Analysis of Event-Related Potentials Elicited During Information Processing Tasks

Thesis submitted in partial fulfillment of the requirements for the degree of

“DOCTOR OF PHILOSOPHY”

By

Dror Lederman

Submitted to the Senate of Ben Gurion University of the Negev

Beer-Sheva

March 2009
Analysis of Event-Related Potentials Elicited During Information Processing Tasks

Thesis submitted in partial fulfillment of the requirements for the degree of

“DOCTOR OF PHILOSOPHY”

By

Dror Lederman

Submitted to the Senate of Ben Gurion University
of the Negev

Beer-Sheva

March 2009

Approved by the advisor, Dr. Joseph Tabrikian: ________________

Approved by the Dean of the Kreitman School of

Advanced Graduate Studies, Prof. Rami Brushtein: ________________
This work was carried out under the supervision of
Prof. Arnon Cohen (R.I.P.) and Dr. Joseph Tabrikian

In the Department of Electrical and Computers Engineering

Faculty of Engineering
Acknowledgements

First of all, I owe a dept of gratitude to my supervisor, Dr. Joseph Tabrikian, with whom I have been working in the past nine years. I had the honor of being his teaching assistant, assistant associate editor and finally his Ph.D. student. Joseph taught me invaluable lessons both on scientific research and on scientific writing.

During the course of my research, I have faced several difficult and critical moments which I would not overcome without Joseph’s constant support, attention and optimism. His critical and enthusiastic attitude towards my research always kept me going. I am grateful to Joseph for his close and continuous supervision and support.

I wish to thank Dr. Izchak Lewkovich and Amit Kam who opened me the door to the wonderful world of teaching and advised me in my first steps in this world.

I would like to thank my colleagues at Ben-Gurion University of the Negev: Amit Kam, Dr. Yaniv Zigel, Dr. Mark Kliger, Morad Dacca, Dr. Vladimir Lapidus, Dr. Igal Bilik and Noam Broshi, for their friendship and continuous help during the process.

I would also like to thank the anonymous reviewer and Prof. Amir Geva, whose constructive and helpful comments helped to significantly improve this dissertation.

Finally, I wish to express my gratitude and love to my family: my wife Ziva, my 9-years old son Lavie and my 4-years old daughter Gaya, for their sweet nature and good behavior, and to my parents and my parents-in-law, without whose loving support (and hours of babysit) this thesis might not have materialized.
While working on this project, Prof. Arnon Cohen, head of the signal processing laboratory at Ben-Gurion University of the Negev, passed away. Prof. Cohen supervised me during my Bachelor studies, Master studies and during the first years of this work. I will never forget his didactic approach and the fruitful time I spent with him in our weekly meetings and during the courses he taught. In fact, I owe much of my professional knowledge to Prof. Cohen.

Prof. Cohen was known world-wide as a researcher who wrote numerous articles, chapters and books, taught thousands of students, and led a vibrant laboratory. By completing this work, which was inspired by his ideas and enthusiasm, we continue his legacy of honest and fair research.
Dedicated to my children,
Lavie and Gaya
Abstract

This work is motivated by the problem of estimation and classification of different electroencephalography (EEG) patterns elicited during various information processing tasks. Since the discovery of Hans Berger in 1929, interest in developing algorithms to process human brain and mind information by analyzing the EEG and ERP signals has been constantly growing. Both the EEG and ERP signals have been shown to be related to many psychological processes involved in attention, memory and motor control. Moreover, both types of signals have been proven to be useful in classification of different mental and cognitive tasks. Consequently, the idea of developing a brain computer interface (BCI), i.e. to use the brain electrical signals in order to control remote devices, has emerged.

Research in this field has faced various engineering and physiological problems. These include, among others, low signal-to-noise ratio (SNR), overlapping spectra between the ERP and the spontaneous EEG, high intrasubjects and intersubjects variability, signals characterized by nonstationarity and nonergodicity and unknown number and location of intracranial sources and unknown intrabrain wave propagation channels. Therefore, it is required to employ appropriate methods for estimating and classifying these signals.

In this work, novel methods for single-trial EEG estimation and classification are developed. First, a model-based approach for multichannel EEG classification is introduced. The approach is based on parallel hidden Markov models (PHMMs) and a maximum-likelihood-based decision rule. The performances of the PHMM classifier are studied using an artificial EEG database and two real databases. The results show that the PHMM algorithm outperforms other existing methods, with an improvement of 2% and 10%, in the classification rates for the two real databases, respectively, comparing to the best reported method.

One of the main obstacles in EEG estimation is the presence of modeling mismatch. In order to cope with this problem, a new estimator is derived based on the minimum mean-square error (MMSE) criterion with constraints on the first- and second-order statistics of the parameters of interest. The constrained MMSE (CMMSE) estimator provides a suboptimal solution, which guarantees that the first- and second-order statistical properties of the estimated signal match those of the parameter of interest. The CMMSE estimator is found to
be robust to signal distribution mismatch, since it incorporates statistical information on the parameters of interest and restricts the solution space accordingly.

The performance of the CMMSE estimator under different mismatch conditions is studied both analytically and via simulations using several examples. It is shown that with no distribution mismatch, the CMMSE performance is slightly lower than the optimal MMSE, while in the presence of signal distribution mismatch, the CMMSE estimator outperforms the MMSE estimator and several other commonly used estimators.

The CMMSE estimator preserves the first- and second-order statistics of the parameter of interest, while higher order statistics might not match to the true statistics. Hence, the idea of the constrained estimator is further extended to an estimator which incorporates constraints on the probability density function (PDF) of the parameter of interest. The statistical constraints are based on Gaussian mixture model (GMM) representation of the PDF. This estimator guarantees that the statistical properties of the estimated signal match those of the parameter of interest. Since a closed form expression for this estimator is difficult to obtain, an approximated estimator, termed as GMM-CMMSE, which is based on a weighted sum of linear CMMSE estimators, is proposed. The performance of the GMM-CMMSE estimator under different mismatch conditions is studied via simulations using several examples. It is shown that with no distribution mismatch, the GMM-CMMSE performance is slightly lower than the optimal MMSE. However, in the presence of signal distribution mismatch, the GMM-CMMSE estimator is superior to the MMSE estimator, and the other tested estimators. In addition, the GMM-CMMSE is easy to implement, since it can be represented by a weighted sum of linear CMMSE estimators.

The CMMSE and GMM-CMMSE estimators are employed for single-trial EEG estimation. Some simulation experiments are performed using real EEG signals acquired from seven subjects while performing five mental tasks. The performance of the CMMSE and the GMM-CMMSE estimators are compared with those of an ICA-based estimator and the Wiener filter. It is shown that the CMMSE and the GMM-CMMSE estimators outperform both the ICA-based estimator and the Wiener filter in most of the illustrated SNRs. Further evaluation of the estimators performances using real databases of various types of response-related EEGs, and the impact of the proposed estimators on the classification rates, is dependent upon acquiring or obtaining an appropriate database, and is therefore a topic for future research. In addition, it is postulated that incorporation of mutual information between different data channels and/or incorporation of inter-channel statistical constraints may improve the perfor-
mances of the CMMSE and GMM-CMMSE estimators.

Beyond the use of the above-mentioned estimators for EEG processing, they may be utilized in many other signal processing applications in which the MMSE estimator is employed, such as Kalman filter, expectation-maximization algorithm, etc. Therefore, these estimators are of great importance in signal processing theory and applications.

Hence, the main contributions of this dissertation are the development of a novel multi-channel classification algorithm based on parallel HMMs, the development of robust constrained estimators, and the implementation of these estimators for single-trial EEG estimation.

**Keywords** Classification, Event-Related Potentials, Electro-encephalogram, Parallel HMMs, Constrained MMSE, Bayesian Estimation, GMM, Statistical constraints.
Contents

1 Introduction 1

1.1 Background ......................................................... 1

1.2 Outline ............................................................ 5

2 A Parallel HMM for the Classification of Multichannel EEG signals 6

2.1 Introduction .......................................................... 6

2.2 HMM-based classification of EEG patterns .............................. 9

2.2.1 Features ............................................................ 9

2.2.2 Classification scheme .............................................. 11

2.2.3 The training algorithm ............................................ 12

2.2.4 The recognition algorithm ....................................... 13

2.3 Experimental results ............................................... 15

2.3.1 Simulation ........................................................ 15
3 Constrained MMSE Estimator for Distribution Mismatch Compensation 26

3.1 Introduction ......................................................... 26
3.2 Constrained MMSE .................................................. 30
3.3 Analytical performance comparison between
the CMMSE and the MMSE ............................................. 35
3.4 Simulation results .................................................... 38
  3.4.1 Example 1 ....................................................... 38
  3.4.2 Example 2 ....................................................... 43
  3.4.3 Example 3 ....................................................... 44
3.5 Conclusions .......................................................... 46

4 Constrained MMSE Estimator for Distribution Mismatch Compensation of
GMM-Distributed Random Vectors 48

4.1 Introduction .......................................................... 48
4.2 PDF constrained MMSE for GMM distributed random vectors .... 50
4.3 Simulation results .................................................... 55
  4.3.1 Example 1 ....................................................... 56
5  CMMSE-based single-trial EEG estimation  
5.1 Introduction  
5.2 GMM-CMMSE-based single-trial EEG estimation  
5.3 Conclusions  

6  Summary and Future Work  
6.1 Summary of main results  
6.2 Further study
Glossary of abbreviations

AERP    Auditory Event-Related Potential
AR      AutoRegressive
BAEP    Brain Auditory Evoked Potentials
BCI     Brain Computer Interface
CD-HMM  Continuous Density Hidden Markov Model
CMMSE   Constrained Minimum Mean Square Error
EEG     ElectroEncephaloGram
EOG     ElectroOculoGram
EM      Expectation-Maximization
EP      Evoked Potential
ERD     Event-Related Desynchronization
ERP     Event-Related Potential
ERS     Event-Related Synchronization
fMRI    Functional Magnetic Resonance Imaging
GMM     Gaussian Mixture Model
HCI     Human Computer Interface
HMM     Hidden Markov Model
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICA</td>
<td>Independent Components Analysis</td>
</tr>
<tr>
<td>KLT</td>
<td>Karhunen-Loève Transform</td>
</tr>
<tr>
<td>LORETA</td>
<td>LOw Resolution brain Electromagnetic Tomography</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Predictive Coefficients</td>
</tr>
<tr>
<td>MEG</td>
<td>MagentoEncephaloGraphy</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Networks</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle Component Analysis</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>PMC</td>
<td>Parallel Model Combination</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>WSS</td>
<td>Wide Sense Stationary</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>MAP</td>
<td>Maximum <em>A-posteriori</em> Probability</td>
</tr>
<tr>
<td>MV</td>
<td>Minimum Variance</td>
</tr>
<tr>
<td>MVDR</td>
<td>Minimum Variance Distortionless Response</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
</tbody>
</table>
List of Figures

2.1  Block diagram of the preprocessing and feature extraction procedure.  . . . . 11

2.2  The PHMM concept. Each of the two EEG/ERP HMMs is constructed of
     $L$ HMMs connected in parallel. The selected model is the one whose HMM
     provides the highest likelihood.  . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2.3  A block diagram of the proposed classification system. Upper part- training
     algorithm in which features are extracted from each signal to be saved in mem-
     ory (middle part) for later processing. Lower part- testing algorithm in which
     the features of the candidate signal are estimated and classified to one of the
     EEG/ERP HMMs based on the ML criterion.  . . . . . . . . . . . . . . . . . . . 14

2.4  An example of a random warping function. $y$ axis- trial frame number axis, $x$
     axis- reference frame number. The graph represents scaling of the trial frames
     axis comparing to the reference frames axis.  . . . . . . . . . . . . . . . . . . . 16

2.5  An example of synthesized EEGs generated from real averaged EEG. The SNR
     in this case is -5dB.  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
2.6 Classification rate as a function of SNR for the artificial EEG database. The feature vector consisted of 10 linear filter bank coefficients, log energy and its derivative.

2.7 PDFs for two-dimensional filter bank coefficients.

2.8 Five states HMMs for the two EEG classes and their transition probabilities.

3.1 Ratio between the RMSE of the constrained and unconstrained estimators, comparing to the theoretical threshold, as a function of $\rho_z$. The SNR is 0dB and $\rho_x = 1$. The theoretical threshold where the CMMSE estimator is expected to outperform the CMMSE estimator, is equal to one.

3.2 Performance comparison of the CMMSE and MMSE estimators under noise variance mismatch, with respect to the unconstrained estimator without noise variance mismatch.

