Brain Signal Analysis in Space-Time-Frequency Domain; An Application to Brain Computer Interfacing

Thesis submitted to the University of Cardiff in candidature for the degree
of Doctor of Philosophy

Kianoush Nazarpour

Center of Digital Signal Processing
Cardiff University
May 2008

UMI Number: U585107

UMI

Dissertation Publishing

ProQuest

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .... (candidate)
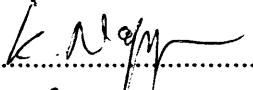
Date ...01-09-08

STATEMENT 1

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit reference. A bibliography is appended.

Signed .... (candidate)
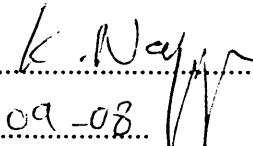
Date ..01-09-08

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photo-copying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed .... (candidate)
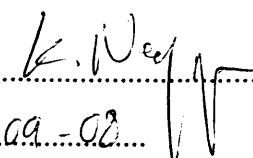
Date ...01-09-08

ABSTRACT

In this dissertation, advanced methods for electroencephalogram (EEG) signal analysis in the space-time-frequency (STF) domain with applications to eye-blink (EB) artifact removal and brain computer interfacing (BCI) are developed. The two methods for EB artifact removal from EEGs are presented which respectively include the estimated spatial signatures of the EB artifacts into the signal extraction and the robust beamforming frameworks. In the developed signal extraction algorithm, the EB artifacts are extracted as uncorrelated signals from EEGs. The algorithm utilizes the spatial signatures of the EB artifacts as priori knowledge in the signal extraction stage. The spatial distributions are identified using the STF model of EEGs. In the robust beamforming approach, first a novel space-time-frequency/time-segment (STF-TS) model for EEGs is introduced. The estimated spatial signatures of the EBs are then taken into account in order to restore the artifact contaminated EEG measurements. Both algorithms are evaluated by using the simulated and real EEGs and shown to produce comparable results to that of conventional approaches. Finally, an effective paradigm for BCI is introduced. In this approach prior physiological knowledge of spectrally band limited steady-state movement related potentials is exploited. The results consolidate the method.

To my magnificent half, Elham.

ACKNOWLEDGEMENTS

STATEMENT OF

ORIGINALITY

I hereby declare that the work described in this thesis was carried out entirely by the author in the Center of Digital Signal Processing, Cardiff School of Engineering, Cardiff University. The thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university. And that to the best of my knowledge, it does not contain any materials previously published or written by any other individual except where due reference is made in the text.

CONTENTS

PUBLICATIONS

11 **K. Nazarpour**, P. Praamstra, R. C. Miall, and S. Sanei, "Brain computer interfacing via steady-state movement related potentials," *IEEE Trans. Biomed. Eng.*, 2008, submitted.

10 **K. Nazarpour**, Y. Wongsawat, S. Sanei, J. A. Chambers, and S. Oraintara, "Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 9, pp. 2221-2231, Sep. 2008.

9 **K. Nazarpour**, H. R. Mohseni, C. W. Hesse, J. A. Chambers, and S. Sanei, "A novel semi-blind signal extraction approach incorporating PARAFAC for the removal of eye-blink artifact from EEGs," *EURASIP J. Advances Sig. Process.*, vol. 2008, Article ID 857459, doi:10.1155/2008/857459.

8 **K. Nazarpour**, P. Praamstra, R. C. Miall, and S. Sanei, "Steady-state movement related brain potentials for BCI," in *Proc. 28th IEEE/EMBS EMBC 2008*, Vancouver, Canada, 2008.

7 **K. Nazarpour**, P. Praamstra, R. C. Miall and S. Sanei, "Steady-state movement related potentials for BCI; an exploratory approach in the STF domain," in *Proc. BCI 2008 Workshop*, Graz, Austria, 2008.

6 T. Tsalaile, S. M. Naqvi, **K. Nazarpour**, S. Sanei, and J. A. Chambers, "Blind source extraction of heart sound signals from lung sound recordings exploiting the periodicity of the heart sound," in *Proc. IEEE/ICASSP 2008*, Las Vegas, USA, 2008.

5 **K. Nazarpour**, Y. Wongsawat, S. Sanei, S. Oraintara, and J. A. Chambers, "A robust minimum variance beamforming approach for the removal of the eye-blink artifacts from EEGs," in *Proc. 27th IEEE/EMBS EMBC 2007*, Lyon, France, 2007.

4 **K. Nazarpour**, S. Sanei, and J. A. Chambers, "A novel semi-blind signal extraction approach incorporating PARAFAC for the removal of the the removal of eye-blink artifact from EEGs," in *Proc. IEEE DSP 2007*, Cardiff, UK, 2008.

3 **K. Nazarpour**, L. Shoker and S. Sanei, "Brain computer interfacing in space-time-frequency domain," in *Proc. BCI 2006 Workshop*, Graz, Austria, 2006.

2 **K. Nazarpour**, S. Sanei, L. Shoker and J. A. Chambers, "EEG signal parallel space-time-frequency decomposition for brain computer interfacing," in *Proc. EUSIPCO06*, Florence, Italy, 2006.

1 L. Shoker, S. Sanei, **K. Nazarpour**, and A. Sumich, "A novel space-time-frequency masking algorithm for extracting propagation in EEGs: with an application to brain computer interfacing," in *Proc. EUSIPCO06*, Florence, Italy, 2006.

List of Figures

List of Tables

List of Abbreviations

AAR   Adaptive Auto-Regressive

BBCI   Berlin BCI

BCG   Ballistocardiogram

BCI   Brain Computer Interfacing

BSE   Blind Source Extraction

BSS   Blind Source separation

CORCONDIA Core Consistency Diagnostic

CSF   Cerebrospinal Fluid

CSP   Common Spatial Patterns

CT   Computed Tomography

EB   Eye-Blink

ECG   Electrocardiography

ECoG   Electrocorticograms

EEG   Electroencephalogram

EMG   Electromyography

EOG	Electro-Oculogram
ERD	Event Related Desynchronization
ERP	Event Related Potentials
ERS	Event Related Synchronization
FDA	Fisher's Linear Discriminant
fMRI	functional Magnetic Resonance Imaging
ICA	Independent Component Analysis
KFD	Kernel Fisher's Discriminant
LCMVB	Linearly Constrained Minimum Variance Beamformer
LDA	Linear Discriminant Analysis
LFP	Local Field Potential
LRP	Lateralized Readiness Potentials
MEG	Magnetoencephalography
MRI	Magnetic Resonance Imaging
OA	Ocular Artifact
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PET	Positron Emission Tomography
RMVB	Robust Minimum Variance Beamforming
RP	Readiness Potentials

SBSE	Semi-Blind Source Extraction
SCBSS	Spatially Constrained Blind Source Separation
SCP	Slow Cortical Potential
SNR	Signal to Noise Ratio
SoI	Source of Interest
ssMRP	Steady-State Movement Related Potential
ssVEP	Steady-State Visual Evoked Potential
STF-TS	Space-Time-Frequency/Time-Segment
STFT	Short-Time Frequency Transform
STF	Space-Time-Frequency
SVM	Support Vector Machine
TALS	Trilinear Alternating Least Squares
TTD	Tübingen Thought Translation Device

LIST OF SYMBOLS

x Scalar

$\mathbf{x}$ Column vector

$\mathbf{X}$ Matrix

$\underline{\mathbf{X}}$ Higher order array

$\arg\min_{\mathbf{x}} J(\mathbf{x})$ The argument x which minimizes the value of the cost function $J(\mathbf{x})$.

$\mathrm{diag}(\mathbf{X})$ Vector holding the diagonal elements of $\mathbf{X}$

$\mathrm{Tr}(\mathbf{X})$ The trace of $\mathbf{X}$, i.e., the sum of the diagonal elements of $\mathbf{X}$

$\mathbf{X} \otimes \mathbf{Y}$ The Kronecker product of $\mathbf{X}$ and $\mathbf{Y}$ where $\mathbf{X}$ is of size $\mathcal{I} \times \mathcal{J}$, is defined as

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & \cdots & x_{1\mathcal{J}}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ x_{\mathcal{I}1}\mathbf{Y} & \cdots & x_{\mathcal{I}\mathcal{J}}\mathbf{Y} \end{bmatrix}$$

$\mathbf{X} \odot \mathbf{Y}$ The Khatri-Rao product. The matrices $\mathbf{X}$ and $\mathbf{Y}$ must have the same number of columns, i.e. $\mathcal{I}$. Then

$$\mathbf{X} \odot \mathbf{Y} = [\mathbf{x}_1 \otimes \mathbf{y}_1 \quad \mathbf{x}_2 \otimes \mathbf{y}_2 \quad \cdots \quad \mathbf{x}_I \otimes \mathbf{y}_I]$$

$\mathbf{X}^\dagger$ The Moore-Penrose inverse of $\mathbf{X}$

$\|\mathbf{X}\|_2$ The l_2 (Euclidian) norm

$\mathbf{I}$ The identity matrix

$\mathbf{x} \perp \mathbf{Y}$ $\mathbf{x}$ is orthogonal to the space spanned by the column of $\mathbf{Y}$,

                                        i.e. $\mathbf{y}_i$s

$$\mathbf{x} \perp \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_I\}$$

$\mathbf{E}_{\mathbf{x}|\mathbf{y}^\perp}$ The projection of $\mathbf{x}$ onto the space $\mathbf{y}^\perp$, i.e.

$$\mathbf{E}_{\mathbf{x}|\mathbf{y}^\perp} = \frac{\mathbf{x}\mathbf{y}'}{\mathbf{y}'\mathbf{x}}$$

$\langle . \rangle$ The vector inner product

Chapter 1

INTRODUCTION

1.1 The periodic nature of the brain

Brain electrical activities were recorded by Hans Berger in 1926 and termed electroencephalogram (EEG) signals. He called the large amplitude rhythm, of about 10 Hz, induced by eye closure in an awake and calm subject, the "alpha" (α) rhythm since he observed it first. During the last century several other brain rhythms have been recorded during various cognitive, motor, or functions of the brain. Discovering the dynamic patterns of brain electrical activity is highly important in understanding the role of the brain in human behavior and cognition. Since Berger's early achievements, there have been mainly three unanswered questions [1]; how are EEG patterns with respect to a certain motor action of behavior generated, why are they oscillatory, and what are their contents? Answering these questions may not be possible without investigating the brain activities by means of fully controlled experiments. To this end, several academic and industrial research centers have introduced techniques and machines to help the neuroscientists.

I, as a researcher, aim at developing effective recording frameworks and

1

signal processing schemes for detecting specific electric events within the brain, relevant to the simple task of voluntary (possibly repetitive) movement of left and right indices. For such processing technique, artifact removed brain signals are necessary. Therefore, a considerable fraction of this dissertation is devoted to two novel artifact removal frameworks.

1.2 Classic windows of the brain

EEGs are commonly defined as brain electrical manifestation recorded from the surface of the scalp. The most common type of bio-potential electrode used for EEG recording is the silver/silver-chloride (Ag/AgCl) electrode. Several electrodes are placed on the scalp using a carefully positioned nylon cap. The conductive gel plays the role of an electrolyte medium between the scalp and the electrodes. Recently active[1] and dry electrodes [2] have been manufactured for EEG recordings. The utilization of the active electrodes has been fully justified, see the first footnote. However, dry electrodes are still in the research and development phase [2]. The electrical activity of brain can also be recorded invasively by using subdural electrodes. Invasive methods enable the measurement of brain potentials directly from its surface. Subdural electrodes are inserted into the scalp in order to record the electrical activity from underneath the dura.[2] Generally, in intracellular recordings, the activations of the individual neurons' are measured using an electrode inserted into a cell. In contrast, the extracellular recordings are made using electrodes places within the brain tissue sufficiently close to neurons. Extracellular recording reflects the activation of several neighboring neurons, i.e. local field potential (LFP).

---

[1]For production history and technical details of the active electrodes visit the BioSemi Instrumentation company website on http://www.biosemi.com/index.htm.

[2]Dura is a membrane which covers the brain.

In order to explore the electro-physiological patterns of the brain various technologies have been established. Essentially, brain imaging techniques are categorized into two groups; those modalities which reflect detailed anatomical views of brain and those which highlight the involving regions during various mental tasks. Currently, existing functional imaging modalities which present high spatial resolutions suffer from rather low temporal accuracies and vice versa. For instance, EEGs which are mainly used to determine functional properties of brain present a very high temporal resolution. However, they suffer from a very low spatial accuracy. Other imaging modalities present much better spatial resolution (in the range of 1 mm^3) but their temporal resolution is not in the acceptable range of 5-10 ms [3]. In the following, several classic brain imaging modalities are reviewed and their spatial and temporal resolutions are described.

In computed tomography (CT), a series of X-ray beams are emitted to the subjects at different angles. Then, cross-sectional images from the body are reconstructed by the computer. The brightness of the reconstructed image is proportional to the density of the tissue. The CT images reflect the anatomical information. The major disadvantage of CT is that the subjects are exposed to the X-ray radiations several times. The spatial resolution of CT is considerably high, however, the temporal resolution is in the range of 50-100 ms.

A common window to the brain is the magnetic resonance imaging (MRI). In MRI, subjects are not exposed to X-ray beams and therefore, this imaging procedure is safer than CT for the subjects [3]. MRI measures changes in the brain magnetic characteristics in response to an external large magnetic field. By MRI, three dimensional (3D) images of the brain can be constructed with high spatial resolution (approximately 1 mm^3). The func-

tional magnetic resonance imaging (fMRI) reflects the reaction of the oxygen molecules against a controlled magnetic field; hence, it highlights brain functional properties indirectly by measuring oxygen consumption. fMRI enjoys a relatively high spatial resolution (between 2 mm^3 to 4 mm^3), however, the low temporal resolution (on average 3 s) can be considered as a drawback.

In Magnetoencephalography (MEG), the tiny (on the order of 100 to 1000 fT (femto Tesla, femto $= 10^{-15}$)) magnetic fields in the brain are recorded. Specifically, these magnetic fields are generated by the ionic currents flowing through the dendrites. High-tech equipments are required to detect the brain magnetic fields. MEG is acquired from up to 300 recording superconducting quantum interference devices (SQUID) located on the scalp. It provides acceptable spatial resolutions, but lower than that of CT and MRI. MEG provides a fairly good temporal resolution and is less sensitive than EEG to the patients' movements. Note that, based on the inherent technical limitations of MEG, localization of radial sources, in contrast to tangential sources, is impossible.

Positron emission tomography (PET) measures the brain metabolism level through monitoring the blood pumped into the brain. Before recording PET, a short lived radioactive substance known as radiotracer is injected to the subject. The radiotracer emits several positrons as it decays. When the positron strikes an electron of an oxygen molecule, two photons are emitted in opposite directions. The positron is then detected and localized by the scanning device. In principle, highly activated brain regions, which consume more energy, transmit larger responses to PET. PET reflects the metabolic processes and, hence, is typically combined with CT or MRI. PET provides relatively lower spatial resolutions than fMRI and suffers from the

poor temporal resolutions (approximately 2 s).

The common imaging methods were briefly reviewed and Fig. 1.1 (adopted from [3]) summarizes the spatial and temporal resolutions of the above techniques.

As pointed out earlier, EEG presents high temporal resolutions and the low spatial accuracy is its drawback. Note that the spatial resolution of EEG can be enhanced by utilizing several (up to 256) scalp electrodes. The physiological principles of EEG will be presented in Chapter 2.

1.3 Aims and objectives of this research

The first aim of this research was to acquire a solid understanding of EEG signals. A prerequisite to achieve the defined goals was to understand the neurophysiological processes that occur in brain before, during, and after a simple finger movement.

During this research, I attended several courses, conferences, and meetings and consulted with two neurologists in order to enhance my basic understanding of brain anatomy and functions.

The objectives of the research are

- understanding the electroencephalogram and being able to interpret EEGs recorded for BCI;

- reviewing the current researches on brain computer interfacing (BCI) and artifact rejection;

- developing effective algorithms for removal of artifacts from EEGs;

- establishing new BCI approaches to detect the brain responses during repetitive finger movements.

Characteristics	Less				More
Ability to measure both cortical and deep structures	NIRS		MEG ERP, EEG	SPECT	PET, MRS, fMRI, phMRI, DTI
Temporal resolution	MRI DTI MRS	phMRI	PET SPECT	fMRI NIRS	EEG, ERP MEG
Spatial resolution	ERP EEG	MEG	SPECT	PET	NIRS, fMRI, DTI MRI, phMRI
Invasiveness of method		ERP EEG NIRS MEG	fMRI MRI DTI MRS	phMRI	PET SPECT
Expense of method		EEG ERP NIRS	fMRI DTI MRI MRS	phMRI	PET SPECT MEG
Ease of use with developmental populations	PET SPECT	phMRI	fMRI	MRI DTI MRS	EEG, ERP NIRS

Figure 1.1. Qualitative approximate rankings on the distinguished characteristics are provided for various imaging modalities. The figure has been adopted from [3].

1.4   Structure of this thesis

Chapter 2 details the classic EEG-based BCI framework. It comprises of a survey on the active BCI research centers, the BCI state of art techniques, and the classic and well established physiological features of the EEGs usually used for BCI.

Chapter 3 presents the first contribution of this dissertation to BCI. In this chapter after reviewing mathematics of the complex wavelet transform and the parallel factor analysis, it will shown that how arrays of the time-varying energy representations of the multi-channel EEGs can be decomposed into the spatial, temporal, and spectral signatures. A space-time-frequency (STF) decomposition of EEG in the Rolandic μ band will be realized to explore the STF dynamics of EEG signals during left and right index movements. These signatures will be classified by a support vector machine (SVM) classifier. Finally, the main reasons the presented approach may not be implemented in real-time BCI will be discussed.

EEGs can be severely contaminated by various artifacts such as eye-blinks (EB), eye- or body- movements, cardiomuscular activities, and power line noise. Essentially, artifacts interfere with the processing algorithms and may cause unacceptable loss of important parts of the recorded signals. Traditionally, the artifact contaminated recordings are excluded before averaging the multi-trial recordings. However, in many cases such as real-time EEG analysis, valid and reliable methods are needed in order to remove the artifacts. Fundamentally, the prime target of a reliable artifact correction algorithm is to remove the artifact with minimum distortion of the underlying EEG activity. Between various artifacts, ocular artifacts (OA), also called electro-oculogram (EOG), are the main source of interference for EEGs. An EOG refers to the potential difference generated by eye ball movement in

its socket or when an EB occurs. EOG may be propagated from the frontal area of the brain toward all the recording electrodes and contaminated the EEG potentials.

In Chapters 4 and 5, two effective preprocessing algorithms will be developed for the removal of EB artifacts from EEGs. They will be based upon blind source extraction (BSE) and beamforming, respectively.

In Chapter 4, after an overview on the existing methods for EB artifact detection and removal, a semi-blind signal extraction (SBSE) algorithm will be developed in order to extract the EB artifact as an uncorrelated signal from EEGs. The proposed method is semi-blind since it utilizes the spatial signatures of the EB artifacts which are identified by the STF model introduced in Chapter 3. The results of the developed artifact removal algorithm are comparable with that of previous studies. In this approach, the spatial signature of EB is estimated in advance and then introduced to the SBSE. There are two important issues in this algorithm; first, the decomposition of the STF model into the spatial, temporal, and spectral signatures is computationally expensive. Second, the performance of this method depends on how unbiased the EB spatial signature is estimated. Chapter 4 will be finished by demonstrating how a slight deviation of the estimated EB signature from the actual signature influences the extraction process and consequently the whole artifact removal procedure.

A hybrid "STF modeling"-"robust minimum variance beamforming" framework will be presented in Chapter 5 in order to address the two aforementioned shortcomings. First, a novel space-time-frequency/time-segment (STF-TS) model for EEGs, specifically designed for identification of the EB artifacts, will be introduced. Upon few necessary conditions for effective decomposition, the resulting spatial signatures will be utilized in a robust

beamforming paradigm to extract and remove the EB artifact. In this re-
gard, the major contribution is in the estimation of the steering vector, that
is, the spatial signature corresponding to the EB artifact, without solving the
conventional EEG forward models. There are two differences between the
approach presented in this chapter and that of the previous chapter. In Chap-
ter 4, the estimation of the steering vector corresponding to the EB artifact
was assumed to be precise. Here, this vector is allowed to deviate by a small
bias vector from its actual value. Second, in Chapter 5, the computational
complexity of the STF model estimation is significantly reduced by using
the STF-TS model.

Chapter 6 of this dissertation will introduce an approach for BCI based
on the steady-state movement related potentials (ssMRP). In order to sur-
pass the two main problems of the slow cortical potentials (SCP)-based
BCI, i.e. the slowness and the low signal to noise ratio (SNR), recent and
established physiological findings on ssMRPs will be exploited. Subjects
participating in this study are asked to tap at the constant rate of 2 Hz. This
provides a lateralized slow potentials over contralateral motor areas in the
frequency range of about 2 Hz. Classification of such potentials is straight-
forward since they are highly localized in space-time-frequency domain.
Two conventional pattern classifiers, namely, the Fisher's discriminant anal-
ysis (FDA) and the kernel FDA (KFD), will be implemented to quantify the
performance of the introduced BCI approach. The potentials and shortcom-
ings of this approach will be investigated.

And finally, in Chapter 7, this dissertation will be concluded and several
suggestions will be made for future research.

Chapter 2

BRAIN COMPUTER

INTERFACING

2.1 Introduction

Brain Computer Interfacing (BCI) typically refers to typing using a computer keyboard, clicking a mouse button, or moving a curser on the monitor by interpreting the EEGs. In other words, EEG-based BCI enables people suffering from severe neurological disabilities (cognitively intact) to interact with their environment by intention rather than physical contact. In BCI, users can concentrate on various tasks[1] and the computer identifies the task.

In [4], it is found that planning and execution of voluntary (self-paced) movements, real or mental imagination of movement, cause a pre-movement attenuation and a post-movement amplification in amplitude of some frequency bands in EEG over the contralateral sensorimotor cortex.

Various medical imaging modalities have shown [4] that the sensori-

---

[1]Although in theory, it is possible to design a BCI system for various mental tasks, in practice it has been limited to certain tasks such as left and right index, tongue, and leg movements [4].

motor cortex is activated during real or imaginary movements. It has also been well known [4,5] that preparation and execution of movements lead to short-lasting and circumscribed attenuation of the Rolandic μ (8-13 Hz) and the central β (14-28 Hz) rhythms known as event related desynchronization (ERD) following by an amplifications rebound phase called (ERS). Due to low spatial resolution of EEGs precise localization of ERD/ERS sources within the brain is rather difficult and entails complex computations. Moreover, the clearest ERD/ERS, to be utilized in BCI, may occur at different frequency bands and different time instances.

2.1.1 Contributing research areas

BCI is an interdisciplinary research topic to which researchers from different fields contribute. An ideal BCI research team would enjoy research experts from the following fields.

- Signal Processing: EEGs are time series. Advanced and preferably real-time signal processing techniques are necessary to detect the task related components of the recorded EEGs.

- Software Engineering: To provide the subject with a feedback of his/her actions, a computer interface is required to record EEG, to apply several (pre-)processing algorithms, and to control the visual feedback process.

- Information Theory: BCI presents a communication channel between the user and the machine. Performance evaluation of this channel in terms of error rate is recommended.

- Machine Learning: To produce a user tunable practical BCI system, machine learning and adaptive techniques would be of importance.

- Neurophysiology: An important issue in BCI research is designing effective EEG recording paradigms. Moreover, solid neurophysiological knowledge is needed to interpret EEG measured during various movements.

- Psychology: The BCI user is often provided with the visual feedback Psychologist can contribute by designing effective feedback training frameworks for BCI in order to explain human motor behavior during BCI.

2.1.2 Organization of this chapter

In this chapter, BCI state of the art will be presented. The existing EEG-based BCI systems in various pioneering research centers will be reviewed. The EEG basic physiology and its BCI relevant rhythms will be reviewed. The advantages and disadvantages of the classic (common) methods and algorithms for BCI in μ and β bands will be discussed. Finally, the application of the readiness potentials (RP) in BCI will be pointed out.

2.2 BCI-State of art

During the last two decades many BCI research groups have been established. An overview of the most successful ones is presented below. In particular, first the results of invasive methods are briefly discussed in section 2.2.1. As of non-invasive methods, five prominent research groups led by Jonathan Wolpaw in Albany (section 2.2.2), Niels Birbaumer in Tübingen (section 2.2.3), Klause-Robert Müller in Berlin (section 2.2.4), Gert Pfurtscheller in Graz (section 2.2.5) and José del R. Millán in Martigny (section 2.2.6) are exemplified.

2.2.1   Invasive methods

So far, most of studies on the invasive methods have been carried out on monkeys.[2] The monkey is well trained to actuate a prosthesis instead of its fixated arms. In the beginning of training procedure, the monkey strives on moving its arm. After a while and learning to operate the prosthesis, the monkey can move the prosthesis by its *thoughts* accurately. It has been reported [7] that the monkeys are able to learn to control an artificial arm if a feedback is provided. Recently, a monkey was also able to use the prosthesis for grasping in such a skilled way that it could pick up and eat fruit [8]. For many years, repeating such experiments with paralyzed but cognitively intact humans for rehabilitation purposes has been a milestone [6].

2.2.2   The Wadsworth BCI

The Wadsworth BCI research team uses the ERD of the μ rhythm for EEG classification during real or imaginary movements. Based on a fixed setup, the user is able to move a cursor toward different targets on the right side of the screen whereby the cursor moves with constant speed from left to right. The movement is controlled by suitable ERD of μ rhythm for which the subjects have to be trained for several weeks [9]. After multiple training sessions with visual feedback, the subjects are able to achieve over 90% hit rates [9].[3] Recently, an approach for simultaneous controlling of a cursor on a screen in vertical and horizontal directions has been presented in [10] where the user controls the device by suitable modulations of μ and β rhythms. The result of this study is reported as having "not reached an ac-

---

[2]There are also very recent studies on realization of invasive BCI methods for paralyzed subjects by positioning a matrix of recording electrodes directly over the motor cortex [6].

[3]A positive hit is considered in the performance if the subject instructed to imagine i.e. the right index movement and he/she produces EEGs which can be labeled as right index movement by a previously trained classifier.

ceptable level of accuracy" when compared to the results of cursor control in 1D [10].

2.2.3   The Thought Translation Device

The Tübingen Thought Translation Device (TTD) [11–13] enables subjects to self-regulate of the slow cortical potentials (SCP) after several feedback training sessions. In TTD, the cursor is controlled vertically through EEG signals and the patients are able to generate binary decisions with an accuracy of up to 90%.

Recently, this team have also reported in [14] that they have investigated the electrocorticogram (ECoG) measurements for BCI purposes. It should the Tübingen group has established a practical BCI system for locked-in patients [11].

2.2.4   The Berlin BCI (BBCI)

In 2000, Berlin BCI (BBCI) started to develop a new approach for BCI which was based on the idea of transferring the training effort from the human to the machine. Therefore, the level of inter-subject variability of the BCI should be reduced by minimizing the level of subject training. Currently, their BCI system uses high dimensional feature vectors extracted from a 128 channels EEG system. By adapting to the specific brain signals by advanced machine learning methods, it has been possible to detect the pre-movement potentials in the healthy subjects [15].

In [16, 17], it was reported that the classification rates of approximately 90% between left and right hand key presses can be achieved. Furthermore, these results were achieved after approximately one hour of data recording showing that long training sessions would not be essential [18].

Based on the imagination of different motor tasks, using a priori neuro-physiological knowledge about the accompanying ERD effects in the μ and β rhythms and negation effects in SCPs, it is possible to present promising results by incorporating advanced machine learning techniques [19].

The BBCI team has also been one of the pioneer groups in designing adaptive spatio-temporal and spatio-spectral filters to maximize the class separability measures utilized in the BCI classification stage [20].

2.2.5   The Graz BCI

The expert user of the Graz BCI machine is able to control a device based on the modulations of the μ or β rhythms of sensorimotor cortex. Note that the Wadsworth BCI system presents the power modulations to the user but the Graz BCI for the first time used the machine adaptation protocols in BCI [4].

In [21], it is reported that the classification accuracies of over 96% can be achieved by adaptive auto-regressive (AAR) modeling of EEGs. Moreover, they have been able to a BCI-based functional electrical stimulation (FES) to restore the functionality of a paralyzed subject's arm [22–24].

A real-time BCI system has been introduced in [25] where the parameters of the classifier are updated adaptively in order to maximize the performance of the BCI machine.

Very recently, the applicability of near infra-read spectroscopy (NIRS)-based techniques for BCI has been investigated [26]. Very interestingly, they have also shown that the transient increase in heart rate due to the subject's movement intention can be used in BCI [27].

2.2.6 The Martigny BCI

In Martigny, several researchers with academic backgrounds in machine learning and artificial intelligence launched an adaptive BCI scheme. In [28], it is suggested that in order to gain higher classification rates, a local neural classifier based on quadratic discriminant analysis would be a suitable choice. In this scheme, three subjects scored an average correct recognition rate of approximately 75% only after few training days. This was obtained by imagination of left or right-hand motions or by relaxation with closed eyes in an asynchronous environment. The mis-classification rate was below 5%. The trained subjects could type with a virtual keyboard, i.e. by choosing a letter approximately every 22 ms. [29].

In [30], the authors showed that three further classes (cube rotation, subtraction and word association) can be added to BCI. They were able to control a robot which moved in an artificial maze with these new classes. The robot could be instructed to turn left, right, or forward [30].

Reviewing the above-mentioned EEG approaches to BCI, an important issue can be concluded [4] that BCI (specifically in the μ or β bands) suffers from the low SNR. A worse case is when the single trial EEGs are analyzed. Note that in single trial analysis the EEG signal itself is not convincingly predictable. Although, the utilization of the advanced machine learning or signal processing algorithms for BCI is appreciated, they would not result in real-time adaptive BCI systems - mainly due to the computational requirements. Acknowledging the fact that the investigation of the brain and its functional features using other imaging modalities such as fMRI or MEG would ultimately increase the knowledge about brain, these modalities may not eventually be applied in real-life BCI since the paralyzed user will not be able to carry the fMRI or MEG machine. Considering recent advance-

ment in EEG recording and processing, an eligible approach for BCI still would be relying on new EEG recording paradigms and possibly using simple and effective EEG feature extraction and classification schemes. Eventually, these approaches would circumvent the limitations of BCI in μ or β bands.

Chapter 6 of this dissertation suggests a potential approach for EEG recording in which the sensorimotor cortex is physiologically constrained to generate (partly time-locked) motor related rhythms. Implantation of such approach in real life applications would be readily possible.

2.3   Principles of electroencephalography

The brain electrical activities have been investigated for about a century. The variation of the surface potential distribution on the scalp reflects functional activities of the underlying brain [31]. These potentials can be recorded by placing a matrix of electrodes on the scalp and recording the differential voltages between pairs of these electrodes. Figure 2.1 shows a sample multi-channel EEG recording of about 3.3 seconds from nine electrodes. In EEG signal processing community the 10-20 recording system [31] is frequently utilized for electrode placement. The 10-20 system is based on a relationship between the location of an electrode and the underlying area of cerebral cortex (the "10" and "20" refer to the 10% and 20% inter-electrode distances) [31]. Fig. 2.2 illustrates the standard electrode placement in the 10-20 system. In BCI, the electrodes in the central regions such as $C3$ and $C4$ are usually used. It has been concluded [31] that the asymmetric distribution of the scalp potentials over central area of the brain can be used to infer the presence of preferential (preparatory) brain activity.

The physiological background knowledge helps with understanding the
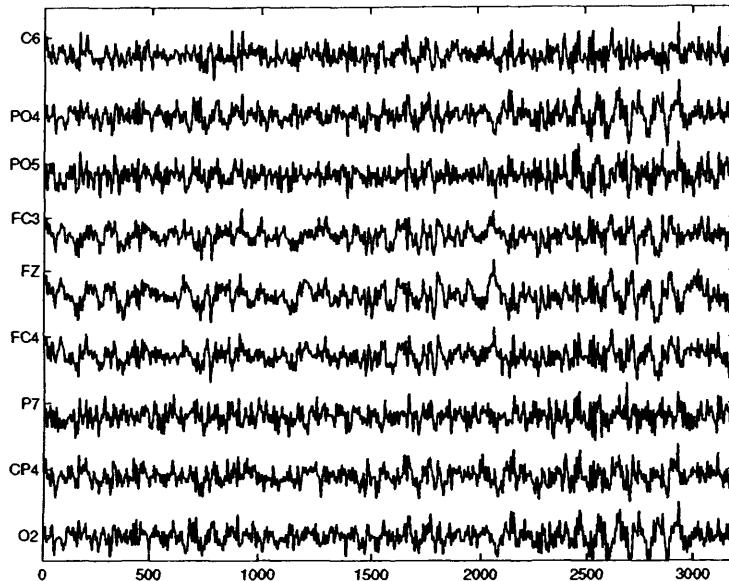
Figure 2.1. A set of EEG recordings.
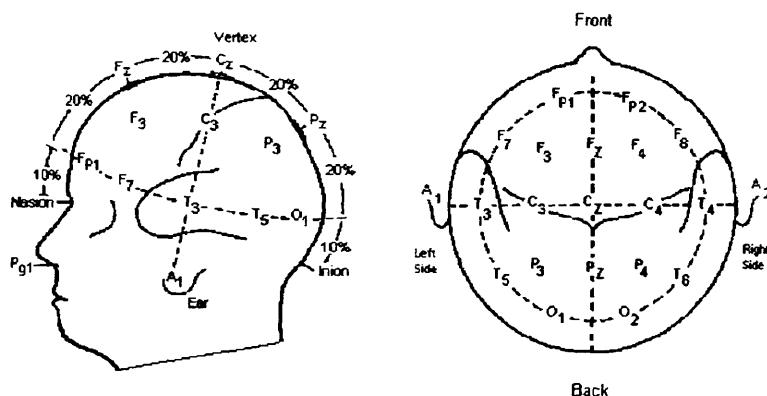
Figure 2.2. The 10-20 electrode positioning standard.

possibilities and challenges in EEG analysis and particularly in BCI. The goal of this section is to introduce the main concepts and features of EEG which are frequently used in BCI.

There are mainly three approaches to BCI. First, BCI based on event related potentials (ERP), described in Section 2.3.1, second, BCI based on analysis of the oscillatory features (rhythmic waves) of EEG, explained in Section 2.3.2, and third, BCI based on lateralized readiness potentials (LRP)[4], addressed in Section 2.3.3. BCI based on ERPs is beyond the scope

---

[4]In the clinical literature, LRPs or simply readiness potentials are considered as ERPs,

of this dissertation; however, BCI based on oscillatory features of EEG and readiness potentials will be explored in Chapters 3 and 6, respectively.

2.3.1 BCI and event related potentials

The temporally locked (linked to external stimulations variations) of the ongoing EEG are called ERPs. As opposed to evoked potentials, ERPs include an early response, due to the primary processing of the stimulus, as well as a late response which is a reflection of higher cognitive processing induced by the stimulus [32]. The P300 wave is a late appearing component of ERPs with a latency of approximately 300 ms elicited by rare or significant stimuli (visual, auditory, or somatosensory). P300 has been widely used for BCI [33].

2.3.2 BCI and oscillatory EEG

EEG contains a fairly wide frequency spectrum. The frequency range of EEG has fuzzy lower and upper bounds. There are as well very low and very high frequency components which to the author's knowledge do not really have any significance in clinical studies or in BCI. Practical EEG spectrum is limited to 0.1 Hz to 100 Hz and in a restricted sense between 0.3 Hz to 70 Hz. This range is usually subdivided into slow, medium, fast and very fast ranges- 0.3-7 Hz, 8-13 Hz, 14-30 Hz and above 30 Hz. These bands are named as below:

- Delta (δ) below 3.5 Hz (usually 0.1-3.5 Hz).

- Theta (θ) 4-7 Hz.

- Alpha (α) 8-13 Hz.

---

however, specifically in BCI literature, they are considered to be different.

- Rolandic Mu (μ) 8-13 Hz.

- Beta (β) above 13 Hz, usually 14-35 Hz.

- Gamma (γ) above 35 Hz unlimited in the upper bound.

For BCI applications the α, the μ, and the β are frequently utilized.

- The α rhythm

  1. Definition: α rhythms occur in the frequency range of 8-13 Hz during wakefulness over the posterior regions of head, generally with higher voltages over the occipital lobe. The amplitude of α rhythm is variable but mostly below 50 μV in adults. It is readily recordable when the subjects close their eyes and sit still. The α rhythm is highly attenuated by subject's visual attention or mental effort.

  2. Wave Morphology: The α rhythm is usually characterized by a rounded or sinusoidal waveform.

  3. Spatial Distribution: The α rhythm is a clear manifestation of the posterior half of the brain and is usually found over occipital, parietal, and posterior temporal regions. The α rhythm may extend into central areas, the vertex, and also the mid-temporal regions. When the brain central region is strongly involved in a task, the α rhythm can be distinguished from the co-existing μ rhythm. In that case, keeping the eyes opened will highly attenuate the α rhythm and the remaining signal would be the μ rhythm. The α rhythm may occasionally extend slightly to the frontal electrodes.

4. Reactivity: The posterior α rhythm is attenuated by an influx of light (eye-opening), other afferent stimuli, or mental activities. The degree of reactivity varies; the α rhythm can be blocked or attenuated. The α attenuation due to tactile or auditory stimuli as well as increased mental activity is usually less pronounced than the attenuation caused by eye-opening.

- The μ rhythm

    1. Definition: The Rolandic (central) μ rhythm is in frequency range of the α activity, but its spatial distribution and physiological significance are different.

    2. Wave Morphology: The μ rhythm has a sharp (spiky) negative peak and rounded positive phase.

    3. Spatial Distribution: μ rhythm is essentially confined to central area of the brain. The μ rhythm which is produced by movement is unilateral at the beginning of movement and then it becomes bilaterally distributed over the cortex.

    4. Reactivity: μ rhythm is bilaterally attenuated by real or imagery movements [31]. However, this attenuation is more pronounced on the motor region contralateral to the side of movement [4,5], prior to the onset of muscular contraction. It is reported [4] that there are delays of 50 ms to 7 seconds (on average 1.5 seconds) at the onset of μ blocking effect after the initiation of the movement of the contralateral finger with the ipsilateral lagging behind the contralateral response.
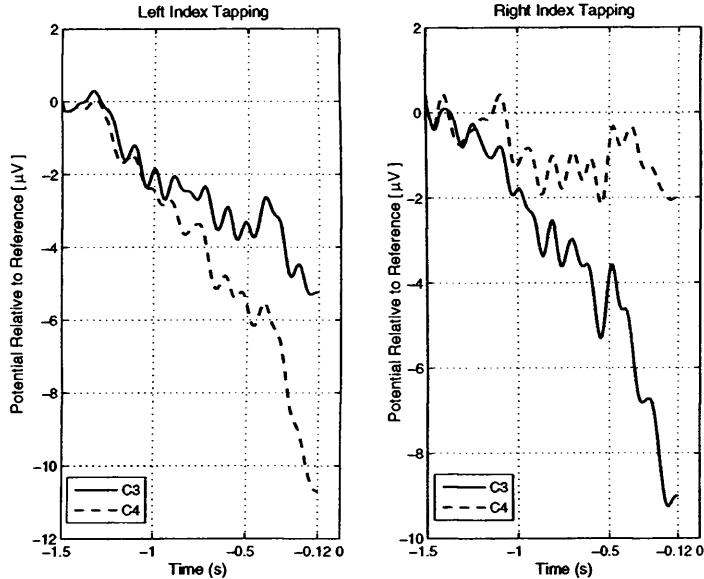
Figure 2.3. Readiness potentials recorded from $C3$ and $C4$ electrodes before left and right finger movement.

2.3.3 Readiness potentials

The readiness potentials (RP)[5] are the transient postsynaptic responses of the main pyramidal peri-central neurons [34]. They lead to negative EEGs over the primary motor cortex areas during motor preparation. The RPs show negative peaks about 120 ms before the movement onset. They are often considered for self-paced (subject-paced) BCI. RPs are spatially localized in the contralateral hemisphere with respect to the moving finger. RPs are easily detectable.

Fig. 2.3 shows the averaged EEG recordings from $C3$ and $C4$ electrodes in a typical BCI scenario of self-paced tapping with left and right fingers. The EEGs used for this demonstration were acquired from the BBCI team in Berlin.[6] The sampling rate was set to 1000 Hz.

For the left finger movement (the left subplots in Figure 2.3), the av-

---

[5]Also called Bereitschaftspotentials in the clinical literature [34].

[6]The dataset was recorded for the BCI competition III. It is available in Berlin BCI webpages $http : //ida.first.fhg.de/projects/bci/competitioniii$.

Right Index Tapping   Left Index Tapping
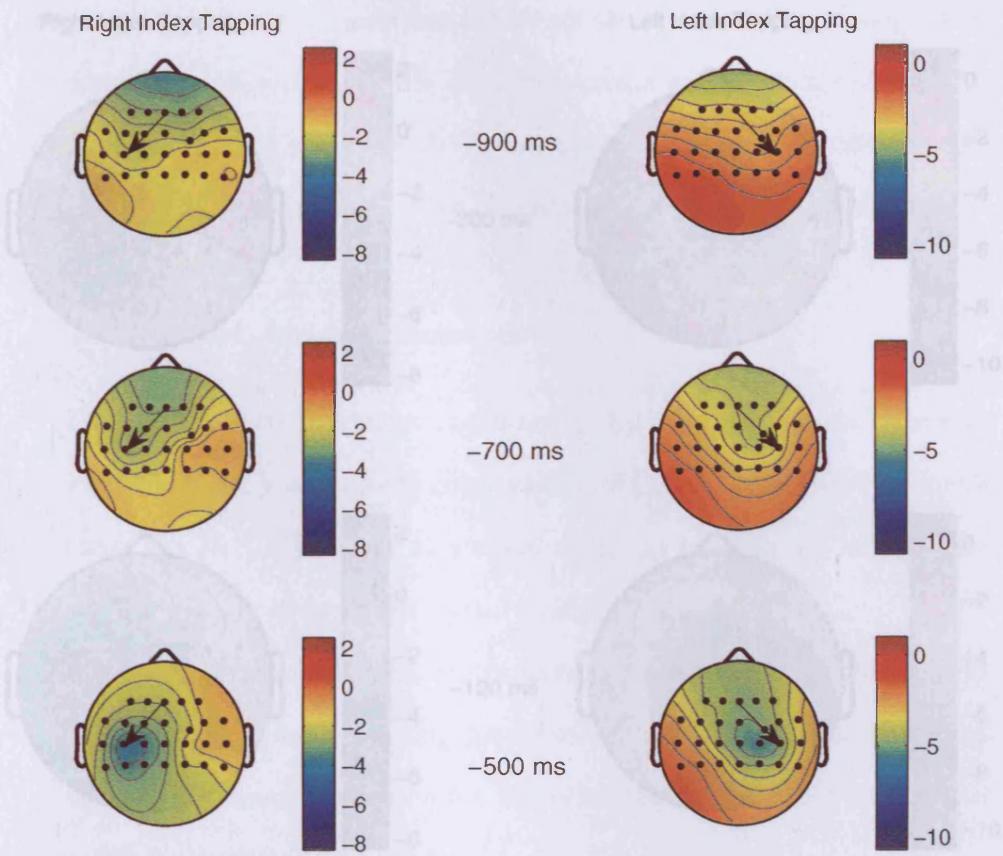
Figure 2.4. Generation of the readiness potentials for left and right index tapping from 900 ms to 500 ms before the EMG onset; see the footnote. The colorbars are in μV.

eraged cortical potential under $C4$ electrode is considerably smaller than that of $C3$ electrode. This shows higher engagement of the contralateral hemisphere than the ipsilateral hemisphere. In the right finger movement case (the right subplots in Figure 2.3), the averaged readiness potential is also asymmetrically distributed where $C3$ presents a more negative potential than $C4$. The higher engagement of the contralateral hemisphere rather than the ipsilateral hemisphere is evident.

