 3, pp. 399–409, March 2003. [Online]. Available: http://dx.doi.org/10.1016/S1388-2457(02)00387-5

[30] B. Obermaier, C. Munteanu, A. Rosa, and G. Pfurtscheller, "Asymmetric hemisphere modeling in an offline brain-computer interface," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 31, no. 4, pp. 536–540, Nov 2001.

[31] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in graz brain-computer interface (BCI) research," *IEEE Trans. on Neural Systems and Rehab. Eng.*, vol. 8, no. 2, pp. 216–219, 2000. [Online]. Available: http://dx.doi.org/10.1109/86.847821

[32] G. Pfurtscheller, C. Neuper, G. Muller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank, "Graz-BCI: state of the art and clinical applications," *IEEE Trans. on Neural Systems and Rehab. Eng.*, vol. 11, no. 2, pp. 1–4, June 2003.

[33] J. Pineda, B. Allison, and A. Vankov, "The effects of self-movement, observation, and imagination on ? rhythms and readiness potentials (rp's): toward a brain-computer interface (BCI)," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 219–222, Jun 2000.

[34] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 993–1002, June 2004.

[35] J. Borisoff, S. Mason, and G. Birch, "Brain interface research for asynchronous control applications," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 160–164, June 2006.

[36] V. Bostanov, "BCI competition 2003-data sets ib and iib: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1057–1061, June 2004.

[37] D. Garrett, D. Peterson, C. Anderson, and M. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 2, pp. 141–144, June 2003.

[38] R. Scherer, G. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 979–984, June 2004.

[39] M. Kaper and H. Ritter, "Generalizing to new subjects in brain-computer interfacing," in *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, vol. 2, Sept. 2004, pp. 4363–4366.

[40] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 1, pp. 24–29, March 2006.

[41] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of Neural Engineering*, vol. 2, no. 4, pp. L14+, December 2005. [Online]. Available: http://dx.doi.org/10.1088/1741-2560/2/4/L02

[42] C. Anderson, J. Knight, T. O'Connor, M. Kirby, and A. Sokolov, "Geometric subspace methods and time-delay embedding for EEG artifact removal and classification," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 142–146, June 2006.

[43] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 4, pp. 457–463, Dec 2000.

[44] S. Chiappa and S. Bengio, "Hmm and iohmm modeling of EEG rhythms for asynchronous BCI systems," in *12th European Symposium on Artificial Neural Networks (ESANN)*, 2004, pp. 199–204.

[45] R. Palaniappan, "Utilizing gamma band to improve mental task based brain-computer interface design," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 3, pp. 299–303, Sept. 2006.

[46] D. Balakrishnan and S. Puthusserypady, "Multilayer perceptrons for the classification of brain computer interface data," in *Bioengineering Conference, 2005. Proceedings of the IEEE 31st Annual Northeast*, April 2005, pp. 118–119.

[47] T. Felzer and B. Freisieben, “Analyzing EEG signals using the probability estimating guarded neural classifier,” *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 11, no. 4, pp. 361–371, Dec. 2003.

[48] K. Tavakolian and S. Rezaei, “Classification of mental tasks using gaussian mixture bayesian network classifiers,” in *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, Dec. 2004, pp. S3/6–9–11.

[49] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2nd ed. Wiley-Interscience, November 2000.

[50] P. Sykacek, S. J. Roberts, and M. Stokes, “Adaptive BCI based on variational bayesian kalman filtering: an empirical evaluation,” *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 5, pp. 719–727, 2004. [Online]. Available: http://dx.doi.org/10.1109/TBME.2004.824128

[51] S. Lemm, C. Schafer, and G. Curio, “BCI competition 2003-data set iii: probabilistic modeling of sensorimotor ? rhythms for classification of imaginary hand movements,” *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1077–1080, June 2004.

[52] S. Solhjoo, A. M. Nasrabadi, and M. Golpayegani, “EEG-based mental task classification in hypnotized and normal subjects,” in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, Jan. 2005, pp. 2041–2043.

[53] Z. Keirn and J. Aunon, “Man-machine communications through brain-wave processing,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 9, no. 1, pp. 55–57, Mar 1990.

[54] G. Barreto, R. Frota, and F. de Medeiros, “On the classification of mental tasks: a performance comparison of neural and statistical approaches,” in *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop*, 29 2004-Oct. 1 2004, pp. 529–538.

[55] P. Sykacek, S. Roberts, and M. Stokes, “Adaptive BCI based on variational bayesian kalman filtering: an empirical evaluation,” *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 5, pp. 719–727, May 2004.

[56] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc IEEE*, pp. 257–286, 1989.

[57] F. Cincotti, D. Mattia, C. Babiloni, F. Carducci, S. Salinari, L. Bianchi, M. Marciani, and F. Babiloni, “The use of EEG modifications due to motor

imagery for brain-computer interfaces," *IEEE Trans. Neural Sys. and Rehab. Eng.*, vol. 11, no. 2, pp. 131–133, June 2003.

[58] S. Solhjoo, A. M. Nasrabadi, and M. Golpayegani, "EEG-based mental task classification in hypnotized and normal subjects," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, Jan. 2005, pp. 2041–2043.

[59] E. Curran, P. Sykacek, M. Stokes, S. Roberts, W. Penny, I. Johnsrude, and A. Owen, "Cognitive tasks for driving a brain-computer interfacing system: a pilot study," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 12, no. 1, pp. 48–54, March 2004.

[60] A. Schloegl and G. Pfurtscheller, "Considerations on adaptive autoregressive modelling in EEG analysis," in *Proceedings of First International Symposium on Communication Systems and Digital Signal Processing*, 1998.

[61] N. Jain and S. Dandapat, "Spectral estimation using constrained autoregressive (car) model," in *Intelligent Sensing and Information Processing, 2005. ICISIP 2005. Third International Conference on*, Dec. 2005, pp. 110–114.

[62] B. Hjorth, "An on-line transformation of EEG scalp potentials into orthogonal source derivations," *Electroencephalography and Clinical Neurophysiology*, vol. 39, no. 5, pp. 526–530, November 1975. [Online]. Available: http://dx.doi.org/10.1016/0013-4694(75)90056-5

[63] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 780–788, 2002.

[64] H. Karabalkan, "Tandem approach for information fusion in audio visual speech recognition," *Master Thesis, Sabanci University*, 2009.

[65] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "Emg and eog artifacts in brain computer interface systems: A survey," *Clinical Neurophysiology*, vol. In Press, Corrected Proof. [Online]. Available: http://dx.doi.org/10.1016/j.clinph.2006.10.019

[66] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *Journal of Neural Engineering*, vol. 3, no. 3, pp. 208–216, 2006. [Online]. Available: http://stacks.iop.org/1741-2552/3/208

[67] C. Brunner, R. Scherer, B. Graimann, G. Supp, and G. Pfurtscheller, "Online control of a brain-computer interface using phase synchronization," *IEEE Trans. Biomedical Engineering*, vol. 53, no. 12, pp. 2501–2506, Dec. 2006.

[68] S. Blanco, H. Garcia, R. Quiroga, L. Romanelli, and O. Rosso, "Stationarity of the EEG series," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 4, pp. 395–399, Jul/Aug 1995.

[69] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001, pp. 282–289.