3.3 Ratio between the RMSE of the CMMSE and MMSE estimators as a function of the variance uncertainty parameters, $\rho_x$ and $\rho_w$.

3.4 Ratio between the RMSE of the CMMSE and MMSE estimators for different SNRs, as a function of the number of samples, when the variances are estimated from another data set.

3.5 Performance comparison of the LMMSE, Minimax difference, Minimax ratio and the CMMSE estimators. The signal variances are estimated based on 10 data samples.
3.6 Performance comparison of the LMMSE, Minimax difference, Minimax ratio and the CMMSE estimators. The parameter of interest is assumed to be Gaussian, while its true distribution is Laplacian.

4.1 An example of the estimated signal distribution for the MMSE, the GMM-MMSE and the GMM-CMMSE estimators, compared to the original distribution of the parameter of interest. The original parameter of interest is a $3^{rd}$-order GMM random process.

4.2 Performance comparison of $3^{rd}$-order GMM-CMMSE, $3^{rd}$-order GMM-MMSE, $2^{nd}$-order GMM-CMMSE, $2^{nd}$-order GMM-MMSE, CMMSE and MMSE, as a function of SNR. The parameter of interest is a $3^{rd}$-order GMM random process.

4.3 Performance comparison of the Minimax difference, Minimax ratio, Plug-in MMSE, CMMSE, $2^{nd}$-order GMM-MMSE and the $2^{nd}$-order GMM-CMMSE estimators. The parameter of interest is a $3^{rd}$-order GMM random process.

5.1 Amplitude distribution of the spontaneous EEG and response-related EEG of channel c3 from the 1st subject.

5.2 An example of the original response-related EEG acquired from channel c3 of the 1st subject, noisy EEG measurements and the estimated response-related EEG.
5.3 RMSE for the estimated response-related EEG using the CMMSE and the 3\textsuperscript{rd}-order GMM-CMMSE estimators in comparison to the Wiener filter and the ICA-based estimator.
# List of Tables

2.1 Comparison of the classification error rates for DB Ia . . . . . . . . . . . . . . 22
2.2 Comparison of the classification error rates for DB III . . . . . . . . . . . . . . 23
2.3 Classification error rates for different feature sets (DB III) . . . . . . . . . . . 23
Chapter 1

Introduction

1.1 Background

Electroencephalography (EEG) is the scientific study and analysis of electrical fields of the brain (topography, polarity, and changes over time) recorded by amplifying voltage differences between electrodes placed near (scalp EEG), upon (cortical EEG), or within (depth EEG) the brain [1]. Systematic research on the EEG can be traced back to 1875, to the work of Richard Caton. In 1929, Hans Berger discovered the alpha rhythm in the human EEG. Ahead of his time, Berger predicted that the EEG would provide a “window on the mind” useful for treating psychiatric disorders [2]. In 1938, EEG pioneers Albert Grass and Frederick Gibbs concluded, “After having made (Fourier) transforms of 300 electroencephalograms, we are convinced that the system not only expresses data in a manner more useful and concise than is possible by present methods, but that in many cases it indicates important changes in the
electroencephalogram which would otherwise remain hidden” [3].

Since the discovery of Berger, interest in developing algorithms to process human brain and mind information through the use of scalp-recorded brain potentials has emerged. Ongoing EEG and Event-Related Potentials (ERPs) have been shown to be related to many psychological processes involved in attention, memory, and motor control. Moreover, both types of signals have been proven to be useful in classification of different mental and cognitive tasks. Consequently, the idea of developing a Brain Computer Interface (BCI), i.e. to use the brain electrical signals in order to control remote devices, has emerged (e.g. [4–6])). In addition, other techniques such as magnetoencephalography (MEG), positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), which provide important information are expensive and complex to operate. Thus, EEG monitoring has become an important tool for the study of cognitive processing in human subjects (e.g. [7,8]).

This research draws together work from a diverse range of disciplines with common interest is the incorporation of electrophysiological information into the human-machine relationship. For this purpose it is required to develop methods for single-trial EEG processing. Various methods have been proposed for single-trial EEG processing. Averaging is the most commonly used method in cognitive studies. The method is very simple, but removes the information related to variations observed in single-trials, resulting in loss of important inter-trial information. Other methods include ensemble of SVMs [9–11], wavelet transform [12,13], Kalman filter [14], particle filtering [15] and spatio-temporal filtering [16], and Independent Component Analysis (ICA) [17–19]. Unfortunately, these techniques fail to estimate single-trial EEGs and
CHAPTER 1. INTRODUCTION

ERPs because of the very low SNR, the non-Gaussian and stochastic nature of the EEG, and the inter-trial variability of the recorded signals.

In this work, novel methods for single-trial EEG estimation and classification are developed. The first part of this dissertation is aimed at analyzing ERPs elicited during information processing tasks. The goal is to develop a model-based approach for ERP estimation and classification, in particular for BCI applications. The approach is based on parallel hidden Markov models (PHMMs) and a maximum-likelihood (ML)-based decision rule. The performances of the proposed algorithm are studied using an artificial database of EEG patterns and two real EEG databases. The results show that the proposed algorithm outperforms other commonly used methods, with an improvement of 2% and 10%, in the classification rates for the first and second databases, respectively, comparing to the best reported method.

The problem of single-trial EEG estimation led to a more general aspect in signal processing—the problem of Bayesian estimation in the presence of signal distribution mismatch. This problem is addressed in the second part of this dissertation. A new estimator is derived based on the MMSE criterion with constraints on the first- and second-order statistics of the parameters of interest. The constrained MMSE (CMMSE) estimator provides a suboptimal solution, which guarantees that the first- and second-order statistical properties of the estimated signal match those of the parameter of interest. The CMMSE estimator is found to be robust to signal distribution mismatch, since it incorporates first- and second-order statistical information on the parameters of interest and restricts the solution space accordingly. The performance of the CMMSE estimator under different mismatch conditions is studied both analytically and
via simulations using several examples. It is shown that with no distribution mismatch, the CMMSE performance is slightly lower than the optimal MMSE, while in the presence of signal distribution mismatch, the CMMSE estimator outperforms the MMSE estimator and several other commonly used estimators.

The idea of the constrained estimator is further extended to an estimator which incorporates constraints on the probability density function (PDF) of the parameter of interest. The statistical constraints are based on Gaussian mixture model (GMM) representation of the PDF. The GMM-based constrained estimator (GMM-CMMSE) provides a suboptimal solution, which guarantees that the statistical properties of the estimated signal match those of the parameter of interest. Since a closed form of this estimator is difficult to obtain, an approximated GMM-CMMSE is proposed. This estimator is based on a weighted sum of linear CMMSE estimators. The approximated GMM-CMMSE estimator is robust to signal distribution mismatch, since it incorporates statistical information on the parameters of interest and restricts the solution space accordingly. The performance of the approximated GMM-CMMSE estimator under different mismatch conditions is studied via simulations. It is shown that with no distribution mismatch, the GMM-CMMSE performance is slightly lower than the optimal MMSE. However, in the presence of signal distribution mismatch, the GMM-CMMSE estimator is superior to the MMSE estimator and several other commonly used estimators. Furthermore, the new estimator is easy to implement since it can be represented by a weighted sum of linear CMMSE estimators.

The CMMSE and GMM-CMMSE estimators are employed for single-trial EEG estima-
Some simulation experiments are performed using real EEG signals acquired from seven subjects while performing five mental tasks. The performance of the CMMSE and the GMM-CMMSE estimators are compared with those of an ICA-based estimator and the Wiener filter. It is shown that the CMMSE and the GMM-CMMSE estimators outperform both the ICA-based estimator and the Wiener filter in most of the illustrated SNRs.

1.2 Outline

The dissertation is arranged as follows: Chapter 2 introduces a parallel HMM-based algorithm for classification of different EEG patterns. Chapter 3 addresses the general problem of Bayesian estimation in the presence of signal distribution mismatch. A novel estimator, CMMSE, is derived based on the MMSE criterion with constraints on first- and second-order statistics of the parameters of interest. The idea of the constrained estimator is further extended in Chapter 4, which presents a CMMSE with constraints on the parameters’ PDF. In Chapter 5 we employ the CMMSE estimator and the GMM-CMMSE estimator for the problem of single-trial EEG estimation. In Chapter 6, the work is summarized, and possible subjects for further investigation are proposed.
Chapter 2

A Parallel HMM for the Classification of Multichannel EEG signals

2.1 Introduction

Automatic classification of single-trial EEG patterns has received increasing attention in a wide range of biomedical engineering recently. EEG provides a potential nonmuscular communication channel for severely disabled persons, such as those suffering from amyotrophic lateral sclerosis (ALS) or locked-in syndrome, since some mental tasks yield distinguishable EEG signals that can be used to control an assistant device, i.e. BCI [20]. Consequently, the idea of developing a BCI, i.e. to use the brain electrical signals in order to control remote devices, has emerged (e.g. [4–6]). On the other hand, other useful techniques such as MEG, PET and fMRI, are slow, expensive and complex to operate. As a result, the EEG signals have
become an important tool for the study of cognitive processing in human subjects (e.g. [7,8]).

The feasibility and reliability of this communication heavily depends on robust and accurate recognition of the EEGs corresponding to respective mental processes. In the last twenty years, considerable amount of work has been invested in developing methods for EEG and ERP analysis. Research in this field has faced various engineering and physiological problems. These include, among others, low SNRs, overlapping spectra between the ERP and the spontaneous EEG, high intrasubjects and intersubjects variability, signals characterized by nonstationarity and nonergodicity, unknown number and location of intracranial sources and unknown intrabrain wave propagation channels.

Various types of signals have been used in BCI research. Among them are Mu rhythm [21], steady-state visual evoked potentials [22], single-trial ERPs based on combining gamma-band power with slow cortical potentials [23], movement-related cortical potentials and event-related (de)synchronization [7,24], and rhythmic macroscopic EEG [25]. Generally, the types of signals used in this field can be categorized into: oscillatory EEG components, slow cortical potentials and event-related potentials.

Classification of various mental tasks based on EEG and ERP analysis, has been extensively investigated. Various methods have been proposed, such as common spatial subspace decomposition [26], auto-regressive with exogenous input (ARX) parametric modeling [5], support vector machine (SVM) [4,9,10] and local temporal common spatial patterns [6]. As none of the methods are optimum, research in this field has been continued.

Most of the research works in this field have been aimed at analyzing the EEGs recorded in
one or two channels, independently of the other channels. Analysis of multichannel EEGs show that different channels reflect different aspects of the brain activity due to the presentation of an external stimulus. Therefore, utilizing the mutual information among the different channels may add crucial interchannel information and consequently improve the classification performance.

In [27], parametric multichannel fusion models were introduced in order to exploit the complementary brain activity information recorded from multiple channels. The decision fusion model combines the independent decisions of each channel classifier into a decision fusion vector and a parameteric classifier is designed to determine the EEG pattern from the discrete decision fusion vector.

In this chapter, a model-based approach for multichannel classification of different EEG patterns is introduced. The approach is based on a parallel combination of HMMs. The performances of the proposed algorithm are studied using an artificial EEG database and two real databases. The results show that the proposed algorithm outperforms other commonly used methods, with an improvement of 2% and 10%, respectively, in the classification rates for the first and second EEG databases, respectively, comparing to the best reported method. The approach is applicable to various types of EEG patterns such as ERPs as well as to any multisensor signals.

This chapter proceeds as follows. Section 2.2 discusses the rationale of the PHMM-based approach and the features employed by the classification system. The results of the classification experiments are summarized in Section 2.3. The conclusions of this chapter appear in
2.2 HMM-based classification of EEG patterns

HMMs have been used extensively and successfully to represent various types of signals. The main idea of using HMMs to represent different EEG patterns relies on the fact that HMM can model the nonstationarity of the EEG signal and provide automatic dynamic time warping, thus overcoming the difficult problem of time and morphology variability of EEGs. It is shown that this temporal dynamic can be well represented by different states of the HMM.

The next subsections present the features employed by the classification algorithm, the classification scheme and the training and recognition algorithms.

2.2.1 Features

The classification scheme is performed in the feature space. The goal of the feature extraction process is to represent the EEG signal in the feature space.