Interestingly, the spatial distribution of RPs changes from approximately 900 ms to 120 ms before the movement onset. This can be characterized by an increasing lateralization of the RPs with larger amplitudes over the con-

Right Index Tapping   Left Index Tapping
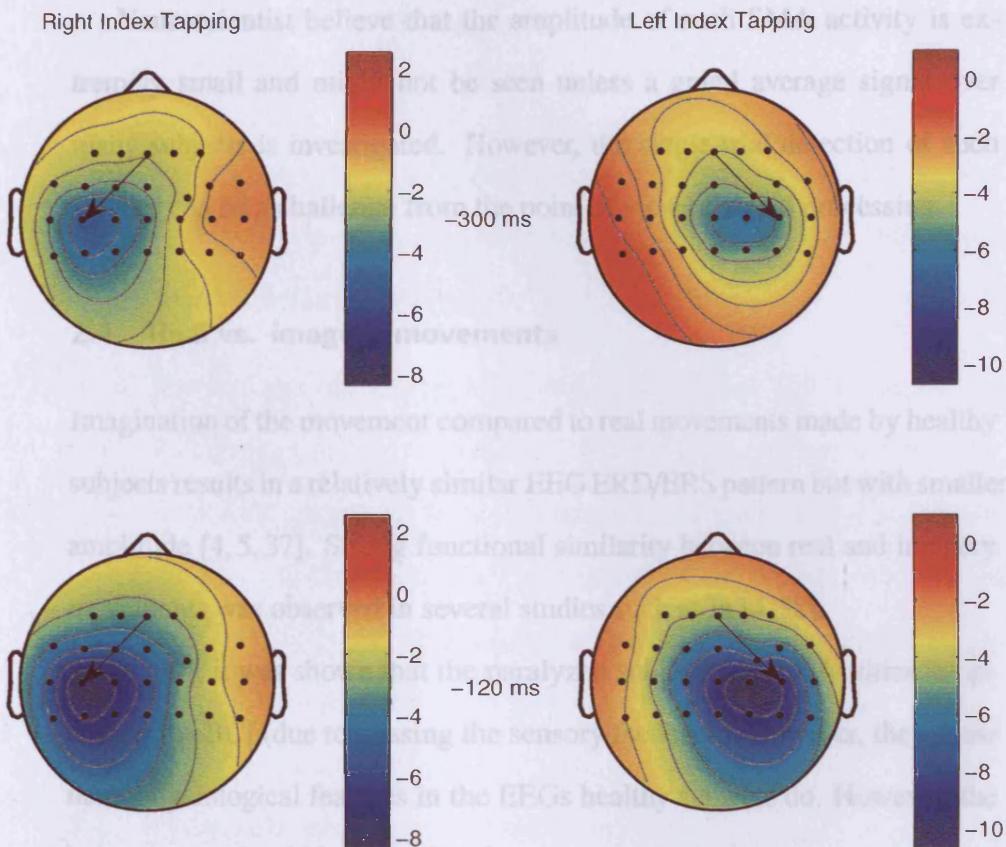
Figure 2.5. Generation of the lateralized readiness potentials for left and right index tapping from 300 ms to 120 ms before the movement onset; see the footnote in the previous page. The colorbars are in μV.

tralateral motor area with respect to ipsilateral hemisphere [34], as shown in Figs. 2.4 and 2.5. The RP signals have been frequently utilized for BCI with various spectrum estimation for preprocessing and feature extraction and machine learning methods for classification [15, 35]. However, to the author's best knowledge, this transition has not been explored for BCI application. For many years, there have been debates in the neuroscience community on whether the RPs in the supplementary motor area (SMA) of brain are generated earlier than the activities in the primary motor areas (M1) or vice versa, see Jahanshahi *et al.* [36] and the references therein for further details.

Neuroscientist believe that the amplitude of such SMA activity is extremely small and might not be seen unless a grand average signal over many subjects is investigated. However, the single trial detection of such activity can be a challenge from the point of view of signal processing.

2.4   Real vs. imagery movements

Imagination of the movement compared to real movements made by healthy subjects results in a relatively similar EEG ERD/ERS pattern but with smaller amplitude [4,5,37]. Strong functional similarity between real and imagery movements was observed in several studies such as in [4,38].

In [37], it was shown that the paralyzed subjects have difficulties in operating the BCI (due to missing the sensory feedback). However, they show neurophysiological features in the EEGs healthy subjects do. However, the below issues should be considered:

1. Facial movements, such as tongue or jaw activities, cause electromyography (EMG) signal interference in the measured EEG. Thus, EEG-controlled BCI for imagination of tongue movements must be well restored from the EMG artifacts in a pre-processing stage. Since the temporal and spectral characteristics of the EMG and EEG signals are similar, this filtering is challenging.

2. It has been shown [1] that EEG-based BCI is highly affected by brain background activities, in particular, by the concentration. During imagination of a movement, the concentration is usually much higher than during a real movements [1]. As the subject believes that the quality of the recorded EEGs is increased with concentration, high amplitude background (movement irrelevant) brain activities are in-

duced into EEGs. These activities are spread over frontal lobe and sensorimotor cortex.

3. Real movements can be easily controlled for their correctness and timing. However, BCI based on imagery movements may suffer from the co-generated EMG signal.

4. It is not easy for the healthy subjects to produce EEGs suitable for BCI during imagery movements [39]. The main problem is that "why should the subject be keen on controlling something by his thought and not with his hands if he/she is healthy?" However, some studies such as in [39] have reported that after multiple training sessions, the imagery movement becomes a skill for the subject.

2.5 Event related de-synchronization

A phenomenon in BCI that has been well investigated for the past 20 years is the EEG de-synchronization [4]. This phenomenon has been first reported by Berger in 1930 [40] and is of interest particularly in the μ and β bands.

Brain signals change from a synchronized (high amplitude) into a de-synchronized (low amplitude) mode and vice versa. Various frequency bands can show different ERD patterns. Changes of oscillatory activity in relation to internally or externally paced events can be either time- and phase-locked (evoked) or time-locked but not phase-locked (induced). The former is called an event related potential (ERP) and the latter is known as the event related (de)-synchronization [5]. The ERD/ERS can be described by the spatiotemporal pattern as well as by the frequency specificity [5].

An important task for developing the BCI based on ERD/ERS is to search for the most reactive frequency components in different electrode

signals. Therefore, effective methods of processing and visualization of significant ERD/ERS patterns in multi-channel EEG and ECoG data are required essentially.

In this chapter, a review on the importance of ERD/ERS in BCI is presented. A conventional method for visualization of ERD/ERS resulted by averaging of the recorded EEGs is then explained.

The α band rhythm demonstrates a relatively widespread attenuation during the perceptual, judgement, and memory tasks. Interestingly, an increase in task complexity increases the magnitude of ERD. It has to be noted that ERD is measured in percentage of power relative to the reference interval. To make sure that the power in the reference interval is in a rest level, the time difference between two consecutive events should be randomized and not be less than several seconds.

It is important to note that α de-synchronization is not a unitary phenomenon. For instance, two distinct patterns of de-synchronization in α band may be observed. Lower α de-synchronization (below 10 Hz) is obtained in response to almost any type of stimulus. It is topographically widespread and probably reflects the general task demands. Upper α de-synchronization (10-13 Hz) is topographically restricted to the motor cortex. It develops during the processing of sensory-semantic information. The degree of synchronization is highly related to semantic memory processes.

In addition to sensory and cognitive processing, a voluntary movement also results in a circumscribed de-synchronization in the upper α and lower β rhythms, localized over the sensorimotor areas [41]. The de-synchronization commences over the contralateral sensorimotor cortex and becomes symmetric after the execution of the movement. It is interesting that the time onset of the contralateral μ de-synchronization is almost identical for the

brisk or slow finger movements, i.e. approximately 2 seconds prior to the movement onset [42].

The index movement is accompanied by a pronounced ERD in the contralateral and a very low amplitude ERD in the ipsilateral hemispheres but the movements of other fingers are preceded by less lateralized ERDs [42]. The contralateral pre-movement ERD is not only similar in the brisk and slow movements but also with single index and hand movements. The circumscribed hand area μ ERD can be found in nearly every subject [41] and [43].

2.6 Post-movement β synchronization

The post-movement β synchronization (PMBS) is another example of the induced oscillatory brain signals. The PMBS displays several interesting features and has been widely utilized in BCI [4,5,41,43].

- The PMBS is a relatively robust phenomenon and found in nearly every subject after finger or foot movement. It is dominant over contralateral primary sensorimotor and has a maximum approximately 600 ms after the movement onset.

- Induced β oscillations recorded from both hemispheres show a lack of bilateral coherence. The β bursts can include frequency components either in a single band or in multiple bands. The reactivity of specific β frequency components is subject dependent. For finger movements, the largest PMBS is found in the 16-21 Hz band and for foot movements, in a slightly higher band of 19-26 Hz [44].

- The PMBS signals produced by the brisk or slow finger movements are of similar amplitudes. This is of specific interest because brisk

and slow movements involve different neural pathways.

- The PMBS is significantly larger in amplitude with hand movement as compared to finger movement. This can be interpreted by noticing the significantly longer movement time for hand as compared to finger movement.

2.7 Interpretation of (de-)synchronization in the α and β bands

Increased cellular excitability in thalamo-cortical systems results in a low amplitude de-synchronized EEGs. Therefore, EEGs can be interpreted as electrophysiological correlates of the activated cortical areas involved in processing of sensory information or production of motor behavior [5].

Involvement of larger neural networks or more cell assemblies results in an increased ERD. Factors contributing to such an enhancement of the ERD can be the increased task complexity and the efficient task performance [45, 46].

Explicit learning of a movement sequence such as key-pressing with different fingers is accompanied by an enhancement of the ERD over the contralateral sensorimotor cortex [5]. These findings strongly suggest that the activity in the primary motor cortex increases after the task has been learned [5].

2.8 Visualization of significant ERD/ERS patterns

The calculation of ERD/ERS time courses can be performed in different ways. Each way requires a number of event related EEG or ECoG trials which are time-locked to a stimulus. Standard ERD/ERS calculation is carried out by bandpass filtering of each trial, squaring the samples, and the

subsequent averaging over trials and sample points. The ERD/ERS is defined as the power decrease (ERD) or increase (ERS) relative to a specific reference interval, i.e. several seconds before the movement onset. Since evoked potentials can mask the induced activities, it is often useful to subtract the mean of the data for each sample before squaring. This procedure assumes that the computed mean values are temporally and spatially uncorrelated with the ERD/ERS phenomenon; although due to non-stationary nature of EEGs such assumption is far from reality. However, since this procedure is widely accepted [4], it is followed in this dissertation. The ERD/ERS phenomenon can be visualized as below

$$y_{ij} = (x_{ij} - \bar{x}_j)^2; \quad A_j = \frac{1}{N-1} \sum_{i=1}^{N} y_{ij} \qquad (2.8.1)$$

$$R = \frac{1}{k} \sum_{j=r_0}^{r_0+k} A_j \qquad (2.8.2)$$

$$ERD_j = \frac{A_j - R}{R} \times 100\% \qquad (2.8.3)$$

where N is the total number of trials, x_{ij} is the j-th sample of the i-th trial of the bandpass filtered data, and $\bar{x}_j$ is the mean of the j-th sample averaged over all bandpass filtered trials. R is the average power in the reference interval $[r_0, r_0 + k]$; for further details see [5].

One of the most critical issues in multi-channel ERD/ERS analysis methods is the appropriate selection of the frequency bands where ERD/ERS happens. Such selection is difficult since it requires a thorough comparison of several ERD/ERS time courses. In order to overcome this problem, a detailed ERD/ERS analysis of partially overlapping frequency bands covering the entire frequency range of interest can be made so that all the resulting ERD/ERS time courses of a channel are combined in one single map. Thus,

an ERD/ERS map is a matrix whose rows correspond to ERD/ERS calcu-
lations for different frequency bands. This map can be easily displayed as
an image giving detailed information about ERD/ERS activity in various
frequency bands.

As shown in [5], a frequency range of 7 Hz to 34 Hz and a frequency res-
olution of 1 Hz should be sufficient to cover the most important ERD/ERS
patterns in the EEG. Pfurtscheller *et al.* in [5] selected different frequency
ranges for different frequency bands; for α frequencies a bandwidth of 2 Hz,
for lower β frequencies a bandwidth of 4 Hz, and for higher β frequencies a
bandwidth of 8 Hz were assigned. These values were chosen in [5] to have
relatively higher resolutions in the α band than in the β band. Note that, the
selection of frequency ranges and bands depends on the application.

Only statistically significant patterns should be displayed to improve the
clarity of the maps. Most statistical techniques assume that the size of the
available set of sample values is sufficiently large. Since the probability dis-
tributions of EEG signals are not known and the number of trials may be
rather limited, especially if only artifact-free trials are considered, the use
of standard statistical methods to calculate the significance of the ERD es-
timates is questionable [47]. Bootstrap techniques are an alternative to the
asymptotic methods [47,48]. The basic idea is to replace the unknown popu-
lation distribution with the known empirical distribution. Various properties
of the estimator such as the confidence intervals are then determined based
on this empirical distribution [47].

2.9 Common spatial patterns

A popular approach to increase a statistical distance between the classes
(left and right movements) in the feature space is the use of the common

spatial patterns (CSP) [49] and [50]. The EEG patterns corresponding to the left and right finger movements are localized within the contralateral motor cortex, therefore, these patterns can be used to identify the effector. CSPs find a common describing factor to both classes as a way to maximally discriminate between the left and right finger movements.

In mathematical terms, the procedure for finding suitable CSPs can be explained as follows [20]. Let $\mathbf{X}^k = (X^k_{n,t})$, $n = 1, \cdots, N$, and $t = t_0, \cdots, T$ denote the (potentially bandpass filtered) EEG recording of the k-th trial, where N is the number of electrodes. Correspondingly, $Y^k \in \{1, 2\}$ represents the class-label of the k-th trial. The two class-covariance matrices are then given as

$$\Sigma_1 = \mathbf{X}^k \mathbf{X}^{k'}_{k:Y^k=1}$$

$$\Sigma_2 = \mathbf{X}^k \mathbf{X}^{k'}_{k:Y^k=2} \qquad (2.9.1)$$

where ['] denotes the transpose operation. The CSP analysis consists of calculating a matrix $\mathbf{W}$ and diagonal matrix $\mathbf{D}$ such that

$$\mathbf{W}\Sigma_1\mathbf{W}' = \mathbf{D},$$

$$\mathbf{W}\Sigma_2\mathbf{W}' = \mathbf{I} - \mathbf{D}. \qquad (2.9.2)$$

This can be accomplished by first, whitening the matrix $\Sigma_1 + \Sigma_2$, that is determining a matrix $\mathbf{P}$ such that

$$\mathbf{P}(\Sigma_1 + \Sigma_2)\mathbf{P}' = \mathbf{I}. \qquad (2.9.3)$$

This decomposition can always be found due to the positive definiteness of $\Sigma_1 + \Sigma_2$ and second, defining $\mathbf{S}_1 = \mathbf{P}\Sigma_1\mathbf{P}'$ and $\mathbf{S}_2 = \mathbf{P}\Sigma_2\mathbf{P}'$, and calculating

an orthogonal matrix $\mathbf{R}$ and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{S}_1' = \mathbf{R}\mathbf{D}\mathbf{R}'. \qquad (2.9.4)$$

$\mathbf{S}_1 + \mathbf{S}_2 = \mathbf{I}$ results in $\mathbf{S}_2' = \mathbf{R}(\mathbf{I} - \mathbf{D})\mathbf{R}'$. Note that the projection given by the p-th row of matrix $\mathbf{R}$ has a relative variance of d_p (p-th element of $\mathbf{D}$) for trials of class 1 and a relative variance of $1 - d_p$ for trials of the class 2. If d_p is close to 1 the filter given by the p-th row of $\mathbf{R}$ maximizes variance for class 1, and since $1 - d_p$ is close to 0, minimizes variance for class 2. The final decomposition that satisfies (2.9.2) can be obtained from

$$\mathbf{W} = \mathbf{R}'\mathbf{P}. \qquad (2.9.5)$$

EEG recordings are projected by

$$\mathbf{Z}^k = \mathbf{W}\mathbf{X}^k \qquad (2.9.6)$$

using the decomposition matrix $\mathbf{W}$ which acts as a spatial filter and projects the transforms the recorded EEGs $\mathbf{X}^k$ to $\mathbf{Z}^k$. Then, the classic feature extraction methods are implemented on the $\mathbf{Z}^k$. Incorporation of common spatial patterns in BCI leads to an enhancement in the class separability measures and ultimately results in higher correct classification rates. For instance, the authors in [20] and [50] have achieved more than 90% classification rate using CSP combined with a linear discriminant (LDA) classifier.

2.10 Preliminary Investigations

Based on the algorithm in [5], the ERD is computed for a subject during left finger imagination. To suppress the effect of noise, 45 trials are averaged
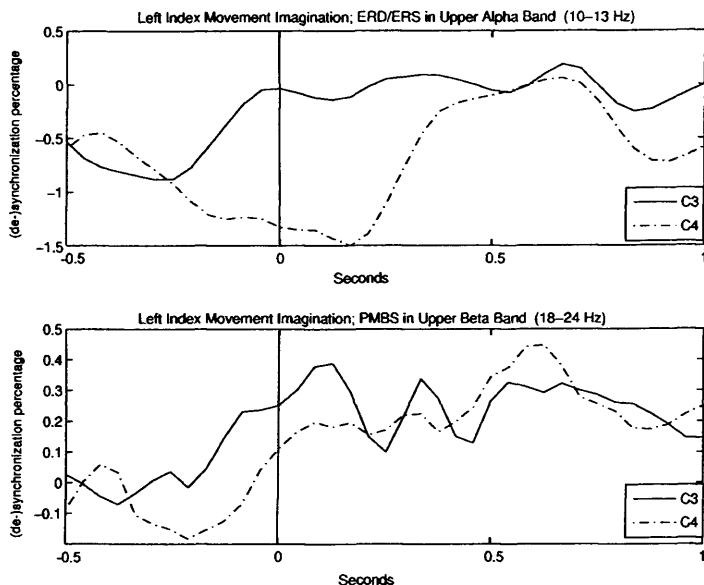
Figure 2.6. ERD for left finger imagination-vertical line shows the onset of imagination.

and the results are demonstrated. The details of EEG data used here will be fully reported in Chapter 3.

Fig. 2.6 demonstrates 1.5 seconds ERD from 0.5 seconds before the imagination onset to 1 second after that. The upper figure shows the upper α band between 10-13 Hz ERD and the lower figure is that of β band. When the subject imagines to move his/her left index, the ERD happens pronouncedly in the right hemisphere at the C4 electrode. In the lower plot, the post-movement β synchronization at C4 is relatively higher than that of C3. The ERD/ERS has been widely used in BCI; for details refers to [5] and [15] and the references therein.

Unfortunately, for all the subjects the clearest ERD/ERS does not always happen at the same electrode, the same time, and the same frequency band. Although it has been shown [4] that ERD often happens at C3 or C4 electrodes in the upper μ band before the motion imagination and lasts for approximately 1.5 seconds after the movement onset, it may not be the case for every subject. The brain background activity which is attenuated

by multi-trial averaging would remain as the main interference in single trial EEG processing. The ERD for 2 subjects for different electrodes have been plotted in Fig. 2.7 and Fig. 2.8. As in Fig. 2.6, the onset of imagination is at time 0 seconds. In Fig. 2.7(a), where the subject imagines left finger motion, it is evident that ERD happens at electrodes $C2$, $C4$, and $C6$ in the contralateral hemisphere. In contrast, as in Fig. 2.7(b), ERD is not pronounced for the same subject during right index imagination. This would degrade the BCI results. Although, still ERD for $C3$ is greater than that for $C4$, however, other electrode signals on the left hemisphere do not show such pattern when compared to signals recorded from the right hemisphere. It should be mentioned that the EEG classification for BCI is mainly performed based on the relative amplitudes of signals recorded from $C3$ and $C4$ electrodes. Fig. 2.8 shows that a relatively acceptable ERD (suitable for classification) happens at another electrodes; for left and right finger imaginations acceptable ERDs manifest at is best at $C6$ and $C1$, respectively. Furthermore in Fig. 2.8(a) and (b), ERD has happened respectively in 10-13 Hz and 7-9 Hz band for left (Fig. 2.8(a)) and right (Fig. 2.8(b)) finger imaginations. Note that, even one subject can show different ERD/ERS patterns in EEGs recorded in different days [4]. Therefore, it can be concluded that ERD is not the best solution to BCI.

Based on the above preliminary investigations, it can be concluded that new measures of brain activity are needed for BCI. These measures should ultimately resolve the inherent problems of ERD/ERS approach.
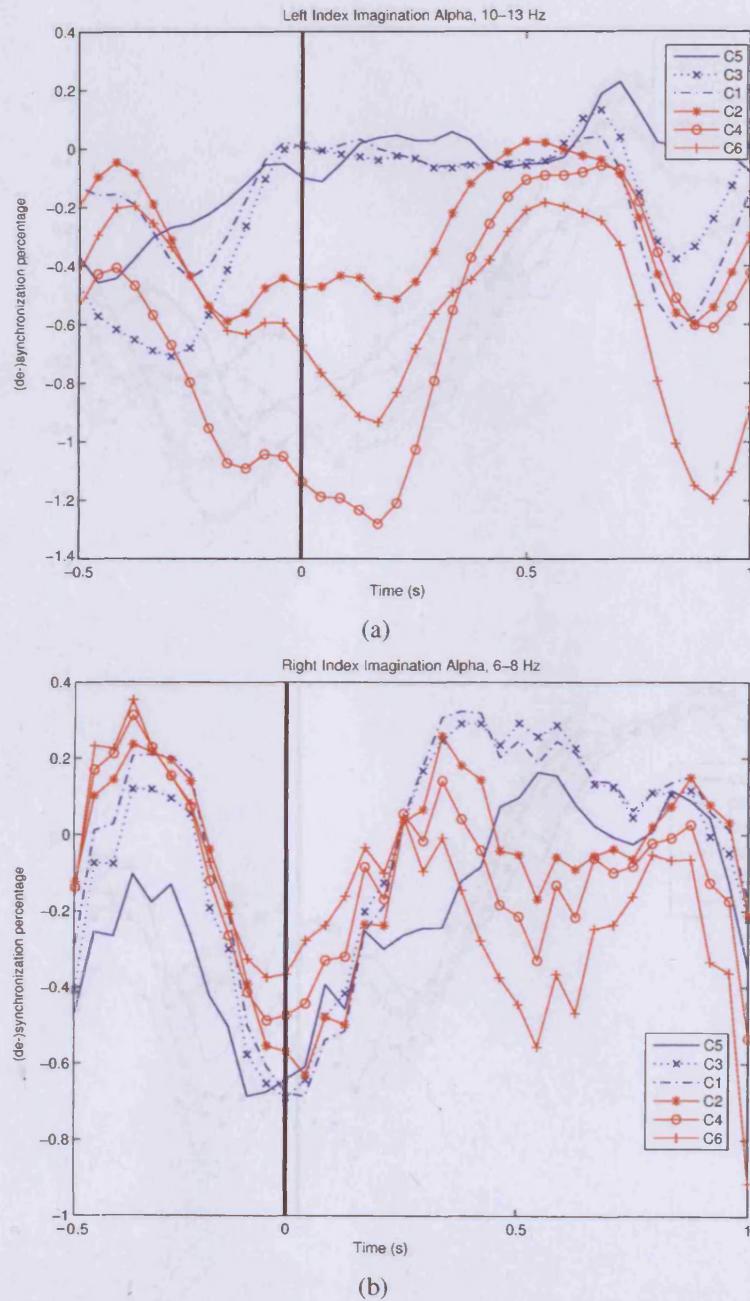
Figure 2.7. ERD for subject three from NIPS2001 BCI dataset during (a) left and (b) right finger imagination.

Chapter 3

PARALLEL FACTOR

ANALYSIS

This chapter introduces a simple parallel factor analysis- (PARAFAC) [51] based approach for BCI where the spatio-tempo-spectral characteristics of the single trial EEGs are jointly considered. First the time-varying single trial EEG spectrums are decomposed into their space, time, and frequency signatures using PARAFAC. Then, the left and right index finger imagery movements are differentiated by interpretation of these signatures. PARAFAC is capable of identifying the ERD/ERS phenomenon, moreover, it can extract the brain background activity. By omitting the factor which does not correspond to the finger movement and utilizing a support vector machine (SVM) classifier, the left and right index movements are distinguished where the feature is the spatial distribution of the movement triggered μ rhythm.

This chapter initially describes the motivation of the developed STF approach for BCI. The construction of the STF models for EEG measurements

is then explained. The sufficient and necessary mathematics of multi-way analysis and PARAFAC are presented next. The results of the approach on real EEG measurements are reported. Finally, discussions and concluding remarks are presented.

3.1 Introduction

Several EEG processing approaches have been addressed to enhance the performance (correct classification rate) of BCI. Most of these studies rely on the temporal or spectral features of the preprocessed EEGs [52]. In [53], the spatial as well as temporal and spectral information have been considered by means of multivariate autoregressive (MVAR) modeling of the multi-channel EEG. An approach based on the analysis of joint space, time, and frequency features of EEG during finger movement is introduced in [54] where the EEG signals are classified with respect to the correlative time-frequency representations (CTFRs) of different channels. Existence of irrelevant potentials over the scalp in parallel with the motion related signals restrains the performance of BCI [55]. Background activity of the brain, motion and ocular artifacts are of such interferences.

PARAFAC-type algorithms have a long history in EEG decomposition. An early study in [56] used the PARAFAC in order to decompose the EEG signals. In [57], PARAFAC was reinstated and termed as "topographic component analysis" and employed to study the event related potentials (ERPs). The phrase "Topographic time-frequency decomposition of the EEG" was adopted in [58] where the distinct activities in the STF domain (called atoms) were simultaneously characterized by their spatial, temporal, and spectral signatures. The authors in [58] extracted physiologically significant activities in the EEG by imposing some mathematical constraints,

e.g. the minimum norm and maximum smoothness conditions. In [59], it was shown that these conditions are not necessary.

It has been established [51] that unique multi-linear decomposition of multi-way arrays is possible using the PARAFAC. In this chapter, PARAFAC is used to decompose EEGs tensors in the space-time-frequency domain for BCI applications and to help with identifying the movement related activity as well as the background brain potentials. The inherent uniqueness of the PARAFAC solution leads to single trial EEG decomposition with minimum *a priori* assumptions [51]. It is worth noting that previous applications of PARAFAC in EEG analysis have only considered the averaged EEGs.

Various BCI studies have established that [4, 5, 54] cortical sensorimotor systems are activated during imagery similar to real motions. It has also been well established that planning and execution of movement leads to a short-lasting amplitude attenuation following by amplification in the μ rhythm (8-13 Hz) called ERD/ERS [4]. Since these brain activities are spatially blurred during propagation from different mediums from the origin to the recording electrode on the scalp [31] (volume conduction), they are spatially smeared. Therefore, their exact localization is rather difficult and entails complex computations. Also, the clearest ERD/ERS, to be utilized in BCI, may occur at different frequency bands and different time points. Index finger movement produces a short-lasting amplitude attenuation followed by amplification of the μ rhythm, mainly in the contralateral motor cortex [4].

3.2   Signal acquisition and pre-processing

The EEG dataset used in this research has been made available by Dr. A. Osman at University of Pennsylvania for *NIPS2001 BCI Workshop*[1]. EEG signals were recorded by 59 channels positioned on the scalp according to the international 10-20 system; sampling rate was set to 100 Hz. A fifth-order Butterworth filter was used for temporal bandpass filtering from 5 to 30 Hz, after baseline removal as introduced in (2.8.3). The baseline removal procedure is carried out in order to cancel out the drift potentials that would appear during the course of signal recording. This drift can be caused by the sub-optimal skin electrode contacts or sweating [4, 31]. The subjects were asked to sit comfortably in front of a computer screen and asked to imagine either left or right finger movement for 180 trials, i.e. 90 for left and 90 for right fingers. Each trial was started with a blank screen displayed for 2 seconds and lasts for 6 seconds; details are shown in Fig. 3.1.

In each trial, the subjects were presented by two highly predictable timing cues for preparation and for execution of movements. During the former, which started at 3.75 seconds and lasted for 250 ms, a letter "L" or "R" (respectively for left or right) appeared on the screen indicating the finger to be moved. The second cue began at 5.0 seconds and displayed an "X" for 50 ms to instruct the user to start the imagery movement. The interested reader is referred to [60] for further details of the temporal structure of data recording. The supplied data files consist of 10 blocks of the synchronized movement experiment recorded from each of 8 subjects with a sampling rate of 100 Hz. While data is available for 8 classes (real or imagined for left, right, both, and none trials), only EEGs corresponding to the left and right index imaginations are available online. In this chapter, signals recorded

---

[1]The dataset is available in $http : //liinc.bme.columbia.edu/competition.htm$.
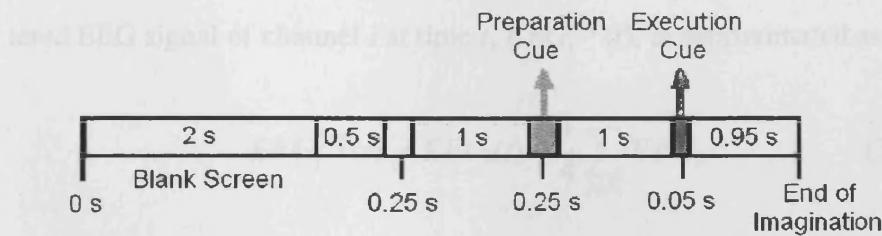
Figure 3.1. Time sequence of each EEG recording epoch.

from the first two subjects for the motor imagination case are considered for the classification. Hence, for each subjects 90 trials, i.e. 45 trials for each index, are considered.

3.2.1 Surface Laplacian filtering

Scalp recorded EEG signals are manifestation of the noisy spatio-temporal superposition of electrical activities originating from different brain regions. In order to accentuate localized activity and reduce electrical diffusion in multi-channel EEG the spatial filtering technique is used. This involves a spherical-spline interpolation procedure [61] in which the measured three dimension locations of the electrodes were used to fit a spline - a continuous surface representing voltage across the scalp. The Laplacian derivative of EEGs at an electrode site is the second derivative of the spline function at that location. Because it is related to rate of change over space; sharp spatial gradients in voltage contribute more to the filtered EEG while gradual changes do less. This amplifies the contribution of nearby electrical sources and diminishes that of distant ones.

Assuming that the distances from a given electrode to its four directional neighboring electrodes are approximately equal, the surface Laplacian fil-

tered EEG signal of channel i at time t, $EEG_i^{Lap}(t)$, is approximated as

$$EEG_i^{Lap}(t) = EEG_i(t) - \frac{1}{4} \sum_{j \in N_i} EEG_j \qquad (3.2.1)$$

where EEG_i is the scalp EEG signal of the i-th channel, and N_i is an index set of the four neighboring channels.

3.3   Continuous wavelet transform

The spectrum of a signal $x(t)$ is given by its Fourier transform [31]

$$x(t) \Leftrightarrow X(f)$$

$$X(f) = \int_{-\infty}^{+\infty} x(t) . \exp(-j2ft)dt. \qquad (3.3.1)$$

However, the Fourier transform reveals temporal characteristics of frequency changes of the signal. This leads to the development of the short-time Fourier transform (STFT) [31]. In STFT, the signal is Fourier transformed within a finite time-window, giving a temporal resolution of the frequency components of the signal.

Unfortunately, the time-window is fixed so it limits the temporal resolutions in the higher frequencies. The wavelet transform (WT) resolves this problem as

$$x(t) \Leftrightarrow X(a, \tau)$$

$$X(a, \tau) = \int_{-\infty}^{+\infty} x(t)\Psi^*(a, \tau, t)dt \qquad (3.3.2)$$

where a and τ represent respectively the scale and shift in the time-frequency domain. A wavelet is a waveform of effectively limited duration and has an
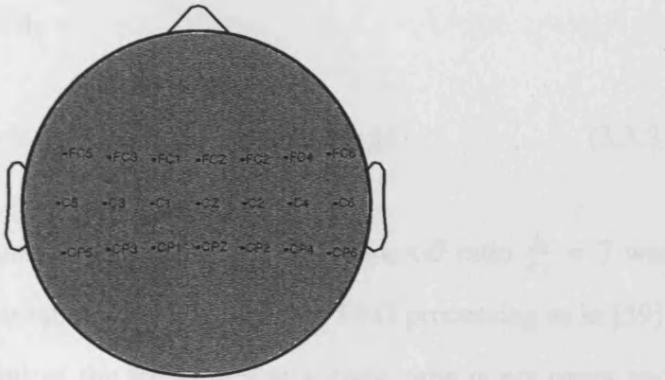
Figure 3.2. EEGs used in this BCI experiment have been recorded from the above 21 electrodes.

average value of zero. Scaling a wavelet simply means stretching or compressing it, and shifting a wavelet means delaying or hastening its onset.

The wavelet analysis is a widely used in EEG noise reduction and feature extraction. Wavelets are separated into continuous and discrete wavelets. A wavelet is called continuous if it can be scaled and shifted to any value; complex Morlet wavelet is an example of continuous wavelets [62].

3.3.1   Complex wavelet transform

Each trial lasts for 6 s as in Fig. 3.1, but not all the time points during this period contain beneficial classification information of the different EEG patterns regarding left- and right index imagination.

The multi-channel wavelet transforms the EEG measurements of *channel* $\times$ *time* into a multi-way arrays of *channel* $\times$ *time* $\times$ *frequency*. To setup a 3-way array, in the present study, the wavelet transform is utilized to provide a time-varying representation of the energy of the signal in μ band using the 21 electrodes over the sensorimotor cortex, i.e. $FC5$, $FC3$, $FC1$, FCz, $FC2$, $FC4$, $FC6$, $C5$, $C3$, $C1$, Cz, $C2$, $C4$, $C6$, $CP5$, $CP3$, $CP1$, CPz, $CP2$, $CP4$, $CP6$, see Fig 3.2.

Following the notation used in [59], the complex *Morlet* wavelets $w(t, f_0)$

can be expressed as [62],

$$w(t, f_0) = A \exp(\frac{-t^2}{2\sigma_t^2}) \exp(2i\pi f_0 t) \qquad (3.3.3)$$

where $\sigma_f = 1/2\pi\sigma_t$, and $A = (\sigma_t \sqrt{\pi})^{-1/2}$. The trade-off ratio $\frac{f_0}{\sigma_f} = 7$ was selected to create the suitable wavelet family for EEG processing as in [59]. Although the estimation of the frequency at a given time is not exact and the whole analysis is slightly influenced by the chosen wavelet, the wavelet analysis is considered a very powerful tool in the analysis of the temporal evolution of the EEG spectrum.

The time-varying energy $E(t, f_0)$ value of $EEG^{Lap}(t)$ at a specific frequency band is computed as the squared norm of the convolution of a complex wavelet of the signal $EEG^{Lap}(t)$

$$\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}} = E(t, f_0) = |w(t, f_0) * EEG^{Lap}(t)|^2 \qquad (3.3.4)$$

where $EEG^{Lap}(t)$ are the Laplacian-filtered multi-channel EEG signals. It is important to notice the classic uncertainties in EEG measurement itself and also in the CWT. The CWT has edge artifacts because the wavelet might not completely be localized in time. It is useful to introduce the cone of influence (COI) [62] to rectify the edge effect problem. The COI can be defined as the area in which the wavelet power caused by a discontinuity at the edge has dropped to $exp(-2)$ of the value at the edge [62]. Conventionally, the EEG measurement error is neglected [5]. In order to surpass the second problem, the time window of interest in BCI is located in the centre of the COI. $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$ is a 3-way matrix indexed by I channels and $\mathcal{J} \times \mathcal{K}$-th component of the estimated energy in the time-frequency domain. The time window from 2.75 seconds to 5.75 seconds for a sampling rate of 100 Hz

and the frequency band from 8 to 13 Hz are chosen. PARAFAC extracts the underlying factors.

The key idea behind this research is in considering the EEGs as the superposition of electro-potentials of the neurons measured by scalp electrodes. EEGs may be represented by using the linear models which are defined in space, time, and frequency, in order to investigate their spatial, temporal, and spectral dynamics. Similar to [59], it is here assumed that distinct local EEG activities (on the scalp) are uncorrelated with the activities of the neighboring areas. EEGs can be modeled as the sum of the distinct components where each distinct component is formulated by the product of its basis in space, time, and frequency domains.

In order to decompose the EEGs into their spatial, temporal, and spectral signatures, the three-way PARAFAC is applied to the three-way EEG data $\underline{X}^{I \times J \times K} \triangleq \underline{X}(1 : I, 1 : J, 1 : K)^2$ where I, J, and K are respectively the number of EEG channels, time instants, and frequency bins. Therefore, $U^{I \times F}$, $D^{J \times F}$, and $S^{K \times F}$ are respectively the spatial, temporal, and spectral signatures of $\underline{X}^{I \times J \times K}$.

While retaining the consistency of formulation, the superscripts may occasionally be dropped to simplify the presentation. The main intention is to keep the notation as simple as possible while clear and helpful. This notation is also kept Chapter 4. However, since Chapter 5 is devoted to a novel STF model of EEGs, a more detailed notation will exploited.

---

[2]Note that the MATLAB matrix notation has been utilized.

3.4 Multi-way arrays

Multi-way arrays are higher-order tensors or multi-dimensional matrices [51].
They are simply sets of data and their elements can be arranged as

$$x_{i_1 i_2 i_3 \ldots i_n} \quad ; \quad i_1 = 1, \cdots, \mathcal{I}_1, i_2 = 1, \cdots, \mathcal{I}_2, i_3 = 1, \cdots, \mathcal{I}_3, \cdots, i_n = 1, \cdots, \mathcal{I}_n$$

$$\underline{\mathbf{X}} \in \mathcal{R}^{\mathcal{I}_1 \times \mathcal{I}_2 \times \mathcal{I}_3 \times \cdots \times \mathcal{I}_n}. \tag{3.4.1}$$

Notice that the vectors and the matrices are two special cases of multi-way
arrays.

3.4.1 Unfolding

The unfolding operation unfolds one of the ways of the multi-way data
onto another. For instance, consider the three-way array $\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}}$ defined
by $x_{ijk}; i = 1, \cdots, \mathcal{I}, j = 1, \cdots, \mathcal{J}, k = 1, \cdots, \mathcal{K}$. Unfolding the third way
of $\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}}$ onto the second way gives

$$\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}} \rightarrow \mathbf{X}^{\mathcal{I} \times \mathcal{J} \mathcal{K}}. \tag{3.4.2}$$

Unfolding the second way of $\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}}$ onto the third way gives

$$\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}} \rightarrow \mathbf{X}^{\mathcal{I} \times \mathcal{K} \mathcal{J}}. \tag{3.4.3}$$

For a three-way array there are six different options of unfolding $\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K}}$
into a matrix as shown in Fig 3.4.1.

The unfolding can be performed sequentially. For instance, a four-way
array can be unfolded into a vector by three consecutive unfolding opera-
tions:

$$\underline{\mathbf{X}}^{\mathcal{I} \times \mathcal{J} \times \mathcal{K} \times \mathcal{L}} \rightarrow \underline{\mathbf{X}}^{\mathcal{I} \mathcal{L} \times \mathcal{J} \times \mathcal{K}} \rightarrow \mathbf{X}^{\mathcal{I} \mathcal{L} \times \mathcal{J} \mathcal{K}} \rightarrow \mathbf{X}^{\mathcal{I} \mathcal{L} \mathcal{J} \mathcal{K}}. \tag{3.4.4}$$
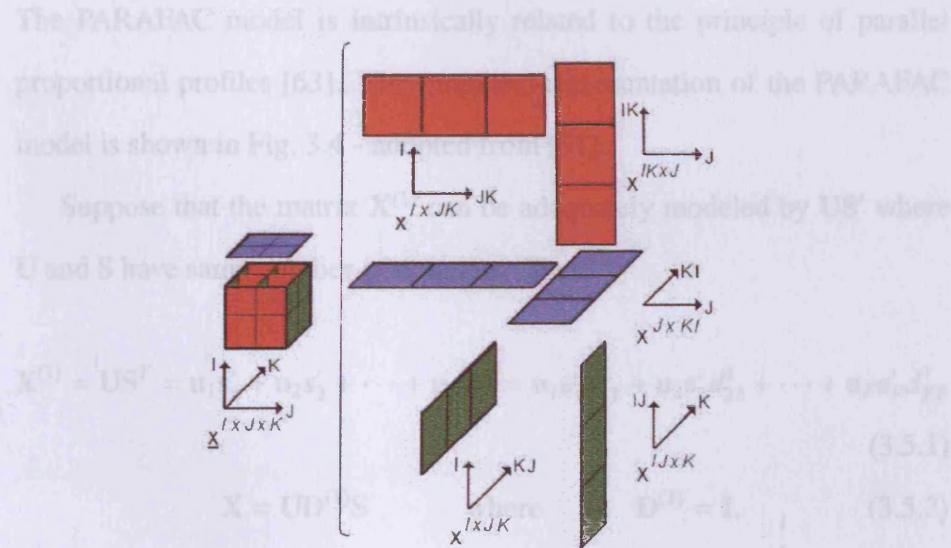
Figure 3.3. The six ways of unfolding the three-way array into a matrix.

Figure 3.4. The graphical representation of the PARAFAC model. The model decomposes the multi-way array into a sum over factor effects pertaining to each dimension. This figure is adopted from [51].

Unfolding multi-way data enables manipulation of the data using normal vector and matrix calculation.

3.5   Parallel factor analysis

The two most used forms of decomposition of multi-way arrays are the PARAFAC and the TUCKER model [63]. Where the PARAFAC decomposition results in interpretable components, the TUCKER model is a convincing multi-linear generalization of the singular value decomposition [63]. Furthermore, the TUCKER model enables evaluation of the PARAFAC by using the so-called core consistency diagnostic (CORCONDIA) measure.

The PARAFAC model is intrinsically related to the principle of parallel proportional profiles [63]. The graphical representation of the PARAFAC model is shown in Fig. 3.4 - adopted from [51].

Suppose that the matrix $\mathbf{X}^{(1)}$ can be adequately modeled by $\mathbf{US}'$ where $\mathbf{U}$ and $\mathbf{S}$ have same number of columns. Thus,

$$\mathbf{X}^{(1)} = \mathbf{US}^T = \mathbf{u}_1\mathbf{s}_1' + \mathbf{u}_2\mathbf{s}_2' + \cdots + \mathbf{u}_F\mathbf{s}_F' = \mathbf{u}_1\mathbf{s}_1'd_{11}^1 + \mathbf{u}_2\mathbf{s}_2'd_{22}^1 + \cdots + \mathbf{u}_F\mathbf{s}_F'd_{FF}^1$$

$$(3.5.1)$$

$$\mathbf{X} = \mathbf{UD}^{(1)}\mathbf{S} \qquad \text{where} \qquad \mathbf{D}^{(1)} = \mathbf{I}. \qquad (3.5.2)$$

If another matrix $\mathbf{X}^{(2)}$ can be described by the same matrices $\mathbf{U}$ and $\mathbf{S}$ but only scaled

$$\mathbf{X}^{(2)} = \mathbf{US}' = \mathbf{UD}^{(2)}\mathbf{S} \qquad (3.5.3)$$

where $\mathbf{D}^{(2)}$ is a diagonal matrix. Suppose that the above procedure can be repeated for F times, i.e, $\mathbf{X}^{(F)} = \mathbf{US}' = \mathbf{UD}^{(F)}\mathbf{S}$. Then the PARAFAC model which was independently proposed by Harshman [56], is expressed by

$$x_{ijk} = \sum_{\lambda=1}^{F} u_{i\lambda}d_{j\lambda}s_{k\lambda} \qquad (3.5.4)$$

where F is the number of components and u, d, and s are elements of $\mathbf{U}, \mathbf{D}$, and $\mathbf{S}$. Due to the symmetry of the components in (3.5.4) the index order of the components does not matter. A general formulation of the model is given as

$$\mathbf{X}^{(i)} = \mathbf{UD}^{(i)}\mathbf{S} \qquad (3.5.5)$$

where

$$\underline{\mathbf{X}} = [\mathbf{X}^1 \quad \mathbf{X}^2 \quad \cdots \quad \mathbf{X}^F]' \qquad (3.5.6)$$

and $\mathbf{D}^i$ is diagonal.

Finally, the common model is expressed in a more compact form by using Kronecker product and its columnwise version, i.e. the Khatri-Rao product. The Kronecker product is computed as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{a}_{11}\mathbf{B} & \cdots & \mathbf{a}_{1m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{m1}\mathbf{B} & \cdots & \mathbf{a}_{mm}\mathbf{B} \end{bmatrix} \tag{3.5.7}$$

and the Khatri-Rao product is given as [63]

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \cdots \quad \mathbf{a}_f \otimes \mathbf{b}_f] \tag{3.5.8}$$

where $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_f]$, $\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_f]$.

Therefore, PARAFAC model is represented as

$$\mathbf{X}^{I \times JK} = \mathbf{U}^{I \times F}(\mathbf{S}^{K \times F} \odot \mathbf{D}^{J \times F})' + \mathbf{E}^{I \times JK} \tag{3.5.9}$$

where $\mathbf{E}^{I \times JK}$ represents the modeling error.

The PARAFAC model can be generalized to higher orders. The higher order PARAFAC is given by

$$x_{i_1 i_2 \ldots i_N} = \sum_{\lambda=1}^{F} u_{i_1\lambda}^{(1)} u_{i_2\lambda}^{(2)} \cdots u_{i_N\lambda}^{(N)} + e_{i_1 i_2 \ldots i_N} \tag{3.5.10}$$

where F is the number of factors. The fourth order PARAFAC model will be investigated in Chapter 5.

3.6   Core consistency diagnostic

The CORCONDIA can be applied to any model that is considered to be a restricted 3-way TUCKER model [51, 63, 64]. According to [64], the

PARAFAC model given in (3.5.9) is a restricted 3-way TUCKER model since

$$\mathbf{X}^{I \times \mathcal{J}\mathcal{K}} = \mathbf{U}\mathbf{G}(\mathbf{S} \otimes \mathbf{D})' \tag{3.6.1}$$

where the elements of $\underline{\mathbf{G}}$ are zero expect those lie on the super-diagonal. The super-diagonal elements are all one. The matrix $\mathbf{G}$ is the matricized version of tensor $\underline{\mathbf{G}}$ on its first dimension, same as $\mathbf{X}^{I \times \mathcal{J}\mathcal{K}}$ for $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$. Thus, the model can be fitted using alternating least squares (ALS) optimization method as in [64] as

$$[\hat{\mathbf{U}}, \hat{\mathbf{G}}, \hat{\mathbf{S}}, \hat{\mathbf{D}}] = \min_{\mathbf{U}, \mathbf{G}, \mathbf{S}, \mathbf{D}} \|\underline{\mathbf{X}}^{I \times \mathcal{J}\mathcal{K}} - \mathbf{U}\mathbf{G}(\mathbf{S} \otimes \mathbf{D})'\|_2^2 \tag{3.6.2}$$

where $\|.\|_2^2$ denotes the squared Euclidian norm.

PARAFAC model is valid if $\underline{\mathbf{G}}$ with elements g_{def} resembles the super-diagonal identity matrix $\underline{\mathbf{T}}$ with elements t_{def}. A resemblance is the COR-CONDIA [51] computed as

$$\text{CORCONDIA} = 100 \times \left(1 - \frac{\sum_{d=1}^{F} \sum_{e=1}^{F} \sum_{f=1}^{F} (t_{def} - g_{def})^2}{\sum_{d=1}^{F} \sum_{e=1}^{F} \sum_{f=1}^{F} g_{def}^2}\right). \tag{3.6.3}$$

From (3.6.3), the PARAFAC model is perfect if the numerator simply becomes zero giving 100% consistency. A core consistency well below 70%-90% indicates that either too many components are used or the model is misspecified [64]. The threshold is set objectively [64]; that is, depending on the application a certain value above 70% is selected by which interpretable factors can be extracted. In [64], the value 85% is suggested as a proper value for diverse applications. Although the CORCONDIA is an effective measure of the number of factors to be included, *Bro* in [64] emphasizes that other measures such as sum of the squared residuals versus

---

Algorithm 1 PARAFAC by ALS

Require: Initial values for maxIter, **D** & **S**

Ensure: iter = 0, ΔSSE > ϵ & SSE$_0$ = 0

    while iter < maxIter & ΔSSE > ϵ **do**

        $\mathbf{Z} = (\mathbf{S} \odot \mathbf{D})$

        $\mathbf{U} = \mathbf{X}^{(I \times \mathcal{J}\mathcal{K})}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{\dagger}$

        $\mathbf{Z} = (\mathbf{S} \odot \mathbf{U})$

        $\mathbf{D} = \mathbf{X}^{(\mathcal{J} \times I\mathcal{K})}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{\dagger}$

        $\mathbf{Z} = (\mathbf{D} \odot \mathbf{U})$

        $\mathbf{S} = \mathbf{X}^{(\mathcal{K} \times I\mathcal{J})}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{\dagger}$

        $\text{SSE}_{iter} = \|\mathbf{X}^{(I \times \mathcal{J}\mathcal{K})} - \mathbf{U}(\mathbf{S} \odot \mathbf{D})'\|_2^2$

        $\Delta\text{SSE} = |\text{SSE}_{iter} - \text{SSE}_{iter-1}|$

    end while

---

the number of factors, inspection of the parameters and cross validation can be taken into account. In this chapter, the conventional value for threshold, 85%, is selected.

3.7   PARAFAC by alternating least squares

The implementation of PARAFAC is based on the alternating least squares (ALS) optimization. It follows quite simple steps; randomly initialize all model parameters, update each parameter by minimizing a cost function with respect to that parameter while keeping all other parameters fixed, continue the above procedure until a minimization criterion is met.

Consider the PARAFAC model defined in (3.5.9); given the cost function as $[\hat{\mathbf{U}}, \hat{\mathbf{S}}, \hat{\mathbf{D}}] = \min_{\mathbf{U},\mathbf{S},\mathbf{D}} \|\mathbf{X}^{I \times \mathcal{J}\mathcal{K}} - \mathbf{U}(\mathbf{S} \odot \mathbf{D})'\|_2^2$, the ALS algorithm for PARAFAC is implemented as in **Algorithm** (1) where $\dagger$ denotes the Moore-Penrose inverse operator and the absolute value operation is shown by $|.|$.

3.8   BCI and parallel factor analysis

Traditionally, decomposition of EEG into its constituent components has been carried out using the independent component analysis (ICA). How-

ever, the assumption that the brain sources are mutually independent is questionable [31]. Suppose that the motion related potentials are synchronous, highly localized, and uncorrelated from the background neuronal activities. This assumption leads to exploit the PARAFAC. The noteworthy distinction of PARAFAC is that the decomposition of multi-way data using PARAFAC is unique without further orthogonality or independence constraints [63]. Multi-channel EEG data are transformed into time-frequency domain by CWT. The increase of dimensionality gives the 2-way array, the matrix of space-time, an extra modality of frequency yields a 3-way array of space-time-frequency. ICA can merely analyze such data by unfolding some modalities into others, reducing the multi-way array again into matrices. The unfolding process makes the interpretation of the results quite doubtful since it removes some specific information endorsed by those modalities. Consequently, rather than unfolding these multi-way arrays into matrices, the data is analyzed using the multi-way PARAFAC model. The main advantage of PARAFAC over ICA is that uniqueness is ensured making it unnecessary to impose constraints such as statistical independence. Using ALS, as the most common method, PARAFAC model parameters are estimated. In ALS, in order to decompose a tensor to parallel factors, a cost function (normally the squared error) is minimized as

$$[\hat{\mathbf{U}}, \hat{\mathbf{S}}, \hat{\mathbf{D}}] = \arg\min_{\mathbf{U},\mathbf{S},\mathbf{D}} \|\mathbf{X}^{I \times \mathcal{J}\mathcal{K}} - \mathbf{U}^{I \times \mathcal{F}}(\mathbf{S}^{\mathcal{K} \times \mathcal{F}} \odot \mathbf{D}^{\mathcal{J} \times \mathcal{F}})'\|_2^2 \qquad (3.8.1)$$

which corresponds to optimizing the maximum likelihood of a Gaussian noise model. The algorithm can be initialized in several ways, e.g. by randomly defining all parameters and stopping when all parameters have converged [63].

3.9 Classification results

Based on the above formulations, the EEG recordings for the first two subjects of the *NIPS2001* competition, see Section 3.2, were classified in a two-class (left and right index imaginations) BCI framework. EEG recordings were firstly bandpass filtered between 5 Hz and 30 Hz, after baseline removal. Then, following [65], the surface Laplacian filtering algorithm was implemented in order enhance the SNR, Section 3.2.1. It is physiologically established [31] that during finger movement, either real or imagery, the motor cortex area is highly involved. Therefore, in order to reduce the computational complexity, the electrodes shown in Fig. 3.2 were selected. Then, the EEG recordings were wavelet-transformed in order to construct the three-way tensors as explained in Section 3.3.1.

The three-way tensors were introduced to PARAFAC and the STF factors for all of the trials were extracted. Examples of such extracted components for left and right finger imagination have been demonstrated in Fig. 3.5 to Fig. 3.8. Note that, in some cases such as Fig. 3.6 and Fig. 3.8, two components were identified based on the CORCONDIA value. Here, the number of components was obtained when the CORCONDIA value was more than 85% [63]. As discussed previously, the factors demonstrating clear ERDs in the contralateral hemisphere were selected. The other factor, if extracted, represented

- the background brain activity; if it was spatially spread all over the scalp or had flat temporal or spectral signatures;

- the eye-blink or other types of ocular artifacts if it was confined to the frontal area;

- the background α rhythm if it was in the occipital area,
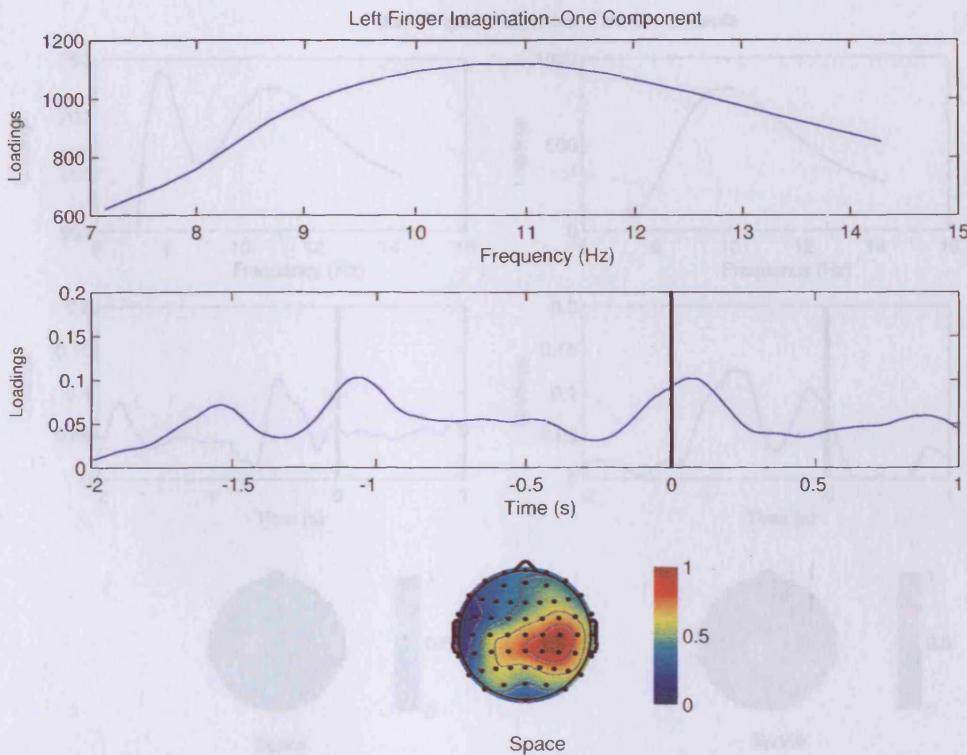
Figure 3.5. Left finger imagination; one component. The upper figure represents the spectral signature of the extracted component. The middle figure shows the temporal signature and the lower figure demonstrates the spatial distribution. The time onset of execution cue is at 0 seconds and shown by the vertical line.

and was eliminated from analysis. Note that, in Figs. 3.5 to 3.8, for the purpose of presentation, all the electrodes were included. However, as mentioned for the classification purposes only the electrodes over the sensory-motor cortex were considered.

3.9.1 Support vector machine classifier

The SVM classifier was utilized to classify the spatial signatures of the selected factors. The goal of an SVM is to find an optimal separating hyperplane (OSH) for a given feature set. The OSH is found by solving the

Figure 3.6. Left finger imagination; two component. The upper figures represent the spectral signatures of the extracted components. The middle figures show the temporal signatures and the lower figures demonstrate the spatial distributions. The time onset of execution cue is at 0 seconds and shown by the vertical lines.

following constrained optimization problem,

$$\min_{z,b,\gamma_i} \left(\tfrac{1}{2}\|z\|_2^2 + C \sum_{i=1}^{l} \gamma_i \right)$$

$$\text{s.t.} \quad q_i(z \cdot g_i - b) + \gamma_i \geq 0 \qquad i = 1, 2, \cdots, l. \qquad (3.9.1)$$

where l is the number of training vectors, $q_i \in \{1, -1\}$ is a class label, and $(\cdot)$ is the dot product. The parameter z determines the orientation of the separating hyperplane, γ_i is the i-th positive slack parameter, and g_i is a vector containing the features $g_i = [f_1(i)f_2(i) \cdots f_{21}(i)]'$.

The 21-elements feature vector g contained the spatial signature of the selected factor. The non-negative parameter C is a penalty term and can be
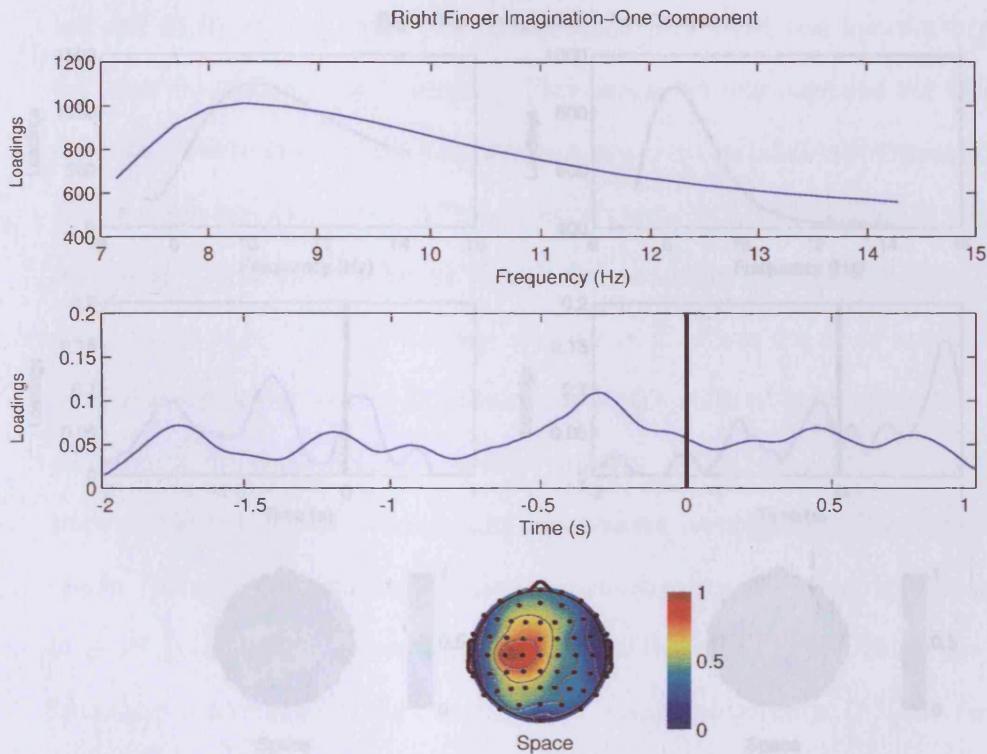
Figure 3.7. Right finger imagination; one component. The upper figure represents the spectral signature of the extracted component. The middle figure shows the temporal signature and the lower figure demonstrates the spatial distribution. The time onset of execution cue is at 0 seconds and shown by the vertical line.

considered as the regularization parameter. It is set by the user. A larger C is equivalent to assigning a higher penalty to the training errors. The optimum value for C is found such that it minimizes the cross validation error. The support vectors (SVs) are the samples from the training data set that fall on the separating hyperplane. It is desirable to have small number of SVs and a more compact classifier. The OSH (generally nonlinear) is then computed as a decision surface of the form

$$f(\mathbf{g}) = \text{sgn}(\sum_{i=1}^{L_s} q_i \alpha_i \mathcal{K}(\mathbf{g}_i^s, \mathbf{g}) + b) \qquad (3.9.2)$$

where $\text{sgn}(\cdot) \in \{\pm 1\}$, $\mathbf{g}_i^s$ are SVs, $\mathcal{K}(\mathbf{g}_i^s, \mathbf{g})$ is the nonlinear kernel function (if
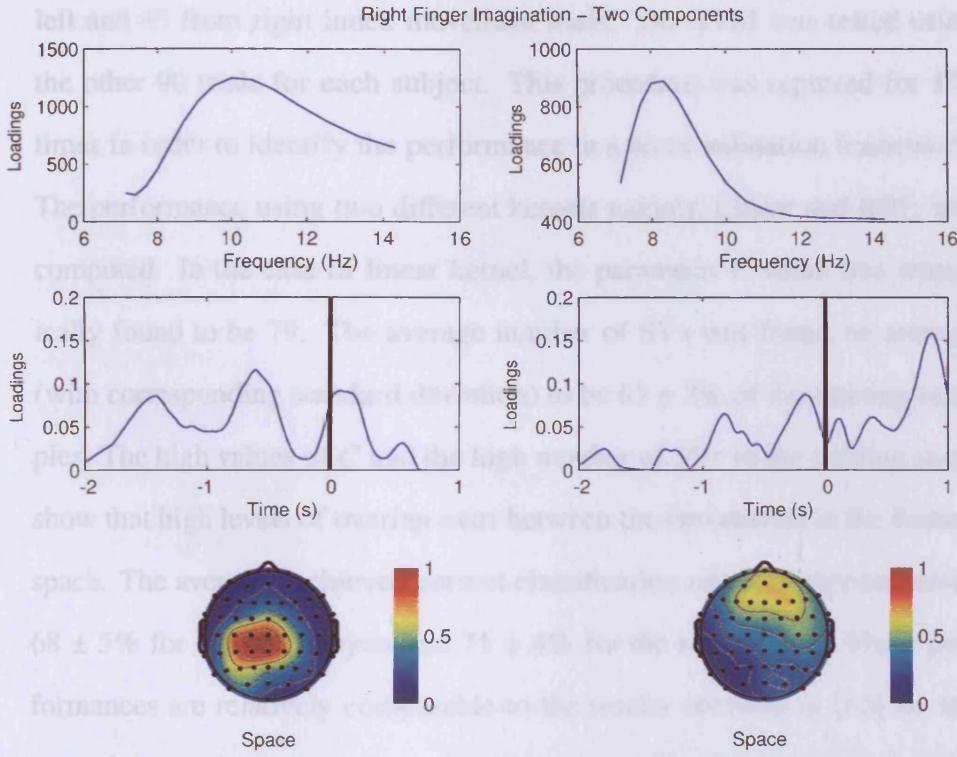
Figure 3.8. Right finger imagination; two component. The upper figures represent the spectral signatures of the extracted components. The middle figures show the temporal signatures and the lower figures demonstrate the spatial distributions. The time onset of execution cue is at 0 seconds and shown by the vertical lines.

$\mathcal{K}(\mathbf{g}_i^s, \mathbf{g}) = \mathbf{g}_i^s \cdot \mathbf{g}$ the SVM is linear), and L_s is the number of support vectors. The kernel for a nonlinear SVM projects the samples to a feature space of higher dimension.

Among nonlinear kernels, the Gaussian radial basis function (RBF) defined as $\mathcal{K}(\mathbf{g}_i^s, \mathbf{g}) = \exp(-|\mathbf{g} - \mathbf{g}_i|^2/2\rho)$ is widely used due to quasi-Gaussian distribution of the data sets with large number of samples. In the RBF kernel, ρ is an adjustable parameter governing the width (variance) of the kernel. In the following, the results of the new hybrid PARAFAC-SVM scheme are described. The PARAFAC decomposes the EEGs and SVM classifies the extracted features.

For training the SVM classifier, 90 trials were randomly chosen; 45 from

left and 45 from right index movement trials. The SVM was tested using the other 90 trials for each subject. This procedure was repeated for 100 times in order to identify the performance in a cross-validation framework. The performance using two different kernels namely, Linear and RBF, was computed. In the case of linear kernel, the parameter C value was empirically found to be 79. The average number of SVs was found on average (with corresponding standard deviation) to be 63 ± 3% of the training samples. The high values of C and the high number of SVs in the training stage show that high levels of overlap exist between the two classes in the feature space. The averaged achieved correct classification rate was approximately 68 ± 5% for the first subject and 75 ± 4% for the second one. These performances are relatively comparable to the results obtained in [65] for the same dataset where an average classification results of approximately 85% has been reported. It is worth noting that the performance of the classifier is highly dependent on the C. For instance, reducing C to 10 can degrade the performance by 20% to 30%, makes it below the chance level.

In order to increase the performance, the RBF kernel was utilized in the SVM classifier. Optimal values for C and ρ which maximize the correct classification rate were found by carrying out a grid search. The search suggested that $\rho = 1$ and $C = 10$ yield the maximum correct classification results. The average number of support vectors was reduced to 57 ± 2% of the training trials when using the RBF kernel. The best achieved averaged classification rate was 76 ± 6% for the first and 79 ± 4% for the second subject.

Based on the above results, it can be concluded that the STF modeling approach is not fully capable of extracting classifiable features. In next the section, the main reasons why the PARAFAC-based BCI can not produce

outstanding results are discussed.

3.10 Discussions

Despite the novelty of the presented method for BCI and its theoretical sim-
plicity, there are few issues that will be raised here. As mentioned earlier,
careful analysis of the brain signals is of great importance for BCI. Although
the STF decomposition of EEG spectrums may be called an effective ap-
proach for BCI, its application in real-world systems is questionable due to
the following problems.

- The main issue in the application of PARAFAC for BCI is the com-
  putational requirements for the continuous wavelet transform. More-
  over, decomposition of multi-way array to their space, time, and fre-
  quency signatures by using the ALS is a time consuming procedure.

- The second concern is the utilization of the STF signatures of EEGs
  for classification purposes. Unfortunately, they suffer from high intra-
  and inter-subject variability which can be observed in the temporal
  signatures.

- Moreover, when the movement related potentials are highly obscured
  in the background brain activity, e.g. the dominant 10 Hz potentials
  generated in the occipital lobe, the PARAFAC method cannot be ef-
  fectively used. In these cases, application of Laplacian transform
  is strongly recommended [61]. However, the Laplacian filtering in-
  creases the computations, which might not be of interest in real-time
  BCI applications.

- High dimensional feature space can lead to a complicated pattern
  recognition problem. Using unsupervised (non-)linear transforms such

as (kernel) principal component analysis (PCA) or supervised ones, i.e. sequential forward selection (SFS) [49], a sufficient number of features which provide acceptable classification results may be selected. Evidently, the selected features should present effective class discriminatory information.

One might argue that, if the application of the STF modeling in single trial EEG processing is such troublesome, there is no use in implementing it. In next two chapters, it will be shown that how it is possible to make better use of PARAFAC and the STF modeling in extraction and removal of the commonly dominant artifact such as eye-blinks. It will be shown that the eye-blink artifacts may be readily identified by using the STF model of the contaminated EEG measurements.

3.11 Conclusions

In this chapter, an approach based on a hybrid PARAFAC-SVM method for a simple BCI paradigm was presented. Within this scheme, an EEG space-time-frequency decomposition in μ band (8-13 Hz) at the preprocessing stage was developed. Using PARAFAC, two (or one) distinct factors in μ band for each EEG trial are extracted. The SVM classifier is utilized to classify the spatial distribution of the movement related factor. The movement related factor is identified by analyzing the spatial, temporal, and spectral signatures of the resulting factors extracted by PARAFAC.

The presented approach provides comparable results to previous studies on the same dataset.[3] However, comparing to [65], the results of PARAFAC-based BCI are lower by approximately 10%. It is worth noting that the

---

[3]Visit http://liinc.bme.columbia.edu/competitionresults.htm for the results of the NIPS2001 BCI competition.

powerful SVM classifier was effectively utilized here, however, the weak

classification performance is attributed to the overlap between the left and

right classes in the feature space identified by PARAFAC.

Chapter 4

BLIND SIGNAL
EXTRACTION AND
PARAFAC

In this chapter, a blind signal extraction (BSE) scheme for removal of the eye-blink (EB) artifacts from EEG signals is developed. In this method, the conventional source extraction algorithm is provided with an estimation of the column of the mixing matrix corresponding to the *point source EB artifact*. Hence, the BSE approach is called semi-blind signal extraction (SBSE). After the extraction of the EB source, the artifact-removed EEGs are reconstructed by deflation [66]. The vector corresponding to the spatial distribution of the EB factor, the *a priori* knowledge, is identified by fitting a space-time-frequency (STF) model to the EEG measurements using PARAFAC. This approach introduces the possibility of incorporating PARAFAC within the blind signal extraction framework for single trial EEG processing applications. Aiming at extracting the EB artifact, the SBSE exploits the spatial as well as temporal prior information during the extraction procedure. Experiments on synthetic and real EEG measurements confirm that the proposed algorithm identifies and removes the EB artifact from the raw EEG measurements.

This chapter begins with a detailed review of the classic methods for EB artifact removal. SBSE algorithm is then developed and its performance is evaluated. The results of the implementation of the proposed method on the synthetic data are compared to that of a recently published approach on artifact removal from EEGs in a spatially constrained blind source separation (SCBSS) framework [67]. The chapter is closed by the implementation of the proposed algorithm on EB contaminated real EEG recordings.

4.1 Introduction

Artifacts are due to (non-)cerebral potentials and contaminate the EEGs. For instance, for a BCI system based on the steady-state movement related potentials (ssMRPs), the steady-state visual evoked potentials (ssVEPs) which are basically cerebral potentials are not desirable. The main problem with such physiologically generated artifacts is that they can be included in the analysis mistakenly. Therefore, there is a critical need to avoid, reject, or remove them. Non-physiological artifacts are mainly produced by the outside world, for instance the 50 Hz power-line noise or changes in electrode impedances which may be avoided by careful filtering.

Physiological artifacts arise from a variety of electro-mechanical sources within the body, e.g. brain itself or heart. Electrocardiography (ECG) artifacts may introduce a rhythmic activity into the EEG signal or the respiration which causes very low frequency artifacts. Moreover, the skin conditions alter during the course of recording and may degrade the quality of recording. For instance, sweating increases the impedance of electrode-skin contact and causes the drift artifacts in EEGs [68, Section 1].

Particularly for BCI, two physiological artifacts, ocular and muscular, are widely investigated [31]. EOG signals are generated by eye-movements

or blinks [31]. Suppressing EBs over a sustained recording course is particularly difficult, mainly due to their amplitude which is on average ten times larger than the average amplitude of the cortical potentials. EBs may contaminate the majority of the electrode signals even those recorded from the occipital lobe. It has become very desirable to remove the EB artifacts without distorting the underlying EEG [31]. In this regard, a reliable and fast, either iterative or batch, algorithm for EB artifact removal is of great interest. The EMG signals generated by head, body, or jaw movements may also cause considerable disturbances in the measurements. Physiological artifacts such as EOGs and EMGs are much more challenging to handle than non-physiological artifacts. Generally, there are different ways of handling the artifacts in BCI systems:

- *Artifact avoidance*: The first is to instruct the subjects to avoid blinking or moving their body during the EEG recording sessions. Although, it might be very advantageous to have the data without any EB or motion artifacts, instructing the subject to avoid blinking or moving may introduce additional mental tasks.

- *Artifact rejection*: Artifact rejection refers to the process of eliminating the artifact contaminated trials after the recording session. It presents important advantages over the artifact avoidance approach. It is easier for the subjects to participate in the experiments and perform the required tasks without concerning about the occurrence of the artifacts. Second, almost all the EEG processing software packages have the "automatic EB rejection" feature. Further monitoring of body movements is usually carried out by manual inspection of EEGs. Artifact rejection is the first choice in clinic. However, it is not applicable in real-time BCI.

- *Artifact removal*: The procedure of identifying and removing artifacts from the recorded brain signals is generally called artifact removal. An artifact removal method should be able to detect and remove the artifacts. Common methods for removing artifacts in EEG signals are summarized here.

Linear filtering: It is easy to remove the artifacts localized in certain frequency bands or certain time periods which do not overlap with that of the neurological phenomenon of interest. For example, a simple lowpass filtering may remove the high frequency EMG artifacts. Also, the highpass filtering can remove the EOG artifacts if one is interested in high frequency components of EEGs. However, the filtering approach may not always remove EOG artifacts without distorting the desired information. Such methods are partly successful in BCI systems that classify features extracted from higher frequency components of the EEG (e.g., μ or β rhythms). However, for BCI systems that depend on low frequency rhythms, this procedure is not applicable since the spectrum of slow cortical potentials and EB artifacts overlap.

Linear combination and regression: The ocular artifacts can be removed from EEGs by using a linear combination of the EB-contaminated EEG signals and an EOG channel as the reference channel. Briefly, this procedure estimates and removes the EOGs in a "least squares" sense. Mathematically, the coefficient B is calculated using (4.1.1), where X_i and Y_i are the measured EOG and EEG potentials, respectively, at time point i.

$$B = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{(X_i - \bar{X})^2} \qquad (4.1.1)$$

The resultant B is then used to correct the EEG according to (4.1.2). $EEG_{filt,i}$ and EEG_i are respectively the restored and measured EEG voltages at time point i, and C is the constant from the least squares equation, i.e., $C = \bar{Y} - [B \times \bar{X}]$

$$EEG_{filt,i} = EEG_i - B \times EOG_i - C \qquad (4.1.2)$$

Since, this technique falls out of this dissertation scope, the interested reader is referred to [69, 70], and [71]. Besides, some general recommendations on the effective use of regressors for EOG removals have been made in [72].

Blind source Separation (BSS): BSS-based methods are used in order to separate EEG recordings into uncorrelated or independent components (sources). Among them, a few components are labeled as artifact in accordance with some criteria. EEGs are then reconstructed after discarding those components [55]. BSS-based methods have been widely used to remove the EOG artifacts in clinical studies. The main advantage of using such methods is that they do not rely on the availability of the reference electrode signals. However, the drawback is that proper criteria should be defined to distinguish between the artifact and non-artifact components. Selection of the artifacts may be carried out by visual inspection or in an automatic framework [55, 73, 74, 74–79].

Principal component analysis: PCA by singular value decomposition (SVD) is used to reversibly decompose epochs of EEGs into multiple linearly independent (temporally and spatially uncorrelated) components. The result of SVD includes the components, expressible

as time series waveforms, and the factors that determine how much each component waveform contributes to each EEG channel. By discarding some of component waveforms from the linear combination, new EEGs may be reconstructed, differing from the original in desired ways, for instance, EOG artifacts may be removed. In [80], an approach has been developed in which the factors that reconstruct the modified EEG from the original are stored as a matrix. This matrix acts as a spatial filter in order to remove EOG and ECG artifacts. However, it has been reported [80] that PCA cannot completely separate the EOGs from EEG signals when they have comparable amplitudes. The implementation of PCA is indeed easy and does not entail considerable computations. This has been the main reason for the popularity of PCA among the engineers and clinical researchers.

In summary, various methods for EB artifact removal from EEGs have been documented that are mainly based on blind source separation [31, Ch. 2] and linear regression [81]. Approaches, such as trial rejection, eye-fixation, EOG subtraction, principal component analysis [71] and robust beamforming [77] have been also documented as having varying success. A hybrid BSS-SVM method for removing the EB artifacts along with a temporally constrained BSS algorithm have been recently developed in [55] and [74]. Moreover, several other methods based on H^∞ [82] adaptive- and spatial filters [80] have been presented in the literature for EB removal. It has been shown [31,55,74,76–79,81,83] that the regression- and BSS-based methods are most reliable approaches despite no quantitative comparison for any reference dataset being available.

Nonstationary EEG signals yield temporal and spatial information about active areas of the brain and have been efficiently exploited for localizing the

EEG sources and the removal of various artifacts from EEG measurements using PCA. In [84], the PCA is utilized to decompose the signals into uncorrelated components where the first component, the component with the highest variance, represents the EB artifact. However, the use of PCA introduces a nonuniqueness due to an arbitrary choice of rotation axes; this nonuniqueness may be resolved by introducing a reasonable constraint. Recently, the ICA has been applied to solve this problem by imposing the statistical independence constraint which is stronger than the orthogonality condition exploited by PCA [85]. Moreover, by using ICA the energy (variance) information is lost and the EB component should be identified manually or in an automatic correction framework [55].

In these conventional methods prior assumptions such as independence, entropy, orthogonality, ortho-normality, non-negativity, and sparsity have been frequently considered in the separation process. However, such mathematical constraints do not usually reflect specific physiological phenomena. In essence, there are two different approaches for incorporating prior information within the semi-blind EEG source separation (extraction); first, the Bayesian method [86] which introduces a probabilistic modeling framework by specifying distributions of the model parameters with respect to prior information. Often the probabilistic approach is too complicated to be implemented, specifically in high density EEG processing; the slow convergence drawback should also be highlighted. The second more feasible approach proposes the expansion of the conventional gradient-based minimization of particular *cost functions* by including rational physiological constraints. Established temporally or spatially constrained BSS algorithms such as [55, 75, 83, 87] are the outcome of the above approach. However, constrained BSS based methods suffer from the extensive computational

requirements (unlike blind source extraction methods [88]) of source separation. Moreover, in CBSS approach, the accuracy of the prior is of great importance. An error in computation (estimation) of the prior might lead to significant degradation of the BSS algorithm. For instance, in an CBSS the prior can be considered to be a column of the mixing matrix corresponding to the source of interest. Then, if this prior is not faithfully estimated, the optimization procedure is initialized with an erroneous column. In [83], it is well investigated that since that column may be kept constant during to the optimization procedure, the error can potentially influence the rows of the demixing matrix. In such cases, robust BSS approaches are needed which can compensate the error induced by the erroneous prior [83].

Simple and straightforward priors, such as the spectral knowledge of the ongoing EEGs or spatial topographies of some source sensor projections, can be exploited in semi-blind EEG processing. In this regard, an interesting work on topographic-time-frequency decomposition is proposed in [58] in which, however, two mathematical conditions on the time-frequency signatures, namely, minimum norm and maximal smoothness are imposed. It has been shown that these conditions may provide a unique model for EEG measurements. Consolidating [58], recently in [59] the STF model of a multi-channel EEG has been introduced by using PARAFAC [51]. Chapter 3, presented the utilization of the STF model in single trial EEG processing for brain computer interfacing [31, 89].

In this chapter, a physiologically inspired semi-blind signal extraction technique for removing the EOG artifacts from single trial multi-channel EEGs is presented. The SBSE method is based on that introduced in [88]. Incorporating the spatio-tempo-spectral signatures of the extracted factor(s), the EB factor is selected and its spatial distribution is exploited in the SBSE

as priori knowledge. The main advantages of the proposed method are as follows:

1. In the BSS- and CBSS-based methods [67, 73–75, 83, 90, 91] identification of the correct number of sources is an important issue. In addition, they require high computational costs. However, the simplicity of the method is due to use of the spatial *a priori* information in order to guarantee that the first extracted source is the EB source. Therefore, there is no need to extract other sources. This reduces the computational requirements significantly. EEGs are then reconstructed in a very fast batch deflation procedure.

2. Unlike the methods presented in [73] and [55], there is no need to compute any objective criteria for distinguishing between EB and spurious peaks in the ongoing EEGs.

3. Unlike the regression-based methods [69], the proposed method does not need any reference EOG channel recordings (typically three channels).

4. It is important to notice that in presented approach there is no need to separate the dataset into training and testing subsets as in [74]. As long as, by using any primitive method an EB artifact is identified, the presented method can be utilized to remove it from EEGs.

This chapter is organized as follows. In Section 4.2, the SBSE method is presented and its performance is compared to that of an existing spatially constrained BSS algorithm presented in [83]. Afterwards, the fundamentals of the PARAFAC are briefly discussed. An effective procedure to identify the spatial signature of the EB relevant factor is introduced. The results

are subsequently reported in Section 4.3, followed by a brief study on the robustness of the SBSE in extracting the EB artifacts in Section 4.4 and concluding remarks in Section 4.5.

4.2 Algorithm Development

The EB contaminated EEG measurements are modeled as linear, instantaneous mixtures. For simplicity, it is frequently assumed that [55, 67, 69, 73–75, 83, 90, 91] the traveling medium between the origin of the electrical activity and the recording electrode, comprising of the brain matter, the cerebrospinal fluid (CSF), the scull, and the skin can be represented by a linear transfer function - although the linearity assumption can be unrealistic [31]. On the other hand, the assumption of instantaneous mixing of EEG sources is widely accepted [31]. The main point is that due to the very high propagation velocity of EEGs, in the range of electromagnetic waves, it is essentially reasonable to assume that EEG sources are mixed instantaneously. Recently, there have been some studies that explore the EEGs by convolutive models [92].

In this thesis, the contaminated EEG measurements are assumed to be generated from $N - 1$ zero-mean real mutually uncorrelated sources and one EB source $s_j(t)$ as

$$\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_j(t), \cdots, s_N(t)]' \qquad (4.2.1)$$

at time instant t where $[\cdot]'$ denotes the transpose operator mixed by an $N \times N$ real full column rank matrix

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N] \qquad (4.2.2)$$

where generally a_i is the i-th column of A and specifically a_j is the column of A corresponding to the EB source s_j. The vector of mixtures at time instant t is given as

$$\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_N(t)]' \qquad (4.2.3)$$

and therefore the mixing system is modeled as

$$\mathbf{x}(t) = \mathbf{As}(t) + \mathbf{n}(t) \qquad (4.2.4)$$

where

$$\mathbf{n}(t) = [n_1(t), n_2(t), \cdots, n_N(t)]' \qquad (4.2.5)$$

is the additive white Gaussian zero-mean noise. It is often assumed [55, 69, 83, 87] that the noise is spatially uncorrelated with the sensor data and temporally uncorrelated. Therefore, the time lagged autocorrelation matrix $\mathbf{R}_k$ can be calculated as

$$\mathbf{R}_k = E[\mathbf{x}(t)\mathbf{x}'(t - \tau_k)] = \sum_{i=1}^{N} r_i(\tau_k)\mathbf{a}_i\mathbf{a}_i' \qquad (4.2.6)$$

for $k = 1, 2, \cdots, K$ where K is the index of the maximum time lag τ_K and $E[\cdot]$ denotes the statistical expectation operator. In (4.2.6), $r_i(\tau_k) = E[s_i(t)s_i(t - \tau_k)]$ is the time lagged autocorrelation value of $s_i(t)$.

4.2.1 Semi-Blind EB Signal Extraction

Blind signal extraction has received much attention in biomedical signal processing due to its potential applicability to a wide range of problems. The second order statistics based BSE methods are widely preferred [74, 88] to higher order statistics based methods since they need shorter data record-

ings. Moreover, multi-channel EEGs are not conventionally sampled at very high frequency rates suitable for higher order statistics based algorithms.

The vector $x(t)$ in (4.2.4), the EEG recording at time t, is a linear combination of the columns of the mixing matrix, the a_is, weighted by the associated sources and contaminated by the sensor noise $n(t)$. Therefore, the most straightforward way to extract the j-th source, the EB artifact s_j, is to project $x(t)$ onto the space in $\mathbb{R}^N$ orthogonal to, denoted by $\perp$, all of the columns of A except a_j, that is, $\{a_1, \cdots, a_{j-1}, a_{j+1}, \cdots, a_N\}$. Two vectors are defined first: p which is orthogonal to the subspace spanned by columns of A expect a_j and q which is equivalent to a_j. By adopting the notation used in [88] and [93], it may be written that

$$y(t)q = E_{q|p^\perp}x(t) \qquad (4.2.7)$$

where $y(t)$ is an estimate of one source, say $s(t)$, and $p^\perp$ denotes the space in $\mathbb{R}^N$ orthogonal to p. In (4.2.7), $E_{q|p^\perp} = \frac{qp'}{p'q}$ represents the projection of q onto the space $p^\perp$, that is the space spanned by the columns of A except a_j. Then, $y(t)$ can be extracted using the spatial filter p as

$$y(t) = p'x(t) \qquad (4.2.8)$$

in which the scalar $\frac{1}{p'q}$ has been omitted and q has been dropped from both sides of (4.2.7). In the second-order statistics based BSE [88], both p and q are unknown. In order to extract one source the following cost function is proposed

$$[\hat{d}, \hat{p}, \hat{q}] = \arg\min_{d,p,q} J_M(d, p, q) \qquad (4.2.9)$$

where

$$J_M(\mathbf{d}, \mathbf{p}, \mathbf{q}) = \sum_{k=1}^{K} \|\mathbf{R}_k\mathbf{p} - d_k\mathbf{q}\|_2^2 \qquad (4.2.10)$$

and $\mathbf{d}$ is a column vector

$$\mathbf{d} = [d_1, d_2, \cdots, d_K]' \qquad (4.2.11)$$

and $\| \cdot \|_2^2$ denotes the squared Euclidean norm.

Multiple time lags instead of only a single time lag are employed. This approach minimizes the chance of the time-lagged considered autocorrelation matrices having duplicate eigenvalues and, hence, leading to failure in the extraction process [55]. The cost function J_M utilized in (4.2.9) exploits the fact that for BSE, $\mathbf{R}_k\mathbf{p}$ should be collinear [94] with $\mathbf{q}$ incorporating the coefficients d_k which provides $\mathbf{q}$ with the proper scaling. The trivial answer for (4.2.9) is $\mathbf{d} = \mathbf{p} = \mathbf{q} = \mathbf{0}$. This solution has been avoided by imposing the condition $\|\mathbf{q}\|_2 = \|\mathbf{d}\|_2 = 1$. Successful minimization of (4.2.9) leads to the identification of $\mathbf{p}$, which extracts the source of interest (SoI) in (4.2.8).

The main advantage of using (4.2.9) for BSE rather than other conventional BSE methods which incorporate higher order statistics [85] is that it is indeed computationally simpler and more effective in extracting the nonstationary sources. However in BSE, it is not possible to tune the algorithm to extract the SoI as first. Therefore, for such an objective, some prior knowledge should be incorporated into the separation process. An auxiliary cost function is defined as below

$$J_{Aux} = \sum_{k=1}^{K} \|b_k\mathbf{q} - \mathbf{q}_{est}\|_2^2 \qquad (4.2.12)$$

where $\mathbf{b}$ is a column vector $\mathbf{b} = [b_1, b_2, \cdots, b_K]'$ and $\mathbf{q}_{est}$ is the prior spatial

information of the EB source, the estimation of q, provided by PARAFAC (Section 4.2.2).

By minimizing J_{Aux} coupled with (4.2.9) in a Lagrangian framework, $J_{tot} = J_M + \eta_q J_{Aux}$, the SoI is effectively extracted as the first estimated source. Moreover, as it will be shown in Section 4.3, incorporation of J_{Aux} in the J_{tot} results in faster minimization. The new cost function is therefore

$$[\hat{b}, \hat{d}, \hat{p}, \hat{q}] = \arg \min_{b,d,p,q} \sum_{k=1}^{K} \left(\|R_k p - d_k q\|_2^2 + \eta_q \|b_k q - q_{est}\|_2^2 \right) \quad (4.2.13)$$

where η_q is the Lagrange multiplier. In (4.2.13), the $b_k, k = 1, 2, \cdots, K$ values are free parameters to scale q and $\|b\|_2 = 1$.

Essentially, there are two approaches in using the spatial priors which vary the degree of freedom of the optimizing process (4.2.13). One can strictly minimize the difference between q and q_{est} regardless of the probable errors in estimation of q_{est}. On the other hand, it is possible to consider that q can deviate from the prior estimated vector q_{est} by an l_2 norm-bounded threshold. In in soft constraining, the estimation bias is often considered as

$$\delta = q - q_{est} \quad (4.2.14)$$

where

$$\|\delta\|_2 < \epsilon \quad (4.2.15)$$

and ϵ is a known positive constant. For the majority of spatially constrained BSS applications [67, 83] and references therein, the latter conservative approach is preferable even when q_{est} is accurately estimated. However, for EB artifact removal from EEGs strictly constrained the extracting algorithms are sufficient since sparsely occurring EBs are the dominant sources

superimposed on the ongoing EEGs. In this chapter, the former approach

has been explored and it has been assumed that the estimation of q_{est} by the

PARAFAC-based STF model is accurate enough.[1] It has been also experi-

mentally found that although the introduction of b in (4.2.13) does not have

any rotational effect in estimation of q, it results in further minimization of

J_{tot}. The interested reader is referred to Chapter 5 and [77, 79] in which a

conservative method for EB artifact removal from the EEGs is realized.

The solution to (4.2.13) is found by alternatively adjusting its param-

eters. The four unknown vectors are iteratively updated by an alternating

least squares (ALS) method until the convergence. Firstly, q, d, and b are

fixed and p is updated. Taking the gradient of J_{tot} with respect to p leads to

an optimal analytical solution for p as

$$\frac{\partial J_{tot}}{\partial \mathbf{p}} = 2 \sum_{k=1}^{K} \mathbf{R}_k (\mathbf{R}_k \mathbf{p} - d_k \mathbf{q}) = \mathbf{0} \qquad (4.2.16)$$

and

$$\mathbf{p} \Leftarrow \mathbf{Q} \left(\sum_{k=1}^{K} d_k \mathbf{R}_k \right) \mathbf{q}; \quad \mathbf{Q} = \left[\sum_{k=1}^{K} (\mathbf{R}_k)^2 \right]^{-1} \qquad (4.2.17)$$

where $a \Leftarrow b$ denotes replacing a by b. Thereafter, p, b, and q are fixed and

d is updated. As in [88], noticing $\|\mathbf{q}\|_2 = 1$, the gradient of J_{tot} with respect

to d_k becomes

$$\frac{\partial J_{tot}}{\partial d_k} = -2 \sum_{k=1}^{K} ((\mathbf{R}_k \mathbf{p})' - d_k \mathbf{q}') \mathbf{q} = 0, \quad k = 1, 2, \cdots, K. \qquad (4.2.18)$$

---

[1]The estimated spatial signatures of EB artifacts using the STF model of EEGs are convincingly similar to results of a previously published PhD thesis [95] on estimation of spatial signatures of EBs using BSS.

The update rule for $\mathbf{d}$ is as

$$\mathbf{d} \Leftarrow \frac{\mathbf{u}}{\|\mathbf{u}\|_2}; \quad \mathbf{u} = [\mathbf{r}_1'\mathbf{q}, \mathbf{r}_2'\mathbf{q}, \cdots, \mathbf{r}_k'\mathbf{q}]' \qquad (4.2.19)$$

where $\mathbf{r}_k = \mathbf{R}_k\mathbf{p}$.