ERPs and other EEG changes associated with external stimulus are often characterized by its time domain components: positive and negative peaks in some latencies of the signal. Generally, the EEG temporal dynamics is one of the most important discriminative features between different types of EEG patterns. Therefore, classification of different EEG patterns has been commonly performed based on the detection of positive and negative peaks at specific latencies. This implies that features, which are based on the signal amplitude, energy and
their derivatives, may be good discriminative features and are therefore employed by the proposed classification algorithm. Other features, which we considered due to their success in many other classification problems, are linear predictive coefficients (LPC), LPC-based features such as LPC-derived cepstrum (LPCC), partial correlation coefficients (PARCOR) and linear filter bank coefficients [28]-[29].

The feature extraction process is performed as follows. First, the signal is divided into quasi-stationary overlapping frames and each frame is multiplied by a Hamming window. Next, the first \( p + 1 \) autocorrelation coefficients are estimated from each windowed frame of the signal, and a vector of LPC coefficients is calculated for each windowed frame using the Levinson-Durbin recursive algorithm [30]. The LPC coefficients are used to calculate other static features such as LPCC, PARCOR and LAR. Other static features, such as mean frame energy and log energy, are estimated directly from the time domain signal. The dynamic features of the signal are estimated using a first-order orthogonal polynomial approximation over a finite length window with the following formula [29]:

\[
\Delta F(l, m) = \frac{\sum_{k=-K}^{K} k F(l - k, m)}{\sum_{k=-K}^{K} k^2}, \quad 1 \leq m \leq Q
\]  

(2.1)

where \( F(l, m) \) is the static feature, \( m \) is the feature index, \( Q \) is the order of the feature set and \( l \) is the frame number. \( K \) was chosen in this work to be 2. \( \Delta F(l, m) \) is referred to as the “Del” of the relevant feature. The Del Del of the static feature is also computed by applying (2.1) on the Del feature.
Fig. 2.1 presents a block diagram of the feature extraction procedure. In this figure, \( N_f \) represents the number of samples in each frame and \( M_f \) represents the number of overlapped samples between consecutive frames.

![Block diagram of the preprocessing and feature extraction procedure.](image)

2.2.2 Classification scheme

The task of the classification system is to automatically distinguish among observations of different classes (types). In supervised classification \textit{a-priori} statistical knowledge of the features of each class is available [31].

Most of the EEG classifiers proposed in previous studies, have been concerned with classification of the data extracted from one or two channels. When using data from multiple channels, recorded from different scalp electrodes, it is essential to allow a different represen-
tation for each channel. According to the PHMM approach this is obtained by assigning a different HMM for each of the class channels, forming a generalized model for each class. The PHMM concept is presented in Fig. 2.2, which refers to the common case of two classes of two EEG/ERP patterns.

![Diagram of PHMM concept](image)

Figure 2.2: The PHMM concept. Each of the two EEG/ERP HMMs is constructed of $L$ HMMs connected in parallel. The selected model is the one whose HMM provides the highest likelihood.

2.2.3 The training algorithm

During the training phase, the supervised classification system is presented with observations of each channel, in each class, and accordingly estimates the parameters of the HMM which represent the channel in the class. The HMM parameters are initially estimated using the
k-means algorithm [28], and then the Baum-Welch algorithm [28]-[29] is employed until a predefined convergence criterion is met.

### 2.2.4 The recognition algorithm

During the recognition phase, the likelihood of an observation sequence to be generated from each one of the HMMs is estimated based on the highest likelihood criterion. The selected model is the one whose HMM provides the highest likelihood. The ML decision rule is therefore given by:

$$m = \arg \left\{ \max_{l=1,\ldots,L} \left( \max_{s=1,\ldots,S} \left( P(O|\lambda_{ls}) \right) \right) \right\}$$

(2.2)

where $l^{th}$ represents the class index, $s^{th}$ represents the index of the HMM within the $l^{th}$ class, and $\lambda_{ls}$ represents the $s^{th}$ HMM of the $l^{th}$ class. $P(O|\lambda_{ls})$, the conditional PDF of the observation vector, $O$, given the model $\lambda_{ls}$, is estimated using the Viterbi algorithm [28].

Fig. 2.3 shows a block diagram of the PHMM-based classification system. The PHMM achieves better representation for each class using several HMMs, one for each channel. The main problem with this approach is that it requires a relatively large training set in order to allow reliable parameter estimation for each channel model. This problem can be solved by employing an adaptation algorithm, where instead of training each channel model using the data of the particular channel, the whole training dataset of the class is used to train each model. Then, the parameters of each one of the channel models are adapted using the channel dataset. In this way, less data are required to estimate each model, because these data are
used for adaptation and not for the initial estimation of the model parameters.

Figure 2.3: A block diagram of the proposed classification system. Upper part- training algorithm in which features are extracted from each signal to be saved in memory (middle part) for later processing. Lower part- testing algorithm in which the features of the candidate signal are estimated and classified to one of the EEG/ERP HMMs based on the ML criterion.
CHAPTER 2. A PHMM FOR THE CLASSIFICATION OF MULTICHANNEL EEGS

2.3 Experimental results

This section presents the experiments carried out in order to evaluate the performance of the PHMM-based approach. The first subsection presents the results obtained in the simulation. The results of the experiments with real EEG data are reported in the second subsection.

2.3.1 Simulation

In order to evaluate the performance of the proposed algorithm under different SNRs (defined by $SNR = \frac{\sigma_x^2}{\sigma_w^2}$, where $x$ and $w$ represent the ERP and the ongoing EEG, respectively) and with different parameters, an artificial ERPs database was created. The basic EEG patterns used in these simulations were formed out of two averaged EEGs from two classes of a real EEG database. Then, each trial was created by performing nonlinear time warping on a real averaged EEG and adding a zero-mean white Gaussian noise with variance $\sigma^2$, such that each trial is a noisy time-scaled version of an averaged EEG. Fig. 2.4 shows an example of the warping function used to create the time-scaled EEG.

The artificial database was divided into training and testing sets, each consists of 500 trials. A left-to-right continuous density HMM with 5 states and 3 Gaussian mixtures per state, with a feature vector consisting of 10 linear filter bank coefficients, log energy and its derivative, was applied for the data classification. Fig. 2.5 presents an example of synthesized ERPs generated in the case of $SNR=-5$dB.

Fig. 2.6 presents the classification rates as a function of SNR. It can be seen that for SNRs
smaller than -5dB, the correct classification rate is around 74%. For SNRs greater than 0dB, the classification rate increases rapidly from about 82% for an SNR of 0dB, to about 91% for an SNR of 5dB.

2.3.2 Real data experiments

Two real EEG databases were used in this work. These databases were published in [32] during the 2003 BCI competition. In this competition, the participant research groups were given a labeled training dataset and an unlabeled testing dataset, and were asked to provide
classification decision on each and every signal in the testing dataset. It should be noted that at the time of performing the experiments reported here, the labels of the testing dataset were already available. However, in order to ensure that the experiments are performed in the same conditions as in the BCI competition, only the labeled dataset was used to adjust the classifier’s parameters, while the classifier performance was evaluated based on the ‘unlabeled’ dataset.

The first database (marked as DB Ia) consists of two classes of EEGs elicited during a task of imagery of upward and downward movements of a computer screen cursor [33]. The
Figure 2.6: Classification rate as a function of SNR for the artificial EEG database. The feature vector consisted of 10 linear filter bank coefficients, log energy and its derivative.

data were acquired from a single healthy subject at the University of Tuebingen, Germany, as described in [34]. The subject was asked to move a cursor up and down on a computer screen, while his/her cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Cortical negativity led to an upward movement of the cursor, while cortical positivity led to a downward movement of the cursor on the screen. Six EEG electrodes were all referenced to the vertex electrode Cz as follows: Channel 1- channels A1-Cz (international 10-20 system (A1-left mastoid)), channel 2- A2-Cz, channel 3- 2cm frontal of C3, channel 4- 2cm parietal of C3, channel 5- 2cm frontal of C4,
CHAPTER 2. A PHMM FOR THE CLASSIFICATION OF MULTICHANNEL EEGS

channel 6- 2cm parietal of C4. The signals were acquired using a 16 bits A/D converter (Computer Boards PCIM-DAS1602) at a sampling frequency of 256Hz. The trials structure was as follows: 6sec inter-trial intervals, task presentation from 0.5sec to 6.0sec and feedback period from 2.0sec to 5.5sec. Only the signals from the feedback phase were provided such that each trial contained 896 samples for each channel. The training labeled data set consists of 268 trials and the validation set consists of 293 trials.

The second database (marked as DB III) consists of two classes of sensimotor EEGs elicited during a feedback-regulated left-right motor imagery task [34].

The data were provided by the Department of Medical Informatircs, Institute for Biomedical Engineering, University of Technology, Graz, Austria. The EEG signals were acquired from a single healthy subject during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left or right hand movements. The order of left and right cues was random. The experiment included 7 runs with 40 trials each. All runs were conducted on the same day with few breaks of several minutes each. The signals were bandpass filtered between 0.5Hz and 30Hz and were sampled at 128Hz. The EEG from three channels (C3, Cz, C4) was acquired using G.tec amplifier and a Ag/AgCl electrodes. Three bipolar EEG channels (anterior '+', posterior '-') were measured over C3, Cz and C4. The signals were sampled at 128Hz and filtered between 0.5 and 30Hz.

The data set consists of 140 labeled and 140 unlabeled trials, with an equal number of left and right hand trials. Each trial has a 9sec duration; after a 3sec presentation period a visual cue (arrow) is presented pointing either to the left or to the right. This is followed by another
6sec period for performing the imaginary task.

In the first classification experiment, a left-to-right PHMM with 6 data channels, 5 states per HMM and 3 Gaussians for each state, were trained using 10 single-trial EEGs. The feature vectors were constructed from 10 linear filter bank coefficients.

Fig. 2.7 shows the two-dimensional-PDF of the first two filter bank coefficients for the two EEG classes of DB III. Although the PDFs of the two classes are different, the overlap between them indicates that more than two filter bank coefficients are needed to reliably distinguish between the two classes in the two dimensional subspace. It should be noted that similar results were obtained for DB Ia.

Fig. 2.8 shows the left-to-right HMMs and the transition probabilities, which were estimated using the Baum-Welch algorithm [28]-[29]. The transition probabilities are denoted by $a_{ij}$ for each couple of states $\{i, j\}$. It can be seen that transitions between successive states occur in non-zero probability.

Tables 2.1 and 2.2 present the classification error rates of the PHMM classifier comparing to the rates obtained using several methods. Detailed description of these methods was provided only for the winning methods and can be found in [23] and [25], for the first and second databases, respectively. Note that only the methods with the four lowest error rates are mentioned in these tables. The results show that the PHMM classifier provides the lowest error rates, with an improvement of 2% and 10%, for databases DB Ia and DB III, respectively, comparing to the best reported method.

The second experiment was aimed at evaluating the classification rates for different feature
Figure 2.7: PDFs for two-dimensional filter bank coefficients.
CHAPTER 2. A PHMM FOR THE CLASSIFICATION OF MULTICHANNEL EEGS

Figure 2.8: Five states HMMs for the two EEG classes and their transition probabilities.

Table 2.1: Comparison of the classification error rates for DB Ia

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>Error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-to-right HMM with 5 states and 3 Gaussian/state (the current work)</td>
<td>10 linear filter bank coefficients, log energy and Δ log energy</td>
<td>9</td>
</tr>
<tr>
<td>Discriminant analysis (Brett and Mensh)</td>
<td>DC-level and features from spectral analysis of high beta power band</td>
<td>11</td>
</tr>
<tr>
<td>Support vector machines (Ka-Min and Chung)</td>
<td>Averaged EEGs amplitudes after downsampling to 25 Hz</td>
<td>12</td>
</tr>
<tr>
<td>Nonlinear support sector Machines (Tzu-Kuo and Huang)</td>
<td>Time series features after linear support vector machines</td>
<td>15</td>
</tr>
</tbody>
</table>

sets. Various combinations of features were considered. Table 2.3 summarizes the results obtained for DB III. The results show that in this case, the best combination of features includes 10 linear filter bank coefficients, log energy and Δlog energy. It is shown that this feature set is much better than a combination of LPC and ΔLPC, for instance, which is still
CHAPTER 2. A PHMM FOR THE CLASSIFICATION OF MULTICHLANNELE EEGS

Table 2.2: Comparison of the classification error rates for DB III

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>Error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-to-right HMM with 5 states and 3 Gaussian/state (the current work)</td>
<td>10 linear filter bank coefficients, log energy and $\Delta$ log energy</td>
<td>1</td>
</tr>
<tr>
<td>Bayes classifier (Lemm et al.)</td>
<td>Morlet-wavelets at 10 and 22 Hz in channels C3 and C4</td>
<td>11</td>
</tr>
<tr>
<td>Linear discriminant analysis (Akash and Narayana)</td>
<td>Ratio of AR spectral power in 4 frequency bands, channels C3 and C4</td>
<td>12</td>
</tr>
<tr>
<td>Several neural networks trained on different time regions (Saffari)</td>
<td>Adaptive AR parameters (no details were given)</td>
<td>17</td>
</tr>
</tbody>
</table>

in common use in BCI applications.