Then, fixing $\mathbf{p}$, $\mathbf{d}$, and $\mathbf{b}$, the vector $\mathbf{q}$ is adjusted. Considering

$$\frac{\partial J_{tot}}{\partial \mathbf{q}} = -2\sum_{k=1}^{K} d_k\mathbf{r}_k - 2\eta_{\mathrm{q}}\sum_{k=1}^{K} b_k\mathbf{q}_{est} + 2(1 + \eta)\mathbf{q} = \mathbf{0}, \qquad (4.2.20)$$

the vector $\mathbf{q}$ is adjustable by

$$\mathbf{q} \Leftarrow \frac{\mathbf{v}}{\|\mathbf{v}\|_2}; \quad \mathbf{v} = \sum_{k=1}^{K}(d_k\mathbf{r}_k + \frac{1}{K}\eta_{\mathrm{q}}b_k\mathbf{q}_{est}). \qquad (4.2.21)$$

For updating $\mathbf{b}$, the rest of the variables are fixed first, i.e., $\mathbf{q}, \mathbf{p}$, and $\mathbf{d}$ and procedure is carried on by minimizing (4.2.13) with respect to b_k as

$$\frac{\partial J_{tot}}{\partial b_k} = 2\eta_{\mathrm{q}}\sum_{k=1}^{K}(b_k - \mathbf{q}_{est}'\mathbf{q}) = 0. \qquad (4.2.22)$$

Thus, $\mathbf{b}$ is updated as

$$\mathbf{b} \Leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|_2}; \quad \mathbf{w} = [\mathbf{q}_{est}'\mathbf{q}, \mathbf{q}_{est}'\mathbf{q}, \cdots, \mathbf{q}_{est}'\mathbf{q}] \qquad (4.2.23)$$

where $\mathbf{b}$ is retained as a vector instead of a scalar to present a consistent formulation.

Finally, in order to solve (4.2.13) for the Lagrange multiplier η_{q}, the vector $\mathbf{e}_i$ is defined as a vector whose elements are all zero except for the i-th component which is one, as

$$\mathbf{e}_i = [0, \cdots, 0, 1, 0, \cdots, 0]', \quad \forall \, i \in \{1, 2, \cdots, K\}. \qquad (4.2.24)$$

Considering that

$$\mathbf{v} = \sum_{k=1}^{K}(d_k\mathbf{r}_k + \frac{1}{K}\eta_q b_k \mathbf{q}_{est}) \qquad (4.2.25)$$

in (4.2.21), η_q can be easily solved updated with respect to the newly estimated $\mathbf{v}$ after each iteration. Therefore, a new value for η_q is assigned as

$$\eta_q = \frac{[\frac{1}{b_i}(\mathbf{v} - \sum_{k=1}^{K} d_k\mathbf{r}_k)]'\mathbf{e}_i}{\mathbf{q}'_{est}\mathbf{e}_i}. \qquad (4.2.26)$$

There are several issues about (4.2.13) worth attention; first is the convergence. In this work, through extensive number of independent runs, it has been observed that (4.2.13) almost always converges. Nevertheless, proving the convexity of (4.2.13) is an interesting study. The second issue regarding (4.2.13) is that minimization of (4.2.13) is computationally simple and effective for extraction of nonstationary sources. Third, during the course of extraction, it is possible to extract the EB source as the first extracted source. The proposed SBSE algorithm decreases the processing time in the real-time applications comparing to the BSS methods where all sources should be extracted first and then the source of interest is identified. And finally, similar to many other CBSS algorithms [67,67], when a low resolution estimate of the spatial signature of the SoI is introduced to (4.2.13), the optimization procedure would not result in the SoI. Note that Section 4.4 explores the robustness of the SBSE method when the prior knowledge is biased.

The performance of the proposed SBSE procedure has been evaluated through a comparison with the SCBSS algorithm proposed in [67,83] for 1000 sets of synthetically mixed analytic sources. Four signal sources, including two sinusoids of 10 Hz (s_1) and 12 Hz (s_2) representing the brain rhythmic waves, a white Gaussian distributed signal as the background brain
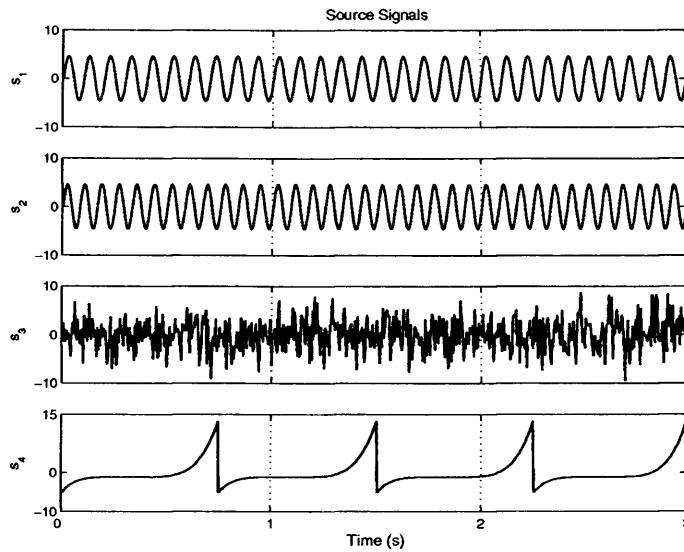
Figure 4.1. Simplified scalp EEG measurements; four synthetic sources, namely, s_1 and s_2 represent brain rhythmic activities, s_3 for background white noise and s_4 the EB artifact source.

activity (s_3), and a spiky source standing for the EB artifact (s_4) have been mixed. The source signals have been selected as such in order to cover the range of sub-Gaussianity to super-Gaussianity. The source waveforms are depicted in Fig. 4.1.

One thousand 4×4 mixing matrices, i.e. **A**, have been drawn from a zero mean and unit variance normal distribution and used for synthesizing artifact s_4 contaminated mixtures. For visualization purposes, only one example is presented where the mixing matrix is as below

$$\mathbf{A} = \begin{bmatrix} 0.6046 & 0.5352 & 0.2296 & 0.1808 \\ -0.5278 & 0.4285 & -0.1983 & -0.6817 \\ -0.2069 & -0.2991 & 0.2856 & -0.3171 \\ -0.5596 & -0.6637 & -0.9091 & 0.6340 \end{bmatrix}.$$

The original artifact contaminated mixtures are plotted in Fig. 4.2 and Fig. 4.3 in solid black. Mixtures x_3 and x_4 are highly affected by the spiky
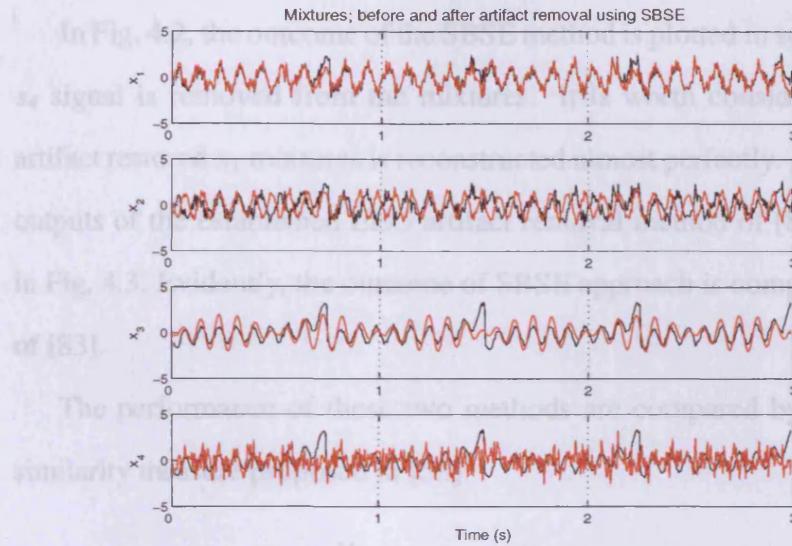
Figure 4.2. Four mixtures of four synthetic sources; the mixed signals in solid black; x_3 and x_4 are highly contaminated by the EB source s_4. The artifact removed mixtures have been also plotted in solid red by using the SBSE.
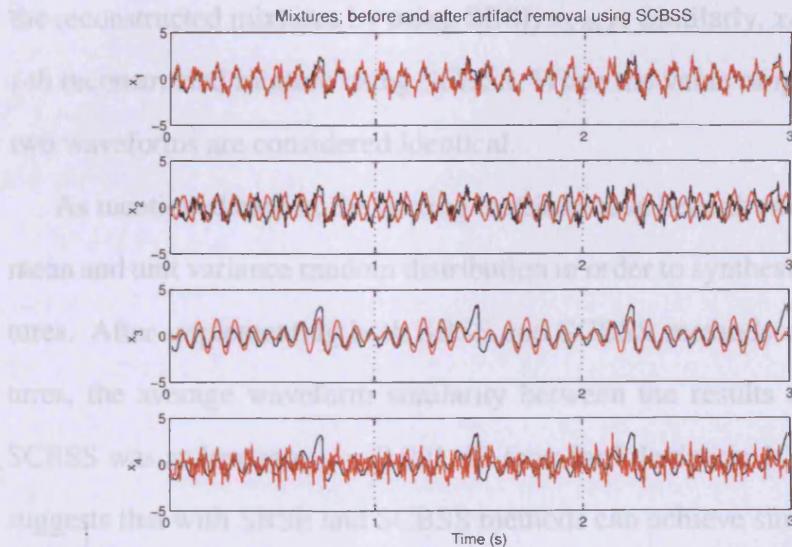
Figure 4.3. Four mixtures of four synthetic sources; the mixed signals in solid black; x_3 and x_4 are highly contaminated by the EB source s_4. The artifact removed mixtures have been also plotted in solid red by using the SCBSS.

source s_4. Here, the objective is to illustrate the results of the proposed

method (Fig. 4.2) with that of [83] (Fig. 4.3) in which SCBSS method based

on FastICA [85] is suggested for EB artifact removal.

In Fig. 4.2, the outcome of the SBSE method is plotted in solid red where s_4 signal is removed from the mixtures. It is worth considering that the artifact restored x_1 mixtures is reconstructed almost perfectly. Moreover, the outputs of the established EEG artifact removal method of [83] are shown in Fig. 4.3. Evidently, the outcome of SBSE approach is comparable to that of [83].

The performance of these two methods are compared by a waveform similarity measure proposed in [55]:

$$\eta_{dB} = 10 \log \left[\frac{1}{M} \sum_{i=1}^{M} \left(1 - E\{\mathbf{x}_{SBSE}(i) - \mathbf{x}_{SCBSS}(i)\} \right) \right]. \qquad (4.2.27)$$

where $M = 4$ is the number if mixtures and $\mathbf{x}_{SBSE}(i)$ is the i-th channel of the reconstructed mixtures by using SBSE $\mathbf{x}_{SBSE}$. Similarly, $\mathbf{x}_{SCBSS}(i)$ is the i-th reconstructed mixture using SCBSS. When the value of η_{dB} is zero, the two waveforms are considered identical.

As mentioned earlier, the mixing matrices have been drawn from a zero mean and unit variance random distribution in order to synthesize 1000 mixtures. After implementing both SBSE and SCBSS methods on these mixtures, the average waveform similarity between the results of SBSE and SCBSS was as low as $\eta_{dB} = 0.001$ dB (standard deviation 10^{-5} dB) which suggests that with SBSE and SCBSS methods can achieve similar results.

As a second measure, the averaged correlation coefficients (CC) [31] between the reconstructed signal signal using both methods for different mixtures were computed. The CC between two discrete random variables x and y over a fixed interval is defined as [31]

$$CC = \frac{|\sum_{i=1}^{w} x(i)y(i)|}{\sqrt{\sum_{j=1}^{w} x^2(j)} \sqrt{\sum_{j=1}^{w} y^2(j)}} \qquad (4.2.28)$$
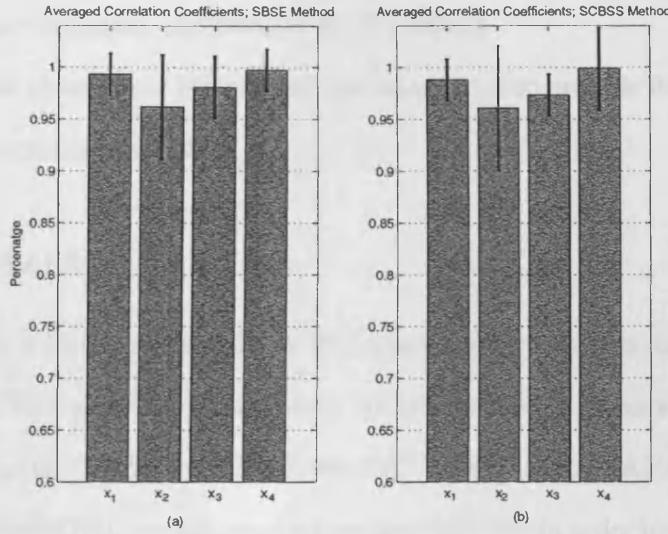
Figure 4.4. The averaged *CC* values between the segments of cleaned mixtures (after removing s_4) and the original mixtures by using SBSE in (a) and SCBSS in (b); the *CC* values of about unity justify that the SBSE method provides similar results as to SCBSS.

where w is the number of time samples. Fig. 4.4 demonstrates the averaged *CC* values between segments of the cleaned mixtures (after removing s_4) and the original mixtures by using the proposed method and that of [67,83]. *CC* values of about unity show that the SBSE method provides similar results as to SCBSS. In the above simulations, it has been assumed that the spatial distribution (signature) of the EB-type source s_4 is estimated in advance. This estimation can be obtained by using PARAFAC as in this dissertation or implementing the BSS method for short segments of mixtures as in [67,83]. This assumption helps to validate the results of the SBSE method comparing to [67, 83] regardless of how accurate various existing methods perform in estimating the spatial vector corresponding to EB artifact.

Moreover, through simulation studies it has been found that the SoI may be identified much faster by using SBSE method rather than the BSE method proposed in [88]. Faster performance elaborates that the incorporation of the auxiliary cost function J_{Aux} into the extraction process significantly up-

grades the performance, see Section 4.3 for details.

Next, it is shown how PARAFAC can be utilized to provide the required *a priori* information for SBSE.

4.2.2 PARAFAC

PARAFAC is a widely accepted tool in extracting disjoint multi-dimensional phenomena with growing applications in food sciences, communications, and biomedicine [51, 59, 76, 77, 89, 96–99]. By exploiting PARAFAC, the EB contaminated EEG measurements are decomposed in order to extract the factor relevant to the EB artifact to be used in SBSE. The resulting spatial signature of the EB-related factor, q_{est}, is exploited in (4.2.13). The spatial signature of this factor corresponds to the level of EB contamination in each electrode and is thereby comparable to the column of the mixing matrix that propagates the point source EB artifact onto the EEG channels. Physiologically, this assumption is rational since EB is attenuated while propagating from the frontal area to the central and occipital areas of the brain.

In this approach, the multi-channel EEG recordings are transformed into the time-frequency domain. This transformation gives the two-way EEG recordings, the matrix of space(channel)-time, an extra dimension, i.e. frequency. and yields a three-way array. In other words, for I EEG channels, the energies of the time-frequency transform for J time instants and K frequency bins are computed. By stacking these I matrices (of size $J \times K$) and adopting the MATLAB matrix notation, the three-way array $\underline{X}^{I \times J \times K} \equiv \underline{X}(1 : I, 1 : J, 1 : K)$ is set up and introduced to PARAFAC.

The PARAFAC is exploited here to decompose the EEG recordings and provide an STF model. As stated in Chapter 3, the key idea is in considering the EEGs as superposition of the electro-potentials of the neurons measured

by positioning the electrodes on the scalp. EEGs can be represented by the linear models which are defined in three domains of space, time, and frequency in order to investigate their spatial, temporal, and spectral dynamics [31,59,76–79,89,98].[2]

Complex Wavelet Transform: To setup a three-way array, the continuous wavelet transform is utilized to provide a time-varying representation of the energy of the signal over all channels, as in Chapter 3. The complex Morlet's wavelets $w(t, f_0)$, with $\sigma_f = 1/(2\pi\sigma_t)$, and $A = (\sigma_t \sqrt{\pi})^{-1/2}$, is used here in which the trade-off ratio (f_0/σ_f) is 7, to create a wavelet family. The time-varying energy $E(t, f_0)$ of a signal at a specific frequency band is the squared norm of the convolution of a complex wavelet of the signal $x(t)$, i.e., $E(t, f_0) = |w(t, f_0) * x(t)|^2$.

Here, the mathematical formulas of Chapter 3 which have been used in this chapter are briefly reviewed. The ordinary factor analysis is expressed as

$$\mathbf{X}^{I \times \mathcal{J}} = \mathbf{U}^{I \times F}(\mathbf{S}^{\mathcal{J} \times F})' + \mathbf{E}^{I \times \mathcal{J}} \qquad (4.2.29)$$

where $\mathbf{U}$ is the factor loading, $\mathbf{S}$ is the factor score, $\mathbf{E}$ is the error, and F is the number of factors. Similarly, PARAFAC for the three-way arrays, $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$, is presented as

$$\mathbf{X}^{I \times \mathcal{J}\mathcal{K}} = \mathbf{U}^{I \times F}(\mathbf{S}^{\mathcal{K} \times F} \odot \mathbf{D}^{\mathcal{J} \times F})' + \mathbf{E}^{I \times \mathcal{J}\mathcal{K}} \qquad (4.2.30)$$

where $\mathbf{D}$ is the factor score corresponding to the second modality. ALS is the most common way to estimate the PARAFAC model [51]. In order to decompose the multi-way array to parallel factors the below cost function is

---

[2]The interested reader is referred to [96,97,100] for further mathematical details of the PARAFAC model, the uniqueness conditions, and its robust iterative fitting which are out of the scope of this thesis.

usually minimized [51]

$$[\hat{\mathbf{U}}, \hat{\mathbf{S}}, \hat{\mathbf{D}}] = \arg\min_{\mathbf{U},\mathbf{S},\mathbf{D}} \|\mathbf{X}^{I \times \mathcal{J}\mathcal{K}} - \mathbf{U}^{I \times F}(\mathbf{S}^{\mathcal{K} \times F} \odot \mathbf{D}^{\mathcal{J} \times F})'\|_2^2. \qquad (4.2.31)$$

Here, $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$ is the three-way array of wavelet energy of multi-channel EEG recordings and $\mathbf{U}^{I \times F}, \mathbf{D}^{\mathcal{J} \times F}$, and $\mathbf{S}^{\mathcal{K} \times F}$ denote respectively the spatial, temporal, and spectral signatures of $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$. The trilinear alternating least squares (TALS) method [101] is used to compute the parameters of the STF model.

4.2.3   The Deflation Method

In order to achieve EB-free EEG recordings, $\mathbf{x}_{filt}(t)$, after the extraction of the EB source $y(t)$ using (4.2.8), the deflation procedure is applied which eliminates the previously extracted signal, $y(t)$, from the recording mixtures $\mathbf{x}(t)$ as

$$\mathbf{x}_{filt}(t) = \mathbf{x}(t) - \widetilde{\mathbf{p}}y(t) \qquad (4.2.32)$$

where, as in [66, Section 5.2.5], $\widetilde{\mathbf{p}}$ can be estimated either adaptively or simply after minimization of the mean square cost function J with respect to $\widetilde{\mathbf{p}}$

$$J(\widetilde{\mathbf{p}}) = E\{\mathbf{x}_{filt}(t)'\mathbf{x}_{filt}(t)\}$$

$$= E\{\mathbf{x}(t)'\mathbf{x}(t)\} - 2\widetilde{\mathbf{p}}'E\{\mathbf{x}(t)y(t)\} + \widetilde{\mathbf{p}}'\widetilde{\mathbf{p}}E\{y^2(t)\}. \qquad (4.2.33)$$

Eq. (4.2.33) results in an efficient batch one-step formula to estimate $\widetilde{\mathbf{p}}$ as

$$\widetilde{\mathbf{p}} = \frac{E\{\mathbf{x}(t)y'(t)\}}{E\{y(t)^2\}} = \frac{E\{\mathbf{x}(t)\mathbf{x}'(t)\}\mathbf{p}}{E\{y(t)^2\}} \qquad (4.2.34)$$

where **p** is computed by (4.2.17). In fact, $\widetilde{\mathbf{p}}$ is an estimation of $\mathbf{a}_j$, the j-th column of **A**, neglecting the arbitrary scaling and the column permutations ambiguities.

In summary, the proposed method consists of the following stages. Given an artifact contaminated EEG data,

1. bandpass filter the EEGs between 1 Hz and 40 Hz,

2. set up the three-way array, i.e. $\underline{\mathbf{X}}^{I \times \mathcal{J} \times \mathcal{K}}$, as stated in Section 4.2.2,

3. execute PARAFAC and select the EB artifact relevant factors as will be described in Section 4.3,

4. exploit the spatial signature of the EB artifact factor in the SBSE cost function (4.2.13),

5. reconstruct the artifact removed EEGs by deflation.

4.3 Results

The SBSE algorithm is applied to real EEG measurements. The database was provided by the School of Psychology, Cardiff University, UK, and contained a wide range of EB. The scalp EEGs were recorded using 25 Ag/AgCl electrodes positioned on the scalp. EEGs were recorded to provide a reference dataset specifically for the purpose of evaluating different artifact removal methods from one healthy subject and contained numerous EBs, eye-movements, and motion artifacts. The sampling rate was set to 200. In order to reduce the computational costs of the PARAFAC modeling, 16 channels were selected out of the above mentioned 25 channels as illustrated in Fig. 4.5.
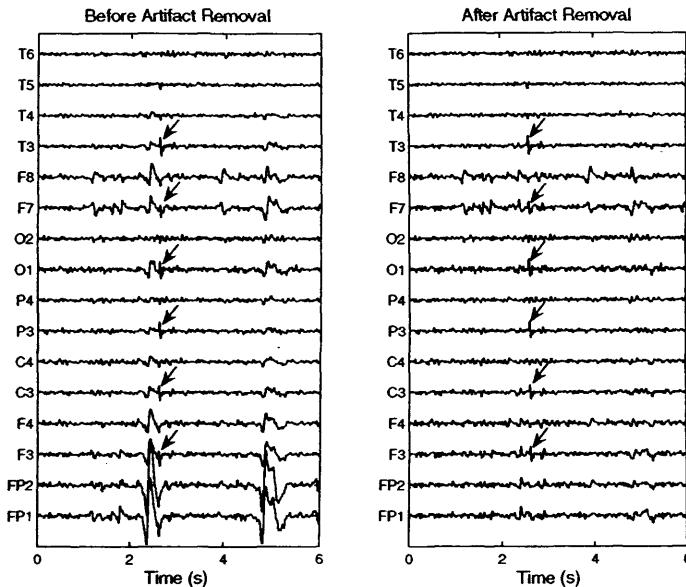
Figure 4.5. The result of the proposed EB artifact removal method for a sample of real EEG signals recorded from the selected 16 electrodes. In the left subplot, the EBs are prominent in the frontal electrodes. In the right subplot, artifact restored EEGs are illustrated. Note the small spike-type signals, indicated by arrows, are precisely retained after the artifact removal.

Each EEG segment was transformed into the time-frequency domain by means of the complex wavelet transform. A frequency band from 2 Hz to 25 Hz with resolution of 0.1 Hz was considered. This three-way array was then introduced to PARAFAC where the number of factors was selected as one or two, as will be highlighted in the following experiments, using the CORCONDIA [64]. PARAFAC was used to identify the most significant factors with CORCONDIA values greater than 85% [64]. Two sample results are demonstrated here in order to show the potential of the presented method.

4.3.1 Experiment 1

The left subplot of Fig. 4.5 shows the EEG measurements contaminated with two EBs at approximate times of two and half and five seconds. The effects of the EBs are evident in the frontal and central electrodes. Imple-
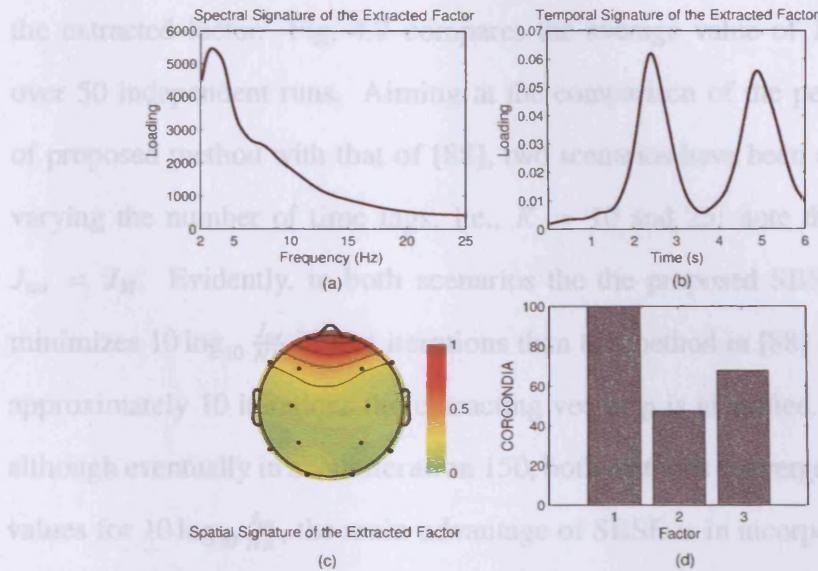
Figure 4.6. The extracted factor by using PARAFAC; (a) and (b) illustrate respectively the spectral and temporal signatures of the extracted factors and (c) represents the spatial distribution of the extracted factor which has been considered as the *a priori* knowledge during extraction procedure. (d) shows that the number of factors F suggested by CORCONDIA to be one since the bars corresponding to $F = 2$ and $F = 3$ are less than the threshold, i.e., 85%.

mentation of PARAFAC on this measurement results in the STF signatures depicted in Fig. 4.6-(a) to (c). Although, there are two EBs, the CORCON-DIA suggests the number of factors F to be one as in Fig. 4.6-(d). This value is rational since both of the EBs are originated from certain vicinity (frontal lobe of the brain) and occupy the same frequency band and there is no significant brain background activity. By using the spatial distribution of the extracted factor as the *a priori* information, the EB artifacts are effectively removed. Notice the resolution of the proposed algorithm; it does not affect the very low amplitude spike-type signals after the first EB, see Fig. 4.5, during the extraction process.

In order to minimize (4.2.13) the initial values of the vectors **b**, **d**, **p**, and **q** were independently drawn from standardized normal distributions $N(0, 1)$, η_q was initialized to 5, and q_{est} was set to the spatial signature of

the extracted factor. Fig. 4.7 compares the average value of $10 \log_{10} \frac{J_{tot}}{NK}$

over 50 independent runs. Aiming at the comparison of the performance

of proposed method with that of [88], two scenarios have been devised by

varying the number of time lags, i.e., $K = 10$ and 25; note that in [88]

$J_{tot} = J_M$. Evidently, in both scenarios the the proposed SBSE method

minimizes $10 \log_{10} \frac{J_{tot}}{NK}$ in less iterations than the method in [88] does; after

approximately 10 iterations the extracting vector $\mathbf{p}$ is identified. Note that

although eventually in about iteration 150, both methods converge to similar

values for $10 \log_{10} \frac{J_{tot}}{NK}$, the main advantage of SBSE is in incorporating the

prior knowledge. Therefore, it is guaranteed that $\mathbf{p}$ extracts the EB source as

the first extracted source. The EB is then removed from the multi-channel

EEG using the batch deflation algorithm in [66]. Note that since SCBSS

uses the FastICA it is not possible to compare its performance with that

of SBSE (or BSE) in terms of "cost function minimization" as in Fig. 4.7.

However in Section 4.3.3, SCBSS and SBSE methods will be compared in

how effectively they can remove the EB artifact. It will be carried out by

comparing the resultant averaged CC values between extracted EB artifact

and the EEGs traces before and after the removal procedure obtained from

each.

4.3.2 Experiment 2

Performance of the SBSE method with similar initial values for another set

of EEGs from the database is demonstrated in Fig. 4.8. In the left subplot,

the truncated 4 seconds of EEG recordings before and after the application

of the EB removal process are plotted. Fig. 4.8-(b) illustrates the averaged

correlation coefficients between the artifact removed channel signals and

the original contaminated ones with their corresponding standard deviations
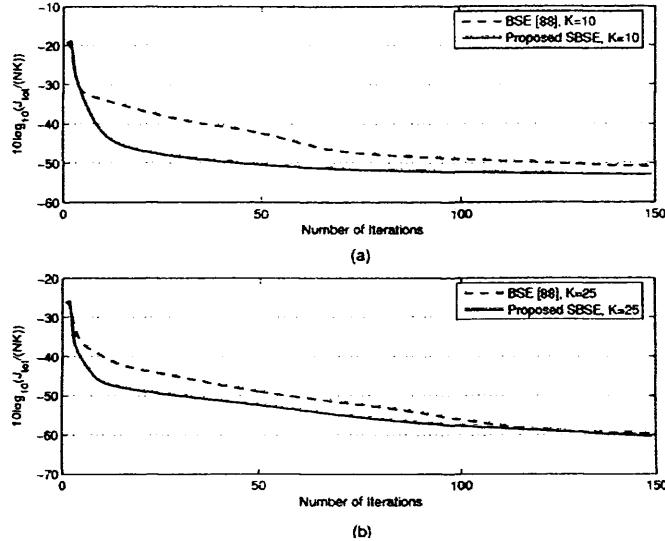
Figure 4.7. The averaged (over 50 independent runs) convergence characteristics, $10\log_{10}\frac{J_{tot}}{NK}$, of the SBSE and conventional BSE are depicted for two values of $K = 10$ in (a) and $K = 25$ in (b). In both subplots the solid and dashed curves correspond respectively to the proposed SBSE and BSE.

over 25 independent runs. As expected, the CC values corresponding to the signals recorded from the frontal electrodes are relatively low showing these signals are significantly restored. However, the values corresponding to other channel signals, i.e. parietal, central, temporal, and occipital, are almost unity demonstrating that the algorithm does not affect clean EEG measurements.

The STF model of this recording is identified using PARAFAC. In contrast to the previous experiments, the CORCONDIA suggests $F = 2$ since PARAFAC identified a significant brain background activity during occurrence of the EB. In this experiment, the CORCONDIA value for $F = 1$ was 100%, for $F = 2$ was 94.6%, and for $F = 3$ was 38%, thus $F = 2$ was selected. Fig. 4.9-(a) to (d) illustrate the estimated signatures of the 16-channel contaminated EEGs. Note that the first component (Factor 1) of the STF model demonstrates the EB-relevant factor since

(a)   (b)
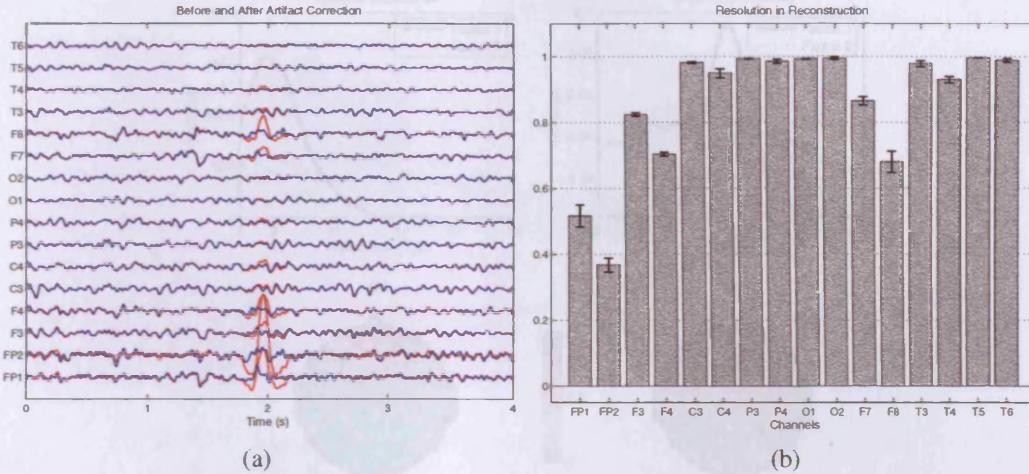
Figure 4.8. The results of the proposed EB artifact removal method for a set of real EEG signals recorded from 16 electrodes; (a) shows the EB contaminated EEGs in red and the artifact corrected EEGs in blue. In the right subplot, the averaged CC values between the artifact corrected channel signals and the original contaminated EEGs with their corresponding standard deviations over 25 independent runs are plotted. CC values corresponding to the frontal channel signals are relatively lower than the values corresponding to other channel signals which are almost unity.

1. It mainly occurs in the frequency band of around 5 Hz while the other factor exists in the entire band and represents the ongoing activity of the brain or perhaps a broadband white noise-like component, Fig. 4.9-(a).

2. The temporal signature of the first factor clearly shows a transient event such as EB while that of Factor 2 consistently exists throughout the course of recording, Fig. 4.9-(b).

3. Unlike Fig. 4.9-(d), in Fig. 4.9-(c), the spatial distribution of the extracted factor is confined to the frontal area which demonstrates the frontal origin of EBs. The other factor shows the background activity of the brain as it spreads all over the topographic map.
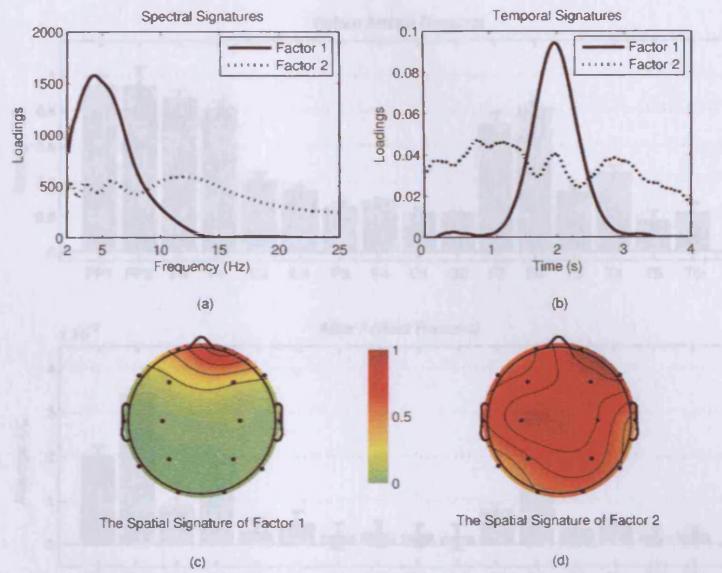
Figure 4.9. The extracted factors by using PARAFAC; (a) and (b) illustrate respectively the spectral and temporal signatures of the extracted factors; (c) and (d) present the spatial distribution of the factors, respectively. Evidently, Factor 1 demonstrates the EB phenomenon as it occurs in frequency band of around 5 Hz (a), it is indeed transient in the time domain (b) and it is confined to the frontal area.

4.3.3   Performance Evaluations

In order to provide a quantitative measure of performance for the proposed SBSE-based artifact removal method, the *CC* values between the extracted EB artifact source and the original and the artifact removed EEGs were computed and plotted in Fig. 4.10. For each of the 20 different artifact contaminated EEGs, the proposed SBSE algorithm was executed. The aforementioned *CC*s for each run were then computed between the extracted EB and the EEGs before and after the artifact removal. These values were subsequently averaged, Fig. 4.10. Furthermore, their corresponding standard deviations have also been reported. For comparison purposes, the same experiment was carried out and the SCBSS [67] was executed instead of the SBSE. The results are depicted in Fig. 4.11. As expected, the *CC* values are significantly decreased by using the proposed method. Note that the re-
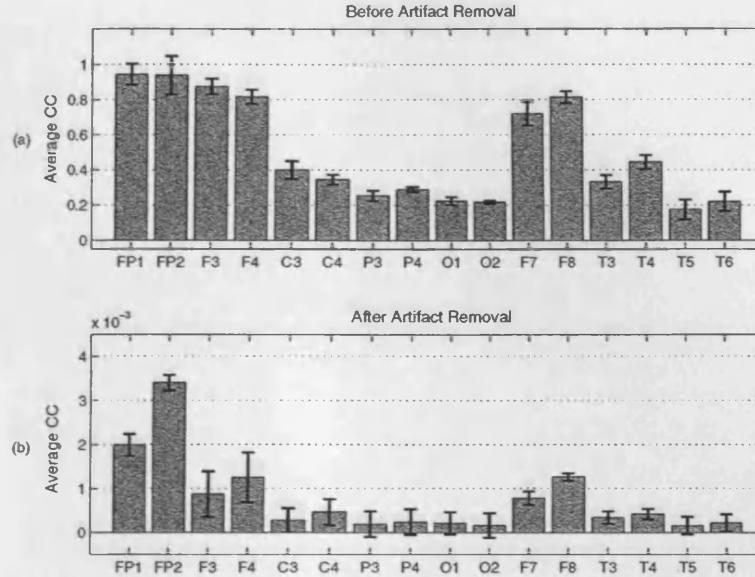
Figure 4.10. The averaged CC values (and their corresponding standard deviations) between the extracted EB and the restored EEGs before and after artifact removal using SBSE for different channels in (a) and (b), respectively. The experiments have been performed for 20 different EB contaminated EEG recordings. Note that the scales are different by a factor of 10^3.

sults of the SBSE and SCBSS methods are quite similar; the differences are not significant. Simulations for 20 EEG measurements demonstrate that the proposed method can efficiently identify and remove the EB artifact from the raw EEG measurements.

As a second criterion for measuring the performance of the overall system, a segment of EEG x_{seg} and a reconstructed EEG $\hat{x}_{seg}$ were selected. Notice that x_{seg} does not contain any artifact. The waveform similarity was computed as in (4.2.27) by

$$\eta_{dB} = 10 \log \left[\frac{1}{M} \sum_{i=1}^{M} \left(1 - E\{x_{seg}(i) - \hat{x}_{seg}(i)\} \right) \right]. \qquad (4.3.1)$$

Again, when the value of η_{dB} is zero, the original and reconstructed waveforms are identical. From the 20 sets of EEGs, the average waveform similarity was as low as $\eta_{dB} = 0.01$ dB (standard deviation 10^{-3} dB). These
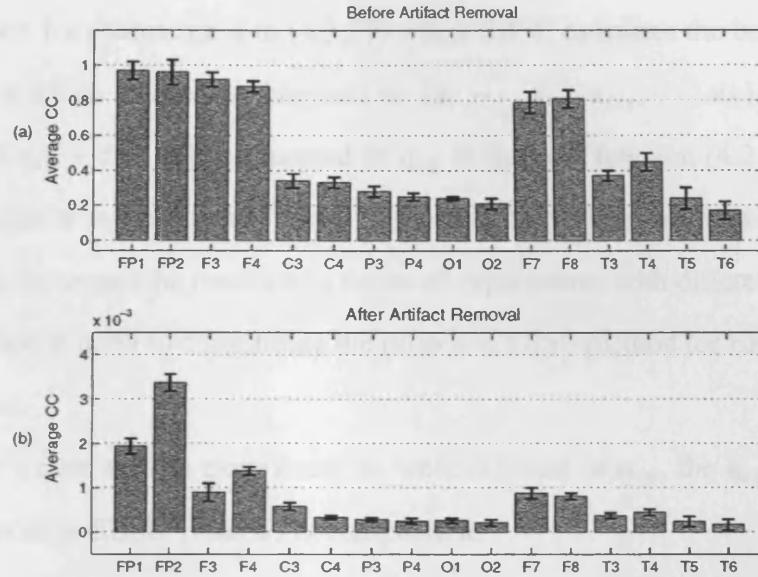
Figure 4.11. The averaged CC values (and their corresponding standard deviations) between the extracted EB and the restored EEGs before and after artifact removal using SCBSS for different channels in (a) and (b), respectively. The experiments have been performed for 20 different EB contaminated EEG recordings. Note that the scales are different by a factor of 10^3.

results suggest that the EEGs can be faithfully restored from EBs.

4.4 The influence of estimation bias on the SBSE performance

As indicated in (4.2.14) and (4.2.15), in soft constrained BSE (or BSS as in [83]) schemes, even if in the estimation of $\mathbf{q}_{est}$ is slightly biased, the optimization algorithm would take that into account and accommodate it during the extraction of the SoI. However, as indicated in Section 4.2.1, in this chapter a hard (in contrast to soft) approach has been followed where the algorithm strictly minimizes the cost function (4.2.13) regardless of the probable errors or biases in $\mathbf{q}_{est}$.

Interestingly, the scenario is not actually as restricted as it seems. That is, if there was a small deviation in the $\mathbf{q}_{est}$ from the actual $\mathbf{q}$, the SBSE accommodates it similar to [83]. The truth lies in the alternating least squares

approach for estimating q in (4.2.21) where SBSE estimates the best set of q and p which are both orthogonal to $\{a_1, \cdots, a_{j-1}, a_{j+1}, \cdots, a_N\}$. Therefore, if $q_{est} + \delta$ is utilized instead of q_{est} in the cost function (4.2.13), the optimization process would result in the originally estimated vector q, i.e. q_{est}. In the sequel the results of a series of experiments with different δs are presented in order to consolidate the proposed SBSE method for EB artifact removal.

Let's start with an experiment in which instead of q_{est}, the $q_{est} + \delta_1$ is introduced to SBSE. Here, δ_1 is computed as

$$\delta_1 = 0.1 \times r. \tag{4.4.1}$$

In (4.4.1), r is a vector of 16 elements drawn form a zero-mean and unit-variance normal distribution. Using (4.4.1), $\|\delta_1\|_2$ is likely to be less than 0.6. Therefore, the SBSE can compensate for the deviation of q_{est} from q and extracts the EB artifact if $\|\delta_1\| < 0.6$ potentially. An example has been provided in Figs. 4.12 and 4.13 where $\|\delta_1\| = 0.503$. In Fig. 4.12-(a), q_{est} obtained by PARAFAC is depicted which should be used in (4.2.13). Fig. 4.12-(b) shows the perturbed q_{est} by δ_1 which has been replaced in (4.2.13) instead of q_{est} and introduced to SBSE. Finally, in Fig. 4.12-(c), the resulting q after the alternative least squares optimization has been illustrated. Indeed, Fig. 4.12-(c) is quite similar to Fig. 4.12-(a).

The result of the artifact removal is depicted in Fig. 4.13. EEG traces in red are the original artifact contaminated recordings. Traces in blue are the resulting restored EEGs using the original estimate on q, i.e. q_{est}. EEG plots in black are the resulting artifact restored EEGs by using the artificially perturbed q_{est}, i.e. $q_{est} + \delta_1$.

(a)   (b)   (c)
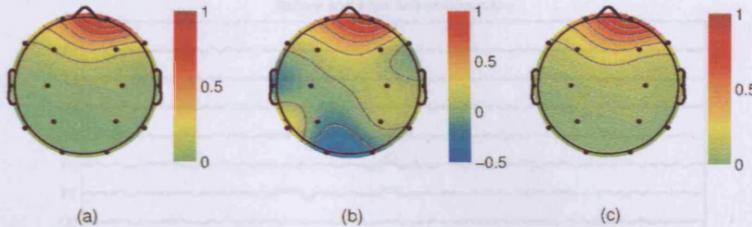
Figure 4.12. In (a), $\mathbf{q}_{est}$ is depicted. (b) shows the deviated $\mathbf{q}_{est}$ by δ_1 which has been put in (4.2.13) instead of $\mathbf{q}_{est}$. (c) illustrates the resulting $\mathbf{q}$ after ALS optimization procedure.

Then, instead of $\mathbf{q}_{est}$, $\mathbf{q}_{est}+\delta_2$ was introduced to SBSE. The vector δ_2 was computed in the same way as δ_1 by keeping the coefficient as 0.1 in (4.4.1), norm $\|\delta_1\| = 0.430$. Since $\mathbf{q}_{est} + \delta_2$, depicted in Fig. 4.14-(b), is significantly different in steering direction from Fig. 4.14-(a), SBSE may not compensate the deviation δ_2. In Fig. 4.14-(a), $\mathbf{q}_{est}$ resulted by PARAFAC is depicted. Fig. 4.14-(b) shows the perturbed $\mathbf{q}_{est}$ by δ_2 which has been replaced in (4.2.13) instead of $\mathbf{q}_{est}$ and introduced to SBSE. Finally, in Fig. 4.14-(c), the resulting $\mathbf{q}$ after the ALS optimization has been illustrated. The vector plotted in Fig. 4.14-(c) did not converge to the vector plotted in Fig. 4.14-(a).

The result of the artifact removal is depicted in Fig. 4.15. Again as Fig. 4.13, the EEG traces in red are the original artifact contaminated recordings. Traces in blue are the restored EEGs using the original estimation of $\mathbf{q}$, i.e $\mathbf{q}_{est}$. However, EEG plots in black do not acceptable performance in artifact removal procedure when $\mathbf{q}_{est} + \delta_2$ is used.