Table 2.3: Classification error rates for different feature sets (DB III)

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 enrg. filter bank + log enrg. + $\Delta$log enrg.</td>
<td>1</td>
</tr>
<tr>
<td>10 enrg. filter bank</td>
<td>3</td>
</tr>
<tr>
<td>10 enrg. filter bank + 10 $\Delta$enrg. filter bank</td>
<td>14</td>
</tr>
<tr>
<td>10 LPC</td>
<td>49</td>
</tr>
<tr>
<td>10 LPC + log enrg. + $\Delta$log enrg.</td>
<td>30</td>
</tr>
<tr>
<td>10 LPC + 10 $\Delta$LPC</td>
<td>56</td>
</tr>
<tr>
<td>10 LPCC</td>
<td>26</td>
</tr>
<tr>
<td>10 LPCC + 10 $\Delta$LPCC</td>
<td>48</td>
</tr>
<tr>
<td>10 LAR</td>
<td>51</td>
</tr>
<tr>
<td>10 PARCOR</td>
<td>53</td>
</tr>
<tr>
<td>10 PARCOR + 10 $\Delta$PARCOR</td>
<td>44</td>
</tr>
<tr>
<td>10 PARCOR + log enrg. + $\Delta$log enrg.</td>
<td>25</td>
</tr>
</tbody>
</table>
2.4 Conclusions

In this chapter, a model-based approach for multichannel EEGs classification is introduced. The approach is based on PHMMs and an ML-based decision rule. The proposed algorithm is studied using an artificial EEG database and two real EEG databases. In the case of artificial EEG database, it was shown that for SNRs smaller than -5dB, the correct classification rate is around 74%. For SNRs greater than 0dB the classification rate increases rapidly with around 82% corrected classification rates for an SNR of 0dB, and around 91% for an SNR of 5dB.

The results show that in this case, the best combination of features includes 10 linear filter bank coefficients, log energy and Δlog energy. It is shown that this feature set is much better than a combination of LPC and ΔLPC, for instance, which is still in common use in BCI applications.

The results show that the PHMM classifier provides the lowest error rates, with an improvement of 2% and 10%, for databases DB Ia and DB III, respectively. It is shown that the PHMM-based classifier with a left-to-right architecture without skips is capable of representing different types of EEGs and that it performs better than other commonly used classifiers. Moreover, a feature set which consists of linear filter bank coefficients with log energy and Δlog energy, is shown to provide the best classification rates, outperforming other feature sets such as LPC and LPC-derived cepstrum.

The results also show that the classification performances strongly depend on SNR. It is therefore required to develop methods for single-trial EEG estimation in order to improve the
classification performances. The next two chapters address the problem of single-trial EEG estimation.
Chapter 3

Constrained MMSE Estimator for Distribution Mismatch Compensation

3.1 Introduction

In the previous chapter, classification of different EEG patterns was addressed. As mentioned above, most of the response-related EEG patterns are embedded in the spontaneous EEG signal and are characterized by low SNR. Since the classification performance strongly depends on SNR, it is required to estimate the response-related EEG signal before classification. It is postulated that \textit{a-priori} knowledge of the statistics of the parameter of interest (i.e. the response-related EEG), or of the nuisance parameter (i.e. the spontaneous EEG) may allow to obtain a reliable estimate of the original signal. Usually, the statistics are not known \textit{a-priori}, but can be estimated from the available data. For instance, the EEG statistics can be
The problem addressed in this chapter, as described follows. It is required to estimate an unknown random vector given an observation vector and a conditional PDF while conserving the statistics of the parameter of interest. The MMSE estimator [38] provides an optimal estimate by means of the mean-square error (MSE). However, it does not guarantee that the statistics of the estimated signal match those of the original signal. Hence, the resulting estimate might be inappropriate even though it satisfies the MMSE criterion. Furthermore, the MMSE requires prior knowledge of the conditional statistics of the parameters of interest given the observations. In many practical applications, the noise statistics are not completely specified or are a-priori unknown and therefore have to be estimated from a finite observation sequence. Even when the signal statistics are known, the MMSE estimator might be difficult to implement, unless some specific distributions are assumed. In such cases, signal distribution mismatch is introduced and consequently, the MMSE estimator performance may significantly degrade.

Several non-Bayesian methods have been proposed to handle the problem of distribution mismatch. One of the commonly used methods is the minimax approach [39]– [40]. According to this approach, the parameter of interest is estimated based on a minimax MSE criterion. The main disadvantage of this approach is that it is based on optimization of the worst case scenario on the expense of other cases.

A Bayesian approach for distribution mismatch compensation has been proposed in [41]. According to this method, the parameter of interest is treated as a random vector and it is
assumed that its covariance matrix is a-priori known up to a specific range of uncertainty. This method is based on minimization of the worst-case MSE ratio regret criterion between the MSE attainable using a linear estimator and the MMSE attainable by optimum linear estimation with a known signal covariance. The main disadvantage of this method is that it is based on the assumption that the estimator is linearly dependent on the observation vector. Furthermore, the method is highly dependent on an appropriate choice of the uncertainty parameters range.

In [42] a constrained maximum a-posteriori probability (MAP) estimator has been proposed. This estimator assumes that the parameter of interest $x$ satisfies $x \in X$, where $X$ is a known convex polytope and that $x$ is a random vector with prior Gaussian distribution. Furthermore, the estimator assumes a linear regression model.

Other estimators proposed in the literature to handle the distribution mismatch problem include the Tikhonov regularization [43], the ridge estimator [44], the Shrunken estimator [45] and the covariance shaping least-squares (LS) estimator [46]. The common idea of these estimators is introducing appropriate design parameters based on a-priori knowledge of the statistics of the parameter of interest and employing the LS estimates. The LS criterion is therefore optimized using such estimators, but the MSE of the resulting estimate might be relatively large. Moreover, the LS solution is often based on a regression model assumption which in many cases does not present the data model.

In this chapter, a CMMSE estimator is derived. The CMMSE estimator incorporates a-priori statistical information on the parameters of interest and restricts the solution space
accordingly. Therefore, assuming that this *a-priori* information is accurate, the sensitivity to noise distribution mismatch is reduced, and therefore the robustness of the estimator to distribution mismatch is greater.

In contrary to most of the estimators proposed in the literature to handle the problem of distribution mismatch, the proposed estimator does not enforce any restrictions on the signals distributions or on the relationship between the measurements and the estimated signal which might degrade the performances. The estimator conserves the first- and second-order statistical properties of the original signal.

The performance of the CMMSE estimator is studied both analytically and via simulations using several examples, in comparison to several other estimators. It is shown that in the absence of mismatch, the performance of the CMMSE is slightly lower than the optimal MMSE performance. However, in the presence of distribution mismatch, the CMMSE estimator significantly outperforms the mismatched MMSE estimator and the other estimators.

This chapter proceeds as follows. Section 3.2 presents the proposed constrained MMSE and its properties. In Section 3.3 an analytical performance comparison between the CMMSE and the MMSE is presented. The performance of the CMMSE is evaluated and analyzed in Section 3.4. The conclusions of the chapter appear in Section 3.5.
3.2 Constrained MMSE

The unbiased property of the MMSE estimator implies that the first-order statistics of the MMSE estimate are equal to the first-order statistics of the parameters of interest. However, the second-order statistics of the MMSE estimator do not necessarily match the true statistics. Consider, for instance, the following scenario: \( z = x + w \), where \( z \) is the observation vector, and the signal vector \( x \sim N(\mu_x, C_x) \) and the noise vector \( w \sim N(\mu_w, C_w) \) are statistically independent. The MMSE estimator of \( x \) from the observation vector \( z \), is given by \( \hat{x}_M = \mu_x + C_x (C_x + C_w)^{-1} (z - \mu_x) \), and the MMSE covariance matrix is given by \( C_M \triangleq \text{cov}(\hat{x}_M) = C_x (C_x + C_w)^{-1} C_x \). Thus, the MMSE criterion does not guarantee that the statistical properties of the estimated signal match those of the original signal. For example, suppose that the elements of \( x \) and \( w \) represent white and colored random processes, respectively. It can be easily verified that the MMSE estimate of \( x \) is a colored random process. Likewise, in the presence of nonstationary noise, the MMSE estimator of \( x \) might provide a nonstationary random process while the original signal, \( x \), is a wide-sense stationary (WSS) random process with a Toeplitz covariance matrix. Hence, the statistical distributions of the resulting MMSE estimate and the original signal might not coincide.

It is therefore suggested to incorporate constraints on the first- and second-order statistics of the parameters of interest in the MMSE solution space, and derive a new estimator, termed as constrained MMSE. Subsequently, the first and second-order statistics of the estimated signal match those of the original signal. This constraint is shown to result in an improvement comparing to the mismatched MMSE while in the absence of distribution mismatch, the
proposed estimator performance is slightly lower than the MMSE estimator.

Consider the problem of estimating an unknown random vector $\mathbf{x}$, given an observation vector $\mathbf{z}$ with the PDF $f_{\mathbf{x}|\mathbf{z}}(\mathbf{x}|\mathbf{z})$. The MMSE estimator of $\mathbf{x}$ is given by:

$$\hat{\mathbf{x}}_M = \mathbb{E}_{\mathbf{x}|\mathbf{z}}(\mathbf{x}).$$

(3.1)

The mean and covariance matrix of the MMSE estimator satisfy $\mathbb{E}_z(\hat{\mathbf{x}}_M) = \mathbb{E}(\mathbf{x}) \overset{\triangle}{=} \mu_\mathbf{x}$ and $C_M = \text{cov}(\hat{\mathbf{x}}_M) = \mathbb{E}_z(\text{cov}(\mathbf{x}|\mathbf{z})) \leq \text{cov}(\mathbf{x}) = C_\mathbf{x}$. We shall assume that $C_\mathbf{x}$ and $C_M$ are invertible.

In order to ensure that both the first- and second-order statistics match those of the parameter of interest, $\mathbf{x}$, a constraint is set on the solution space, such that:

$$\mu_\hat{x} \overset{\triangle}{=} \mathbb{E}_z(\hat{x}) = \mu_\mathbf{x},$$

(3.2)

$$C_\hat{x} \overset{\triangle}{=} \text{cov}(\hat{x}) = C_\mathbf{x}.$$  

(3.3)

As shown in the following, the first order statistical constraint given in (3.2) is redundant and can be omitted. Minimization of the MSE with these constraints can be performed using Lagrange-multipliers:

$$G \overset{\triangle}{=} \mathbb{E}_{\mathbf{x},\mathbf{z}}(||\mathbf{x} - \hat{\mathbf{x}}||^2) + \text{tr} (\Lambda (C_\hat{x} - C_\mathbf{x})) + \gamma^T (\mu_\hat{x} - \mu_\mathbf{x}) \rightarrow \min_x$$

(3.4)
CHAPTER 3. CMMSE ESTIMATOR

where \( \Lambda \) and \( \gamma \) are real matrix and vector of Lagrange-multipliers, respectively and \( T \) denotes the transpose operator. It is therefore required to minimize \( G \) with respect to \( \hat{x} \) and then apply the statistical constraints in order to find \( \Lambda \) and \( \gamma \). \( G \) can be rewritten as:

\[
G = E_x \left[ E_{x|z} \left[ (x - \hat{x})^H (x - \hat{x}) \right] \right] + \text{tr} \left( \Lambda \left( E_x \left[ (\hat{x} - \mu_{\hat{x}})(\hat{x} - \mu_{\hat{x}})^H \right] - C_x \right) \right) + \gamma^T (E_x (\hat{x}) - \mu_x),
\]

(3.5)

where the superscript \( H \) denotes the Hermitian operator. \( G \) can be further simplified and rewritten as:

\[
G = E_x \left[ E_{x|z} \left[ (x - \hat{x})^H (x - \hat{x}) \right] + (\hat{x} - \mu_{\hat{x}})^H \Lambda (\hat{x} - \mu_{\hat{x}}) - \text{tr}(\Lambda C_x) + \gamma^T (\hat{x} - \mu_x) \right].
\]

(3.6)

In order to minimize \( G \) with respect to \( \hat{x} \) it is required to minimize the intern term of \( E_{x|z}[\cdot] \) in (3.6) for every \( z \):

\[
\tilde{G}(z) \overset{\triangle}{=} E_{x|z} \left[ (x - \hat{x})^H (x - \hat{x}) \right] + (\hat{x} - \mu_{\hat{x}})^H \Lambda (\hat{x} - \mu_{\hat{x}}) - \text{tr}(\Lambda C_x) + \gamma^T (\hat{x} - \mu_x) \rightarrow \text{min}.
\]

(3.7)

Minimization of \( \tilde{G}(z) \) with respect to \( \hat{x} \) is obtained by equating its gradient to zero:

\[
- E_{x|z}(x - \hat{x}_{CM}) + \Lambda (\hat{x}_{CM} - \mu_{\hat{x}}) + \gamma = 0,
\]

(3.8)
where \( \hat{x}_{CM} \) is the solution to this constrained minimization problem. Hence:

\[
(I + \Lambda)(\hat{x}_{CM} - \mu) = \hat{x}_M - \mu - \gamma,
\]

where \( \hat{x}_M = \mathbb{E}_{x|z}(x) \) is the unconstrained MMSE estimator and \( I \) denotes the identity matrix.