It can be concluded that the SBSE presents a robust performance when $\mathbf{q}_{est}$ is perturbed by a norm bounded small deviation. That is, the direction of the vector $\mathbf{q}_{est}$ should not be changed significantly. Therefore, the bias should be fairly distributed over the elements of $\mathbf{q}_{est}$. Since a normalized version $\mathbf{q}_{est}$ is used in the formulations, it is unlikely that SBSE does not compensate it.
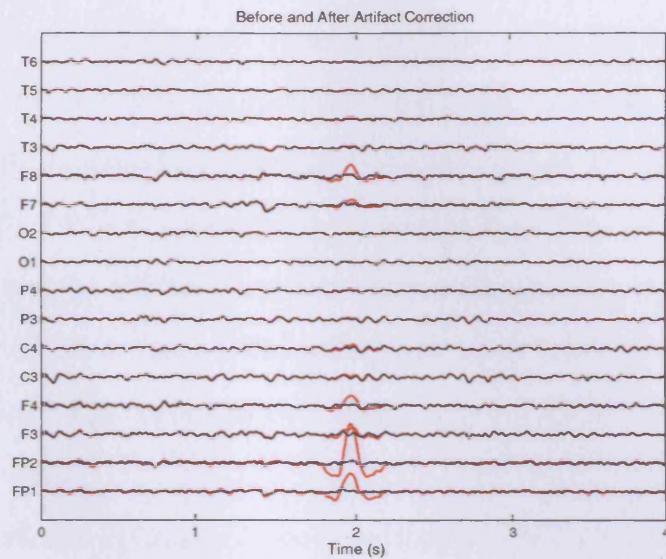
Figure 4.13. The result of the artifact removal from EEGs depicted in Fig. 4.8(a). EEG traces plotted in red color are the original artifact contaminated signals. EEGs in blue color are the resulting artifact removed signals using $\mathbf{q}_{est}$. Traces in black are the resulting artifact restored EEGs by using $\mathbf{q}_{est} + \boldsymbol{\delta}_1$ instead of $\mathbf{q}_{est}$.
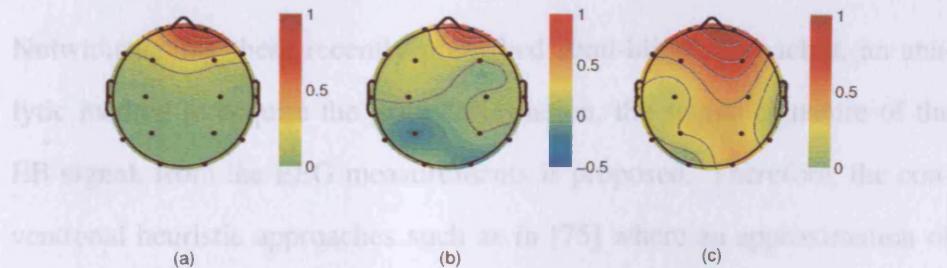
Figure 4.14. In (a), $\mathbf{q}_{est}$ is depicted. (b) shows the deviated $\mathbf{q}_{est}$ by $\boldsymbol{\delta}_2$ which has been put in (4.2.13) instead of $\mathbf{q}_{est}$. (c) illustrates the resulting $\mathbf{q}$ after ALS optimization procedure.

4.5   Concluding Remarks

It is generally accepted that the EB artifact can be removed from EEGs by using the BSS- and regression based methods for respectively multi-channel EEGs data with (or without) the reference EOG electrodes. However, nowadays this challenging topic is often solved by a semi-blind method rather than in a totally blind signal processing framework [55,67,75–79,83].
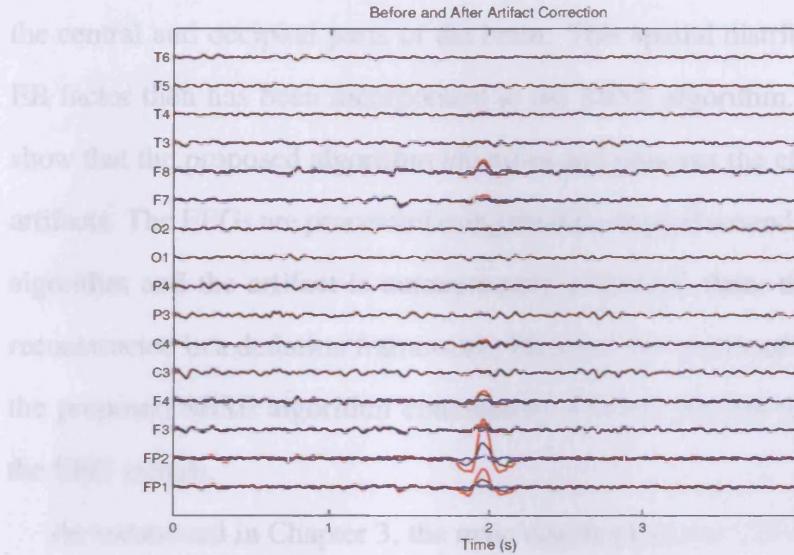
Figure 4.15. The result of the artifact removal from EEGs depicted in Fig. 4.8(a). EEG traces plotted in red color are the original artifact contaminated signals. EEGs in blue color are the resulting artifact removed signals using q_{est}. Traces in black are the resulting of the unsuccessful artifact removal procedure by using $q_{est} + \delta_2$ instead of q_{est}.

Notwithstanding these recently published semi-blind approaches, an analytic method to acquire the prior information, the spatial signature of the EB signal, from the EEG measurements is proposed. Therefore, the conventional heuristic approaches such as in [75] where an approximation of the temporal structure of the EB source signal is included in ICA is not followed here. The presented method is computationally simpler than the SCBSS method in [67, 83] since there is no need to estimate all the columns of the mixing matrix **A** in (4.2.4).

The vector of spatial distribution of the EB factor has been identified using PARAFAC. The vector of spatial signature of the EB factor resulted by the STF modeling of EEGs is utilized as an estimation of the column vector of the mixing matrix **A** that projects the EB source onto the EEGs. This assumption is rational since the EB can be considered as a strong point source which is merely attenuated while propagating from frontal area to

the central and occipital parts of the brain. This spatial distribution of the EB factor then has been incorporated to the SBSE algorithm. The results show that the proposed algorithm identifies and removes the effect of blink artifacts. The EEGs are processed using the time-lagged second-order SBSE algorithm and the artifact is autonomously extracted; then, the EEGs are reconstructed in a deflation framework. Based on the results of this method, the proposed SBSE algorithm consistently removes the EB artifacts from the EEG signals.

As mentioned in Chapter 3, the main drawback of the STF-based methods is their computational complexity. However, due to the dominance of the EB artifact over the background EEG activities the selection of the EB relevant factor could be quite easy. Therefore, the next chapter is devoted to a robust while computationally simple EB artifact removal from EEGs where an approximation to the original STF model is first computed and then the estimated spatial signature of the EB relevant factor is introduced to the signal extraction stage.

Chapter 5

ROBUST MINIMUM VARIANCE BEAMFORMING AND PARAFAC

In this chapter, an approach for the removal of EB artifacts from EEGs based on a novel STF model and the robust minimum variance beamforming (RMVB) [100] is proposed. In this method, the beamformer is provided with an estimation of the steering vector corresponding to the *point source EB artifact*. The artifact-removed EEGs are subsequently reconstructed by deflation. The vector corresponding to the spatial distribution of the EB factor is identified using the STF decomposition of EEGs. In order to reduce the computational complexity present in the estimation of the STF model using the three-way PARAFAC, the time domain is subdivided into a number of segments; a four-way array is then set to estimate the space-time-frequency-time/segment (STF-TS) model using a four-way PARAFAC. The correct number of the factors is estimated by using a novel CORCONDIA-based measure. Subsequently, the STF-TS model is shown to approximate closely the classic STF model with significantly lower computational requirements. The results confirm that the proposed algorithm effectively identifies and removes the EB artifacts from EEGs.

5.1   Introduction

Spatial filtering or simply "beamforming" have been widely used in com-
munications and radar signal processing applications [102]. Besides, in re-
cent years, beamforming methods have also been widely utilized in and
customized for brain signal processing, e.g. MUSIC, RAP-MUSIC, and
FINE methods [103]. Genuinely, the source (dipole) localization has been
the main application of beamforming in EEG analysis [103–106] where one
takes the advantage of high-dimensional EEG recordings and designs the
beamformers so that they pass brain electrical activities originating from
a specific location while attenuating other activities emanating from other
locations. Note that *preferably* these interfering sources should not be spa-
tially or temporally correlated with the SoI.

Theoretically, the variance (energy) of the filter output is equivalent
to the strength of the electrical signal coming from the location of inter-
est. Beamforming has also been very recently utilized in extraction and
localization of the spatially confined sources of interest [107, 108]; for in-
stance in localization of the ERPs.[1] However, to the best of my knowl-
edge, beamforming-based methods have not been specifically considered
in extraction and removal of the EB artifacts from the EEG recordings.[2]
This is understandable since these schemes suffer a significant performance
degradation when the array response vector for the SoI (EB in this case) is
not exactly known [110–113]. The problem arises when the methods used
in [110–114] deal with the electromagnetic waves of known propagation
pattern arriving at mostly spatially linear uniform (rarely non-uniform or
sparse [113]) arrays of receivers. However, in EEG analysis, although the

---

[1]The so-called transformation can be either a linear [102] or non-linear [109] combina-
tion EEGs recorded from multiple spatially distributed electrode signals.

[2]Although in some cases such as [107], the ocular artifact removal is a by-product.

10/20 electrode positioning standard is usually followed, the electrodes are positioned on the subject's scalp manually. This causes major uncertainties about the accuracy of the electrode locations. Therefore, one is always confronted with an *ad-hoc* configuration of electrodes which affects the steering vectors of propagating brain sources. In order to overcome these uncertainties, a method for solving the forward problem has been introduced in [115] by which in [104], the localization of the brain electrical sources has been solved. In [104], the steering vector corresponding to each grid point within the brain toward the scalp electrodes is found and then linearly constrained minimum variance beamformers (LCMVBs) are solved for these grid points to localize the electrical sources. This approach is promising, however it suffers from the complex computations occurred while solving the forward problem [115] for a fine grid.

In this regard, the contribution of this chapter is in the estimation of the steering vector corresponding to the EB artifact regardless of the conventional forward solutions to EEGs. Since, the temporally sparse occurring EB is the dominant source in the ongoing EEGs, this estimation is trustworthy and could be utilized in the beamforming procedure to remove the EB effect from the EEGs. The beamforming approach can identify and extract the EB artifact due to its independence from the EEGs [78]. The presented method is based on the RMVB [114], where the spatial *a priori* knowledge of the mixing process obtained by PARAFAC[3] is exploited as an estimation of the steering vector corresponding to the EB source.

The major advantage of the proposed method is that unlike the respective regression- and BSS-based methods presented in [81] and [55], it needs neither the reference EOG channel recordings nor any objective criterion

---

[3]The interested reader is referred to [96,97,100] for further mathematical details of the PARAFAC model, the uniqueness conditions, and its robust iterative fitting.

for distinguishing between EB and spurious peaks in the ongoing EEGs. The computational complexity in the estimation of the STF model using PARAFAC is reduced by sub-dividing the time domain into a number of segments. A four-way array is then set up to estimate the STF-TS model of the data. Subsequently, the STF-TS model results in the classic STF model.

It is also interesting to notice that in this approach there is no need to separate the dataset into training and testing subsets to tune the parameters. As long as, by using any primitive method, it is ensured that an EB artifact has happened, the presented method can be utilized to remove it.

There are two major differences between the approach followed in this chapter and what has been proposed in the previous chapter. Firstly, assuming the estimation of the steering vector corresponding to the EB artifact is precise, in Chapter 4, this vector has been used in a SBSE. Moreover, here, the steering vector corresponding to the EB source is not estimated by using the ordinary STF model. In contrast, as the second contribution of this chapter, by introducing the STF-TS model, the computational complexity occurred while estimating the STF model is significantly reduced. Note that during the estimation of the STF-TS model, there is trade off between the computational requirements and the proper unbiased estimation of the aforementioned steering vector. The bias is compensated by using the RMVB.

This chapter is organized as follows. In Section 5.2, the minimum variance beamforming used for EEG/MEG source localization and the robust version of that method are briefly reviewed. Then, the spatial signature of the STF-TS model is introduced as an estimation of the array response vector corresponding to the EB artifact. Afterward, the proposed STF-TS based STF model estimation methodology is described. The results are sub-

sequently reported in Section 5.3, followed by concluding remarks and discussions in Section 5.4.

5.2 Algorithm Development

Suppose N zero-mean real and mutually uncorrelated *geometrically stationary* sources $\mathbf{s}(t) = [s_1(t), s_2(t), \cdots, s_N(t)]'$, where $[\cdot]'$ denotes the vector transpose and t the discrete time index, are mixed by an $N \times N$ full column rank matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_N]$ where $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$. The vector of time mixture samples $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_N(t)]'$ is given as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \tag{5.2.1}$$

where $\mathbf{v}(t) = [v_1(t), v_2(t), \cdots, v_N(t)]'$ is the additive white Gaussian zero-mean noise which is assumed to be spatially uncorrelated with the sensor data and temporally uncorrelated. The sources are presumed to be uncorrelated. Therefore, the time lagged symmetrized autocorrelation matrix $\mathbf{R}_{xx}^k$ can be calculated as

$$\mathbf{R}_{xx}^k = E[\mathbf{x}(t)\mathbf{x}'(t - \tau_k)] = \sum_{i=1}^{N} r_i(\tau_k)\mathbf{a}_i\mathbf{a}_i' \tag{5.2.2}$$

for $k = 1, 2, \cdots, K$, where K is the maximum number of time lags, τ_K, and $E[\cdot]$ denotes the statistical expectation operator. In (5.2.2), $r_i(\tau_k) = E[s_i(t)s_i(t - \tau_k)]$ is the time lagged autocorrelation value of $s_i(t)$. The vector $\mathbf{x}(t)$ in (5.2.1) is a linear combination of the columns of the mixing matrix and weighted by the associated source and contaminated by the noise $\mathbf{v}(t)$.

5.2.1 Linearly Constrained Minimum Variance Beamformer

The LCMVB procedure has been effectively exploited for EEG/MEG source localization within the brain [104]. Fundamentally, the LCMVB method is based on electrical brain source and head electromagnetic and also geometric modeling in which the underlying neural activity and the distribution of potentials (electric, magnetic, or both) measured at the surface are related to each other. Here, the basic aspects of the minimum variance based EEG processing and source localization methods [104] are briefly reviewed.

The assumption that the brain sources can be modeled by current dipoles is the key concept for relating the surface measurements to the underlying brain activities. Although there are some recent techniques that deal with distributed (linear [116] or planar [117]) source modeling, it is widely accepted that the dipole source modeling reasonably satisfies the necessities in EEG/MEG processing. The relationship between dipole models and the surface recordings is obtained as follows.

Suppose x is an $N \times 1$ vector of the recorded potentials over N electrodes at a given time instant, presumingly, associated with a single dipole source. If the location of this source is represented by a 3×1 vector such as q, then $x = H(q)m(q)$ where the elements of $m(q)$ of 3×1 dimensions, are the x, y, and z components of the dipole moment at the time instant when x is recorded and the columns, of the $N \times 3$ transfer matrix $H(q)$ are the solutions to the forward problem.

Therefore, the first column of $H(q)$ is the signal recorded at the electrodes due to a dipole source at location q having the unity moment in the x direction and zero moment in y and z directions. Respectively, the second and third columns represent the potential due to sources with unity moment in y and z directions. This kind of modeling the data may be applied to elec-

tric, magnetic, or combined electro-magnetic measurements. The elements of $H(q)$ depend on the recording modality, i.e. EEG or MEG. What it physically means is that the material and geometrical properties of the medium in which the sources are recorded solve the forward problem.

Assuming the potential propagation medium is linear, the potential at each sensor over the scalp is superposition of the potentials from many active neurons underneath that sensor. Suppose that x is generated when L dipole sources are active at different locations, i.e. $q_i, i = 1, 2, \cdots, L$ superimposed by measurement noise n. Then

$$x = \sum_{i=1}^{L} H(q_i) m(q_i) + n \qquad (5.2.3)$$

The LCMVB is an example for the spatial filtering concept which refers to the identification of sources based on their spatial locations. A "narrowband" spatial filter passes signals originating from a small "passband" volume while attenuating those originating from other locations which exactly follow the terminology used in the temporal or spectral domain. In the present application, the spatial samples are elements of the data vector and the spatial filter is implemented as a weighted combination of these samples. The ultimate goal is to design a set of spatial filters where each filter allows signals originating from a specified location within the brain passing while attenuating signals from other locations. Therefore, monitoring the variance at the output of each spatial filter may provide an estimate of the distribution of brain neural activity. This concept has been well established in [102, 104] as below.

The signal emitting from each grid location in the brain consists of a three component dipole moment. Hence, three spatial filters for each lo-

cation are needed; one for each component of the dipole moment. Let the spatial filter for the narrowband volume element Q_0 centered at location q_0 be an $N \times 3$ matrix $\mathbf{W}(q_0)$ and the three component filter output $\mathbf{y}$ be the inner product of $\mathbf{W}(q_0)$ and $\mathbf{x}$.

$$\mathbf{y} = \mathbf{W}'(q_0)\mathbf{x} \qquad (5.2.4)$$

An ideal narrowband spatial filter satisfies

$$\mathbf{W}'(q_0)\mathbf{H}(q) = \begin{cases} \mathbf{I}, & \text{if } q = q_0 \\ \mathbf{0}, & \text{if } q \neq q_0, q \in \Omega \end{cases} \qquad (5.2.5)$$

and Ω represents the volume of the brain. If (5.2.5) is satisfied, then in the absence of noise ($\mathbf{n} = \mathbf{0}$) the filter output is $\mathbf{y} = \mathbf{m}(q_0)$ which is the dipole moment at the considered grid point. Minimizing the variance of the filter output $\mathbf{y}$ constrained to (5.2.5) would solve the source localization problem. This can be put into mathematical formulations as

$$\min_{\mathbf{W}(q_0)} \text{Tr} \, \Sigma_y \quad \text{s.t. } \mathbf{W}'(q_0)\mathbf{H}(q_0) = \mathbf{I} \qquad (5.2.6)$$

where Tr is the trace operator and Σ_y is the covariance of the output signals.

Considering (5.2.4), (5.2.6) can be re-written as

$$\min_{\mathbf{W}(q_0)} \text{Tr} \, (\mathbf{W}'(q_0)\Sigma_x\mathbf{W}(q_0)) \quad \text{s.t. } \mathbf{W}'(q_0)\mathbf{H}(q_0) = \mathbf{I}. \qquad (5.2.7)$$

This classic minimization problem may be easily solved by using the Lagrange multipliers method as in [104]

$$J(\mathbf{W}(q_0), \mathbf{L}) = \text{Tr}\left\{\mathbf{W}(q_0)\Sigma_x^{-1}\mathbf{W}(q_0) + 2(\mathbf{W}(q_0)'\mathbf{H}(q_0) - \mathbf{I})\mathbf{L}\right\} \qquad (5.2.8)$$

where L is a 3×3 matrix of the Lagrange multipliers. The solution to (5.2.8) is computed as

$$W(q_0) = \left[(H'(q_0)\Sigma_x^{-1}H(q_0) \right]^{-1} H'(q_0)\Sigma_x^{-1}. \qquad (5.2.9)$$

Using (5.2.9) in (5.2.4) gives an estimate of the moment at location q_0. Considering that the variance of y coming from location q_0 is an indication for the existence of a source at q_0, after some straightforward algebraic manipulations, it can be shown that

$$\widehat{\text{Var}}(q_0) = \text{Tr}\left\{ \left[(H'(q_0)\Sigma_x^{-1}H(q_0) \right]^{-1} \right\}. \qquad (5.2.10)$$

In order to localize the brain sources, the variance as a function of grid location within the volume of the brain is estimated. This is accomplished by evaluating (5.2.10) as a function of q_0. Regions demonstrating large variances show substantial neural activities, while on the other hand regions with small variance can be considered inactive. In [118], an enhanced LCMV beamforming technique is also developed where a statistical maximum contrast criterion is exploited for MEG source localization.

The above formulations can effectively localize the brain sources. However, they have not been utilized for EB artifact removal applications. The main reasons in my opinion are that, first, if one wants to use LCMV based methods he/she has to solve the forward problem and compute H which does not seem to be rational where the EB artifact can be easily removed by other techniques, see Chapter 4. Second, even if H is estimated, it presents acceptable accuracies mainly for grid points within the brain matter and not over the outer marginal grid points. Therefore, researchers have neglected LCMV based methods after the work of Ille *et al.* in [91, 119] where a

spatially constrained ICA has been developed for EEG restoration from artifacts.

Therefore, there has been a need for easier methods than the classic LCMV of [104] based on the forward model estimation of [115] for EB artifact removal. Chapter 4 provided a theoretically and conceptually simple and straightforward solution to that problem by estimation of one column of **H** which corresponds to the EB source. However, the main problem is the computational requirements occurred in the estimation of the steering vector. In this chapter, first, a robust version of the LCMV based on [113] is presented. Then, a fast algorithm for the estimation of the STF model of EEGs contaminated with the EB artifacts is developed.

5.2.2 Robust Minimum Variance Beamformer

The most straightforward way to extract the j-th source is to project $\mathbf{x}(t)$ onto the space orthogonal to, denoted by $\perp$, all of the columns of **A** except $\mathbf{a}_j$, i.e., $\{\mathbf{a}_1, \cdots, \mathbf{a}_{j-1}, \mathbf{a}_{j+1}, \cdots, \mathbf{a}_N\}$. Since $\mathbf{a}_j$ performs as the steering vector of the j-th source, by defining a vector as a *spatial filter*, $\mathbf{w}_j$, it may be written [78] as

$$y(t) = \mathbf{w}_j'\mathbf{x}(t) \tag{5.2.11}$$

where $y(t)$ is an estimation of the source $s_j(t)$ corresponding to $\mathbf{a}_j$. The spatial filter can be determined by applying the unit-gain constraint, $\mathbf{w}_j'\mathbf{a}_j = 1$ and by minimizing the variance of the filter output $y(t)$ [104]. However, in practice, the steering vector $\mathbf{a}_j$ is not always known [110-113]. Hence, the approach based on the theoretically rigorous worst-case performance optimization, recently developed in [113], is used here in order to compensate

for the deviation vector, δ, of $\hat{\mathbf{a}}_j$ from the actual steering vector $\mathbf{a}_j$, i.e.,

$$\delta = \mathbf{a}_j - \hat{\mathbf{a}}_j \tag{5.2.12}$$

Note that δ is l_2 norm-bounded by some known positive constant ϵ. Denoting

$$\mathbf{R} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{R}_{\mathbf{xx}}^{k} \tag{5.2.13}$$

as outlined in [114, Chapter 2], the beamformer is obtained by minimizing

$$\min_{\mathbf{w}_j} \left(J_c = \mathbf{w}_j' \mathbf{R} \mathbf{w}_j \right) \quad \text{s.t.} \quad \min_{\|\delta\|_2 \le \epsilon} (|\mathbf{w}_j' \hat{\mathbf{a}}_j + \mathbf{w}_j' \delta| - 1) \tag{5.2.14}$$

where $|.|$ denotes the absolute value operator. Equivalently [120], the above optimization process may be rewritten as

$$\min_{\mathbf{w}_j} \left(J_c = \mathbf{w}_j' \mathbf{R} \mathbf{w}_j \right) \quad \text{s.t.} \quad |\mathbf{w}_j' \hat{\mathbf{a}}_j - \epsilon \|\mathbf{w}_j\|_2| = 1. \tag{5.2.15}$$

Following the Lagrange multiplier method, J_c is differentiated with respect to $\mathbf{w}_j$ and set it to zero. Afterward,

$$\mathbf{R}\mathbf{w} + \lambda\epsilon \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|_2} = \lambda \hat{\mathbf{a}}_j. \tag{5.2.16}$$

After dropping the scalar λ (it can be merged in ϵ) the spatial filter can be computed as [120]

$$\mathbf{w}_j = \left[\mathbf{R} + \frac{\epsilon}{\rho} \mathbf{I} \right]^{-1} \hat{\mathbf{a}}_j \tag{5.2.17}$$

where $\rho \triangleq \|\mathbf{w}_j\|$ and $\mathbf{I}$ denotes the identity matrix. In (5.2.17), the main concern in estimating $\mathbf{w}_j$ is to have an estimation of ρ which may be determined

by using the following procedure. Eigenvalue decomposition of **R**, i.e.,

$$\mathbf{R} = \mathbf{U}\Xi\mathbf{U}' \tag{5.2.18}$$

results in the $N \times N$ unitary matrix **U** whose columns are the unit norm eigen-vectors of **R**, and Ξ is the diagonal matrix of the real positive eigenvalues of **R**, with elements ξ_i where

$$\xi_1 \geq \xi_2 \geq \cdots \geq \xi_N > 0. \tag{5.2.19}$$

By defining

$$\Psi(\rho) \triangleq \Xi + \frac{\epsilon}{\rho}\mathbf{I} \tag{5.2.20}$$

and following the procedure suggested in [120], it may be written as

$$\|\mathbf{U}\Psi^{-1}(\rho)\mathbf{U}'\hat{\mathbf{a}}_j\|_2^2 - \rho^2 = \|\Psi^{-1}(\rho)\mathbf{g}'\|_2 - \rho^2 = 0 \tag{5.2.21}$$

where $\mathbf{g} = [g_1, g_2, \cdots, g_N]' = \mathbf{U}'\hat{\mathbf{a}}_j$. Introducing [114, Chapter 2]

$$f(\rho) \triangleq \|\Psi^{-1}(\rho)\mathbf{g}'\|_2 - \rho^2 = \sum_{i=1}^{N}\left[\frac{|g_i|}{\epsilon + \rho\xi_i}\right]^2 - 1 = 0, \tag{5.2.22}$$

it is shown that the necessary and sufficient condition for (5.2.22) to have a unique real positive solution for ρ is that the norm of the mismatch vector is upper bounded by the norm of the estimated signal steering vector, i.e., $\|\delta\|_2 = \epsilon < \|\hat{\mathbf{a}}_j\|_2$. Considering $\|\mathbf{g}\|_2 = \|\hat{\mathbf{a}}_j\|_2$ and (5.2.22), the upper bound of $f(\rho)$ is achieved as

$$f(\rho) < \frac{\sum_{i=1}^{N}|g_i|^2}{(\epsilon + \rho\xi_N)^2} - 1 = \frac{\|\hat{\mathbf{a}}_j\|_2^2}{(\epsilon + \rho\xi_N)^2} - 1 \triangleq f_{max}(\rho). \tag{5.2.23}$$

Algorithm 2 Estimating $\mathbf{w}_j$

Require: $\rho_{max}, f(\rho)$.

Ensure: $i = 0$, using the binary search method, find $\rho_i \in (0, \rho_{max})$ if $f(\rho_i) > 0$ & $f(\frac{13}{12}\rho_i) < 0$.

Ensure: the stopping value κ, Maximum iterations I and $i \leftarrow i + 1$.

    while $|f(\rho_i)| < \kappa$ & $i < I$ **do**

$$\rho_i = \rho_{i-1} - \frac{f(\rho_{i-1})}{\nabla_\rho f(\rho_{i-1})}$$

    end while

Ensure: Using (5.2.17) compute $\mathbf{w}_j$ where $\rho = \rho_0 = \rho_i$.

Note that $f(\rho)$ and $f_{max}(\rho)$ are both decreasing functions of ρ and the root of $f(\rho)$, say ρ_0, is positive. Hence [114, Chapter 2]

$$0 < \rho_0 < \rho_{max} = \frac{\|\hat{\mathbf{a}}_j\|_2 - \epsilon}{\xi_N}. \tag{5.2.24}$$

Therefore, the problem of estimating ρ and consequently the spatial filter, $\mathbf{w}_j$, can be solved within an iterative scheme as in **Algorithm 2** in which $\nabla_\rho f(\rho_{i-1})$ is the derivative of $f(\rho)$ with respect to ρ at $\rho = \rho_{i-1}$.

5.2.3 PARAFAC and STF Modeling

By exploiting PARAFAC, the factor relevant to the EB artifact to be used within the beamforming procedure is extracted. The resulting spatial signature of the EB-related factor is exploited to formulate (5.2.17). Importantly, it has been considered that the spatial signatures of this factor are directly related to the level of EB contamination for each electrode. This assumption is rational since EB may be considered as a strong point source which is only attenuated while propagating from the frontal area to the central and occipital parts of the brain. Hence, the column of the mixing matrix $\mathbf{A}$, i.e. $\hat{\mathbf{a}}_j$, corresponding to the EB source, is estimated by PARAFAC and used in (5.2.17). Hereafter, the novel approach for estimating the STF model of EEGs using the proposed STF-TS model is introduced.

In order to decompose the EEGs into spatial, temporal, and spectral signatures, the three-way PARAFAC is applied to the three-way EEG data

$$\underline{\check{Y}}^{N\times F\times T} \triangleq \underline{\check{Y}}(1:N, 1:F, 1:T) \tag{5.2.25}$$

where N, F, and T are respectively the number of EEG channels, frequency bins and time instants. Note that in this chapter in contrast to the previous chapters, the EEGs are modeled as $\underline{\check{Y}}^{N\times F\times T}$ rather than $\underline{\check{Y}}^{N\times T\times F}$. However, since the three-way tensor of EEGs is constructed in the STF domain, the term "STF" is kept rather that using "SFT".

Therefore, $\check{A}^{N\times M}$, $\check{C}^{F\times M}$, and $\check{D}^{T\times M}$ are respectively the spatial, spectral and temporal signatures of $\underline{\check{Y}}^{N\times F\times T}$ where their elements are denoted as $\check{a}(n,m)$, $\check{c}(f,m)$, $\check{d}(t,m)$. While retaining the consistency of formulation, the superscripts may occasionally be dropped in order to simplify the presentation.

The STF model is presented as:

$$\underline{\check{Y}}^{N\times F\times T} = \underline{\hat{\check{Y}}} + \underline{\check{E}}^{N\times T\times F} \tag{5.2.26}$$

where

$$\underline{\hat{\check{Y}}} = \sum_{m=1}^{M} \check{a}(n,m)\check{c}(f,m)\check{d}(t,m) \tag{5.2.27}$$

is an estimation, denoted by ($\hat{}$), of $\underline{\check{Y}}^{N\times T\times F}$, M stands for the maximum possible number of factors, and $\underline{\check{E}}^{N\times T\times F}$ is the three-way array of the residue of the model which is mostly omitted for brevity.

In order to find M, the known CORCONDIA measure [64] is customized. The signatures $\check{A}$, $\check{C}$, and $\check{D}$ can be estimated by using the alternating least

squares algorithm where the cost function is

$$[\hat{\mathbf{A}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}] = \arg\min_{\check{a},\check{c},\check{d}} \|\check{\underline{\mathbf{Y}}} - \hat{\underline{\mathbf{Y}}}\|_2^2. \tag{5.2.28}$$

Intuitively, the spatial signature $\check{\mathbf{A}}$ obtained from the STF model represents the weighting parameters of the inter-channel correlation among the time-frequency representations of each channel. However, in order to mitigate the high computational cost occurring in using STF with three-way PARAFAC [51], in the sequel, a novel method for estimating the STF model is introduced. The strategy is based on the *divide and conquer* philosophy where, as will be detailed later on, instead of calculating the model signatures from the original data, these signatures are estimated by joining the weighted versions of their *local* temporal signatures.

5.2.4 STF-TS Modeling

For a long-term EEG measurement, the calculations of both the time-frequency transform and STF-based PARAFAC are computationally intensive. Therefore, aiming at reducing this computational complexity, the time domain is divided into a number of segments. After that, the time-frequency transform is applied [78] individually to each segment forming a four-way array. The four-way array

$$\underline{\mathbf{Y}}^{N \times S \times F_s \times T_s} \triangleq \underline{\mathbf{Y}}(1:N, 1:S, 1:F_s, 1:T_s) \tag{5.2.29}$$

is set up where N is the channel index and S is the maximum time/segment index. The energies of the time-frequency transform for T_s time instants and F_s frequency bins are then computed. PARAFAC is then applied to the four-way array. This may be formulated in the same way as in [51] where $\mathcal{A}^{N \times M}$

is the spatial signature, $\mathcal{B}^{S \times M}$ is the temporal/segment signature, $C^{F_s \times M}$ is the

spectral signature, and $\mathcal{D}^{T_s \times M}$ is the temporal signature with matrix elements

denoted respectively as $a(n, m)$, $b(s, m)$, $c(f_s, m)$, and $d(t_s, m)$. Hence,

$$\underline{\mathbf{Y}}^{N \times S \times F_s \times T_s} = \underline{\hat{\mathbf{Y}}} + \underline{\mathcal{E}}^{N \times S \times F_s \times T_s} \qquad (5.2.30)$$

where

$$\underline{\hat{\mathbf{Y}}} = \sum_{m=1}^{M} a(n, m)b(s, m)c(f_s, m)d(t_s, m) \qquad (5.2.31)$$

and $\underline{\mathcal{E}}^{N \times S \times F_s \times T_s}$ is the negligible four-way residual of the model array. In

order to find the model, the following cost function is used

$$[\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{C}, \hat{\mathcal{D}}] = \arg \min_{a,b,c,d} \|\underline{\mathbf{Y}} - \underline{\hat{\mathbf{Y}}}\|_2^2. \qquad (5.2.32)$$

By decomposing the multi-channel EEGs using the STF-TS model, the

number of free parameters P_4, i.e., the number of elements that has to be

estimated by PARAFAC, is

$$P_4 = M(N + S + F_s + T_s), \qquad (5.2.33)$$

while the number of free parameters of the STF model P_3 is as high as

$$P_3 = M(N + F + T). \qquad (5.2.34)$$

Evidently, when T is large, $P_4 \ll P_3$. This means that less parameters

need to be estimated and therefore the computational complexity of the

PARAFAC algorithm is reduced. Here, it is shown how to estimate the

signatures of the STF model using the signatures of the STF-TS model. In

this chapter, the TALS method [101] is used to compute the parameters of

the STF model-trilinear model. Similarly a customized quadlinear version of the trilinear ALS (TALS) is used to compute the parameters of the STF-TS model, i.e. (5.2.32). By using the STF-TS model, the poor convergence of TALS can be avoided by selecting the appropriate size (number) of segments S.

According to (5.2.30), the temporal signatures of the long-term EEGs are estimated by cascading all S segments of the temporal signatures $\mathcal{D}$ which are weighted by their corresponding time/segment signatures $\mathcal{B}$. In order to effectively estimate the STF model from the STF-TS model, the suggested number of segments S and the number of components M should maximize the CORCONDIA value as

$$[S, M] = \arg \max_{S} \left\{ \arg \max_{M} [\text{CORCONDIA}(\underline{Y}, \mathcal{A}, \mathcal{B}, C, \mathcal{D})] \right\}. \quad (5.2.35)$$

The main concept behind (5.2.35) is that by decomposing $\underline{Y}$ to as many as M possible factors for the STF model, it is firstly guaranteed that the correct number of factors for STF is achieved and then, the process of temporal segmentation is carried out. In other words, since the ultimate goal of the STF-TS model is to approximate the STF model, M should be identified for the STF model using the conventional approach of [51] before adjusting S to maximize the CORCONDIA criterion for the STF-TS model.

When the residual is considered negligible, the STF model (5.2.26) can be written in a matrix form as

$$\underline{\check{Y}}_{N \times F \times T} = \check{D} \Sigma_{\check{A}_n} \check{C}', \quad (5.2.36)$$

where $\Sigma_{\check{A}_n}$ is the diagonal matrix with the n-th row of $\check{A}$ as its diagonal elements, $n = 1, 2, \cdots, N$. Similarly, the STF-TS model (5.2.30) is written

in matrix form as

$$\underline{Y}_{N \times S \times F_s \times T_s} = \mathcal{D} \Sigma_{\mathcal{B}_s} \Sigma_{\mathcal{A}_n} C'$$

(5.2.37)

where $\Sigma_{\mathcal{A}_n}$ is a diagonal matrix with the n-th row of $\mathcal{A}$ as its diagonal elements, $n = 1, 2, \cdots, N$. Similarly, $\Sigma_{\mathcal{B}_s}$ is a diagonal matrix with the s-th row of $\mathcal{B}$ as its diagonal elements for $s = 1, 2, \cdots, S$. According to (5.2.36) and (5.2.37), $\check{D}$ for the STF model can be estimated by the scaled version of $\mathcal{D}$ from the STF-TS model as

$$\check{D} \approx [\mathcal{D} \Sigma_{\mathcal{B}_1}, \cdots, \mathcal{D} \Sigma_{\mathcal{B}_S}]'.$$

(5.2.38)

In addition, in order to simultaneously achieve acceptable estimates of the temporal and spectral signatures, the following condition should be addressed:

$$\frac{1}{f_0} \leq \frac{L}{S} \leq T_{int}$$

(5.2.39)

where L is the length of the EEG in seconds and S is the number of segments. T_{int} is the time interval that allows the temporal signatures to have smooth envelopes. The fundamental frequency, f_0, is defined as the frequency of the first peak in the frequency spectrum of filtered EEGs. Bearing in mind that as long as

$$\frac{1}{f_0} \leq \frac{L}{S}$$

(5.2.40)

the spectral signatures are reconstructed faithfully, it has been empirically found that for various EEG recordings, in order to achieve smooth reconstructions for the temporal signatures, T_{int} should take values between 0.7-0.9 seconds. After few simple mathematical manipulations (5.2.39) can be
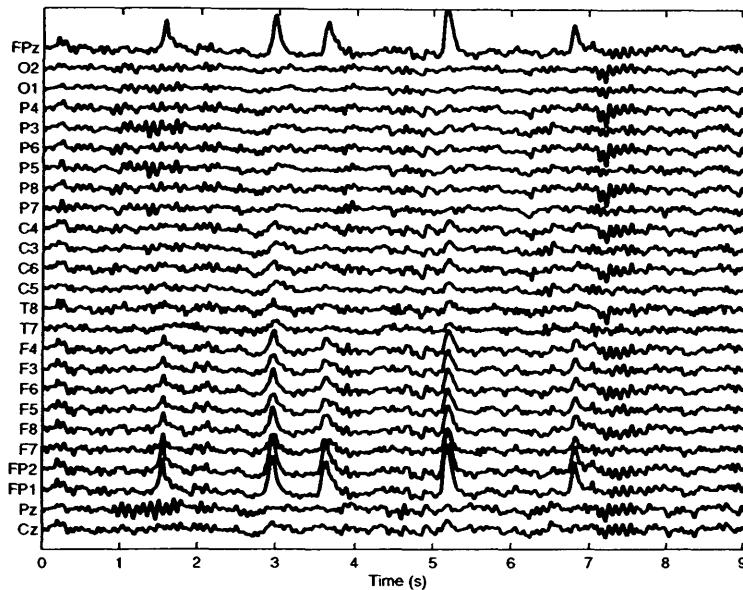
Figure 5.1. A set of real EB contaminated EEG recordings.

easily written as

$$\frac{L}{T_{int}} \leq S \leq f_0 L \quad \& \quad S, \frac{f_s L}{S} \in \mathbb{Z}^+ \qquad (5.2.41)$$

where f_s is the sampling rate, and $\mathbb{Z}^+$ represents the set of positive integers. Here, as explained in Section 5.3, f_0 is set to two Hz since the EEG measurements have been bandpass filtered between 2 to 30 Hertz. If the above conditions are taken into account, the spectral signature $\check{C}$ is also well approximated by C, while the spatial signature $\check{A}$ is approximately equal to $\mathcal{A}$. Here, it is indicated that the acceptable values for S mainly depend on the different terms in (5.2.41), namely, the sampling rate, the length of the data under study, and also the selection of T_{int} which are totally subjective.

From the original available data set, an EB contaminated segment of the EEG of 9.2 seconds length, i.e. 1820 sample points (see Fig. 5.1) is selected. The STF model of EEG recordings of Fig. 5.1 has been shown in Fig. 5.2 where according to the second row of Table 5.1, two factors can be extracted if $S = 1$, i.e. $M = 2$. Evidently, the first components (Factor 1) of the STF model demonstrate the EB-relevant factor since

Table 5.1. The computed CORCONDIA percentage values for different S and M corresponding to the EEG segment in Fig. 5.5. "-" denotes that CORCONDIA does not converge for that specific S and M.

Model	M = 1	M = 2	M = 3
STF	100.0	**98.4**	-
STF-TS (S=10)	100.0	*54.1*	-
STF-TS (S=13)	100.0	<u>**33.5**</u>	-
STF-TS (S=14)	100.0	-11.9	-

Table 5.2. The computed CORCONDIA percentage values for different S and M corresponding to the EEG segment in Fig. 5.5. "-" denotes that CORCONDIA does not converge for that specific S and M.

Model	M = 1	M = 2	M = 3
STF	100.0	**98.9**	46.2
STF-TS (S=10)	100.0	-13.4	-
STF-TS (S=12)	100.0	-5.0	-
STF-TS (S=15)	100.0	-8.6	-
STF-TS (S=18)	100.0	<u>**17.1**</u>	-

1. It mainly occurs in the frequency band of around 5 Hz while the other factor exists in the entire band and represents the ongoing activity of the brain or perhaps a broadband white noise-like component, Fig. 5.2-(a).

2. The temporal signature of the first factor definitely shows a transient phenomenon such as EB while that of Factor 2 consistently exists during the course of the EEG segment, Fig. 5.2-(b).

3. Unlike Fig. 5.2-(d), in Fig. 5.2-(c), the spatial distribution of the extracted factor is confined to the frontal area, which clearly demonstrates the effect of EB. The other factor shows the background activity of the brain as it spreads all over the scalp.

For STF-TS modeling, it is considered $L = 9.2$ and $T_{int} = 0.9$ seconds

and $f_s = 200$ in (5.2.41). Therefore, the initial candidates for S are 13 and 14. Although in (5.2.41) the lower bound for S is $\frac{L}{T_{int}} = 10.22$, $S = 10$ is intentionally included in the analysis in order to demonstrate the accuracy of (5.2.41). The CORCONDIA values for $M = 2$ and $S = 10, 13$, and 14 have been calculated and shown in Table 5.1. Here, as in (5.2.35), the maximum CORCONDIA value for maximum M and S should be selected. Apparently, disregarding (5.2.41), the best CORCONDIA candidate in Table 5.1 is 54.1 for $M = 2$ and $S = 10$. As plotted in Fig. 5.3, an acceptable decomposition was not achieved for $S = 10$, although it presents the maximum CORCONDIA. Evidently, none of the six signatures, i.e. two spectral, two temporal, and two spatial signatures, have been estimated correctly. Note that due to the leakage from the dominant EB factor to the brain activity factor during decomposition, there is a considerable similarity in their spectral and spatial signatures. The temporal signatures are also misidentified. Therefore, it is concluded that not only the CORCONDIA value is important but also S should fulfil the inequalities and conditions of (5.2.41). In practice, such mis-modelings can be avoided by carefully testing the marginal value of S, i.e. 10 in this experiment, or a proper selection of the T_{int}.

Therefore, the next candidate is selected, i.e. $S = 13$ for which the CORCONDIA value is 33.5. The results of the EEG STF-TS modeling for $M = 2$ and $S = 13$ have been plotted in Fig. 5.4 where it is illustrated how well the STF model is approximated by the STF-TS model. Factor 1 stands for the EB factor while again Factor 2 shows the brain background activity. In the sequel, the spatial signature of Factor 1 is used in the beamforming stage.

Note that, the acceptable values for S mainly depend on various terms of (5.2.41). For instance, depending on an specific application, if one selects
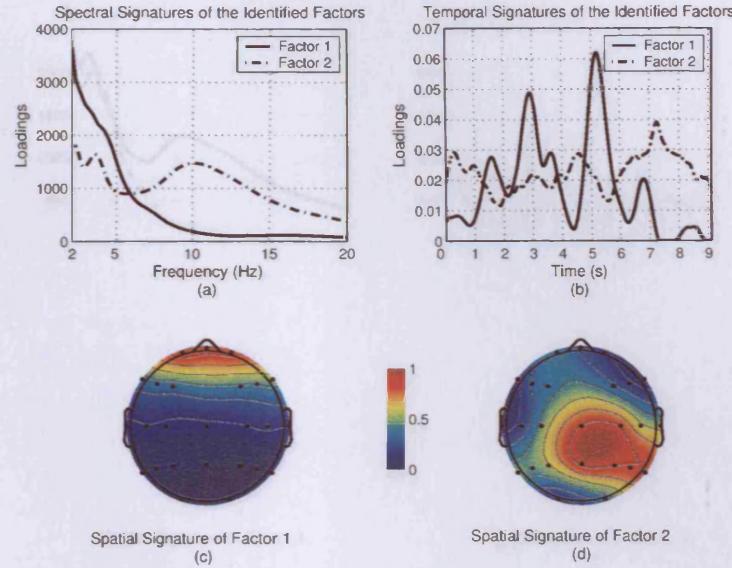
Figure 5.2. The extracted factors by using STF decomposition of the EEG recording in Fig. 5.1; (a) and (b) illustrate respectively the spectral and temporal signatures of the extracted factors, (c) and (d) represent the spatial distribution of the factors, respectively. Evidently, Factor 1 demonstrates the EB phenomenon since it occurs in a frequency band of around 5 Hz (a), it is indeed transient in the time domain (b) and it is confined to the frontal area.

the length of the data to be 4.2 seconds and the sample rate to be 1000 Hz, with $f_0 = 2$ and $T_{int} = 0.9$s, then (5.2.35) should be solved for $S = 5, 6, 7,$ and 8. However, again it is suggested that it is not likely to have an acceptable decomposition for the smallest value of S, i.e. 5. Therefore, in order to avoid such cases, the solution is achieved setting $T_{int} = 0.8$s and computing (5.2.41) for $S = 6, 7,$ and 8. The computed CORCONDIA percentage values for different S and M corresponding to the EEG segment plotted in Fig. 5.5 are reported in Table. 5.2. The bold value **98.9** demonstrates that $M = 2$ is the correct number of factors for the STF model. Estimating the CORCONDIA for $M = 2$ and $S = 10, 12, 15,$ and 18 and selecting the maximum fitness, the bold underlined value **17.1**, shows the right selection of the S and M for the STF-TS model. In summary, the presented method consists of the following steps. Given an artifact contaminated EEG data:
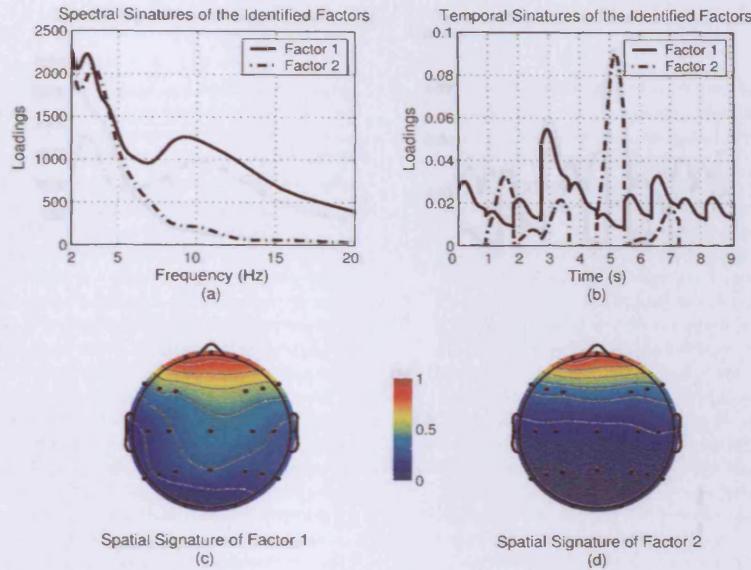
Figure 5.3. The extracted factors by using STF-TS decomposition of the EEG recording in Fig. 5.1 when $M = 2$ and $S = 10$. Regarding Fig. 5.2, none of the six signatures, i.e. two spectral, two temporal, and two spatial, have been estimated correctly. Note that due to the leakage from the dominant EB factor to the brain activity factor while decomposition, there is a considerable similarity in their spectral and spatial signatures. The temporal signatures are also misidentified.

1. bandpass filter the EEGs between 2 Hz and 30 Hz,

2. set up the four-way array, i.e., $\underline{\mathbf{Y}}_{N \times S \times F_s \times T_s}$, as stated in section 5.2.4,

3. execute the four-way PARAFAC and select the EB artifact relevant factor as described in section 5.3,

4. exploit the spatial signature of the EB artifact factor as $\hat{\mathbf{a}}_j$ and execute the beamforming procedure,

5. reconstruct the artifact removed EEGs by deflation.

5.3 Simulation Results

The procedure is applied to real EEG measurements. The dataset was provided by the School of Psychology, Cardiff University, UK. It represents a
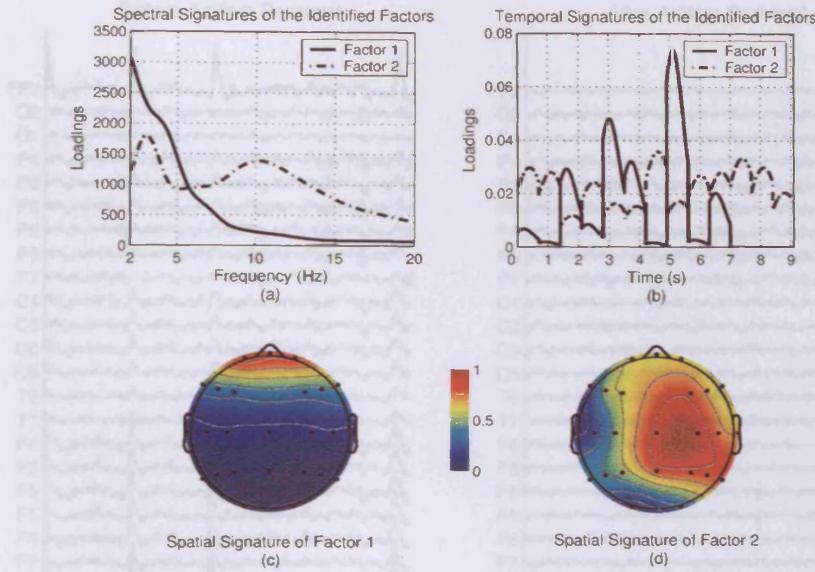
Figure 5.4. The extracted factors by using STF-TS modeling by $M = 2$ and $S = 13$. Interestingly, as expected, the spectral and spatial signatures of the extracted components are very similar to those of Fig. 5.2 and the temporal signatures effectively identify the transient EBs.

wide range of EBs and, therefore, gives a proper evaluation of the method. The scalp EEG was obtained using 25 Silver/Silver-Chloride electrodes placed at locations defined by the conventional 10-20 system [31]. The data were sampled at 200 Hz, and bandpass filtered with cut-off frequencies of 2 Hz and 30 Hz. Twenty real highly EB contaminated EEG recordings, each 9 seconds long have been artifact removed by using this method. The performance of the algorithm can be observed by comparing the EEGs obtained at the electrodes in Fig. 5.5-(a) and the same segment of data after being processed by the proposed algorithm in Fig. 5.5-(b).

In what follows, a detailed comparison between the results of STF modeling using the two mentioned approaches in Section 5.2, i.e., direct three-way PARAFAC (Section 5.2.3), see Fig. 5.6 and the STF modeling by using the STF-TS model of EEGs (Section 5.2.4), see Fig. 5.7, is provided. Averaged CORCONDIA values for three independent runs with differ-
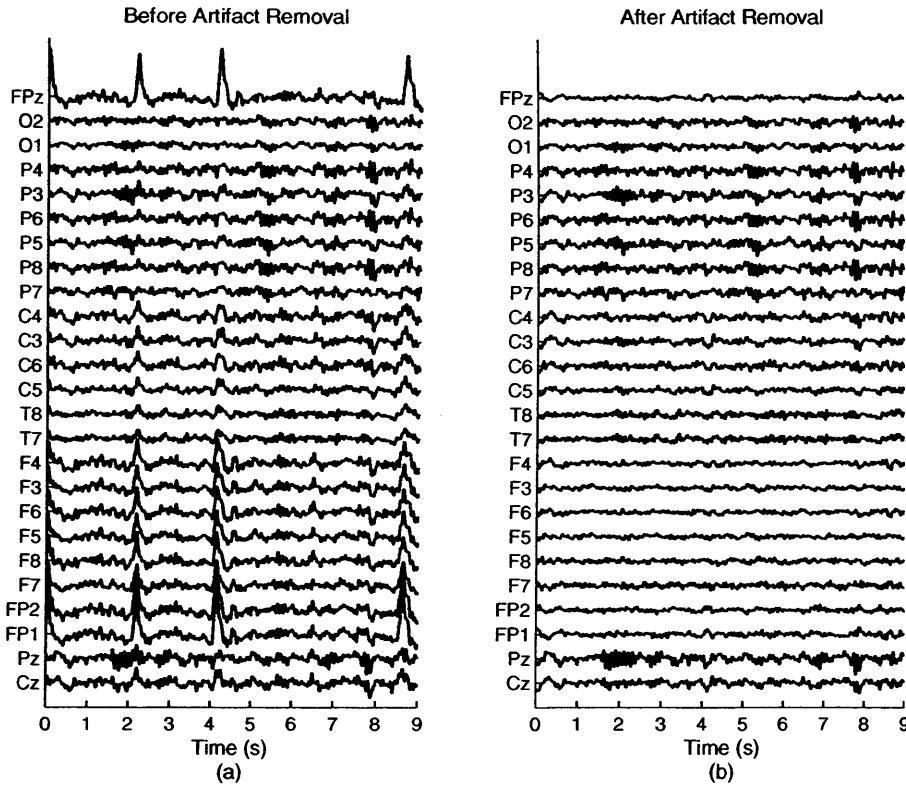
Figure 5.5. The results of the proposed EB artifact removal method for a set of real EEGs. The left subplot depicts highly EB contaminated EEGs before artifact removal while in the right subplot the segment of EEGs after being corrected for EB artifact is illustrated.

ent initialization as detailed in [51] have been computed for methods of STF and STF-TS modeling. In Fig. 5.6, the number of components M is selected as $M = 2$ according to the computed CORCONDIA value, i.e, 98.4% whereas the CORCONDIA for the proposed STF-TS model was 17.1% when the number of segments was $S = 18$ [see (5.2.35) and Table 5.2].

Figs. 5.7-(a) to -(d) illustrate respectively the estimated spectral, temporal, and spatial signatures of the under study EEGs. The results of the STF-TS model in comparison to that of the STF model, i.e. Fig. 5.6, demonstrate the reliability of the STF-TS modeling, since both methods result in approximately the same signatures and, as expected, the STF-TS method is a faster
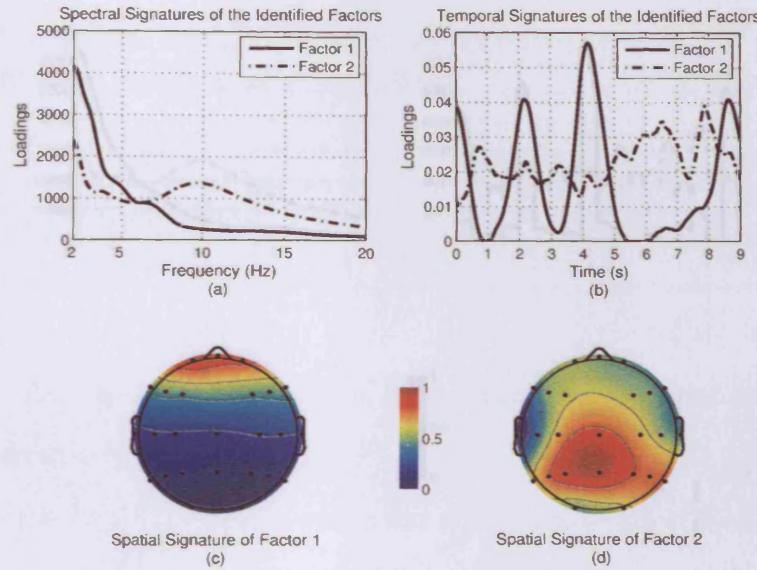
Figure 5.6. The extracted factors by using STF modeling; (a) and (b) illustrate respectively the spectral and temporal signatures of the extracted factors, (c) and (d) represent the spatial distribution of the factors, respectively. Factor 1 demonstrates the EB phenomenon.

algorithm. Small deviations in spectral and temporal signatures of the STF model using STF-TS are negligible, since they are merely utilized to identify the EB relevant factor. Moreover, experimentally it has been found that due to the fact that the EB factor is the dominant factor, it is always effectively identified, if the conditions in (5.2.41) are met; any probable deviation only perturbs the signatures of the background EEG activities.

By using the STF model, the parallel factors of the three-way array of size $N \times F \times T$ have to be calculated. This process takes a longer period of time due to the calculations of more free parameters P_3 as compared to the P_4 values with the STF-TS model. The first row of Table 5.3 shows that the number of free parameters is greatly reduced by using the STF-TS model, where the size of the three-way $\underline{\check{\mathbf{Y}}}_{N \times F \times T}$ for the STF model is $25 \times 1800 \times 180 = 4010$ parameters to be estimated, and the size of the four-way $\underline{\mathbf{Y}}_{N \times S \times F_s \times T_s}$ for the STF-TS model is $25 \times 18 \times 180 \times 100$, i.e.
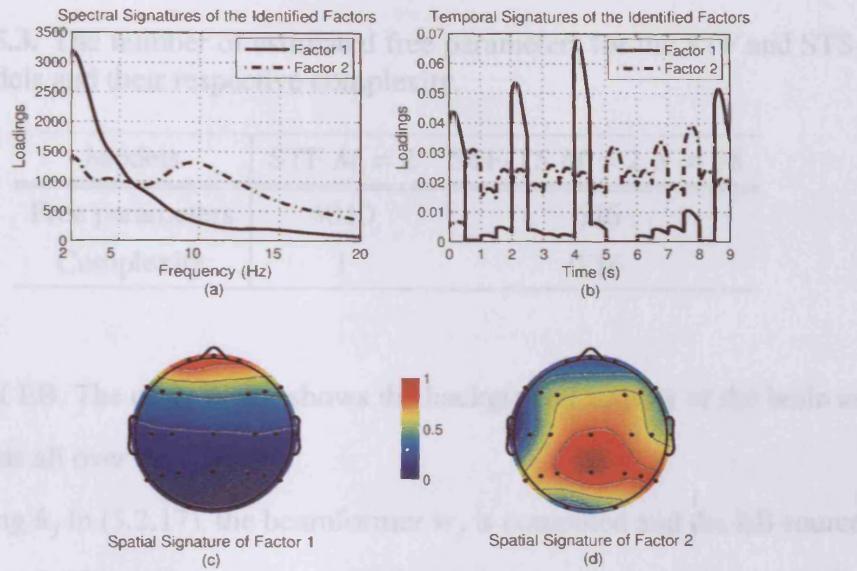
Figure 5.7. The extracted factors by using STF-TS modeling; (a) and (b) illustrate respectively the spectral and temporal signatures of the extracted factors and (c) and (d) represent the spatial distributions of those extracted factors. Interestingly, as expected, the spectral and spatial signatures of the extracted components are very similar to those of Fig. 5.6 and the temporal signatures effectively track the transient EBs of the ongoing EEGs.

646 parameters to be estimated. Consequently, the second row of Table 5.3 illustrates the relative calculation time of the STF and STF-TS models. For the EEGs used in this experiment, the relative calculation time of the STF-TS model, presuming that the calculation time of the STF model compared to the method proposed in [77] is 1, is 0.16.

At this stage, the spatial signature of the EB artifact relevant factor is of interest - to be used in the RMVB algorithm as an approximation to $\mathbf{a}_j$, i.e. $\hat{\mathbf{a}}_j$. The first components (Factor 1) of both STF models resulted from the two approaches demonstrate the EB-relevant factor since it mainly occurs in the frequency band of around 5 Hz and its temporal signature shows a transient phenomenon. Moreover, unlike in Fig. 5.6- and Fig. 5.7-(d), in Fig. 5.6- and Fig. 5.7-(c), the spatial distribution of the extracted factor, to be used as $\hat{\mathbf{a}}_j$, is confined to the frontal area, which clearly demonstrates the

Table 5.3. The number of estimated free parameters for the STF and STS-TS models and their respective complexity.

Models	STF $M = 2$	STF-TS $M = 2, S = 18$
Free parameters	4010	646
Complexity	1	0.16

effect of EB. The other factor shows the background activity of the brain as it spreads all over the scalp.

Using $\hat{a}_j$ in (5.2.17), the beamformer w_j is computed and the EB source is extracted. The artifact removed EEGs are then reconstructed by using the batch deflation method [66, pp. 192]. You are referred to Section 4.2.3 for further details.

In order to provide a quantitative measure of performance for the proposed artifact removal method, the correlation coefficient (CC) between the extracted EB artifact source and the original EEGs and the artifact removed EEGs are computed in a similar way to the previous chapter, see Fig. 5.8.

The values reported in Fig. 5.8 have been computed as follows. For each of the 20 different EB artifact contaminated EEGs, the proposed method is executed. The aforementioned CCs for each run were then computed between the extracted EB and the EEGs before and after the artifact removal. These values have subsequently been averaged and shown in Fig. 5.8. Furthermore, their corresponding standard deviations have also been reported. As expected, the CC values have been significantly decreased by using the proposed method. Simulations for 20 EEG measurements demonstrate that the proposed method can efficiently identify and remove the EB artifact from the raw EEG measurements.

As a second criterion for measuring the performance of the overall sys-
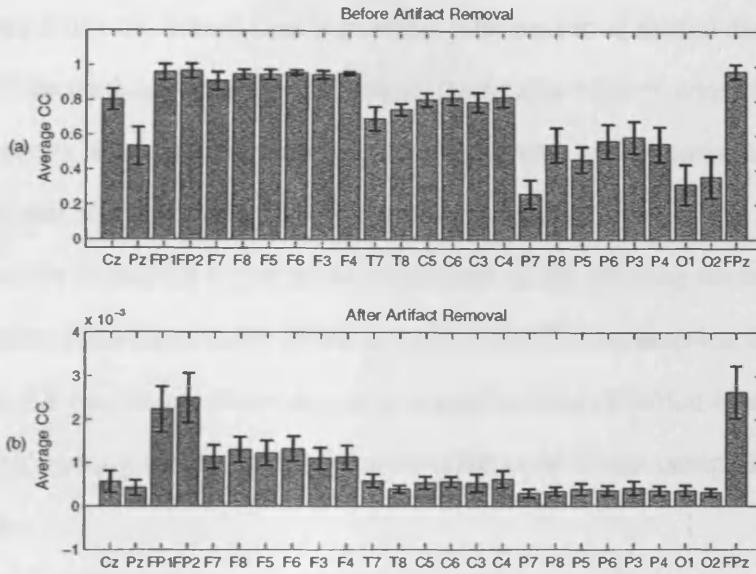
Figure 5.8. The averaged *CC* values (and their corresponding standard deviations) between the extracted EB and the restored EEGs before and after artifact removal of different channels in (a) and (b), respectively. The experiments have been performed for 20 different EB contaminated EEG recordings. Note that the scales are different by 10^3.

tem, a segment of EEG was selected, called x_{seg} and the reconstructed EEG $\hat{x}_{seg}$ which does not contain any artifact, and measured the waveform similarity by

$$\eta_{dB} = 10 \log \left[\frac{1}{M} \sum_{i=1}^{M} \left(1 - E\{x_{seg}(i) - \hat{x}_{seg}(i)\} \right) \right]. \tag{5.3.1}$$

When the value of η_{dB} is zero, the original and reconstructed waveforms are identical. From the 20 sets of EEGs, the average waveform similarity was as low as $\eta_{dB} = 0.008$ dB (standard deviation 10^{-3} dB). These results suggest that the observations have been faithfully reconstructed.

5.4 Concluding Remarks

A robust method for removing EOG from EEG recordings by employing the robust minimum variance beamforming method was presented to allow for the deviation of the estimate of the steering vector corresponding to the

EB source from the actual steering vector. The vector of spatial distribution of the EB factor has been identified using the proposed four-way PARAFAC which enjoys much less computational complexity in comparison with the conventional STF modeling using the three-way PARAFAC [76]. The spatial signature of the EB factor as an estimation of the steering vector that introduces the EB source to the EEGs is exploited. This assumption is rational since the EB can be considered as a strong point source which is attenuated while propagating from the frontal area to the central and occipital parts of the brain.

The approach can be also implemented in the conventional paradigms of adaptive training of a global steering vector from the training data set and use it in removing the artifacts from the test data. However, there are three important issues in implementing the adaptive machine learning schemes for artifact removal as detailed below.

1. EB artifacts can be very different in terms of the amplitude and how they contaminate other channel signals; they may contaminate the EEG recordings from the frontal electrodes or nearly all the recordings even those recorded from the electrodes in the occipital area. Due to these diverse artifact strengths, one faces with different steering vectors. This diversity makes the learning of the optimum steering vector from the training set rather difficult and the training procedure may suffer from a poor generalization while implementing on the test data.

2. Although this method has been tested only for EB artifact removal, its potential for removal of the eye-movements (vertical/lateral movements) and saccade artifacts may be investigated. Even if, it is possible to identify a general steering vector for EB artifacts, for the

eye-movement artifacts, especially the lateral ones, the corresponding steering vectors show very different patterns of inter-session and intra-subject variability. Therefore, one may not find a single steering vector for removal of the eye lateral movement artifacts. In removing saccade artifacts, the situation can be worse depending on the angular speed of the eye which may reach up to 1000 degree per second and also the temporal pattern of saccade which lasts to approximately 200 milliseconds. Thus, the method has been developed for EB artifact removal on a trial by trial basis. Note that the problem of the EB and eye-movement artifact removal has been well solved in [69]. However, this approach would not be effectively applicable without using the extra EOG electrodes.

3. The online implementation of the presented method is very easy. As shown in Table 5.3, the estimation of the STF-TS model is fairly straightforward. If the algorithm is expected to work in the recording session, i.e. in the clinical examinations and mainly for fast reviewing purposes, the STF-TS modeling can be only estimated for the first few segments and then it introduces, for instance, the averaged vector of the estimated steering vectors to the robust beamformer. The beamformer will relatively compensate for the deviations of averaged steering vector of the recent EBs from that of the new EBs and extracts the artifact. This approach can also be regarded as a learning paradigm where the learning process is simply an averaging operation. However, in the offline analysis, a steering vector for EB artifact is identified for each set of contaminated EEG recordings.

The results show that the proposed method extracts and removes the effect of blinking artifacts from EEGs. The EEGs are processed using the

RMVB algorithm and the artifact is extracted; then, the EEGs are reconstructed in a deflation stage. The proposed framework consistently removes the EB artifacts from the EEG signals.

Chapter 6

STEADY-STATE MOVEMENT RELATED POTENTIALS AND BCI

Considering the limitations of the presented BCI approach in Chapter 3 and the experiences achieved in Chapters 4 and 5, a new paradigm for BCI in STF domain is introduced in this chapter. Here, in contrast to other conventional approaches such as ERD/ERS analysis in the μ and β frequency bands, the brain capabilities in controlling repetitive finger movements are investigated for BCI. The neurological aspects of such repetitive movements are reviewed and explored for real EEG recordings both in averaged and single trial modes. Finally, the applicability of two classic classifiers, i.e. Fisher's discriminant analysis (FDA) and kernel FDA (KFD) for BCI purposes are investigated.

6.1 Introduction

Movement-related brain electrical activities have been studied for many years by means of the readiness potentials (RP) [36]. The RP is typically recorded during the performance of the voluntary movements. The con-

struction of the lateralized readiness potentials (LRP) exploits the contralaterally dominant distribution of movement-related brain activity preceding the movements [34]. It consists of a subtraction of potentials recorded ipsilateral to the side of movement from potentials recorded contralaterally, followed by averaging of these difference potentials associated with left and right hand movements [121]. The LRP has proven to be a useful tool in cognitive and clinical neuroscience [122, 123]. Most of the clinical researches on LRP, see [36], have investigated the temporal and spatial characteristics of this spectrally band limited potential in the healthy subjects and patients.

In the worldwide BCI community, there are several research groups that investigate RPs for BCI [15, 18]. As briefed in Chapter 3, the detection of asymmetric EEG potentials during temporally discrete finger tapping is straightforward. In [15, 18], various methodological and mathematical procedures have been tested and correct classification results of up to 95% have been achieved.

6.2   Lateralized Readiness Potentials (LRP)

Pioneering studies on Bereitschaftspotentials (RPs) such as in [124] showed that the brain activity can be detected prior to an overt movement and that the nature of this activity depends on the nature of the impending movement [36, Ch. 14]. The Bereitschaftspotentials begin several hundred milliseconds before the movement onset and as the time for the movement approaches, the scalp asymmetric distribution depends on the effector. This observation gave rise to the LRP measure. In [125], it has been shown that the time at which the brain activity becomes asymmetric is closely related to the time at which the subjects knew whether a right or left finger response would be required. Kutas et al. in [125] have concluded that the asymmetric

distribution of the scalp potentials reflects the mental preparation to execute a specific motor act.

Based on these findings, researches in [126] and [127] independently [36] reasoned that under certain circumstances, the presence of asymmetries could be used to infer the presence of preferential preparatory brain activity. The procedures to derive a measure of asymmetric movement-related brain activity were derived which yielded a measure now referred to as the LRP.[1]

6.2.1   Deriving the LRP

Both [126] and [127] recognized that the asymmetries in electrical brain activity can be observed for a variety of reasons, only some of which involve preparation for movement. The problem of isolating the movement-related contribution to the asymmetries was solved almost similarly in [126] and [127], see the two-stage subtraction and subtraction-averaging sequence methods in [128] and in [129], respectively. For instance, in the subtraction-averaging method, first, the trials are sorted into two groups, namely, left and right finger movement classes. Then, the EEGs from two lateral electrodes placed over left and right motor cortices ($C3$ and $C4$) are averaged separately for the two groups of trials. Refer to Fig. 6.1 for their current source density transformed traces. As shown in each plot, there is an asymmetry between the averages for $C3$ and $C4$ for both finger movements. That is, for the left finger movements, the averaged activity at $C4$ is more negative than that at $C3$ and the converse is the case for the right finger movement. This procedure may be continued by doing the further second subtraction, i.e. the subtraction of the asymmetry waveform for right finger movement from that of the left finger movement. The outcome of the second subtraction is

---

[1]It was originally called "Corrected Motor Asymmetry" in [127].
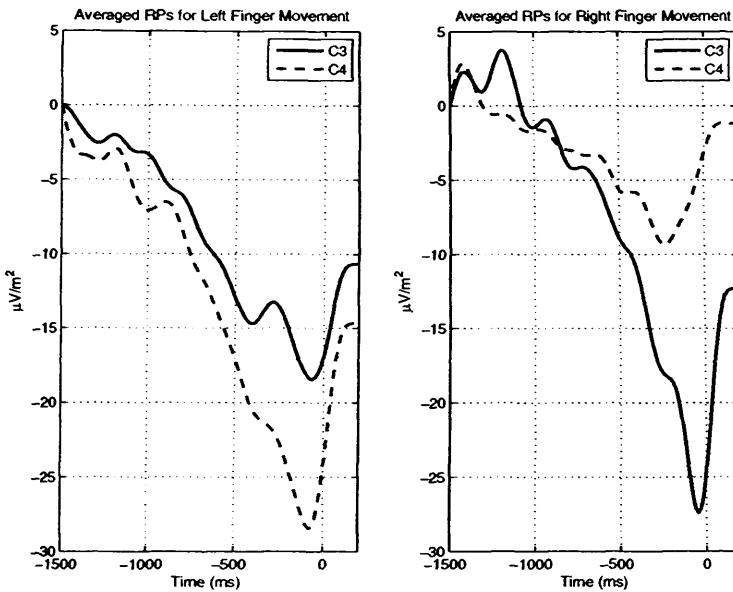
Figure 6.1. The averaged CSD transformed RPs from C3 and C4 during left and right finger movement. In left finger movement the signal at C4 is more negative than that at C3 and the converse is the case for right finger movement. The unit $\mu V/m^2$ denotes the second order spatial derivative of the RPs computed by CSD transform.

of great importance in psychophysiologic studies, see [130] and references therein. However, in BCI investigations the second stage is often relaxed.[2]

Figs. 6.2 and 6.3 respectively show the LRPs computed for the left and right finger movements by using the subtraction-averaging method. Note that, in these two figures, the deference between the average of RPs recorded from an electrode, C3, and the averaged RPs recorded from the corresponding electrode on the other hemisphere, C4, is denoted by C3 − C4. For instance in Fig. 6.2, a positive trend can be seen in the central subplot which demonstrates that the averaged RPs recorded from C4 is more negative than that from C3. LRPs in other subplots also demonstrate similar patterns. Fig. 6.3 reports the asymmetric distribution of RPs during right finger movement.

---

[2]Refer to [36, Ch. 14] for further details.
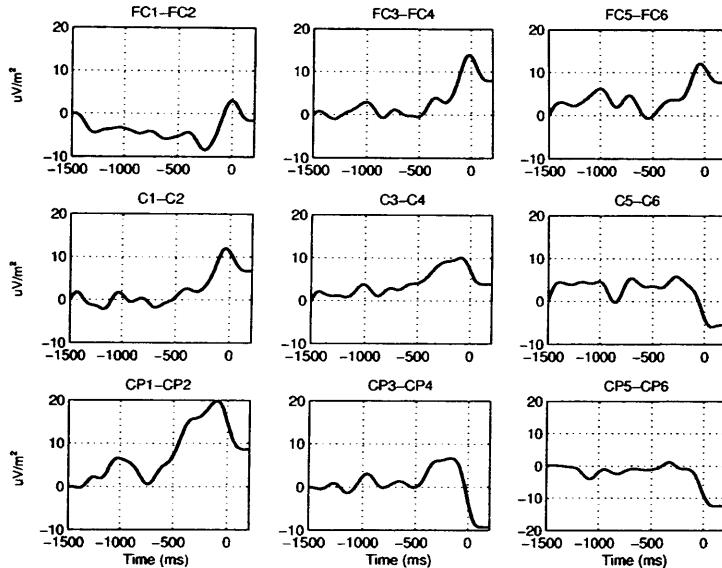
Figure 6.2. In left finger movement; the deference between the average of RPs recorded from electrodes on the left and right hemispheres is denoted for instance by $C3 - C4$. Note that in the central subplot, i.e. $C3 - C4$, a positive trend is seen which demonstrates that the RP recorded from $C4$ has been more negative than that of $C3$. LRPs in other subplots also demonstrate the same pattern although in $CP5-CP6$ the LRP violates the general pattern.

6.2.2 BCI using LRP

Among various BCI approaches, it is fair to state that the BCI based on LRP has been shown to be the most effective one - merely in terms of classification results [15]. Notice that, it has been documented that RPs are readily recordable from almost all the subjects [36]. Therefore, LRP-based BCI seems to be a suitable option both in terms of detection and processing. However, in this research, it has been concluded that although similar to many published works, acceptable BCI performances can be achieved using LRPs, the applicability of such method in real life rehabilitation problems is questionable.

The EEGs shown in previous figures were recorded from one subject in
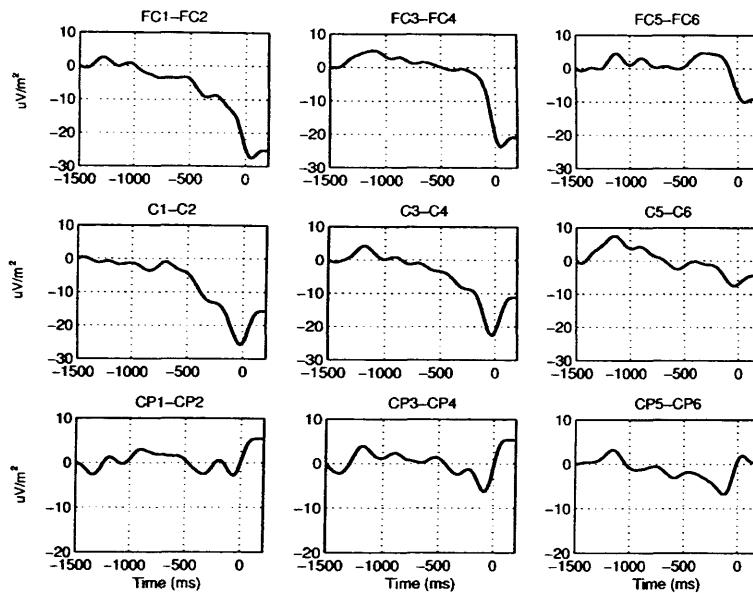
Figure 6.3. In right finger movement; the deference between the average of RPs recorded from electrodes on the left and right hemispheres is denoted for instance by $C3 - C4$. Note that in the central subplot, i.e. $C3 - C4$ a negative trend is seen which states that the RP recorded from $C3$ has been more negative than that of $C4$. LRPs in other subplots consolidates almost similar decreasing pattern.

the Behavioral Brain Sciences Centre, School of Psychology, The University of Birmingham where 240 trials of self-paced finger movement were collected, sampling rate was set to 512 Hz. The subject was asked to sit in front of a screen. Each trial started with a 2 second blank gray screen when the subject was allowed to blink. Then, a fixation cross "+" appeared in the center of the screen. After one and half seconds, "+" was replaced with the go signal "X" which informed the subject that he was allowed to move his finger after some seconds.

Neurological studies have shown [122, 123] that in order to have clear readiness potentials the interval between two subsequent motor actions should not be less that 5 seconds. Therefore, in contrast to [15], the subject was not asked to respond immediately after "X" appeared. He was instructed to carry out the the motor task after some seconds, for instance 5 seconds.
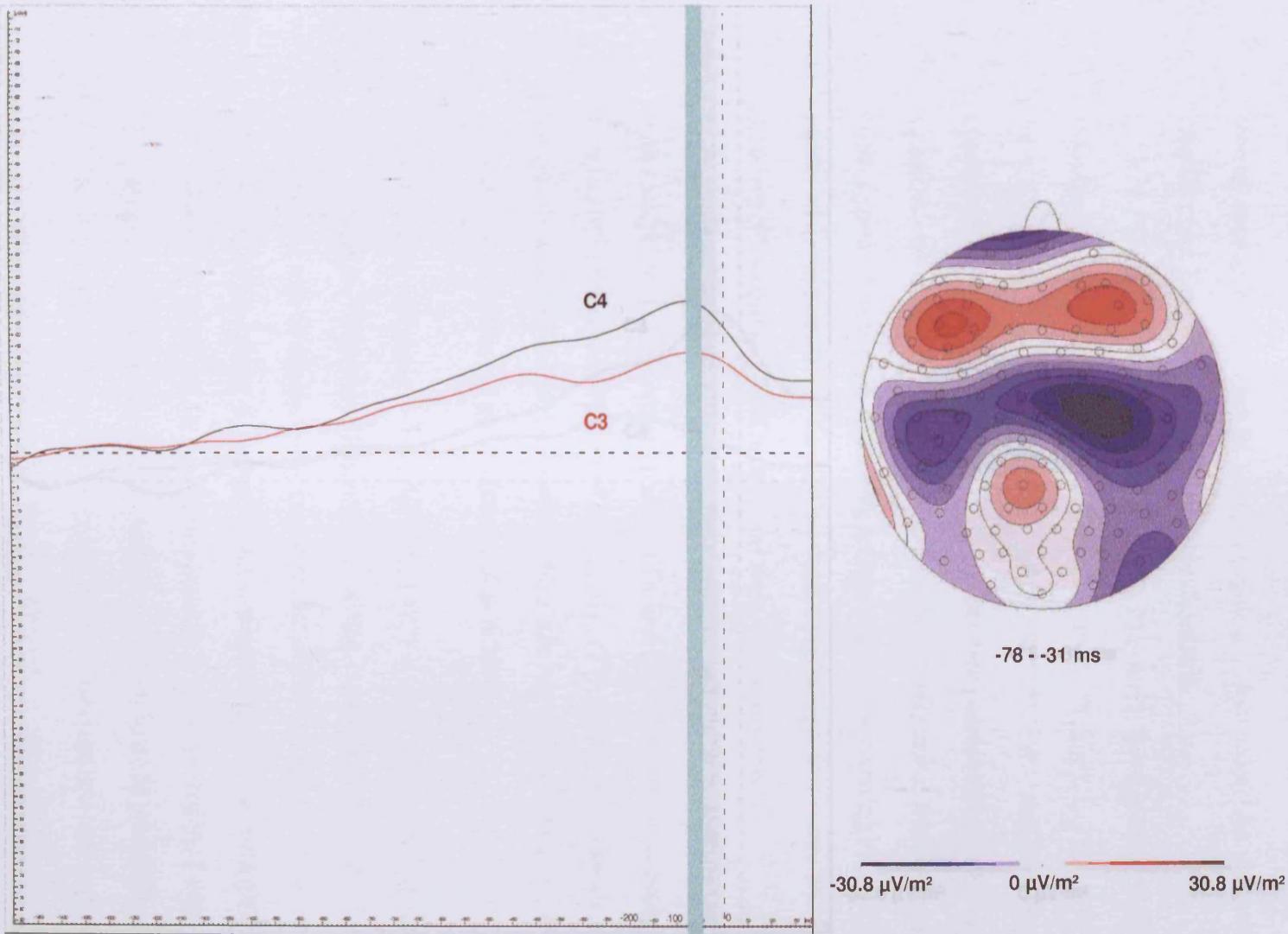
Figure 6.4. Topographic maps of the LRPs prior to the left finger movement. The averaged movement time instant (-78 ms to -31 ms) is shown by green bar.

C3

C4

-78 - -31 ms

-32.6 μV/m²   0 μV/m²   32.6 μV/m²
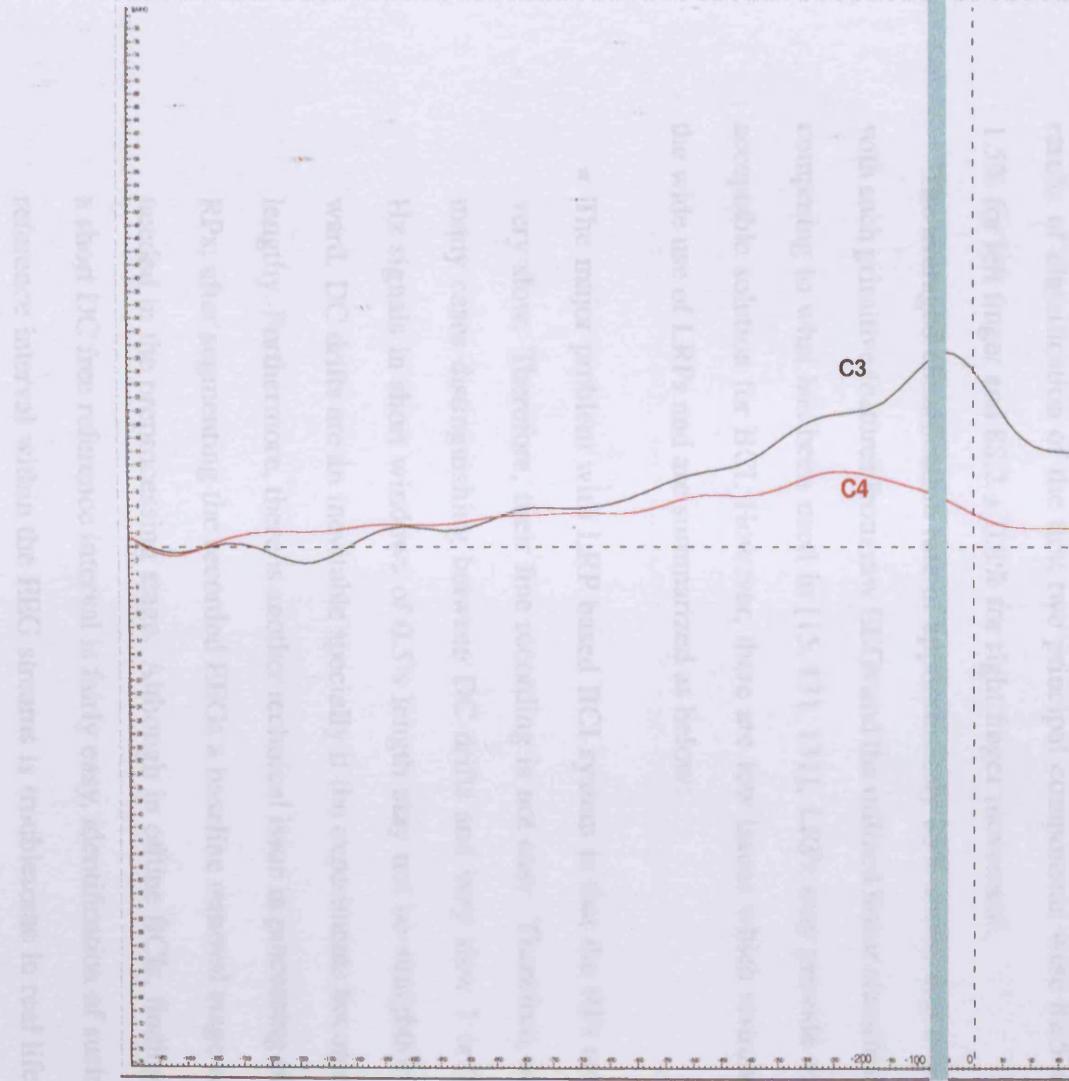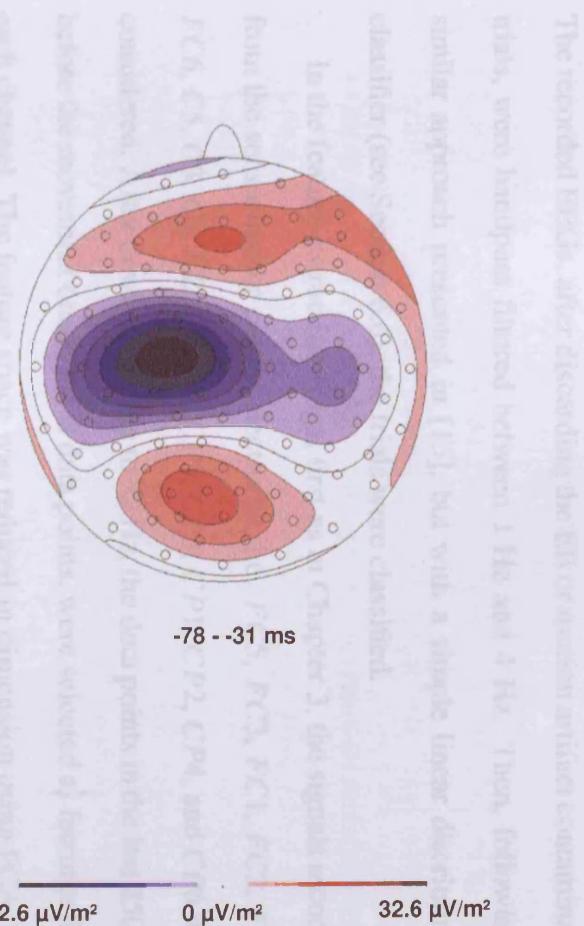
Figure 6.5. Topographic maps of the LRPs prior to the right finger movement. The averaged movement time instant (-78 ms to -31 ms) is shown by green bar.

The recorded EEGs, after discarding the EB or motion artifact contaminated trials, were bandpass filtered between 1 Hz and 4 Hz. Then, following a similar approach presented in [15], but with a simple linear discriminant classifier (see Section 6.6), the trials were classified.

In the feature extraction stage, first as in Chapter 3, the signals recorded from the sensorimotor cortex electrodes, i.e. $FC5$, $FC3$, $FC1$, $FC2$, $FC4$, $FC6$, $C5$, $C3$, $C1$, $C2$, $C4$, $C6$, $CP5$, $CP3$, $CP1$, $CP2$, $CP4$, and $CP6$ were considered. After down-sampling to 32 Hz, the data points in the last 250 ms before the movement onset, i.e. 8 data points, were selected as features from each channel. The feature space was reduced in dimension using PCA. The results of classification of the first two principal components were $82.5 \pm 1.5\%$ for left finger and $88.2 \pm 0.8\%$ for right finger movement.

The averaged classification rate of approximately 85% shows that even with such primitive features from raw EEGs and the utilized linear classifier, comparing to what has been used in [15, 131, 131], LRPs may provide an acceptable solution for BCI. However, there are few issues which restrain the wide use of LRPs and are summarized as below:

- The major problem with LRP based BCI system is that the RPs are very slow. Therefore, their fine recording is not easy. Therefore, in many cases distinguishing between DC drifts and very slow 1 or 2 Hz signals in short windows of 0.5 s length may not be straightforward. DC drifts are an inevitable specially if the experiments become lengthy. Furthermore, there is another technical issue in processing of RPs; after segmenting the recorded EEGs a baseline removal stage is needed in the preprocessing stage. Although in offline BCIs, finding a short DC free reference interval is fairly easy, identification of such reference interval within the EEG streams is troublesome in real life

applications.

- The second problem arises from the nature of the RPs. In order to record clear RPs, the interval between each two finger movements should be several seconds. Although in few literature such as [15], in order to realize a fast LRP-based BCI machine this interval has been reduced to 0.5 s, the classification performance has been significantly degraded. In order to have an output for each single subject's decision, the subject has to fixate for at least a couple of seconds which can be stressful.

- Another problem with LRP-based BCI is that, EEGs are markedly affected by the low frequency motion and EOG artifacts, within the range of 2 to 6 Hz. Therefore, in contrast to the μ rhythm-based BCI systems[3] in which even EB contaminated EEGs trials can be included in the classification stage, all the noisy segments should be discarded.

Noticing the aforementioned reasons, a BCI based on the brain steady-state movement related potentials (ssMRP) is proposed here. This approach would solve almost all the above problems and provide high classification rates. It does not burden much computational load and therefore the real-time implementation is plausible. By considering the steady-state movement related rhythms there is almost no need to remove the baseline and the EB artifacts do not interfere much.

---

[3]As defined in Chapter 2, μ rhythms are movement related brain potentials which lie in the spectral range of 8-13 Hz.

6.3 Brain Steady-State Potentials for BCI

The main issue in all previous BCI systems is to detect and classify the single trial EEG recordings. That is, the subject is asked to move (imagine the movement of) a finger and the recorded EEGs corresponding to that movement is classified by the BCI machine. However, the number of decisions per second is not high enough in such BCI structure as compared to normal brain performance. For instance, as reported in [15], in the classic ERD/ERS based machines, the best achieved performance is approximately 85% when taps are carried out every 2 s. In order to extract clear and interpretable ERD/ERS the inter tap interval should be several seconds. Therefore, the error in single trial ERD/ERS classification increases if taps are carried out at faster paces. Results in [15] quantified that the error would increase up to 27% when the subject taps every 0.5 s. Further results in [15] demonstrated that if the LRP features are used instead of the ERD/ERS features, a similar pattern could be seen; the classification error increased from 5% to 19% when inter tap interval decreased from 2 s to 0.5 s.

In this section, the conventional BCI systems based on steady-state brain potentials are reviewed. In such systems, the subject is exposed to a repetitive cue, i.e. a rhythmic flashing LED or audio beeps, instead of a single cue in each recording trial.

BCI systems based on steady-state visual evoked potentials (ssVEP) have become increasingly important in the past few years [132, 133]. In ssVEP-based BCI, the signal processing unit estimates the spectrum of the recorded EEGs, usually from two or three electrodes in the occipital area. The pattern classification unit labels the recordings to various classes. The classification is based on finding the strongest peak in the spectrum and assigning that trial to the class of the flashing cue with the same frequency.

Middendorf *et al.* in [134] and [135] used such BCI method for controlling a flight simulator. In that study, the average accuracies up to 92% were achieved with a decision time of 2.1 s. In another recent work, Cheng *et al.* [136] demonstrated a 12-class ssVEP-based BCI system. By ssVEP-based system, multiple classes BCI systems may be realized without extensive subject training. This is mainly because that the users do not have to concentrate on simulation of different motor actions and only have to shift the gaze toward the cue in the movement direction of interest which less demanding, though it needs intact eye muscle control. Moreover, real-time implementation of such BCI system with visual feedback is readily possible [137].