By substitution of (3.9) in the constraint in (3.2), one obtains:

\[
\gamma = 0.
\]

This result implies that the constraint in (3.2) is redundant, since the MMSE estimator satisfies the first order statistical constraint and the addition of the second-order statistical constraint does not affect the first constraint.

Employing the second-order statistical constraint presented in (3.3), results in the following equation for \( \Lambda \):

\[
(I + \Lambda) C_x (I + \Lambda)^H = C_M,
\]

where \( C_M \) is the covariance matrix of the MMSE estimator, \( C_M \triangleq \mathbb{E}_z \left( (\hat{x}_M - \mu)(\hat{x}_M - \mu)^H \right) \).

Since \( C_x \) and \( C_M \) are symmetric positive-semidefinite matrices, using Cholesky factorization [47], they can be decomposed as \( C_x = D_x D_x^H \) and \( C_M = D_M D_M^H \), respectively, where \( D_x \) and \( D_M \) are invertible matrices. Therefore, (3.11) can be rewritten as follows:

\[
[(I + \Lambda) D_x] [(I + \Lambda) D_x]^H = D_M D_M^H.
\]
Clearly, there is no unique solution to this equation. One possible solution can be obtained by choosing $\Lambda = D_M D_x^{-1} - I$, $D_M = C_M^{1/2}$ and $D_x = C_x^{1/2}$, which yields:

$$\Lambda = C_M^{1/2} C_x^{1/2} - I. \quad (3.13)$$

Then, the solution to the minimization problem is obtained by substitution of (3.13) in (3.9) and using (3.2):

$$\hat{x}_CM = C_x^{1/2} C_M^{-1/2} (\hat{x}_M - \mu_x) + \mu_x. \quad (3.14)$$

It can easily be verified that the CMMSE estimate is unbiased:

$$E_{x,z}(\hat{x}_CM) = E_{x,z} \left( C_x^{1/2} C_M^{-1/2} (\hat{x}_M - \mu_x) + \mu_x \right) = \mu_x, \quad (3.15)$$

and that it satisfies the second-order statistical constraint:

$$C_{CM} = E_{x,z} \left( (\hat{x}_CM - \mu_x)(\hat{x}_CM - \mu_x)^H \right) =$$

$$E_{x,z} \left( \left( C_x^{1/2} C_M^{-1/2} (\hat{x}_M - \mu_x) \right) \cdot \left( C_x^{1/2} C_M^{-1/2} (\hat{x}_M - \mu_x) \right)^H \right) =$$

$$E_{x,z} \left( C_x^{1/2} C_M^{-1/2} C_M C_M^{-1/2} C_x^{1/2} \right) = C_x.$$

It should be noted that (3.14) indicates that the constrained estimator is obtained by whitening of the MMSE estimator and then “recoloring” it using the a-priori signal covariance matrix while conserving the first-order statistics of the MMSE estimator. Therefore, considering the example mentioned above, of two additive independent random processes $x$ and $w$, it is clear
that in contrary to the MMSE estimator, the CMMSE estimator conserves the second-order statistical properties of the original signal. Thus, if \( x \), for instance, is WSS and has a Toeplitz covariance matrix \( C_x \), then the CMMSE estimate has a Toeplitz covariance matrix and is guaranteed to be WSS.

### 3.3 Analytical performance comparison between the CMMSE and the MMSE

In this section, the theoretical performance of the CMMSE estimator is compared to the MMSE in the presence of distribution mismatch. Let \( \tilde{f} \) and \( f \) denote the presumed and the true PDFs, respectively. The MMSE estimation error is given by:

\[
\epsilon_M = x - \tilde{x}_M, \tag{3.16}
\]

where \( \tilde{x}_M \) is the MMSE estimator, with the presumed PDF, i.e. \( \tilde{x}_M = E_{x|z,f}(x) \). Therefore, the MSE of the estimation matrix, assuming a PDF function \( \tilde{f} \) is given by:

\[
E_{\epsilon_M} \triangleq E_{x,z,f}(\epsilon_M\epsilon_M^H) = E_{x,z,f} \left[ (x - \tilde{x}_M)(x - \tilde{x}_M)^H \right]. \tag{3.17}
\]
The MSE matrix can be rewritten as follows:

\[
\mathbf{E}_{\epsilon_M} = \mathbf{E}_{x,z,f} \left[ (x - \hat{x}_M - (\tilde{x}_M - \hat{x}_M)) (x - \hat{x}_M - (\tilde{x}_M - \hat{x}_M))^H \right] = \\
\mathbf{E}_{x,z,f} \left[ (x - \hat{x}_M) (x - \hat{x}_M)^H + (\tilde{x}_M - \hat{x}_M) (\hat{x}_M - \hat{x}_M)^H - \\
(\tilde{x}_M - \hat{x}_M) (x - \hat{x}_M)^H - (x - \hat{x}_M) (\tilde{x}_M - \hat{x}_M)^H \right].
\] (3.18)

Given that \(\hat{x}_M = \mathbf{E}_{x,z,f}(x)\), the expectation over the last two terms is equal to zero. Consequently:

\[
\mathbf{E}_{\epsilon_M} = \mathbf{E}_{x,z,f} \left[ (x - \hat{x}_M) (x - \hat{x}_M)^H \right] + \mathbf{E}_{x,z,f} \left[ (\tilde{x}_M - \hat{x}_M) (\hat{x}_M - \hat{x}_M)^H \right]. 
\] (3.19)

When no mismatch occurs, the last term of (3.19) equals zero and therefore \(\mathbf{E}_{\epsilon_M} = \mathbf{E}_{x,z,f} \left[ (x - \hat{x}_M) (x - \hat{x}_M)^H \right]\).

The MSE matrix for the CMMSE in the presence of modeling mismatch can be derived in the same manner. Let

\[
\epsilon_{CM} = x - \tilde{x}_{CM},
\] (3.20)

where \(\tilde{x}_{CM}\) denotes the constrained MMSE estimator with the presumed PDF, i.e. \(\tilde{x}_{CM} = \mathbf{C}_x^{1/2} \mathbf{C}_M^{-1/2} (\tilde{x}_M - \mu_x) + \mu_x\). Then, the MSE matrix is calculated as follows:

\[
\mathbf{E}_{\epsilon_{CM}} \overset{\Delta}{=} \mathbf{E}_{x,z,f}(\epsilon_{CM} \epsilon_{CM}^H) = \mathbf{E}_{x,z,f} \left[ (x - \tilde{x}_{CM})(x - \tilde{x}_{CM})^H \right] = \\
\mathbf{E}_{x,z,f} \left[ (x - \hat{x}_M) (x - \hat{x}_M)^H + (\tilde{x}_{CM} - \hat{x}_M) (\hat{x}_M - \hat{x}_M)^H - \\
(\tilde{x}_{CM} - \hat{x}_M) (x - \hat{x}_M)^H - (x - \hat{x}_M) (\tilde{x}_{CM} - \hat{x}_M)^H \right]. 
\] (3.21)
Similar to (3.18), the expectation over the last two terms is equal to zero and consequently:

$$E_{\epsilon_{CM}} = E_{x|z,f} \left[ (x - \hat{x}_M) (x - \hat{x}_M)^H \right] + E_{z,f} \left[ (\hat{x}_{CM} - \hat{x}_M) (\hat{x}_{CM} - \hat{x}_M)^H \right]. \quad (3.22)$$

The last term in (3.22) is positive semi-definite and therefore when no mismatch occurs $E_{\epsilon_{CM}} > E_{\epsilon_M}$, i.e. the CMMSE performs worse than the MMSE, as expected. On the other hand, the CMMSE outperforms the MMSE when:

$$E_{z,f} \left[ (\hat{x}_{CM} - \hat{x}_M) (\hat{x}_{CM} - \hat{x}_M)^H \right] < E_{z,f} \left[ (\hat{x}_M - \hat{x}_M) (\hat{x}_M - \hat{x}_M)^H \right]. \quad (3.23)$$

Let $\tilde{\mu}_x$, $\tilde{C}_x$ and $\tilde{C}_M$ denote the expectation of the parameter of interest, the covariance matrix of the parameter of interest and the covariance matrix of the MMSE estimator, respectively, evaluated given the presumed PDF $\tilde{f}$. Then, substituting (3.14) in (3.23) yields the following condition:

$$E_{z,f} \left[ (\tilde{C}_x^{1/2} \tilde{C}_M^{-1/2} (\tilde{x}_M - \tilde{\mu}_x) + \tilde{\mu}_x - \tilde{x}_M) \left( \tilde{C}_x^{1/2} \tilde{C}_M^{-1/2} (\tilde{x}_M - \tilde{\mu}_x) + \tilde{\mu}_x - \tilde{x}_M \right)^H \right] < E_{z,f} \left[ (\hat{x}_M - \hat{x}_M) (\hat{x}_M - \hat{x}_M)^H \right]. \quad (3.24)$$

The condition can be rewritten as follows:

$$\tilde{C}_x^{1/2} \tilde{C}_M^{-1/2} \tilde{R}_M \tilde{C}_M^{-1/2} \tilde{C}_x^{1/2} - \tilde{C}_x^{1/2} \tilde{C}_M^{-1/2} \tilde{P}_M - \tilde{P}_M^H \tilde{C}_M^{-1/2} \tilde{C}_x^{1/2} < \tilde{R}_M - \tilde{P}_M - \tilde{P}_M^H. \quad (3.25)$$
where
\[
\tilde{R}_M = E_{z_f} \left( (\tilde{x}_M - \mu_x) (\tilde{x}_M - \mu_x)^H \right)
\] (3.26)
and
\[
\tilde{P}_M = E_{z_f} \left( (\tilde{x}_M - \mu_x) (\tilde{x}_M - \mu_x)^H \right).
\] (3.27)

3.4 Simulation results

In order to evaluate the performance of the CMMSE, some simulation experiments were conducted based on the following scenario:

\[
z_n = x_n + w_n, \quad n = 1, \ldots, N,
\] (3.28)

where \(\{x_n\}\) and \(\{w_n\}\) are statistically independent, zero-mean white random processes, with variances \(\sigma_x^2\) and \(\sigma_w^2\), respectively. \(\{z_n\}\) is the measurements vector with variance \(\sigma_z^2 = \sigma_x^2 + \sigma_w^2\).

The performances of the estimators are evaluated in terms of root-mean-square error (RMSE).

For each SNR, defined as \(SNR = \frac{\sigma_x^2}{\sigma_w^2}\), 1000 realizations, each of \(N = 1000\) samples, were used to evaluate the RMSE.