However, since ssVEPs are directly coupled to eye gaze position, the BCI would be limited only to recovering eye position on direction of attention. The objective here is therefore, to introduce a high performance BCI machine based on brain steady-state finger *movement* related potentials (ssMRP), independent of eye-movement.

6.4 Brain's MRPs are elicited during repetitive finger movements?

Timing of one's voluntary movements with respect to an external event has been investigated in sensorimotor timing studies [138–140]. In several studies the subjects synchronize their finger taps with the regular auditory pacing signals [138–140]. Interestingly, the subjects usually feel the exact synchrony between the taps and the pacer. In contrary, the taps lead over the pacer by approximately 20-60 ms which is called the negative asynchrony and is a stable behavioral phenomenon [141].

Neuroelectric and neuromagnetic investigations have demonstrated that the execution of simple unimanual repetitive finger movements is associ-

ated with brain activities within the Rolandic fissure of the contralateral hemisphere corresponding to the primary sensorimotor cortex called ssM-RPs [142, 143]. This activity is temporally locked to the movement onset [142, 143]. It has been shown that [142, 143] the activity within the primary sensorimotor cortex may be decomposed into three marginally different components. First, approximately 100 ms before movement onset, a localized activity in the primary motor cortex (M1) occurs. Further activity arises around the tap onset [144]. The corresponding source is localized within the primary somatosensory cortex (S1) and may represent the neuromagnetic correlate of feedback due to finger movements. However, this source could not be detected by other studies which might be simply due to slightly different analysis procedures from that described in [144]. Finally, the post-movement activity is detected approximately 100 ms after the movement onset. MEG studies reveal that the corresponding dipole is localized within the primary somatosensory cortex but inferior to the first S1 source [143].

6.4.1   How ssMRP may be used for BCI?

Unfortunately, the above mentioned recorded ssMRPs does not provide much higher performance comparing to the classic EEG features for BCI. Again, the low SNR is a major problem and there exists spatial and spectral uncertainties similar to what mentioned in the previous chapters. Therefore, in this section an approach is proposed in order to effectively record ssMRPs and utilize them in a BCI machine. Note that ssMRP-based BCI generally follow the standards of ssVEP-based BCI. However, the main difference is that the ssMRPs are movement related potentials.

The main idea here is that if the subject moves (or imagines the move-

ment of) his/her fingers in synchrony to a flashing rhythm, EEGs are modulated with the similar frequency as of the flashing cue. The synchronization is stronger on the contralateral hemisphere to the moving finger. That is, if the frequency of the flashing cue is set to 2 Hz and the subject repetitively moves his right index with a 2 Hz pace, EEGs with considerable 2 Hz frequency components may be recorded on the contralateral hemisphere, for instance from $C3$ electrode. Similar procedure may be followed for the left index movement. First, the averaged EEGs over approximately 100 trials are investigated in order to observe if a relatively strong 2 Hz frequency component exist. Then, the single trial EEGs are classified. Interestingly, almost all the methods and applications developed for BCI using ssVEP, see [137] and the references therein, may be considered for this novel scheme.

The introduced approach is novel mainly due to the following issues:

- Recording and interpretation of ssMRPs is much easier than processing the RPs. In [15, 131, 145], several *ad hoc* and probably non-generalizable approaches methods for RP analysis have been introduced. Moreover, although in [15], Blankertz *et al.* investigated the effect of fast tapping, they have considered each tap individually in contrast to the proposed continuous framework.

- In contrast to the ssVEPs-based BCI, the proposed approach is much easier to implement since instead of having two or four almost close frequencies in the α band, i.e. 10, 11, 12, or 13 Hz, for different motions, only one frequency is utilized. In other words, instead of exploiting the spectral disparity in the signal processing unit of the BCI machine, one frequency has been considered but the spatial signature of the recorded signal refer to the subject's desired output. The sim-

pler coordinates of the proposed system make the training procedure much faster.

- In terms of computational complexity, also the proposed BCI approach is pretty straightforward. All one has to do is to estimate the variances of the bandpass filtered EEGs and classify them into two classes.

In the next section, the recording procedure and the preprocessing stage for investigating the averaged EEGs recorded during repetitive right and left finger movements will be described. Afterwards, the utilized classification techniques for single trial EEG analysis for BCI purposes will be covered.

6.5   EEG Recording and Pre-Processing

two right-handed healthy individuals participated in the experiment; both gave informed consent. No one had any previous BCI experience.

EEG recording and pre-processing: The experiment was run in a quiet, normally illuminated room. The participants were seated comfortably in an armchair with the forearms placed on the armrests of the chair. Two force transducers were attached to the armrests, on top of which the participants hold their index fingers of each hand. The stimuli were presented in white against a grey background on a 17 inch monitor at a resolution of 800×600. The viewing distance was 100 cm.

Each subject first underwent a practice block of 20 trials. The main recording session was comprised of eight blocks, each contained 40 trials, resulting in 320 trials for further analysis. The block and trial structures are illustrated in Fig. 6.7 and Fig. 6.8. Each trial lasted 7 seconds which includes one second for initial fixation and another 6 seconds for EEG recording

during the motor task. In the first second of each trial a fixation cross was shown in the center of the screen. Afterward, while the cross was kept constant in the center, two flashing "X"s appeared at left and right sides of the cross for 6 seconds; each was 10 cm away from the center. Flashing frequency was set to 2 Hz. The participants were instructed to tap on force sensor under left or right index finger at a constant rate of 2 Hz synchronous to the flashing cues. The rest interval between trials was approximately one and half seconds, randomly changing so that the subjects would not guess the start of next trial. The choice between right or left finger tapping was made freely by the participants in each trial. However, they were asked to be fair between right and left responses.

The main reason for showing the flashing cues was to give the subjects a 2 Hz pace. Equidistant visual cues on either side from the center should not cause development of any asymmetric potentials over the motor cortex. The subjects were asked to maintain fixation on the central cross during the course of tapping. This approach was adopted in order to attenuate undesired ssVEPs. Force transducers were utilized instead of conventional response switches in order to provide a setup in which the subjects did not actually press any switch, only performed repetitive tapping, which maintained the continuity of the repetitive finger movement.

EEG potentials were recorded continuously with 128 active Ag/AgCl scalp electrodes, see Fig. 6.6, relative to an (off-line) averaged left and right mastoid reference. The electrodes were placed according to the 10 – 5 system [146], see Fig. 6.9 which has been taken from [147], using a carefully positioned nylon cap. The eye-movements and eye-blinks were monitored by bipolar horizontal and vertical electro-oculogram (EOG) derivations. EEG and EOG signals were amplified with a bandpass of 0-128
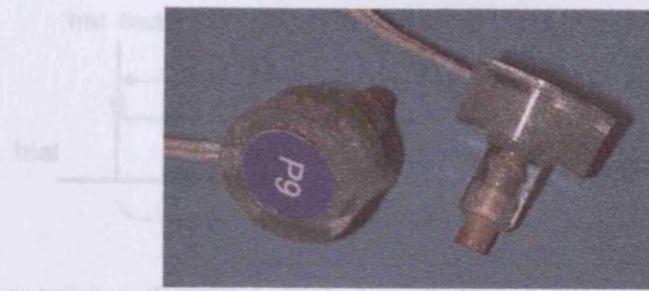
Figure 6.6. Active electrodes used for EEG measurement.

Hz using BioSemi Active-Two amplifier, and sampled at 512 Hz. EEG preprocessing was performed off-line using BrainVision Analyzer software (Brain Products GmbH). Continuous EEG recordings were off-line segmented in epochs from 0-6 s after trial onset. Individual trials containing eye-movement and eye-blink artifacts were rejected before analysis (on average 7% from each subject). Preprocessed EEG segments were later exported to Matlab for further analysis, i.e. single trial feature extraction and classification.

Fig. 6.7 shows the structure of the recording protocol where the experiment consists of 8 sessions. In each session, 40 trials are recorded; 320 trials in total. Each trial, see Fig. 6.8 starts with 7 seconds for recording the EEG and approximately one and half seconds between trial rest period. The length of rest block has considered randomly changing so that the subjects may not guess the start of the next trial. The subjects are allowed to blink during the rest period. This approach has been tested and shown very effective in cognitive neuroscience studies, see [122].

The flat type active electrodes produced by BioSemi used in this project are easily attachable to the skin with paste. As mentioned earlier in section 6.2, this feature is important for recording the slow cortical potentials when EEGs can be very easily contaminated by very low frequency drifts [148]. And secondly, the exploitation of the force transducers with a
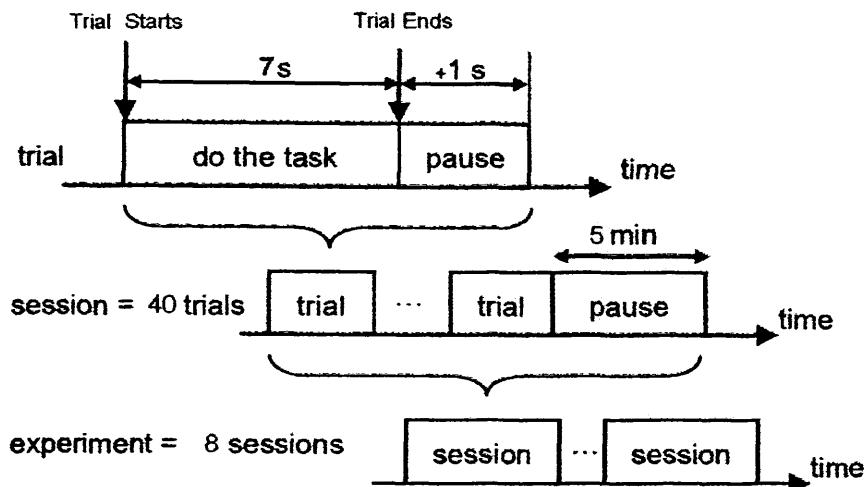
Figure 6.7. Temporal structure of the protocol.

Each Trial - 7 seconds		
1 s Fixation +	6 s Fixation and Tapping	
	X + X	

Figure 6.8. Temporal structure of each trial.

miniature precision strain gage amplifier, see Fig. 6.10, instead of the conventionally used response switches, Fig. 6.11.   The signals produced by the force transducers which make the strain gage active is processed by the BioSemi ActiveTwo AD-box similar to other active sensors. The advantages are that all influences of the connection cable including cable interference and temperature drifts are entirely eliminated. The Active strain gage contains a low-noise, low-power, zero-drift, differential amplifier.

The main motivation behind using this type of force transducers for ssMRP-based BCI was that since the subject does not actually press the button the continuity of the repetitive finger movement is maintained. Notice that force transducers have been widely used in neuroscience studies [138, 149] but not in BCI.
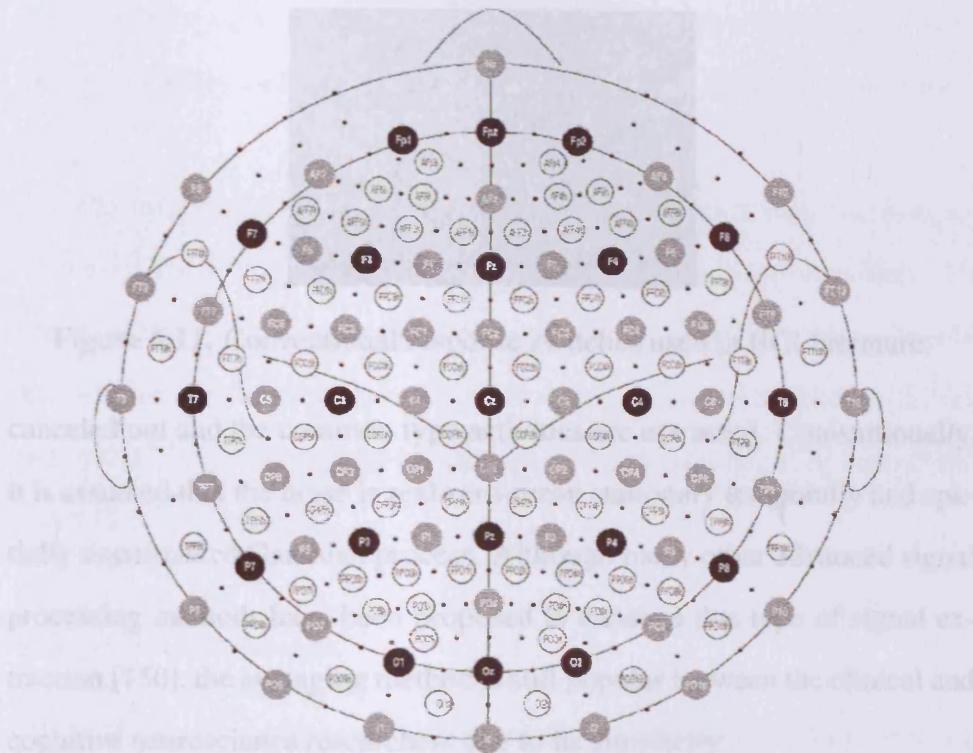
Figure 6.9. The 10 − 5 system. Total number of points is 329 including 12 points, likely lying on the eyes (shown in gray italics). Black open circles indicate the 10 − 20 positions, gray open circles indicate additional positions introduced in the 10 − 10 system. This figure has been adopted from [147].
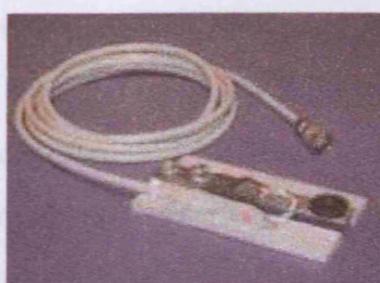
Figure 6.10. The force transducers utilized in recording the ssMRP.

6.5.1   Topographic Analysis of the Averaged EEG Recordings

Conventionally, in EEG signal analysis community before single trial analysis, averaged EEGs over multiple trials are investigated. This will help better visualization of the phenomenon of interest in the time domain due to the fact that by averaging, irrelevant brain potentials will be automatically

Figure 6.11. Conventional response switches used in BCI literature.

canceled out and the common type activities are extracted. Conventionally, it is assumed that the noise is real zero-mean stationary temporally and spatially uncorrelated Gaussian process. Although many other advanced signal processing methods have been proposed to enhance this type of signal extraction [150], the averaging method is still popular between the clinical and cognitive neuroscience researchers due to its simplicity.

The averaged EEGs recorded from the first subject during repetitive left and right finger movement have been shown respectively in Fig. 6.12 and Fig. 6.13. The subject performed left finger movements in 148 trials and in another 172 trials right finger has been moved. Notice that due to the common artifacts, i.e. EBs and motion artifacts, approximately 7% of the trials have been discarded from the averaging procedure.

Of primary interest was the steady-state movement-related potential developing by rhythmic tapping. Therefore, averaged bandpass filtered (1.5-2.5 Hz) EEGs recorded during repetitive left and right finger movement trials were used to visualize ssMRPs in the time domain. The respective topographic maps, Fig. 6.12 and Fig. 6.13, show rapid development of a lateralized signal over contralateral sensorimotor cortex whose polarity alternates every 250 ms, i.e. at 2 Hz during left and right index rhythmic movements. In the left index movement case, see Fig. 6.12, a strong 2 Hz rhythm is observed on the contralateral hemisphere. In Fig. 6.13, the aver-

aged ssMRPs for right finger movements are depicted where a rapid development of a lateralized rhythmic signal over the left hemisphere is evident.

The results of the averaged EEG analysis have been very encouraging. However, for BCI purposes, single trial EEGs should be classified. For stepping into the ultimate stage of a simple BCI machine, i.e. classifier design, first the theoretical bases of two conventional classifiers have been presented. The results of the single trial EEG classification for each classifier will be presented.

6.6 Classifier Design

For BCI purposes, designing reliable while fast and simple classifiers is of great importance. Conventionally, the linear discriminant analysis (LDA) based on the Fisher ratio [49], is utilized for BCI, see [18]. First in this section, the basic mathematics of the Fisher's LDA (FDA) is reviewed. Kernel Fisher discriminant (KFD) [151] is also another successful approach to classification where a kernel transformed version of the LDA is taken into account.

The comparison of different classifiers is not of the main interest of the next sections. Therefore, any classifier may be replaced with the utilized ones. Here, the researcher aims at showing that the performance of the proposed BCI scheme is mainly due to its classifiable feature vectors and not the classifiers. Evidently, higher classification rates may be achieved by using advanced classifiers.
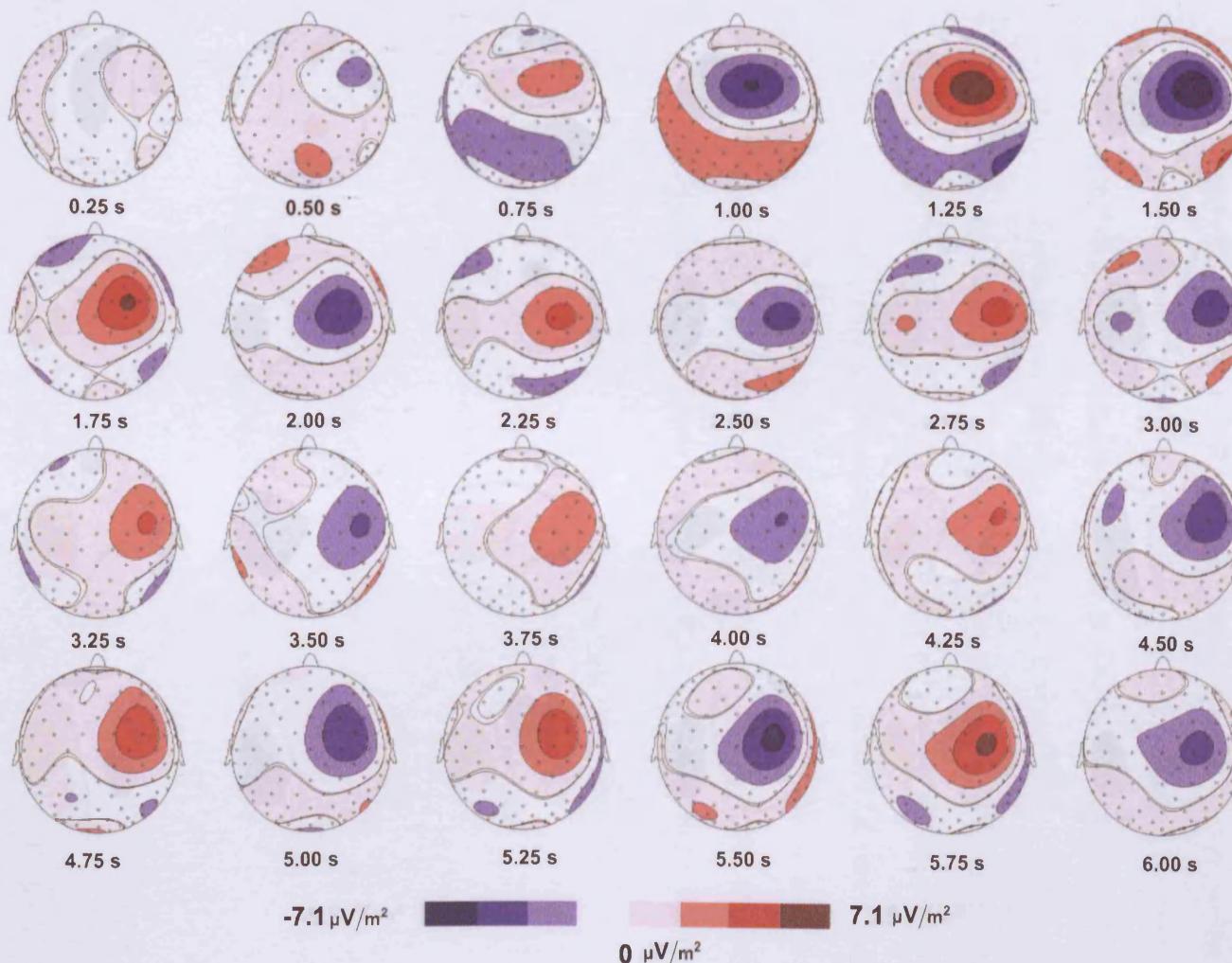
Figure 6.12. Averaged pre-processed EEGs during repetitive left finger movement for a single representative participant. Topographical maps have been depicted in consecutive 0.25 s time windows. The top-left map illustrates the averaged EEGs over 0 and 0.25 seconds time window and the bottom right ones present those of the last 0.25 seconds window, i.e 5.75 to 6 seconds. Notice the rapid development of the lateralized 2 Hz signal on the contralateral left hemisphere.
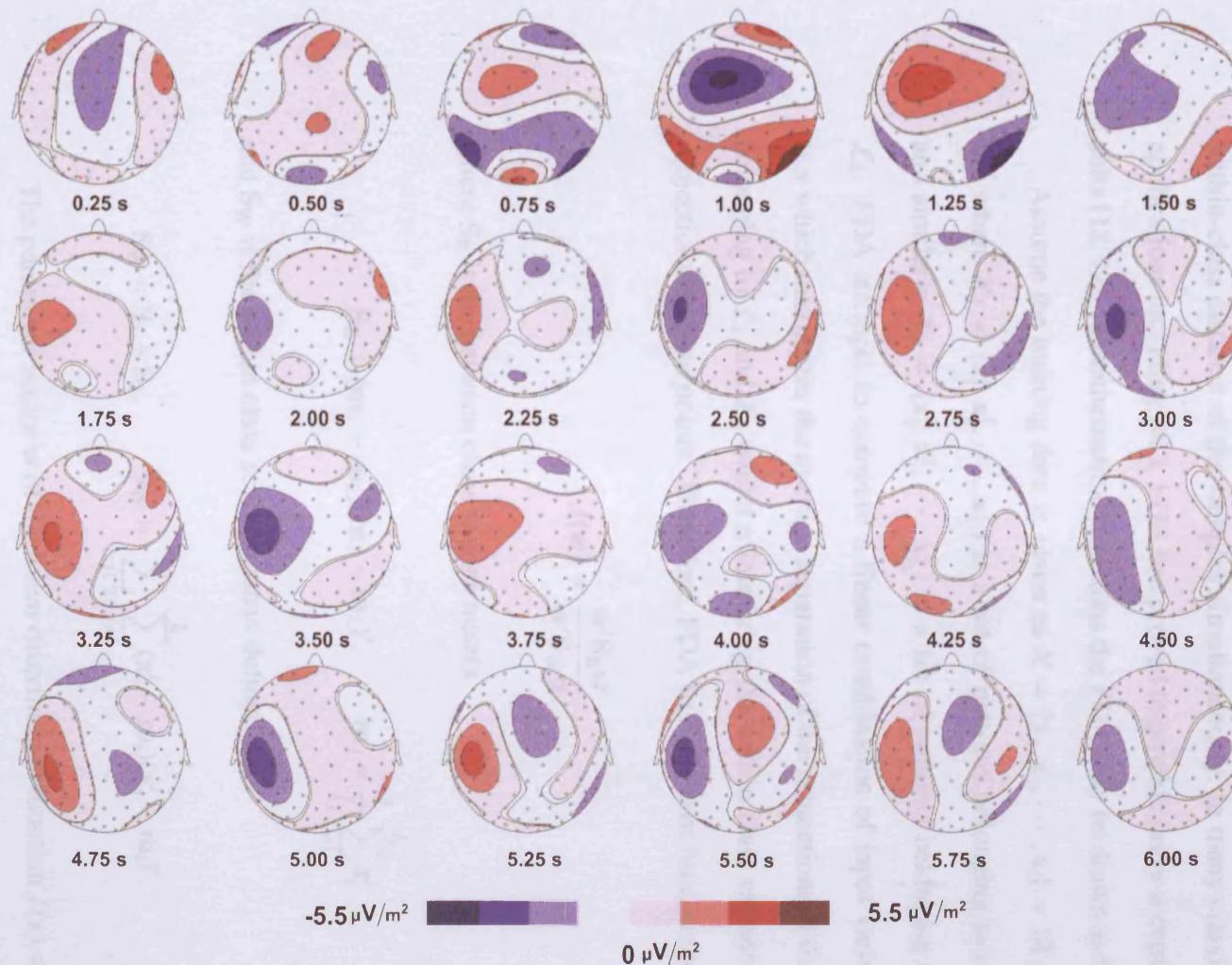
Figure 6.13. Averaged pre-processed EEGs during repetitive right finger movement for a single representative participant. Topographical maps have been depicted in consecutive 0.25 s time windows. The top-left map illustrates the averaged EEGs over 0 and 0.25 seconds time window and the bottom right ones present those of the last 0.25 seconds window, i.e 5.75 to 6 seconds. Notice the rapid development of the lateralized 2 Hz signal on the contralateral right hemisphere.

6.6.1 Fisher Linear Discriminant Analysis

Fisher linear discriminant [49] is a conventional classifier in which the main motive is to maximize the between-class distance while minimizing the within-class distance of the samples simultaneously. In many classification applications including BCI, FDA has proved to provide fairly acceptable results [18,49]. In mathematical notations the FDA may be drawn as follows:

Assume the training data is given as $\mathcal{X} = \{x_1, x_2, \cdots, x_l\} = \{\mathcal{X}_1, \mathcal{X}_2\} \subset \mathbf{R}^N$ where $\mathcal{X}_1 = \{x_1^1, x_2^1, \cdots, x_{l_1}^1\}$ is a set of patterns belonging to class $\mathcal{L}_1$ and similarly $\mathcal{X}_2 = \{x_1^2, x_2^2, \cdots, x_{l_2}^2\}$ is a set of patterns belonging to class $\mathcal{L}_2$. FDA attempts to compute a linear combination of input variables as w.x which maximizes the average separation of the projections of the points belonging to $\mathcal{L}_1$ and $\mathcal{L}_2$ whilst minimizing the within class variance of the projections to those points. Therefore, FDA uses the cost function below

$$J(\mathbf{w}) = \frac{\mathbf{w'S_Bw}}{\mathbf{w'S_Ww}} \qquad (6.6.1)$$

where $\mathbf{S_B}$ is the between class scatter matrix

$$\mathbf{S_B} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)', \qquad \mathbf{m}_j = \frac{1}{l_j}\sum_{i=1}^{l_j} \mathbf{x}_i^j \qquad (6.6.2)$$

and $\mathbf{S_W}$ is the within class scatter matrix defined as

$$\mathbf{S_W} = \mathbf{S}_1 + \mathbf{S}_2, \qquad \mathbf{S_W} = \sum_{i \in \{1,2\}} \sum_{j=1}^{l_i} (\mathbf{x}_j^i - \mathbf{m}_i)(\mathbf{x}_j^i - \mathbf{m}_i)'. \qquad (6.6.3)$$

The parameter vector w of the linear discriminant function $f(\mathbf{x}) = \langle \mathbf{w.x} \rangle + b$, where $\langle . \rangle$ denotes the inner product operator, is determined in order to

maximize the class separability criterion (6.6.1).

$$\mathbf{w} = \arg\max_{\mathbf{w}'} J(\mathbf{w}') = \arg\max_{\mathbf{w}'} \frac{\langle \mathbf{w}' . \mathbf{S_B}\mathbf{w}' \rangle}{\langle \mathbf{w}' . \mathbf{S_W}\mathbf{w}' \rangle}. \qquad (6.6.4)$$

The bias b of the linear rule is determined as

$$\langle \mathbf{w}.\mathbf{m}_1 \rangle + b = -(\langle \mathbf{w}.\mathbf{m}_2 \rangle + b). \qquad (6.6.5)$$

The classical solution of the problem in (6.6.4) is

$$\mathbf{w} = \mathbf{S_W}^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \qquad (6.6.6)$$

Also, (6.6.4) can be reformulated in the quadratic programming (QP) framework as

$$\mathbf{w} = \arg\min_{\mathbf{w}'} \langle \mathbf{w}' . \mathbf{S_W}\mathbf{w}' \rangle,$$

subject to

$$\langle \mathbf{w}'.(\mathbf{m}_1 - \mathbf{m}_2) \rangle = 2. \qquad (6.6.7)$$

This approach may be found useful specially when the matrix inversion in (6.6.6) is hard to compute.

6.6.2 Kernel Fisher Discriminant Analysis

In real-life classification problems, only linear discriminant analysis would not be enough. In [152], an advanced class of nonlinear FDAs using kernel tricks, called KFD, has been introduced. Extensive empirical comparisons [153] have shown that KFD is comparable to other kernel-based classifiers, such as support vector machines. The kernel Fisher discriminant which is the nonlinear extension of the linear FDA is summarized as fol-

lows:

Fisher's linear discriminant is a fixed feature space $\mathcal{F}(\phi : \mathcal{X} \rightarrow \mathcal{F})$, induced by a positive definite *Mercer* kernel $\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbf{R}$ where $\mathcal{K}(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}).\phi(\mathbf{x}')$ [154]. The kernel matrices for the entire data set, **K**, and for each class, i.e. $\mathbf{K}_1$ and $\mathbf{K}_2$ are defined as

$$\mathbf{K} = [k_{ij} = \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^l \qquad (6.6.8)$$

and

$$\mathbf{K}_i = [k_{jk}^i = \mathcal{K}(\mathbf{x}_j, \mathbf{x}_k^i)]_{j,k=1}^{j=l,k=l_i} \qquad (6.6.9)$$

The reproducing kernel theory [155] indicates that **w** may be written in an expanded form as

$$\mathbf{w} = \sum_{i=1}^l \alpha_i \phi(\mathbf{x}_i). \qquad (6.6.10)$$

In this research, Gaussian kernel with a width of 1 is considered. The objective function in (6.6.1) may also be written such that the data $\mathbf{x} \in \mathcal{X}$ appear only within products, giving

$$J(\alpha) = \frac{\alpha' \mathbf{M} \alpha}{\alpha' \mathbf{N} \alpha}, \qquad (6.6.11)$$

where $\alpha = \{\alpha_i\}_{i=1}^l$ and $\mathbf{M} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)'$. In (6.6.11), $\mathbf{m}_i = \mathbf{K}_i \mathbf{u}_i$ in which $\mathbf{u}_i$ is a column vector containing l_i elements with a common value of l_i^{-1} and

$$\mathbf{N} = \sum_{i \in \{1,2\}} \mathbf{K}_i (\mathbf{I} - \mathbf{U}_i) \mathbf{K}_i'. \qquad (6.6.12)$$

I is the identity matrix and $\mathbf{U}_i$ is a matrix with all elements equal to l_i^{-1}. The coefficients α of expansion (6.6.11) are computed by the leading eigenvector (the eigenvector corresponding to the largest eigenvalue) of $\mathbf{N}^{-1}\mathbf{M}$. Note

that since **N** is likely to be singular, or at best ill-conditioned, a regularized solution, i.e. **N**$_\mu$, is usually considered as **N**$_\mu$ = **N** + μ**I**. Here, μ is the regularization constant. Finally, in the KFD method, i.e. $f(\mathbf{x}) = \mathbf{w}.\phi(\mathbf{x}) + b$, the bias b may be computed as [152, 153]

$$b = -\alpha \frac{l_1 \mathbf{m}_1 + l_2 \mathbf{m}_2}{l} \qquad (6.6.13)$$

In the next section, the two aforementioned classifiers are implemented in the BCI problem and their efficiencies are investigated.

6.7 Results

In this section, the classification results of the FDA and KFD classifiers for the BCI experiment are presented. These classifiers are exploited for classification of different spatial and temporal feature vectors. Five different combinations of electrodes have been considered. Moreover, the EEGs are divided temporally into ten segments in order to investigate the effect of elapsing time on the synchronization. Results are presented and using ssMRPs for BCI is evaluated.

For this analysis, although two advanced artifact removal methods have been developed in the previous chapters, the highly EB contaminated trials (about 7% of the trials) were discarded from the analysis and the remaining trials have been considered for the feature extraction and classification stages. EEGs recorded from the scalp electrodes have been finely bandpass filtered between 1.5-2.5 Hz and their energies are calculated and introduced to the classifiers.

For each classifier, various feature sets have been computed in order to effectively investigate the potentials of ssMRP-based BCI scheme. First,

(a) Selected 45 out of the   (b) Electrode subset I   (c) Electrode subset II
128 channels.

(d) Electrode subset III   (e) Electrode subset IV   (f) Electrode subset V
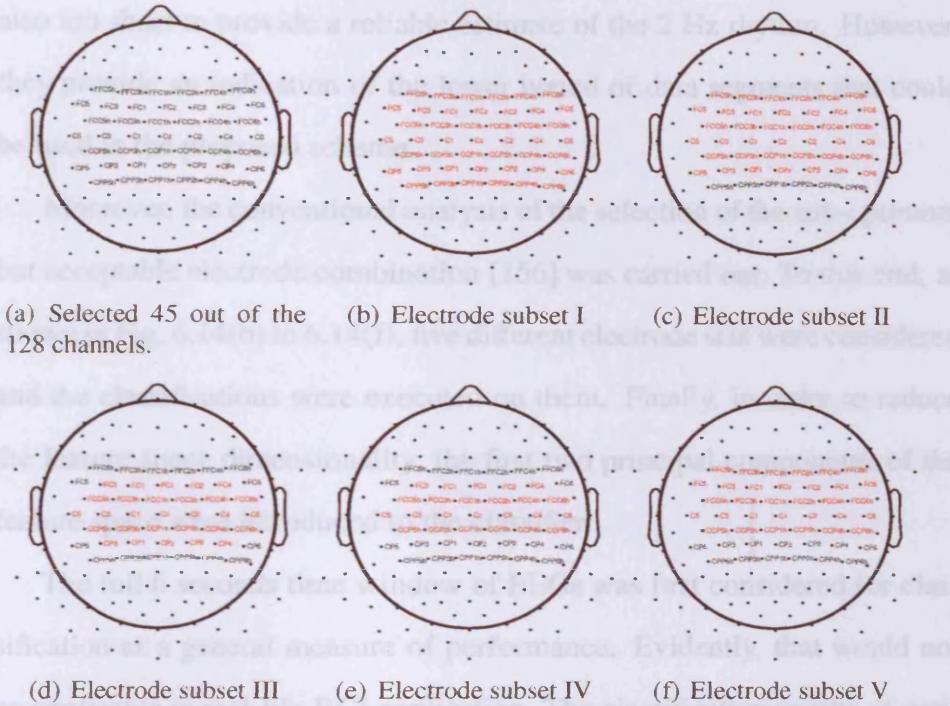
Figure 6.14. Electrode subset selection; (a) shows the primarily selected electrodes for the EEG recording and (b)-(f) demonstrate the five different considered electrode subsets.

EEGs are segmented into ten overlapping windows; comprising of three early windows, i.e. 0-0.5 s, 0-1.5 s, and 0-2.5 s and seven windows of 3 seconds. The latter seven windows have 83% temporal overlap, i.e. 0-3 s, 0.5-2.5 s, 1-4 s, 1.5-4.5 s, 2-5 s, 2.5-5.5 s, and finally 3-6 s. Then, the energy features are computed from multi-channel EEGs.

Note that the first three windows, namely 0-0.5 s, 0-1.5 s, and 0-2.5 s, were considered to investigate the approximate the time needed for the subjects to select and initiate tapping synchronous to the 2 Hz flashing stimulus. It is expected that a relatively poor classification performance would be achieved during the selection and initiation phases in each trial. This might be attributed to the arbitrary difference between the taping and visual cue appearance onsets. The performance would eventually increase after first few seconds. Note that first two time windows, i.e. 0-0.5 s and 0-1.5 s, are

also too short to provide a reliable estimate of the 2 Hz rhythm. However, they provide an indication of the lower bound of data segments that could be used in the proposed scheme.

Moreover, the conventional analysis of the selection of the sub-optimum but acceptable electrode combination [156] was carried out. To this end, as shown in Fig. 6.14(b) to 6.14(f), five different electrode sets were considered and the classifications were executed on them. Finally, in order to reduce the feature space dimensionality, the first two principal components of the feature space were introduced to the classifier.

The full 6 seconds time window of EEGs was first considered for classification as a general measure of performance. Evidently, that would not be applicable in real-life BCI application. The classification results of various shorter time windows and electrode combinations using FDA and KFD classifiers for subject one are summarized respectively in Table 6.1 and Table 6.2. Table 6.3 and Table 6.4 also report the classification results for subject two.[4] In addition, the performance of the KFD classifier was similar to that of FDA which implies linear separability of the feature vectors and suggests the computationally simpler linear classifier is adequate.

Robust classification results of above 88% and 96%, respectively, for the first and seconds subjects show the potentials of the method. Although here, ssMRP-based BCI has been reported only for two subjects, the outcome is consistent with the expectations. A closer look at Table 6.1 to Table 6.4 reveals that, first, the maximum classification rates for almost all the electrode selections are observed in the middle intervals, i.e. 1.5-4.5 seconds. The steady incremental trends of classification rates in the early segments of the trials and the decreasing trend in the late ones are also considerable

---

[4]In Tables 6.1 to 6.4, TW and EC denote stand for time window and electrode combination, respectively.

which might be interpreted as: in the very first seconds of each trial the subjects try to learn the 2 Hz pace rather than focusing on the movement. It would cause a slight degradation in their performances.[5] Therefore, the first one or two taps could be out of phase with the cue in some trials. It would consequently influence the separability of the extracted features.

Interestingly, in both subjects, the steady incremental trends of classification rates in early segments, 0-0.5 s, 0-1.5 s, and 0-2.5 s, are evident. In the very first seconds of each trial, the subjects attempt to adopt the correct 2 Hz pace which causes activities from areas of the brain other than the contralateral motor cortex resulting in slight degradation in BCI classification performance. Although the subjects had a short 5 minutes training block before actual recording, they still reported afterwards they had to attend to the pace or the onset of each trial.

When they gained the pace, taps were carried out at (almost) the right frequency which leads to classification performances of up to 94% for the first and 98% for the second subject. As the trial lasts its final seconds, i.e. after 5 seconds, the correct classification rates slightly decreased. This might be due to the subjects' anticipation of the end of trials: the duration of each trial was fixed. The averaged EEGs recorded from two representative electrodes over the motor cortex, $C3$ and $C4$ show that potential difference between the contralateral and ipsilateral hemispheres follow similar trend, see figure Fig. 6.15. Fig. 6.16 illustrates the classification results corresponding to the electrode subset I for the two subjects.

---

[5]The subjects have had a training session before the actual data recording, when they are asked to tap at 2 Hz pace. However, presumably, initiation of the movement with a certain pace takes a short while.

Table 6.1. Averaged classification results (and their corresponding standard deviations) for subject one in a two class BCI problem using FDA. The results are in [%] and bold values show the maximum performance achieved.

TW vs. EC	I	II	III	IV	V
6 s	97.9 ± 1.4	98.1 ± 1.5	96.9 ± 2.2	93.1 ± 5.6	97.6 ± 1.6
0 - 0.5 s	84.7 ± 15.0	85.9 ± 11.6	87.5 ± 15.4	85.4 ± 13.7	84.2 ± 12.4
0 - 1.5 s	74.2 ± 16.3	82.3 ± 10.2	82.3 ± 12.4	82.0 ± 11.1	89.3 ± 8.8
0 - 2.5 s	88.0 ± 8.0	91.0 ± 7.1	80.2 ± 14.5	87.2 ± 10.2	93.1 ± 4.5
0 - 3 s	89.9 ± 6.8	91.8 ± 6.5	85.9 ± 9.9	**90.6 ± 7.1**	**94.8 ± 4.0**
0.5 - 3.5 s	87.2 ± 8.7	87.9 ± 10.4	76.3 ± 11.6	82.8 ± 8.3	91.1 ± 6.4
1 - 4 s	91.3 ± 6.6	91.6 ± 5.6	89.6 ± 7.5	89.2 ± 6.2	94.0 ± 5.9
1.5 - 4.5 s	**93.1 ± 5.0**	**95.7 ± 3.0**	**92.0 ± 6.2**	55.0 ± 17.3	94.5 ± 3.9
2 - 5 s	89.9 ± 7.2	91.6 ± 6.0	77.6 ± 12.2	72.0 ± 16.5	88.1 ± 9.3
2.5 - 5.5 s	88.2 ± 9.0	93.5 ± 4.9	83.0 ± 11.7	71.6 ± 17.8	85.2 ± 9.5
3 - 6 s	89.6 ± 6.1	95.1 ± 1.5	85.7 ± 9.5	63.1 ± 18.3	90.8 ± 7.1

Table 6.2. Averaged classification results (and their corresponding standard deviations) for subject one in a two class BCI problem using KFD. The results are in [%] and bold values show the maximum performance achieved.

TW vs. EC	I	II	III	IV	V
6 s	97.3 ± 3.2	98.4 ± 1.6	96.0 ± 3.1	93.1 ± 5.7	98.2 ± 2.2
0 - 0.5 s	89.9 ± 8.3	90.9 ± 7.5	**91.1 ± 7.2**	89.0 ± 8.2	90.6 ± 7.4
0 - 1.5 s	74.1 ± 18.0	83.3 ± 11.9	84.8 ± 11.0	86.5 ± 10.5	89.6 ± 10.5
0 - 2.5 s	86.5 ± 12.1	90.7 ± 90.7	82.2 ± 12.3	87.5 ± 7.5	92.4 ± 7.7
0 - 3 s	89.6 ± 8.6	92.8 ± 11.8	85.2 ± 10.7	**89.4 ± 11.8**	**94.8 ± 5.6**
0.5 - 3.5 s	88.8 ± 9.3	88.1 ± 10.1	79.8 ± 16.6	83.2 ± 14.3	92.1 ± 8.1
1 - 4 s	89.8 ± 8.1	91.4 ± 7.6	87.7 ± 8.9	87.5 ± 9.7	93.0 ± 4.0
1.5 - 4.5 s	**93.4 ± 6.8**	**95.6 ± 3.8**	90.6 ± 8.9	57.6 ± 15.5	94.8 ± 5.7
2 - 5 s	89.4 ± 7.5	92.8 ± 8.3	79.2 ± 12.9	72.8 ± 20.2	88.2 ± 11.9
2.5 - 5.5 s	89.5 ± 10.9	92.1 ± 8.9	89.3 ± 11.3	71.0 ± 16.2	84.7 ± 13.0
3 - 6 s	89.3 ± 9.4	91.8 ± 5.8	88.5 ± 9.4	64.0 ± 21.4	90.5 ± 9.3

Table 6.3. Averaged classification results (and their corresponding standard deviations) for subject two in a two class BCI problem using FDA. The results are in [%] and bold values show the maximum performance achieved.

TW vs. EC	I	II	III	IV	V
6 s	99.2 ± 0.2	98.3 ± 0.6	98.3 ± 0.6	98.2 ± 0.8	98.2 ± 0.7
0 - 0.5 s	64.1 ± 10.6	55.8 ± 10.6	57.8 ± 10.4	59.0 ± 10.9	52.7 ± 13.1
0 - 1.5 s	95.0 ± 2.5	92.4 ± 3.7	89.5 ± 5.2	86.0 ± 4.8	90.9 ± 4.4
0 - 2.5 s	97.0 ± 1.4	94.1 ± 2.5	93.2 ± 3.3	92.1 ± 4.0	92.1 ± 3.5
0 - 3 s	97.1 ± 1.4	94.8 ± 2.8	96.2 ± 2.3	94.6 ± 2.8	93.0 ± 2.6
0.5 - 3.5 s	95.9 ± 1.5	96.4 ± 1.5	95.1 ± 2.8	94.0 ± 2.8	96.6 ± 1.8
1 - 4 s	97.1 ± 1.3	96.6 ± 1.1	95.5 ± 1.5	93.9 ± 2.6	96.1 ± 1.9
1.5 - 4.5 s	98.5 ± 0.5	98.2 ± 0.7	97.7 ± 1.1	98.0 ± 1.1	97.9 ± 0.8
2 - 5 s	**98.6 ± 0.5**	98.68 ± 0.5	**98.6 ± 0.5**	**98.6 ± 0.5**	**98.6 ± 0.5**
2.5 - 5.5 s	98.2 ± 0.6	97.0 ± 1.0	97.2 ± 1.0	97.3 ± 0.8	96.8 ± 1.1
3 - 6 s	97.3 ± 1.1	**98.8 ± 1.2**	96.7 ± 1.3	97.7 ± 0.9	98.2 ± 0.7

Table 6.4. Averaged classification results (and their corresponding standard deviations) for subject two in a two class BCI problem using KFD. The results are in [%] and bold values show the maximum performance achieved.