3.4.1 Example 1

In the first example, \(\{x_n\}\) and \(\{w_n\}\) are real white Gaussian random processes. In order to introduce variances mismatch, the following variances are presumed \(\tilde{\sigma}_x^2 = \sigma_x^2 \rho_x^2\) and \(\tilde{\sigma}_z^2 = \sigma_z^2 \rho_z^2\),
where $\rho_x$ and $\rho_z$ are real scalars. The MMSE and CMMSE estimators from (3.1) and (3.14), respectively, are given by:

$$\tilde{x}_{nM} = \left(\frac{\hat{\sigma}_x}{\hat{\sigma}_z}\right)^2 z_n,$$

and

$$\tilde{x}_{nCM} = \frac{\hat{\sigma}_x}{\hat{\sigma}_z} z_n.$$  \hfill (3.30)

The MSE of the mismatched MMSE and the mismatched CMMSE are given by:

$$E_{z;f}(\epsilon_{CM}^2) = \frac{\sigma_x^4}{\sigma_z^2} \rho^4 - 2 \frac{\sigma_x^4}{\sigma_z^2} \rho^2 + \sigma_x^2,$$

and

$$E_{z;f}(\epsilon_{CM}^2) = \frac{\sigma_x^2}{\sigma_z^2} \rho^2 - 2 \frac{\sigma_x^2}{\sigma_z^2} \rho + \sigma_x^2,$$

respectively, where $\rho \overset{\Delta}{=} \frac{\rho_x}{\rho_z}$. Furthermore, it can be easily shown that (3.26) and (3.27) reduce to $\tilde{R}_M = \frac{\hat{\sigma}_x^4}{\sigma_z^2} \sigma_x^2$ $\rho^4$ and $\tilde{P}_M = \frac{\hat{\sigma}_z^2}{\sigma_x^2} \sigma_z^2 = \frac{\sigma_x^4}{\sigma_z^2} \rho^2$, respectively. Then, according to (3.25) the CMMSE estimator outperforms the MMSE estimator when the following condition is met:

$$\rho^3 - \left(2 + \frac{\sigma_z^2}{\sigma_x^2}\right) \rho + 2 \frac{\sigma_z^2}{\sigma_x^2} > 0.$$

(3.33)

It can be verified that when no mismatch is present, i.e. when $\rho = 1$, this condition is never met. Therefore, as expected, with no mismatch the CMMSE performs worse than the MMSE, unless no noise is present, in which case the performances of the two estimators coincide. Fig. 3.1 presents the ratio between the RMSE of the CMMSE and the MMSE estimators as a
function of $\rho_z$, with $\rho_x = 1$ and SNR of 0dB. For comparison, a binary function indicating the regions where the condition in (3.33) is satisfied is also plotted. It can be seen that the experimental results match the theoretical performance condition and that the CMMSE outperforms the MMSE when $\rho_z < \approx 0.5$ and $\rho_z > \approx 1.3$, respectively.

Figure 3.1: Ratio between the RMSE of the constrained and unconstrained estimators, comparing to the theoretical threshold, as a function of $\rho_z$. The SNR is 0dB and $\rho_x = 1$. The theoretical threshold where the CMMSE estimator is expected to outperform the CMMSE estimator, is equal to one.

Fig. 3.2 presents a comparison between performances of the MMSE and the CMMSE with and without distribution mismatch as a function of SNR. In this case, the uncertainty in the variance is given in terms of $\rho_w$, such that: $\tilde{\sigma}_w^2 = \sigma_w^2 \rho_w^2$. For SNRs greater than 0 dB, it can
be observed that with no mismatch ($\rho_w = 1$) the performance of the CMMSE approaches the performance of the MMSE. In the presence of noise variance mismatch ($\rho_w = 3$) the CMMSE outperforms the MMSE for all the considered SNRs. Therefore, in this case, the CMMSE successfully compensates for the variance mismatch.

![Figure 3.2](image.png)

**Figure 3.2:** Performance comparison of the CMMSE and MMSE estimators under noise variance mismatch, with respect to the unconstrained estimator without noise variance mismatch.

Fig. 3.3 presents the ratio between the RMSE obtained by the CMMSE and MMSE estimators for SNRs: 5, 10, 15, 20 dB, as a function of the uncertainty parameters $\rho_x$ and $\rho_w$. As expected, when no mismatch is present ($\rho_w = 1$), the CMMSE performance is slightly lower than the MMSE at all SNRs. As the mismatch between the presumed and the true variances
increases, the performance of the CMMSE improves comparing to the (mismatched) MMSE. For instance, at an SNR of 10 dB and $\rho_x = 1$, when the mismatch parameter is increased beyond a certain level ($\rho_w \approx 1.75$), the CMMSE outperforms the mismatched MMSE. Fig. 3.3 also shows that even when a mismatch between the presumed and the true statistics of the parameter of interest is present, the CMMSE may outperform the MMSE although the constraints in the CMMSE estimate are based on prior knowledge of the statistics of the parameter of interest.

Figure 3.3: Ratio between the RMSE of the CMMSE and MMSE estimators as a function of the variance uncertainty parameters, $\rho_x$ and $\rho_w$. 
3.4.2 Example 2

In the second example, \( \{x_n\} \) and \( \{w_n\} \) are white Gaussian random processes, and the variances mismatch is introduced by assuming that the variance of the noisy signal, \( z \), is \textit{a-priori} unknown and is estimated from the prior observations. Due to the limited number of samples for data variance estimation, variance mismatch is introduced.

Fig. 3.4 presents the ratio between the RMSE of the CMMSE and the MMSE estimators as a function of the number of samples for data variance estimation, for SNRs: 5, 10, 15, 20 dB. The figure shows that when the true variance is used, the performance of the CMMSE is slightly lower comparing to the MMSE (ratio greater than 1). However, in the presence of variance mismatch, the CMMSE compensates for the mismatch and outperforms the MMSE. For instance, the CMMSE outperforms the MMSE (ratio smaller than 1) for SNR of 10 dB.

Fig. 3.5 presents a comparison between the CMMSE and several Bayesian estimators: “plug-in” MMSE estimator, matched to the statistics, which are estimated based on the available data of length \( N = 10 \), minimax difference regret MMSE estimator [39], and minimax ratio regret MSE estimator [41]. It can be noticed that for SNRs between -20 dB and -7.5 dB, the CMMSE outperforms the minimax difference regret and ratio regret minimax estimators, but its performance is worse than those of the plug-in estimator and the optimal MMSE estimator. For SNRs between -5 dB and 5 dB, the performances of the minimax difference regret, ratio regret minimax and the CMMSE are similar, and are better than those of the plug-in estimator. For SNRs between 5 dB and 20 dB, the CMMSE provides the best performance excluding the optimal MMSE estimator.
### Figure 3.4: Ratio between the RMSE of the CMMSE and MMSE estimators for different SNRs, as a function of the number of samples, when the variances are estimated from another data set.

#### 3.4.3 Example 3

In this example, \( \{w_n\} \) is a zero-mean, white Gaussian process with variance \( \sigma_w^2 \) and \( x_n \) is a zero-mean white Laplacian-distributed random process with variance \( \sigma_x^2 \). Since the MMSE is difficult to implement in case of a Laplacian-distributed process, the linear MMSE (LMMSE) is implemented. In other words, a Gaussian distribution is assumed and therefore distribution mismatch is introduced. Fig. 3.6 presents the performances of the CMMSE, LMMSE, minimax difference regret MMSE, and minimax ratio regret MMSE estimators in terms of RMSE. It can
Figure 3.5: Performance comparison of the LMMSE, Minimax difference, Minimax ratio and the CMMSE estimators. The signal variances are estimated based on 10 data samples.

It can be observed that for SNRs between -20dB and -5dB, the CMMSE outperforms the minimax difference regret and minimax ratio regret estimators, and its performance is slightly worse than those of the LMMSE. For SNRs between -5dB and 5dB the performances of the estimators are approximately the same, with a slight advantage of the CMMSE and minimax difference regret estimators over the ratio regret minimax and the LMMSE. For SNRs higher than 5dB the CMMSE clearly outperforms the other estimators.
3.5 Conclusions

In this chapter, an MMSE-based estimator, which is robust to distribution mismatch is derived. The proposed estimator conserves the first- and second-order statistical properties of the original signal. In addition, in contrary to most of the estimators proposed in the literature, the proposed estimator does not enforce any restrictions on the signals distributions or on the relationship between the measurements and the estimated signal, which might degrade the performances. The CMMSE estimator is easy to implement since it does not require any
parameters adjustment.

The performance of the CMMSE estimator is studied both analytically and via simulations using several examples, in comparison to several other estimators. It is shown that in the absence of mismatch, the performance of the CMMSE is slightly lower than the optimal MMSE performance. However, in the presence of distribution mismatch above a certain level, the CMMSE estimator outperforms the mismatched MMSE estimator and the other tested estimators. Thus, the CMMSE provides a simple and efficient way to compensate for distribution mismatch and to improve the estimation performance. It should be noted, however, that the CMMSE preserves the first- and second-order statistical properties of the signal of interest, while higher-order statistics might not match the true statistics. Therefore, in the next chapter we consider a more general framework is considered.
Chapter 4

Constrained MMSE Estimator for
Distribution Mismatch Compensation
of GMM-Distributed Random Vectors

4.1 Introduction

In the previous chapter, the problem of Bayesian estimation in the presence of distribution mismatch was presented and the CMMSE concept was introduced. The CMMSE estimator was shown to be robust to modeling mismatch. However, this estimator preserves only the first- and second-order statistical properties of the signal of interest while higher order statistics might not match the true statistics.

In this chapter, we present a more general framework, which incorporates constraints on
the entire statistics of the signal of interest. A new estimator, which is based on the MMSE criterion with constraints on the PDF of the parameter of interest, is proposed. The statistical constraints are based on GMM representation of the PDF. This estimator can be calculated numerically, but not analytically. Hence, an approximated estimator, termed as GMM-CMMSE, is proposed. The GMM-CMMSE is found to be robust to signal distribution mismatch, since it incorporates statistical information on the parameters of interest and restricts the solution space accordingly.

The performance of the GMM-CMMSE estimator under different mismatch conditions is studied via simulations using several examples. It is shown that with no distribution mismatch, the GMM-CMMSE estimator performance is slightly lower than the optimal MMSE. However, in the presence of signal distribution mismatch, this estimator outperforms the MMSE estimator, the CMMSE estimator and several other commonly used estimators. The estimator approximately conserves the PDF of the original signal and consequently yields a more robust estimator.

This chapter proceeds as follows. Section 4.2 presents the PDF-constrained estimator concept. Section 4.3 presents the results of several simulation experiments. Summary and conclusions are given in Section 5.3.
4.2 PDF constrained MMSE for GMM distributed random vectors

In the previous chapter it was shown that the first- and second-order statistics of the signal estimated with the CMMSE match those of the original signal. Furthermore, it was shown that the CMMSE is able to compensate for the distribution mismatch problem and therefore it outperforms the MMSE and some other estimators proposed in the literature, such as the minimax estimator and the ratio-regret minimax estimator. A PDF constrained MMSE estimator is proposed in this chapter. The idea is to impose constraints not only on the first- and second order statistics but on the PDF of the parameters of interest. These constraints allow to obtain a more robust estimator than the MMSE estimator and the CMMSE estimator.

The conditional PDF is approximated by a mixture model. The most commonly used mixtures for PDF representation are Gaussian mixtures. In [50], it has been shown that any density function can be closely approximated by a finite Gaussian mixture.

The use of GMMs achieves two major goals. Firstly, if the joint PDF of the parameters of interest and the observations is represented by a GMM, it is easy to show that the parameters of interest can be represented by a GMM of the same order. Therefore, it is easy to incorporate statistical constraints on the MMSE solution space by imposing constraints on the GMM parameters of the signal of interest. Secondly, under the GMM assumption, the conditional expectation employed by the CMMSE, can be represented by a linear combination of weighted sum of linear MMSE (LMMSE) estimators [51], which are easy to implement.
The problem is formulated as follows. Consider the problem of estimating an unknown random vector $x$, given an observation vector $z$, with the conditional PDF $f_{x|z}(x|z)$. Both $x$ and $z$ are presumably jointly GMM-distributed.

**Definition:** The vectors $x$ and $z$ are jointly GMM if their joint distribution can be written as

$$
\begin{bmatrix}
  x \\
  z
\end{bmatrix} \sim \text{GMM} \left( \begin{bmatrix}
  \alpha_{xj} \\
  \mu_{xj}
\end{bmatrix}, \begin{bmatrix}
  C_{xxj} & C_{xzj} \\
  C_{zxj} & C_{zzj}
\end{bmatrix} ; j = 1, \ldots, J \right), \quad (4.1)
$$

where $C_{xzj}$ is the cross covariance matrix of $x$ and $z$, for the $j^{th}$ Gaussian. The marginal distributions of $x$ and $z$ are given by:

$$
\begin{align*}
  f_x(x) &= \sum_{j=1}^{J} \alpha_j \Phi(x; \theta_{xj}) \\
  f_z(z) &= \sum_{j=1}^{J} \alpha_j \Phi(z; \theta_{zj}),
\end{align*}
$$

respectively, where $\Phi(x; \theta_{xj})$ denotes a complex Gaussian PDF and $\theta_{xj}$ contains the mean vector $\mu_{xj}$ and the covariance matrix $C_{xj}$.

It is proposed to minimize a cost function based on the MSE criterion: $E_{x,z}(\|x - \hat{x}\|^2)$, with constraints on the GMM parameters of the signal of interest. Let $\eta \triangleq [\eta_1, \ldots, \eta_M]^T$ be a hidden random vector indicating the generated Gaussian, where $\eta = \eta_j$ indicates that the
data vector is generated by the \( j^{th} \) Gaussian. The PDF of \( \eta \) is given by:

\[
f_{\eta}(\eta) = \sum_{j=1}^{J} \alpha_j \delta(\eta - \eta_j),
\]

where \( \delta(\cdot) \) denotes the dirac delta function. Then, the statistical constraints can be written as follows. The first order statistical constraint for each Gaussian is given by:

\[
\mu_{\hat{x}_j} \triangleq E_{z|\eta=\eta_j}(\hat{x}) = \mu_{x_j}, \quad j = 1, \ldots, J.
\]

The second order statistical constraint for each Gaussian is given by:

\[
C_{\hat{x}_j} \triangleq E_{z|\eta=\eta_j}((\hat{x} - \mu_{\hat{x}_j})(\hat{x} - \mu_{\hat{x}_j})^H) = C_{x_j}, \quad j = 1, \ldots, J.
\]

The constrained minimization problem is therefore formulated as follows:

\[
E_{x,z,\eta}(\|x - \hat{x}\|^2) + \sum_{j=1}^{J} \left[ tr \left( \Lambda_j \left( C_{\hat{x}_j} - C_{x_j} \right) \right) + \gamma_j^T \left( \mu_{\hat{x}_j} - \mu_{x_j} \right) + \kappa_j(\alpha_{\hat{x}_j} - \alpha_{x_j}) \right] \rightarrow \min, \quad \hat{x}
\]

where \( \Lambda_j, \gamma_j \) and \( \kappa_j \) are real matrix, vector and scalar, respectively, of Lagrange multipliers.
The Lagrangian can be rewritten as follows:

\[
E_z E_{\eta|x} E_{x|z,\eta} (\|x - \hat{x}\|^2) + \sum_{j=1}^{J} \left[ tr \left( \Lambda_j (C_{x_j} - C_{\hat{x}_j}) \right) + \gamma_j^T (\mu_{\hat{x}_j} - \mu_{x_j}) + \kappa_j (\alpha_{\hat{x}_j} - \alpha_{x_j}) \right] = \\
E_z \left\{ \sum_{j=1}^{J} \left[ P(\eta = \eta_j|x) E_{x|z,\eta} \left( (x - \hat{x})^H (x - \hat{x}) \right) \right] \right\} + \\
\sum_{j=1}^{J} \left\{ E_{x|\eta = \eta_j} \left[ tr \left( (\hat{x} - \mu_{\hat{x}_j})^H \Lambda_j (\hat{x} - \mu_{x_j}) \right) + \gamma_j^T (\hat{x} - \mu_{x_j}) \right] \right\} + \\
\sum_{j=1}^{J} \left\{ \kappa_j (\alpha_{\hat{x}_j} - \alpha_{x_j}) - tr (\Lambda_j C_{x_j}) \right\} = \\
E_z \left\{ \sum_{j=1}^{J} \left[ P(\eta = \eta_j|x) E_{x|z,\eta} \left( (x - \hat{x})^H (x - \hat{x}) \right) + \\
\frac{f_{\hat{x}}(z|\eta = \eta_j)}{f_{\bar{x}}(z)} \left( (\hat{x} - \mu_{\hat{x}_j})^H \Lambda_j (\hat{x} - \mu_{x_j}) + \gamma_j^T (\hat{x} - \mu_{x_j}) \right) + \\
\kappa_j (\alpha_{\hat{x}_j} - \alpha_{x_j}) - tr (\Lambda_j C_{x_j}) \right] \right\},
\tag{4.8}
\]

where \( P(\eta = \eta_j|z) \) is the conditional probability of the mixture component indicator \( \eta = \eta_j \), given \( z \), which is derived from the Bayes rule [48]:

\[
P(\eta = \eta_j|z) = \frac{f_{\bar{x}}(z|\eta = \eta_j) P(\eta = \eta_j)}{f_{\bar{x}}(z)} = \frac{f_{\bar{x}}(z|\eta = \eta_j) \alpha_j}{f_{\bar{x}}(z)} = \frac{\alpha_j \Phi(z; \theta_j)}{\sum_{j' = 1}^{J} \alpha_{j'} \Phi(z; \theta_{j'})}.
\tag{4.9}
\]
Therefore, using (4.9) the Lagrangian can be rewritten as follows:

\[
E_z \left\{ \sum_{j=1}^{J} \left[ \frac{f_z(z|\eta = \eta_j)}{f_z(z)} \left[ \alpha_j E_{x|z,\eta} \left( (x - \hat{x})^H (x - \hat{x}) \right) + \right. \right. \right. \\
\left. \left. \left. (\hat{x} - \mu_{x_j})^H \Lambda_j (\hat{x} - \mu_{x_j}) + \gamma_j^T (\hat{x} - \mu_{x_j}) \right] + \kappa_j (\alpha_{x_j} - \alpha_{x_j}) - tr(\Lambda_j \Sigma_{x_j}) \right] \right\}. \tag{4.10}
\]

In order to minimize the Lagrangian, it is sufficient to minimize the internal term of (4.10) for each \( z \):

\[
\tilde{G}(z) \triangleq \sum_{j=1}^{J} \left[ \frac{f_z(z|\eta = \eta_j)}{f_z(z)} \left[ \alpha_j E_{x|z,\eta} \left( (x - \hat{x})^H (x - \hat{x}) \right) + \right. \right. \right. \\
\left. \left. \left. (\hat{x} - \mu_{x_j})^H \Lambda_j (\hat{x} - \mu_{x_j}) + \gamma_j^T (\hat{x} - \mu_{x_j}) \right] + \kappa_j (\alpha_{x_j} - \alpha_{x_j}) - tr(\Lambda_j \Sigma_{x_j}) \right] \right. \tag{4.11}
\]

\( \tilde{G}(z) \) can be minimized with respect to \( \hat{x} \) by equating its derivative to zero:

\[
\sum_{j=1}^{J} \frac{f_z(z|\eta = \eta_j)}{f_z(z)} \left[ -\alpha_j E_{x|z,\eta} (x - \hat{x}) + \Lambda_j (\hat{x} - \mu_{x_j}) + \gamma_j \right] = 0. \tag{4.12}
\]

Equation (4.12), with the constraints given in (4.5) and (4.6), can be solved only numerically. Therefore, an approximated solution, termed GMM-CMMSE, which is a weighted sum of linear CMMSE estimators, is proposed:

\[
\hat{x}_{CM} = \sum_{j=1}^{J} P(\eta = \eta_j|z) \hat{x}_{CM|\eta = \eta_j}. \tag{4.13}
\]
where $\hat{x}_{CM|\eta=\eta_j}$ is the CMMSE for the $j^{th}$ Gaussian. This solution is based on the fact that given the mixture component indicator, $\eta = \eta_j$, the vectors $x$ and $z$ are jointly Gaussian, resulting in the linear CMMSE estimator. By substituting (3.14) in (4.13) the following estimator is obtained:

$$
\hat{x}_{CM} = \sum_{j=1}^{J} \left\{ P(\eta = \eta_j|z)C_{x_j}^{1/2}C_{M_j}^{-1/2} \left( E_{x|z,\eta=\eta_j}(x) - \mu_{x_j} \right) + \mu_{x_j} \right\}.
$$

(4.14)

Thus, the proposed solution is obtained by whitening of the MMSE estimator for each Gaussian, and then “recoloring” it using the a-priori Gaussian covariance matrix, while conserving the Gaussian mean. It should be noted that in practice, $P(\eta = \eta_j|z)$ is not given a-priori and needs to be estimated from the available data using eq. (4.9). In the following section, the performance of the proposed estimator is studied using several examples.

### 4.3 Simulation results

In order to evaluate the performance of the GMM-CMMSE estimator, some simulation experiments were conducted based on the following scenario:

$$
z_n = x_n + w_n, \quad n = 1, \ldots, N,
$$

(4.15)

where $w_n$ is a zero-mean, white Gaussian process with variance $\sigma_w^2$ and $x_n$ is a 3$^{rd}$-order GMM random process with variance $\sigma_x^2$. \{z_n\} is the measurements vector with variance $\sigma_z^2 = \sigma_x^2 + \sigma_w^2$. 

The performances of the estimators are evaluated in terms of RMSE. For each SNR defined as $SNR = \frac{\sigma_x^2}{\sigma_w^2}$, 1000 realizations, each of $N = 1000$ samples, were used to evaluate the RMSE.

4.3.1 Example 1

The purpose of this example is to demonstrate the preservation of the statistical properties by the GMM-CMMSE estimator. In this example, the presumed model order is equal to the true one and therefore no modeling mismatch is introduced. Fig. 4.1 presents the distribution of the estimated signal for the MMSE, the GMM-MMSE and the GMM-CMMSE estimators, compared to the original distribution of the parameter of interest. The figure shows that the GMM-CMMSE estimate distribution is the closest to the original distribution.

4.3.2 Example 2

In this example, modeling mismatch is introduced by assuming a lower GMM order than the true order. The performance of the GMM-CMMSE estimator is compared with the performance of the GMM-MMSE estimator, for a $1^{st}$-order GMM (Gaussian), a $2^{nd}$-order GMM and a $3^{rd}$-order GMM. In the last case, the presumed GMM order equals the true one and therefore no modeling mismatch is present. It is expected that in this case, the performance of the CMMSE estimator will be worse than the MMSE estimator, which is the optimal estimator. In the first two cases, the presumed GMM order is smaller than the true one and therefore mismatch is introduced. In these cases, the CMMSE estimator is expected to outperform the MMSE estimator.
Figure 4.1: An example of the estimated signal distribution for the MMSE, the GMM-MMSE and the GMM-CMMSE estimators, compared to the original distribution of the parameter of interest. The original parameter of interest is a 3rd-order GMM random process.

Fig. 4.2 presents the RMSE of the compared estimators as a function of the SNR. It can be seen that as expected, when no modeling mismatch exists (3rd-order GMM), the performance of the GMM-CMMSE estimator is slightly worse than the performance of the GMM-MMSE estimator. However, in the presence of mismatch, the GMM-CMMSE estimator significantly outperforms the GMM-MMSE estimator for SNRs greater than 2.5dB.

Fig. 4.3 presents a comparison between several Bayesian estimators: minimax difference regret MMSE estimator [39], minimax ratio regret MSE estimator [41] and the GMM-CMMSE estimator for the scenario described above. It should be noted that these estimators, excluding
the GMM-CMMSE estimator, are difficult to implement unless a linear model is assumed.

It can be noticed that for SNRs between -20dB and -5dB, the GMM-CMMSE estimator outperforms the minimax difference regret and ratio regret minimax estimators, but its performance is worse than those of the plug-in estimator and the CMMSE estimator. For SNRs between 7.5dB and 20dB, the GMM-CMMSE estimator significantly outperforms the other estimators.
4.4 Conclusions

In this chapter, a PDF constrained MMSE estimator for GMM distributed random vectors is proposed. In contrary to most of the estimators proposed in the literature, the proposed estimator does not assume any relationship between the measurements and the estimated signal which might degrade the performance.

The performance of the proposed estimator is studied via simulations, and its performance is compared to several other estimators. The capability of the GMM-CMMSE estimator to
preserve the statistical distribution of a 3\textsuperscript{rd}-order GMM signal, in contrary to the other tested estimators, is demonstrated via simulations. It is shown that in the absence of mismatch, the performance of the GMM-CMMSE estimator is slightly lower than the optimal MMSE performance. However, in the presence of distribution mismatch, the GMM-CMMSE estimator significantly outperforms the mismatched MMSE estimator and the other estimators. The GMM-CMMSE estimator outperforms the MMSE in the presence of mismatch between the presumed and the true statistics of the parameter of interest. Thus, the GMM-CMMSE estimator provides a simple and efficient way to compensate for distribution mismatch and to improve the estimation performance. In addition, the new estimator is easy to implement, since it can be represented by a linear combination of weighted sum of linear CMMSE estimators.
Chapter 5

CMMSE-based single-trial EEG estimation

5.1 Introduction

The problem of single-trial EEG estimation was introduced in Chapter 2. A PHMM-based classifier was presented and evaluated using an artificial database and two real databases. The results show that the classification performances strongly depend on SNR. It is therefore clear that estimation of the single-trial EEG signals prior to their classification, may significantly improve the classification rates.

Averaging is the most commonly used method in cognitive studies. Averaging is based on the assumption that the EEGs invariant activity pattern is perfectly time-locked to the stimulus and is superimposed in the independent stochastic background EEG signals [52].
However, averaging removes the information related to variations observed in single-trials, resulting in loss of important inter-trial information. Furthermore, averaging causes a trade-off between the potential accuracy and the achieved speed of communication. Therefore, interest in analyzing single-trial EEGs and ERPs has evolved. Various methods have been proposed for single-trial EEG and ERP analysis including ensemble of SVMs [9–11], wavelet transform [12, 13], Kalman filter [14], particle filtering [15] and spatio-temporal filtering [16]. Unfortunately, these techniques fail to estimate single-trial EEGs and ERPs because of the very low SNR, the non-Gaussian and stochastic nature of the EEG, and the inter-trial variability of the recorded signals.

In the recent years one of the commonly used methods for single-trial EEG estimation has been ICA, which was first applied to ERP analysis by Makeig et al. in 1997 [17]. The method was further improved in [18] and in [19], in which P300 waves were enhanced based on ICA subspace projections. ICA is a technique for separating linear and instantaneous mixtures of random variables into a set of unobserved independent sources. The idea behind the ICA approach is performing blind source separation on the recorded EEG signal based on higher-order statistics, i.e. by estimating the source signals from highly correlated EEG signals based on the statistical independence criterion, regardless of the physical location or configuration of the source generators. It thus can be used both for artifacts removal and ERPs extraction. ICA separates the problem of source identification (“What”) from that of source localization (“Where”). The ICA does not attempt to perform source localization. Instead, it attempts to find the scalp topography of each source and the time course of its
activation. The ICA is based on information-maximization criterion which takes into account higher-order statistical information about the distribution of the input vectors (concurrent field measurements at many spatial locations). One of the disadvantages of ICA is that once the independent components are obtained, the identity of each component is unknown. In order to correlate between the output component and the signals of concern (usually the evoked signals), additional post-processing steps have to be employed.

It should be noted that it is difficult to test the reliability and objectivity of the above mentioned models and techniques, given the fact that the source signals (as well as their locations and number) are undetermined. Consequently, estimation performances may be evaluated and compared across methods by means of EEGs classification rates or by using artificially generated signals.

In this chapter, an implementation of the CMMSE and the GMM-CMMSE estimators for single-trial EEG estimation is proposed and evaluated using an artificial EEG database, which is generated from real EEG signals. The performances of the CMMSE and GMM-CMMSE estimators are compared with those of the Wiener filter and an ICA-based estimator. For the simulations, a Matlab implementation of the fast ICA algorithm was used. The code can be found at http://www.cis.hut.fi/projects/ica/fastica/.
5.2 GMM-CMMSE-based single-trial EEG estimation

The statistical properties of the spontaneous EEG signal and response-related EEG signals are significantly different. While the spontaneous EEG can be often considered as a short-time stationary signal, response-related EEG signals are highly non-stationary. Due to these differences, the CMMSE and the GMM-CMMSE estimators, which utilize \textit{a-priori} knowledge of the signal statistics are expected to yield better performances than other methods which do not make any use of \textit{a-priori} statistical information. In our case, the statistics are \textit{a-priori} unknown and therefore need to be estimated from the available data, for example, based on the pre-stimulus epoch, while assuming that the statistics of the spontaneous EEG component do not change rapidly over the pre-stimulus and in-stimulus periods.

In order to evaluate the performance of the CMMSE and GMM-CMMSE estimators in estimation of single-trial EEG patterns, an artificial EEG database was generated from real EEG signals. Ongoing EEG signals, each of 10sec long with a sampling frequency of 250Hz, were recorded from seven subjects while performing five mental tasks: a baseline task, a multiplication task, a letter task, a rotation task and a counting task. The following channels were recorded: c3, c4, p3, p4, o1, o2 and EOG. The subjects performed ten trials of each task on two different days. More details on this database can be found in [53].

The data used in this chapter include a baseline task and a rotation task acquired from channel c3 of the first subject. The first 100msec of the rotation task were used to estimate the statistics of the spontaneous task. Noisy EEG signal were generated using a composition
of the EEG of the rotation task and the EEG of the baseline task for SNRs between -15dB and 15dB. In this case, the SNR is defined as: $SNR = \frac{E(x^2)}{E(w^2)}$, where $x$ represents the response-related EEG and $w$ represents the spontaneous EEG, and the signals energies are computed over the entire signals.

Fig. 5.1 presents the amplitude distribution for the spontaneous EEG and response-related EEG, where the distributions are estimated from the entire 10sec (a total of 2500 samples) of channel $c3$ of the first subject. It can be seen that the distributions of both signals are approximately Gaussian with slightly different means and variances.

Figure 5.1: Amplitude distribution of the spontaneous EEG and response-related EEG of channel $c3$ from the 1st subject.
An example of a rotation task EEG, the synthesized noisy EEG and the response-related EEG, estimated using a 3\textsuperscript{rd}-order GMM-CMMSE, is presented in Fig. 5.2.

![Figure 5.2: An example of the original response-related EEG acquired from channel c3 of the 1\textsuperscript{st} subject, noisy EEG measurements and the estimated response-related EEG.](image)

Fig. 5.3 presents the RMSE obtained for the CMMSE and the 3\textsuperscript{rd}-order GMM-CMMSE estimators, comparing to the ICA-based estimator and the Wiener filter, for SNRs between -15dB and 15dB. In order to implement the Wiener filter, the first 100msec of the rotation task was used to estimate the statistics of the spontaneous task, as was done for the CMMSE and the GMM-CMMSE estimators. The figure shows that for SNRs between -15dB and -10dB, the Wiener filter outperforms the other estimators. For SNRs between -10dB and -2dB, both the
CMMSE estimator and the 3rd-order GMM-CMMSE estimator outperform the Wiener filter and the ICA-based estimator. The ICA-based estimator outperforms the other estimators for SNRs between -2dB and 7.5dB. For SNRs greater than 10dB, the CMMSE and GMM-CMMSE estimators outperform the Wiener filtering and the ICA-based estimator, with a very slight advantage of the GMM-CMMSE estimator over the CMMSE estimator. It should be mentioned that contrary to the ICA-based estimator, the CMMSE and GMM-CMMSE estimators, do not incorporate the inter-channel information. Utilizing the mutual information among the different channels may add crucial interchannel information and consequently improve the estimation performance.

The similarity between the performances of the GMM-CMMSE estimator and the CMMSE estimator is expected since as previously mentioned, both signals are approximately Gaussian-distributed. However, the GMM-CMMSE estimator is expected to outperform the CMMSE in other types of signals such as ERPs, whose distribution varies in time.

5.3 Conclusions

In this chapter, an application of the constrained estimators for single-trial EEG estimation was considered. The performances of the CMMSE and GMM-CMMSE estimators were studied using real EEG signals, which were artificially composed to generate noisy EEG signals as a function of the SNR. In most of the illustrated SNRs, the CMMSE and GMM-CMMSE estimators were shown to outperform the Wiener filter and an ICA-based estimator. Further
Figure 5.3: RMSE for the estimated response-related EEG using the CMMSE and the 3\textsuperscript{rd}-order GMM-CMMSE estimators in comparison to the Wiener filter and the ICA-based estimator.

evaluation of the estimators performances using real databases of various types of response-related EEGs, and the impact of the proposed estimators on the classification rates, is a topic for future research, once an appropriate database is collected.
Chapter 6

Summary and Future Work

6.1 Summary of main results

This chapter summarizes the algorithms which were introduced in this work and the results obtained in the simulation experiments.

A model-based approach for multichannel ERPs classification was introduced in Chapter 2. The approach is based on PHMMs and an ML-based decision rule. The proposed algorithm is studied using an artificial ERP database and two real ERP databases. In the case of artificial ERP database, it was shown that for SNRs smaller than -5dB, the correct classification rate is smaller than 68%. For SNRs greater than 0dB the classification rate increases rapidly with around 82% corrected classification rates for an SNR of 0dB, and around 91% for an SNR of 5dB. The results show that the PHMM classifier provides the lowest error rates, with an improvement of 2% and 10% for databases DB Ia and DB III, respectively, comparing to the
best method reported in the literature.

It was shown that the PHMM-based classifier with a left-to-right architecture without skips is capable of representing different types of ERPs, and that it performs better than other commonly used classifiers. Furthermore, it was shown that a feature set which consists of linear filter bank coefficients with log energy and ∆log energy provides the best classification rates, outperforming other feature sets such as LPC and LPC-derived cepstrum.

In Chapter 3 the general problem of Bayesian estimation in the presence of signal distribution mismatch was addressed. A novel estimator, termed CMMSE, was derived based on the MMSE criterion with constraints on the statistics of the parameters of interest. The CMMSE estimator conserves the first and second-order statistical properties of the original signal. The performance of this estimator was studied both analytically and via simulations using several examples, in comparison to several other estimators such as the minimax difference regret MMSE estimator and the minimax ratio regret MMSE estimator. It was shown that in the absence of mismatch, the performance of the CMMSE is slightly lower than the optimal MMSE. However, in the presence of distribution mismatch above a certain level, the CMMSE estimator outperforms the mismatched MMSE estimator and the other tested estimators. Thus, the CMMSE provides a simple and efficient way to compensate for distribution mismatch and to increase the robustness to modeling mismatch.

The idea of the constrained estimator was further extended in Chapter 4, in which a PDF-constrained MMSE estimator for GMM distributed random vectors was introduced. This estimator conserves the statistical properties of the original signal. The estimator can
be calculated numerically, but an analytical solution is difficult to obtain. Therefore, an approximated estimator, termed as GMM-CMMSE, which is based on a weighted sum of linear CMMSE estimators, was proposed. It was shown that in the absence of mismatch, the performance of the GMM-CMMSE estimator is slightly lower than the optimal MMSE performance. However, in the presence of distribution mismatch, the GMM-CMMSE estimator significantly outperforms the mismatched MMSE estimator and some other commonly used estimators. The GMM-CMMSE outperforms the MMSE in the presence of mismatch above a certain level between the presumed and the true statistics of the parameter of interest. The GMM-CMMSE is a simple and efficient way to compensate for distribution mismatch and to improve the estimation performance.

In Chapter 5, the CMMSE and the GMM-CMMSE estimators were utilized for single-trial EEG estimation. Some simulation experiments were performed using real EEG signals. The results show that for a small range of SNRs, both the CMMSE and the GMM-CMMSE estimators outperform the Wiener filter and the ICA-based estimator. In order to evaluate the performance of the proposed estimators with EEG signals, they should be applied on real EEG measurements acquired during various cognitive and mental tasks, such that the impact of these estimators on the HMM classification performances can be studied. Based on the results obtained in the above-mentioned simulation experiments, it is expected that the CMMSE and the GMM-CMMSE estimators will improve the classification performances.
6.2 Further study

During the research work, several ideas for further study were raised. In the subject of ERP estimation, an algorithm to improve EEGs/ERPs averaging based on HMM may be considered. The idea is that the HMM-based segmentation may allow to perform time synchronization between different ERPs prior to averaging them, so as to overcome the difficult problem of EEGs/ERPs time invariance, and obtain a reliable estimate via averaging. An HMM-based EEGs/ERPs averaging algorithm was developed in this work and was partially tested on artificial EEG database. The idea should be further developed and studied. It should be noted that current state-of-the-art techniques in EEG and ERP analysis, such as the LORETA software, use simple ERP averaging and it is postulated that HMM-based EEG/ERP averaging may improve the analysis reliability and performances.

Considering the general field of Bayesian estimators, there are various algorithms, which could benefit from the incorporation of the CMMSE and GMM-CMMSE estimators. First, utilizing CMMSE and GMM-CMMSE estimators for EEG pattern estimation was shown to be successful and is expected to improve the PHMM classification performances. In order to do that, an appropriate database of response-related EEG with pre-stimulus continuous EEG should be acquired such that the pre-stimulus epochs will be used to estimate the EEG GMM.

Another idea which was raised during this work, was the incorporation of mutual information between different data channels and/or incorporation of inter-channel statistical constraints, in order to improve the performances of the CMMSE and GMM-CMMSE esti-
mators.

Other algorithms, such as the EM algorithm, Kalman filters and adaptive beamformers, which are based on the conditional expectation may benefit from an improved and robust estimation if the CMMSE estimator or GMM-CMMSE estimator will be used instead of the MMSE estimator.
Bibliography


[32] B. Blankertz, “BCI competition final results,”