TW vs. EC	I	II	III	IV	V
6 s	99.3 ± 0.2	98.9 ± 0.5	98.3 ± 0.7	98.4 ± 0.8	98.5 ± 0.6
0 - 0.5 s	62.1 ± 9.5	55.6 ± 10.6	56.2 ± 10.4	53.8 ± 10.9	52.2 ± 10.4
0 - 1.5 s	96.1 ± 1.1	94.9 ± 2.6	92.5 ± 3.0	86.7 ± 5.2	90.1 ± 3.5
0 - 2.5 s	97.2 ± 0.6	94.5 ± 2.1	94.4 ± 2.6	98.3 ± 0.5	98.4 ± 0.5
0 - 3 s	97.5 ± 0.8	96.9 ± 1.9	97.1 ± 1.6	94.7 ± 2.8	93.4 ± 2.9
0.5 - 3.5 s	97.1 ± 1.3	97.1 ± 1.2	95.9 ± 1.5	95.1 ± 2.6	97.0 ± 1.4
1 - 4 s	97.8 ± 1.2	96.1 ± 1.4	95.1 ± 2.2	93.5 ± 2.7	96.3 ± 2.1
1.5 - 4.5 s	**98.5 ± 0.6**	98.0 ± 1.0	97.4 ± 1.2	97.6 ± 0.9	97.9 ± 1.0
2 - 5 s	98.3 ± 0.5	**98.3 ± 0.5**	**98.4 ± 0.5**	**98.3 ± 0.5**	**98.4 ± 0.5**
2.5 - 5.5 s	98.0 ± 0.7	96.5 ± 1.3	96.7 ± 1.2	98.0 ± 1.5	97.1 ± 1.5
3 - 6 s	97.4 ± 0.9	95.9 ± 2.0	96.4 ± 1.8	98.1 ± 1.2	97.3 ± 1.3
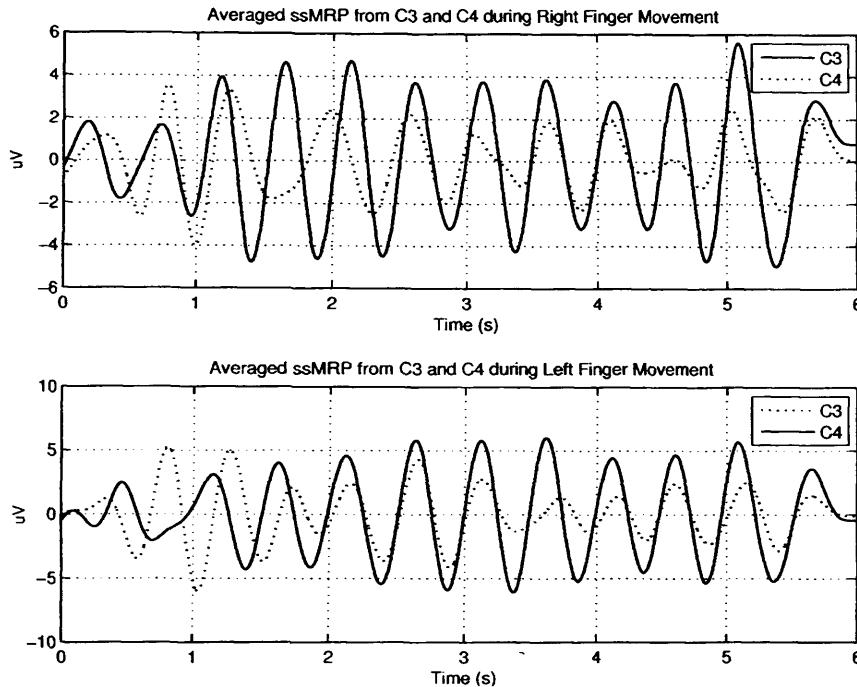
Figure 6.15. Averaged EEGs recorded from C3, and C4 (subject one) demonstrate that the difference potential between the contralateral and ipsilateral hemispheres follow similar trend to that of classification results in Table 6.1 and Table 6.2.

In the top subplot of 6.15, at the beginning of the trial the difference between C3 and C4 are not considerable. After approximately one second, that difference increases which leads to higher classification rates. Lastly, at the end of the trial around time 5 seconds, again this difference decreases. In the bottom subplot of 6.15 similar patterns can be observed.

6.8 On the suitability for Real-time implementation

The ssMRP-based BCI system described here would inherently satisfy classic crucial conditions for real-time rehabilitation [157]: *accuracy* is necessary to fulfill the user's intent; *minimal and simple training* allows users to readily acquire the skills needed to operate a BCI machine; and finally *a low response time* places a time constraint on the recording and processing EEG
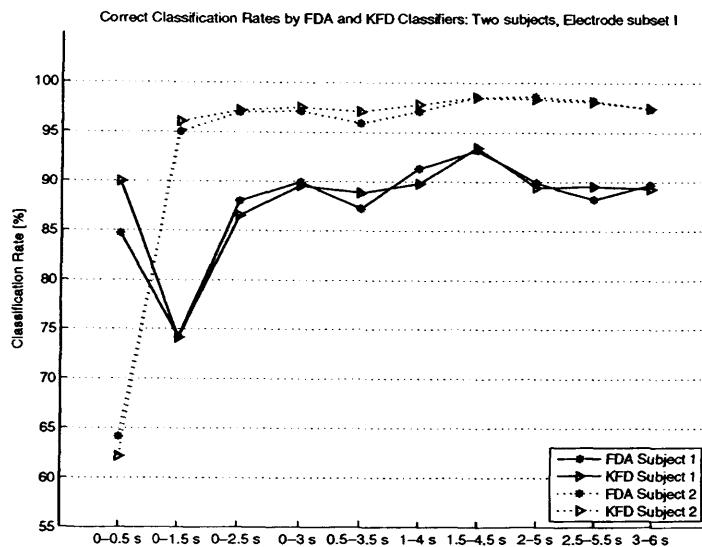
Figure 6.16. The correct classification rates of the temporally segmented EEGs recorded from two subjects. FDA and KFD classifiers are implemented to classify EEG measurements using electrode subsets I.

data. Delays in the control system should appear insignificant to the users, ideally less than 300-500 ms depending on application.

In order to test ssMRPs in an artificial real-time BCI paradigm, the first 200 trials (100 segment from each class) recorded from subject two where were selected. The preprocessing EEGs were concatenated assuming that left and right finger movements had been carried out either alternatingly or pairwise alternatingly. Following similar feature extraction protocol in Section 6.7, each EEG segment was temporally sub-segmented into 24 overlapping windows resulting in overall 4405 sub-segments. The length of each window was 3 s with 2.75 s overlap with the previous sub-segment. The feature vectors were extracted as the energy of 2 Hz rhythm from 45 electrode recordings in each sub-segment.

For classification purposes, if the sub-segments lay fully in one segment, it's label was considered similar to that of the segment. For sub-segments spanning over two segments, the label was considered similar to the segment with which it overlapped most. In case of equal overlap level, a class label
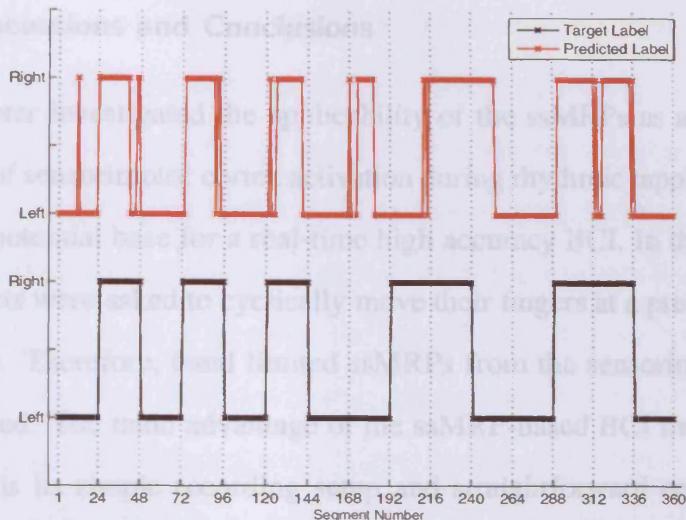
Figure 6.17. Simulated "Real-time" classification results for a set of cascaded EEG segments.

of the former sub-segment was assigned.

First five principal components of each extracted features from sub-segment were classified by a FDA classifier trained using all of the sub-segments. Bearing in mind that a full BCI application would additionally use adaptive feature extraction and pattern classification algorithms, we are confident that further development and refinement of the simple and static classifier we have used will be able to work with shorter data segments in real time. Fig. 6.17 illustrates the results of classification of 384 consecutive sub-segments where the actual and predicted class labels are respectively depicted in black and red. As expected, apart from errors that occurred in classification of the sub-segments crossing segment transitions, the BCI output in red effectively follows the trace in black. However, occasionally, some within segment sub-segments are also mis-classified. For instance, the failures in correct classification of sub-segments 173 to 185, or 240 to 253 may be attributed to either classifier sub-optimality or inevitable error in classification of EEGs recorded in those specific time windows.

6.9 Discussions and Conclusions

This chapter investigated the applicability of the ssMRPs as a continuous measure of sensorimotor cortex activation during rhythmic tapping and suggested a potential base for a real-time high accuracy BCI. In this approach the subjects were asked to cyclically move their fingers at a pre-determined frequency. Therefore, band limited ssMRPs from the sensorimotor cortex are detected. The main advantage of the ssMRP-based BCI over other approaches is its simple recording setup and straightforward computations. Comparing to BCI machines based on RPs, using ssMRP for BCI would not be difficult for the subjects since in each trial they are actively involved in the experiment, rather than waiting for several seconds before the exertion of a single discrete movement.

ssMRP-based BCI is simpler than ssVEP-based BCI systems in terms of subject training time and signal analysis. In the proposed method, advanced spectrum estimation algorithms are not necessary since instead of frequency separation of ssVEP, merely one frequency is dealt with. In other words, instead of exploiting the spectral disparity in the signal processing unit of BCI machine, the topographical distribution of scalp EEG signals in the frequency band of interest is used to identify the effector.

The large number of equally important electrodes utilized in this research should not be regarded as a hindrance towards end-product BCI. Evidently, data driven common spatial patterns (CSP) [49] and their variations e.g. [20] which could increase the performance can also be used here.

There are certain differences between our recording setup and a previous studies such as [15] where the participants could alternate between right and left fingers in consecutive short intervals. That is, the subject could decide whether to move the left or right index on each tap. In [15], the partici-

pants were not provided with rhythmic visual cues and therefore a series of discrete movements were carried out rather than rhythmic continuous movements. Actual number of taps per hand and the inter-hand transition matrix had to be computed objectively and shown on the screen of the experimenter; the taps on the keyboard were extracted as markers into the stream of EEG measurements. Therefore, for instance if the tap rate or the number of alteration deviated from pre-instructed rate, the subject was informed so those could be compensated accordingly. Notice that in some specific applications, such as typing on the keyboard, BCI system should ideally be able to distinguish between left and right finger movement in each individual discrete tap. However, for many other applications such as navigating a wheelchair, a smooth and continuous BCI output is demanded. We believe our protocol will lead to more reliable and less variable BCI output classification.

The only problem seen in ssMRP-based BCI is the performance degradation when the subject reaches the end of the trials. This problem should be considered in real-time as well as offline applications. In a real life application, there is no need to have such long trials. They can easily be as short as few seconds in order have maximum performance. In offline applications, the easiest solution should be to allow the trials to have random lengthes instead of a fixed six second. For instance, the length of trials may be considered four or five seconds and a small standard deviation.

Chapter 7

CONCLUSIONS AND

FURTHER RESEARCH

In this dissertation, a number of new methods for brain signal analysis in the STF domain with application to EB artifact removal from EEGs were developed. These methods were evaluated using simulated and real EEGs and shown to produce comparable results to that of previous studies. Furthermore, a simple and novel paradigm for BCI application was developed which exploits the prior physiological knowledge of spectrally band limited ssMRPs.

Chapter 2 comprehensively reviewed the EEG-based BCI state of art, the conventional and established thoughts and methods for BCI realization. Chapter 3, presented the details of the first contribution of this research to BCI where after briefing some mathematical definitions, it was shown that how the tensor of the time-varying energy of the multi-channel EEGs in the wavelet domain can be decomposed into spatial, temporal, and spectral signatures of EEGs. By STF decomposition of EEGs in the μ band, EEG dynamics during left and right index imagery movements were investigated. The spatial signature of the movement related factor for each trial was introduced to the SVM classifier. The discussion in the last part of Chapter 3 demonstrated that the computational complexity of the STF modeling limits

the use of such BCI in real-time applications.

EEGs can be severely contaminated by various artifacts. Two novel pre-processing algorithms were developed in Chapters 4 and 5 for the removal of EB artifacts from the EEGs. In Chapter 4, after an extensive overview on the currently available methods for EB artifact removal, an effective semi-blind signal extraction algorithm was developed to identify and remove the EB artifacts. In that proposed approach, for the spatial signatures of the EB artifacts were identified by using the STF model and utilized in the extraction stage. Fundamentally, there are two important points that should be considered in using such method. First, the decomposition of the EEGs to their STF signatures is computationally intensive and second, the performance of the SBSE method depends on the proper estimation of the EB spatial signature.

In order to relax the above mentioned two conditions, a hybrid "STF modeling"-"robust minimum variance beamforming" framework was presented in Chapter 5. A novel STF-TS model for EEGs, specifically designed for decomposition of the EB contaminated EEGs, was proposed. The spatial signatures were then exploited in a robust beamforming paradigm to extract and remove the EB artifact. Therefore, the major contribution of this chapter can be summarized in the estimation of the steering vector corresponding to the EB artifact without computing the conventional forward solutions.

Chapter 6 of this dissertation introduced a novel BCI framework based on the ssMRPs. In order to surpass the low SNR slow cortical potentials, recent physiological findings on ssMRPs were adopted. Unlike the conventional methods in the BCI field, brain capabilities in controlling voluntary repetitive finger movements were taken into account. The neurological aspects of the phenomenon of interest was reviewed and investigated for real

EEG recordings in averaged and single trial modes. Finally, the applicability of the two classic classifiers for such BCI scheme were investigated.

7.1 Future Works

This research has developed several high performance algorithms for brain signal processing and may be further expanded. In particular, answering the below questions can be an initiation for future investigations.

- Do STF signal processing approaches consistently outperform the conventional time-space or time-frequency methods in brain signal analysis?

- The SBSE method can exploit spatial signatures of EB artifacts and use them as prior knowledge. How can other prior information such as temporal or spectral signatures of EB artifacts be incorporated in SBSE? What are other applications for such SBSE algorithm?

- How the STF-TS model can be extended to other STF models?

- How the ssMRP based BCI may be extended? Is it possible to realize a motor imagery BCI machine based on ssMRPs?

In reply to the first question, one might argue that although the STF models seem to be very effective, their implementation on the brain signals is merely recommended when a phenomenon of interested sparsely occurs at least in one of those domains. Therefore, as observed in Chapter 3, the PARAFAC based BCI does not noticeably outperform other conventional BCI approaches. However, for instance in Chapters 4 and 5, the STF modeling of EB artifact has resulted in acceptable performances since temporal signatures of EBs are sparse.

An interesting avenue for future research is to examine further applications of the SBSE method in other non-stationary biomedical signals. For instance, a sequential algorithm has been recently developed for a class of periodic signals, namely ECGs, in [158]. In this work, the ECGs are presumed to have fixed period and the major issue of heart rate variability has not been taken into account. Therefore, the semi-blind extraction of quasi-periodic signals can be realized if the period of the signal of interest could be instantaneously estimated and incorporated in the extraction procedure. Aiming at that, the most straightforward approach is to minimize the extraction cost function (4.2.9), instead of in some early time lags, at certain lags where abrupt non-stationarities are observed in the autocorrelation values in a similar way as in [159–161].

As a potential application, recently, simultaneous recording of the EEGs and fMRI has attracted many signal processing and computational neuroscience researchers. There are several types of artifact in EEGs recorded within the MRI machine. For instance, slight movement of subjects' scalps and electrode leads in the high static magnetic field causes ballistocardiac pulse artifact or ballistocardiogram (BCG) artifact. The pioneering algorithms in BCG removal have been generally based on the subtraction of the average heartbeat waveform within a predetermined interval from EEGs. However, the inaccuracy in heartbeat detection and the variability of heartbeat waveform induce serious errors. Extended versions of the procedure presented in Chapter 4 may be applied to remove these artifacts. For instance, the heart beat variability uncertainties may be rectified by combining several autocorrelation matrices at approximate time instants when the heart beats.

Furthermore, following the approach presented in Chapters 5, as for the time domain, the frequency or space domain may also be segmented and the STF model with the additional extensions such as frequency/segment (STF-FS) or space/segment (STF-SS). For instance, for STF-SS model, all channels of EEGs are first equally divided into groups in the space domain to form a three dimensional array of space, time, and space/segment domains. Then, the time-frequency transform is applied to each channel to form the fourth dimension. Thus, the four-way PARAFAC may be applied to extract the signatures of this four dimensional array. The STF-SS model can be mathematically formulated the same way as the STF-TS model except that the selected domain to be segmented is space. However, the main problem using various sub-models of STF model is in model fitting stage. Regarding the basic ALS method as a stepping stone, there have been few approaches designed for tensor signal processing, e.g. [96, 97], which can be adopted for STF fitting. More importantly, future works may include finding the performance bounds for the STF modeling.

Another perspective of this research lies in the extension of Chapter 6 by investigating the potential of ssMRPs for motor imagery BCI. For instance, imagination of the repetitive tapping might enable a user to control some distinct features for BCI pattern classification unit as well as the case of real movement. To this end, standard protocols for ssMRP recordings should be devised. In each trial, a primary short period would be needed to train the subject using the correct pace. That is, the subjects would be exposed to rhythmically flashing cues and asked to tap the tapping synchronous to them. After a few seconds, the flashing cues would disappear and subjects would be asked to imagine the taps with the learnt frequency. The recorded EEGs should then be sent to the feature extraction and classification units.

Similar to other motor imagery BCI systems, the major issue with imagined ssMRP is the low SNR. However, since the subjects imagine the tapping with a pre-determined frequency, the SNR of ssMRPs should still be much higher than in conventional ERD/ERS paradigms.

Another interesting research may be the study of brain functional connectivity versus brain default networks during rhythmics motor output for BCI purposes. For instance, the band limited signals of frequencies approximately similar to that of the rhythmic cue signal over various cortical regions define a spatiotemporal signature which can be used to detect motor activity. Therefore, a two-stage STF approach may be carried out in order to first distinguish between the resting brain and active brain by analyzing the spectral signatures of the recorded potentials. Then, a classifier decides between right or left finger movements by investigating the spatio-temporal signatures.

BIBLIOGRAPHY

[1] G. Buzsáki, *Rhythms of the Brain*. Oxford University Press Inc., 2006.

[2] F. Popescu, S. Fazli, Y. Badower, B. Blankertz, and K.-R. Müller., "Single trial classification of motor imagination using 6 dry EEG electrodes," *PLoS ONE*, vol. 2, no. 7, pp. 1643–1650, 2007.

[3] B. J. Casey and M. de Haan, "Introduction: new methods in developmental science," *Developmental Sci.*, vol. 5, pp. 265–267, 2002.

[4] G. Pfurtscheller and F. H. L. da Silva, *Event-Related Desynchronisation, Handbook of Electroencephalogr. Clin. Neurophysiol.* Revised Series, Elsevier, Amsterdam, 1999.

[5] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, 1999.

[6] P. Shenoy, K. J. Miller, J. G. Ojemann, and R. P. N. Rao, "Generalized features for electrocorticographic BCIs," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 273–280, 2008.

[7] A. B. Schwartz, "Motor cortical activity during drawing movement: population representation during sinusoid tracing," *Jour. Neurophysio.*, vol. 70, pp. 28–36, 1993.

[8] J. P. Donoghue and J. N. Sanes, "Motor areas of the cerebral cortex," *Clin. Neurophysio.*, vol. 11, pp. 382–396, 1994.

[9] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth Center," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 8, pp. 222–226, 2000.

[10] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," in *National Academy of Sciences of the United States of America*, vol. 101, pp. 17849–17854, 2003.

[11] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, pp. 297–298, 1999.

[12] N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The though translation device," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 8, pp. 190–193, 2000.

[13] A. Kübler, B. Kotchoubey, T. Hinterberger, N. Ghanayim, J. Perelmouter, M. Schauer, C. Fritsch, E. Taub, and N. Birbaumer, "The thought translation device: a neurophysiological approach to communication in total motor paralysis," *Exp. Brain Res.*, vol. 124, pp. 223–232, 1999.

[14] T. N. Lal, T. Hinterberger, G. Widman, M. Schroder, N. J. Hill, W. Rosenstiel, C. E. Elger, B. Scholkopf, and N. Birbaumer, "Methods towards invasive human brain computer interfaces, In L. K. Saul, Y. Weiss, and L. Bottou, editors," *Advances in Neural Information Processing Systems*, vol. 17, pp. 737–744, 2005.

[15] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, "The Berlin Brain Computer Interface: EEG-Based communication without subject training," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 14, pp. 147–152, 2006.

[16] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curio, "Boosting bit rates and error detection forthe classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 11, pp. 127–131, 2003.

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

[18] G. Dornhege, *Increasing Information Transfer Rates for Brain-Computer Interfacing*. PhD thesis, Univ. Potsdam, Potsdam, Germany, 2006.

[19] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 993–1002, 2004.

[20] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 1541–1548, 2005.

[21] B. O. Peters, G. Pfurtscheller, and H. Flyvbjerg, "Automatic differen-

tiation of multichannel EEG signals," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 111–116, 2001.

[22] G. R. Müller-Putz, C. Neuper, R. Rupp, C. Keinrath, H. Gerner, and G. Pfurtscheller, "Event-related beta EEG changes during wrist movements induced by functional electrical stimulation of forearm muscles in man," *Neurosci. Lett.*, vol. 340, pp. 143–147, 2003.

[23] G. R. Müller-Putz, R. Scherer, G. Pfurtscheller, and R. Rupp, "EEG-based neuroprosthesis control: A step towards clinical practice," *Neurosci. Lett.*, vol. 382, pp. 169–174, 2005.

[24] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'Thought'- control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neurosci. Lett.*, vol. 351, pp. 33–36, 2003.

[25] C. Vidaurre, A. Schögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully online adaptive BCI," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 1214–1219, 2006.

[26] G. Bauernfeind, R. Leeb, S. Wriessnegger, and G. Pfurtscheller, "Development, set-up and first results for a one-channel near-infrared spectroscopy system," *Biomedizinische Technik = Biomedical engineering*, vol. 53, pp. 36–43, 2008.

[27] G. Pfurtscheller, R. L. D. Friedman, and M. Slater, "Centrally controlled heart rate changes during mental practice in immersive virtual environment: A case study with a tetraplegic," *Int. J. Psychophysiol.*, vol. 351, pp. 33–36, 2003.

[28] J. del R. Millán and J. Mourino and M. Franzé and F. Cinotti and M. Varsta and J. Heikkonen and F. Babiloni, "A local neural classifier for the recognition of EEG patterns associated to mental tasks," *IEEE Neur. Net.*, vol. 13, pp. 678–686, 2002.

[29] J. del R. Millán and J. Mourino, "Asynchronous bci and local neural classifiers: An overview of the adaptive brain interface project," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 11, pp. 1214–1219, 2003.

[30] J. del R. Millán and F. Renkens and J. Mourino and W. Gerstner, "Non-invasive brain-actuated control of a mobile robot by human EEG," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 1214–1219, 2004.

[31] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Wiley, New York, 2007.

[32] E. Donchin and M. G. H. Coles, "Is the P300 a manifestation of context updating?," *Behavioral and Brain Sciences*, vol. 11, pp. 355–372, 1988.

[33] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, "P300-based brain computer interface mouse with genetically-optimised analogue control," *IEEE Trans. Neur. Sys. Rehab. Eng.*, pp. 51–61, 2008.

[34] R. Q. Cui, D. Huter, W. Lang, and L. Deecke, "Neuroimage of voluntary movement: Topography of the Bereitschaftspotential, a 64-channel DC current source density study," *NeuroImage*, vol. 9, no. 1, pp. 124–134, 1999.

[35] Y. Li, X. Gao, H. Liu, and S. Gao, "Classification of the single trial electroencephalogram during finger movement," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1019–1025, 2004.

[36] M. Jahanshahi and M. Hallet, eds., *The Bereitschaftspotential Movement-Related Cortical Potentials*. Kluwer Academic, 2003.

[37] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, and G. Schalk, "The Wadsworth center brain-computer interface (BCI) research and development program," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 11, pp. 1–4, 2003.

[38] R. Beisteiner, P. Hollinger, G. Lindinger, W. Lang, and A. Berthoz, "Mental representations of movements. Brain potentials associated with imagination of hand movements," *Electroencephal. Clin. Neurophysio.*, vol. 96, pp. 183–193.

[39] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysio.*, vol. 113, pp. 767–791, 2002.

[40] H. Berger, "Uber das Elektrenkephalogramm des Menschen II," *Psychol. Neurol.*, vol. 40, pp. 160–179, 1930.

[41] G. Pfurtscheller and A. Aranibar, "Evaluation of event related desynchroniation (ERD) preceding and following self-paced movement," *Electroencephal. Clin. Neurophysio.*, vol. 46, pp. 138–146, 1979.

[42] J. Stancak and G. Pfurtscheller, "Event-related desyncronization of central beta-rhythms during brisk and slow self-paced finger movments of dominant and nondominant hand," *Cognitive Brain Res.*, vol. 4, pp. 171–183, 1996.

[43] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement," *Electroencephal. Clin. Neurophysio.*, vol. 72, pp. 250–258, 1989.

[44] C. Neuper and G. Pfurtscheller, "Post-movement synchronization of beta rhythms in the EEG over the cortical foot area in man," *Neuroscience Lett.*, vol. 219, pp. 17–20, 1996.

[45] F. Boiten, J. Sergeant, and R. Geuze, "Event-realted desynchronization: the effects of the energetic and computational demands," *Electroencephal. Clin. Neurophysiol.*, vol. 82, pp. 302–309, 1992.

[46] L. Defebvre, J. L. Bourriez, A. Destee, and J. D. Guieu, "Movement realted desynchronization pattern preceding voluntary movement in untreated Parkinson's disease," *J. Neurology, Neurosurgery and Psychiatry*, vol. 60, pp. 307–312, 1996.

[47] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data," *Clin. Neurophysiol.*, vol. 113, pp. 43–47, 2002.

[48] A. M. Zoubir and D. R.Iskander, *Bootstrap Techniques for Signal Processing.* Cambridge University press, 2004.

[49] K. Fukanaga, *Introduction to Statistical Pattern Recognition.* Academic, 2nd ed., 1990.

[50] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 8, pp. 441–446, 2000.

[51] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics Intell. Lab. Syst.*, vol. 38, no. 2, pp. 149–171, 1998.

[52] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An

EEG-based brain-computer interface for cursor control," *Electroencephal. Clin. Neurophysiol.*, vol. 78, pp. 252–259, 1991.

[53] G. Ginter, K. J. Blinowska, M. Kaminski, and P. J. Durka, "Phase and apmplitude analysis in time-frequency space-application to voluntry finger movement," *J. Neuro. Method.*, vol. 110, pp. 113–124, 2001.

[54] T. Ebrahimi, J.-M. Vesin, and G. Garcia, "Brain-computer interface in multimedia communication," *IEEE Sig. Process. Mag.*, vol. 20, pp. 14–24, 2003.

[55] L. Shoker, S. Sanei, and J. A. Chambers, "Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm," *IEEE Sig. Process. Lett.*, vol. 12, no. 10, pp. 721–724, 2005.

[56] R. A. Harshman, "Foundation of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis," Tech. Rep. 16, pp. 1-84, UCLA, CA, USA, 1970.

[57] J. Mokes, "Decomposing event-related potentials: a new topographic components model," *Biol. Psychol.*, vol. 26, pp. 199–215, 1998.

[58] T. Koenig, F. Marti-Lopez, and P. A. Valdés-Sosa, "Topographic time-frequency decomposition of the EEG," *NeuroImage*, vol. 14, pp. 383–390, 2001.

[59] F. Miwakeichi, E. Martinez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchia, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, pp. 1035–1045, 2004.

[60] A. Osman and R. Albert, "Time-course of cortical activation during overt and imagined movements," in *Cognitive Neuroscience Annual Meeting*, (New York), Mar. 2001.

[61] F. Perrin, J. Pernier, O. Bertrand, and J. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephal. Clin. Neurophysio.*, vol. 72, pp. 184–187, 1989.

[62] A. Teolis, *Computational Signal Processing with Wavelets*. Birkhauser, 1998.

[63] R. Bro, *Multi-way Analysis in the Food Industry: Models, Algorithms and Applications*. PhD thesis, University of Amsterdam (NL) and Royal Veterinary and Agricultural University, 2001. Toolbox Available [Online] at http://www.models.kvl.dk/users/rasmus/.

[64] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.

[65] B. Kamousi, Z. Liu, and B. He, "Classification of Motor Imagery Tasks for Brain-Computer Interface Applications by means of Two Equivalent Dipoles Analysis," *IEEE Trans. Neural Sys. Rehab. Eng.*, vol. 13, no. 1, pp. 166–171, 2005.

[66] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing-Learning Algorithms and Applications*. Wiley, New York, 2005.

[67] C. W. Hesse and C. J. James, "The fastICA algorithm with spatial constraints," *IEEE Sig. Process. Lett.*, vol. 12, no. 11, pp. 792–795, 2005.

[68] F. H. L. da Silva, W. S. van Leeuwen, and A. Rémond, eds., *Clinical Application of Computer Analysis of EEG and Other Neurophysiological Signals*. Elsevier Science Ltd, 1987.

[69] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, no. 2, pp. 342–353, 2005.

[70] R. J. Croft and R. J. Barry, "EOG correction: A new perspective.," *Electroencephal. Clin. Neurophysiol.*, vol. 107, pp. 387–394, 1998.

[71] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey," *Clin. Neurophysiol.*, vol. 118, pp. 480–494, 2007.

[72] R. J. Croft and R. J. Barry, "EOG correction: Which regression should we use?," *Psychophysiology*, vol. 37, pp. 123–125, 2000.

[73] K. H. Ting, P. C. W. Fung, C. Q. Chang, and F. H. Y. Chan, "Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only," *Med. Eng. Phys.*, vol. 28, pp. 780–794, 2006.

[74] L. Shoker, S. Sanei, W. Wang, and J. A. Chambers, "Removal of eye blinking artifact from electro-encephalogram, incorporating a new constrained blind source separation algorithm," *Med. Biol. Eng. Comp.*, vol. 43, pp. 209–295, 2005.

[75] C. J. James and O. J. Gibson, "Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 9, pp. 1108–1116, 2003.

[76] K. Nazarpour, S. Sanei, and J. A. Chambers, "A novel semi-blind signal extraction approach incorporating PARAFAC for the removal of the removal of eye-blink artifact from EEGs," in *Proc. DSP07*, pp. 127–130, Cardiff, UK, July 2007.

[77] K. Nazarpour, Y. Wongsawat, S. Sanei, J. A. Chambers, and S. Oraintara, "Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming," in *Proc. IEEE EMBC07*, pp. 6211–6214, Lyon, France, Aug. 2007.

[78] K. Nazarpour, Y. Wongsawat, S. Sanei, J. A. Chambers, and S. Oraintara, "Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming," *IEEE Trans. Biomed. Eng.*, 2008, in press.

[79] K. Nazarpour, H. R. Mohseni, C. W. Hesse, J. A. Chambers, and S. Sanei, "A novel semi-blind signal extraction approach incorporating PARAFAC for the removal of the the removal of eye-blink artifact from EEGs," *EURASIP Advances Sig. Process.*, doi:10.1155/2008/857459, 2008.

[80] T. D. Lagerlund, F. W. Sharbrough, and N. E. Busacker, "Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition," *Clin. Neurophysiol.*, vol. 14, no. 1, pp. 73–82, 1997.

[81] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clin. Neurophysiol.*, vol. 118, pp. 98–104, 2007.

[82] S. Puthusserypady and T. Ratnarajah, "H^{∞} adaptive filters for eye blink

artifact minimization from electroencephalogram," *IEEE Sig. Process. Lett.*, vol. 12, no. 12, pp. 816–819, 2005.

[83] C. W. Hesse and C. J. James, "On semi-blind source separation using spatial constraints with applications in EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2525–2534, 2006.

[84] A. C. Soong and P. Z. Koles, "Principal-component localization of sources of the background EEG," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 59–67, 1995.

[85] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, New York, 2001.

[86] C. Phillips, J. Mattout, M. D. Rugg, P. Maquet, and K. J. Friston, "An empirical Bayesian solution to the source reconstruction problem in EEG," *NeuroImage*, vol. 24, no. 4, pp. 997–1011, 2005.

[87] M. A. Latif, S. Sanei, J. A. Chambers, and L. Shoker, "Localization of abnormal EEG sources using blind source separation partially constrained by the locations of known sources," *IEEE Sig. Process. Lett.*, vol. 13, pp. 117–120, 2006.

[88] X. Li and X. Zhang, "Sequential blind extraction adopting second-order statistics," *IEEE Sig. Process. Lett.*, vol. 14, no. 1, pp. 58–61, 2007.

[89] K. Nazarpour, S. Sanei, L. Shoker, and J. A. Chambers, "Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing," in *Proc. EUSIPCO 06*, Florence, Italy, Sep. 2006.

[90] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ICA," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, 2005.

[91] N. Ille, *Artifact Correction in Continuous Recording of the Electro- and Magnetoencephalogram by Spatial Filtering*. PhD thesis, Univ. Mannheim, Mannheim, Germany, 2001.

[92] M. Dyrholm, S. Makeig, and L. K. Hansen, "Model selection for convolutive ICA with an application to spatio-temporal analysis of EEG," *Neural Computation*, vol. 19, pp. 934–955, 2007.

[93] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Trans. Sig. Process.*, vol. 42, no. 6, pp. 1413–1424, 1994.

[94] H. S. M. Coxeter and S. L. Greitzer, *Collinearity and Concurrence*. Washington, DC: Math. Assoc. Amer., 1967.

[95] L. Shoker, *Signal Processing Algorithms for Brain Computer Interfacing*. PhD thesis, Cardiff University, 2006.

[96] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. Sig. Process.*, vol. 48, pp. 810–823, 2000.

[97] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Sig. Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[98] Y. Wongsawat, S. Oraintara, and K. R. Rao, "Reduced complexity space-time-frequency model for multi-channel EEG and its applications," in *Proc. IEEE ISCAS*, pp. 1305–1308, New Orleans, USA, 2007.

[99] T. Acar, Y. Yu, and A. P. Petropulu, "Blind MIMO system estimation

based on PARAFAC decomposition of higher order output tensors," *IEEE Trans. Sig. Process.*, vol. 54, no. 11, pp. 4156–4168, 2006.

[100] S. A. Vorobyov, Y. Rong, N. D. Sidiropoulos, and A. B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Sig. Process.*, vol. 53, no. 8, pp. 2678–2689, 2005.

[101] Y. Rong, S. A. Vorobyov, A. B. Gershman, and N. D. Sidiropoulos, "Blind spatial signature estimation via time-varying user power loading and parallel factor analysis," *IEEE Trans. Sig. Process.*, vol. 53, no. 5, pp. 1697–1710, 2005.

[102] B. D. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, 1988.

[103] L. Ding and B. He, "Spatio-temporal EEG source localization using a three dimentional subspace FINE approach in a realistic geometry in homogeneous head model," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 9, pp. 1732–1739, 2006.

[104] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filter," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867–879, 1997.

[105] D. Gutierrez, A. Nehorai, and A. Dogandzic, "Performance analysis of reduced-rank beamformers for estimating dipole source signals using EEG/MEG," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 840–844, 2006.

[106] H. Liu and P. H. Schimpf, "Efficient localization of synchronous EEG source activities using a modified RAP-MUSIC algorithm," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 652–661, 2006.

[107] D. Chdeyne, A. C. Bostan, W. Gaetz, and E. W. Pang, "Event-related beamforming: A robust method for presurgical functional mapping using MEG," *Clin. Neurophysiol.*, vol. 118, no. 8, pp. 1691–1704, 2007.

[108] L. Spyrou and S. Sanei, "A new beamforming approach for the localization of event related potentials," in *Proc. EUSIPCO07*, pp. 2489–2493, Poznań, Poland, Sep. 2007.

[109] S. Chen, K. Labib, and L. Hanzo, "Clustering-based symmetric radial basis function beamforming," *IEEE Sig. Process. Lett.*, vol. 14, no. 9, pp. 589–592, 2007.

[110] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A bayesian approach to robust adaptive beamforming," *IEEE Trans. Sig. Process.*, vol. 48, no. 2, pp. 386–398, 2000.

[111] Y. C. Eldar, A. Nehorai, and P. S. La Rosa, "An expected least-squares beamforming approach to signal estimation with steering vector uncertainties," *IEEE Sig. Process. Lett.*, vol. 13, no. 5, pp. 288–291, 2006.

[112] O. Besson, A. A. Monakof, and C. Chalus, "Signal waveform estimation in the presnce of uncertainties about the steering vector," *IEEE Trans. Sig. Process.*, vol. 52, no. 9, pp. 2432–2440, 2004.

[113] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Sig. Process.*, vol. 51, no. 2, pp. 313–324, 2003.

[114] J. Li and P. Stoica (Eds.), *Robust Adaptive Beamforming*. Wiley, 2005.

[115] J. C. Mosher, R. Leahy, and P. S. Lewis, "EEG and MEG: forward solutions for inverse methods," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 3, pp. 245–259, 1999.

[116] I. S. Yetik, A. Nehorai, C. H. Muravchik, and J. Haueisen, "Line-source modeling and estimation with magnetoencephalography," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 839–851, 2005.

[117] I. S. Yetik, A. Nehorai, C. H. Muravchik, J. Haueisen, and M. Eiselt, "Surface-source modeling and estimation using biomagnetic measurements," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1872–1882, 2006.

[118] Y.-S. Chen, C.-Y. Cheng, J.-C. Hsieh, and L.-F. Chen, "Maximum contrast beamformer for electromagnetic mapping of brain activity," *IEEE Trans. Biome. Eng.*, vol. 53, no. 9, pp. 1765–1774, 2006.

[119] N. Ille, R. Beucker, and M. Scherg, "Spatially constrained independent component analysis for artifact correction in EEG and MEG," *NeuroImage*, vol. 13, p. S159, 2001.

[120] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Z.-Q. Luo, "Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver," *IEEE Trans. Sig. Process.*, vol. 53, no. 1, pp. 295–305, 2005.

[121] M. G. H. Coles, H. G. O. M. Smid, M. K. Scheffers, and L. J. Otten, *Mental chronometry and the study of human information processing*, ch. Electrophysiology of mind: event-related brain potentials and cognition by M. D. Rugg and M. G. H. Coles, pp. 86–131. Oxford: Oxford University Press, 1995.

[122] P. Praamstra, D. F. Stegeman, A. R. Cools, and M. W. I. M. Horstink, "Reliance on external cues for movement initiation in Parkinson's disease; evidence from movement-related potentials," *Brain*, vol. 121, pp. 167–177, 1998.

[123] P. Praamstra, D. F. Stegeman, M. W. I. M. Horstink, and A. R. Cools, "Dipole source analysis suggests selective modulation of the supplementary motor area contribution to the readiness potential," *Electroencephal. Clin. Neurophysiol.*, vol. 98, pp. 468–477, 1996.

[124] H. H. Kornhuber and L. Deecke, "Hirnpotentialänderungen bei wilkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente ptentiale," *Pflügers Archiv*, vol. 284, pp. 1–17, 1965.

[125] M. Kutas and E. Donchin, "Preparation to respond as manifested by movement-related brain potentials," *Brain Research*, vol. 202, pp. 95–115, 1980.

[126] M. G. H. Coles and G. Gratton, *Cognitive psycophysiology and the study of states and processes*, ch. Energetic and Human Information Processing by G. R. J. Hockey and A. W. K. Gaillard and M. G. H. Coles (Eds.), pp. 409–424. Dordrecht, The NetherlandsL Martinus Nijhof, 1986.

[127] H. G. O. M. Smid, G. Mulder, and L. J. M. Mulder, *The continuous flow model revisited: Perceptual and central motor concepts*, ch. Current Trends in Event-Related Potentials Research (EEG Suppl. 40) by R. Johnson, Jr., J. W. Rohrbaugh, and R. Parasuraman (Eds.), pp. 270–278. Amsterdam: Elsevier, 1987.

[128] R. D. Jong, M. Wierda, G. Mulder, and L. J. M. Mulder, "Use of par-

tial information in responding," *J. Experimental Psychology: Human Perception and Performance*, vol. 14, pp. 682–689, 1988.

[129] G. Gratton, M. G. H. Coles, E. J. Sirevaag, C. W. Eriksen, and E. Donchin, "Pre- and post-stimulus activation of response channels: A psychophysiological analysis," *J. Experimental Psychology: Human Perception and Performance*, vol. 14, pp. 331–344, 1988.

[130] M. Eimer, "The laterized readiness potential as an on-line measure of selective response activation," *Behavior Research Method, Instruments, and Computer*, vol. 30, pp. 146–156, 1998.

[131] R. Krepki, B. Blankertz, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface (BBCI): towards a new communication channel for online control in gaming applications," *Journal of Multimedia Tools and Applications*, vol. 33, no. 1, pp. 73–90, 2007.

[132] D. Regan, *Human Brain Electrophysiology: Evoked Potentials and Evoked Magentic Fields in Science and Medicine*. Amsterdam, The Netherlands: Elsevier, 1989.

[133] E. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. B. Reilly, and G. Mc-Darby, "Brain-computer interface based on the steady-state VEP for immersive gaming control," in *Proc. 2nd Int. Brain-Computer Interface Workshop and Training Course, vol. 49, Ergänzungsband Biomed Techn*, pp. 63–64, Graz, Austria, Sep. 2004.

[134] G. R. McMillan, G. L. Calhoun, M. S. Middendorf, J. H. Schnurer, D. F. Ingle, and V. T. Nasman, "Direct brain interface utilizing selfregulation of steady-state visual evoked response (SSVER)," in *Proc. Rehabilita-*

tion Eng. Assistive Technol. Soc. of North Amer. 18th Annu. Conf. (RESNA), pp. 693–695, USA, 1995.

[135] M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones, "Brain computer interfaces based on the steady-state visual-evoked response," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, pp. 211–214, 2000.

[136] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, 2002.

[137] P. Martinez, H. Bakardjian, and A. Cichocki, "Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm," *Computational Intelligence and Neuroscience*, doi:10.1155/2007/94561, 2007.

[138] T. W. Boonstra, A. Daffertshofer, M. Breakspear, and P. J. Beek, "Multivariate time-frequency analysis of electromagnetic brain activity during bimanual motor learning," *NeuroImage*, vol. 36, pp. 370–377, 2007.

[139] G. Aschersleben and W. Prinz, "Delayed auditory feedback in synchronization," *J. Mot. Behav.*, vol. 29, pp. 35–46, 1997.

[140] G. Aschersleben, "Temporal control of movements in sensorimotor synchronization," *Brain Cogn*, vol. 48, pp. 66–79, 2002.

[141] G. Aschersleben, P. Stenneken, J. Cole, and W. Prinz, *Timing mechanisms in sensorimotor synchronization*, ch. Common Mechanisms in Perception and Action: Attention and Performance by W. Prinz and B.Hommel (Eds.), vol. XIX, pp. 227–244. University Press, Oxford, 2002.

[142] C. Gerloff, C. Toro, N. Uenishi, L. G. Cohen, L. Leocani, and M. Hallett, "Steady-state movement-related cortical potentials: a new approach to assessing cortical activity associated with fast repetitive finger movements," *Electroencephal. Clin. Neurophysiol.*, vol. 102, pp. 106–113, 1997.

[143] C. Gerloff, N. Uenishi, T. Nagamine, T. Kunieda, M. Hallett, and H. Shibasaki, "Cortial activation during fast repetitive finger movements in humans: steady-state movement-related magnetic fields and their cortical generators," *Electroencephal. Clin. Neurophysiol.*, vol. 109, pp. 444–453, 1998.

[144] B. Pollok, K. Muller, G. Aschersleben, A. Schnitzler, and W. Prinz, "The role of the primary somatosensory cortex in an auditorily paced finger tapping task," *Exp. Brain Res.*, vol. 156, pp. 111–117, 2004.

[145] R. Krepki, G. Curio, B. Blankertz, and K.-R. Müller, "Berlin brain-computer interface - the HCI communication channel for discovery," *Int. J. Hum. Comp. Studies, Special Issue on Ambient Intelligence*, vol. 65, pp. 460–477, 2007.

[146] R. Oostenveld and P. Praamstra, "The five percent electrode system for highresolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, pp. 713–719, 2001.

[147] V. Jurcak, D. Tsuzuki1, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, pp. 1600–1611, 2007.

[148] P. A. Pope, A. Holton, S. Hassan, D. Kourtis, and P. Praamstra, "Cortical control of muscle relaxation: A lateralized readiness potential (LRP) investigation," *Clin. Neurophysiol.*, vol. 118, pp. 1044–1052, 2007.

[149] P. A. Pope, P. Praamstra, and A. M. Wing, "Force and time control in the production of rhythmic movement sequences in Parkinson's disease," *European J. Neuroscience*, vol. 23, pp. 1643–1650, 2006.

[150] H. R. Mohseni, K. Nazarpour, E. L. Wilding, and S. Sanei, "A wavelet based particle filtering approach for estimation of single trial event-related potentials," *IEEE Trans. Biomed. Eng.*, in preparation.

[151] B. Schölkopf and A. J. Smola, *Learning with Kernels*. The MIT Press, MA, 2002.

[152] S. Mika, G. Ratsch, and J. Weston, "Fisher discriminant analysiswith kernels," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pp. 41–48, Piscataway, NJ, USA, 1999.

[153] S. Mika, *Kernel Fisher Discriminants*. PhD thesis, University of Technology, Berlin., 2002.

[154] N. Cristianini and J. S. Taylor, *An Introduction to Support Vector Machines (and other kernel-based learning methods)*. Cambridge University Press, Cambridge, UK, 2000.

[155] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[156] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns," *Clin. Neurophysiol.*, vol. 115, pp. 2744–2753.

[157] K. Englehart and B. Hudgins, "A robust, real time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, pp. 848–854, 2003.

[158] M. G. Jafari, W. Wang, J. A. Chambers, T. Hoya, and A. Cichocki, "Sequential blind source separation based exclusively on second-order statistics developed for a class of periodic signals," *IEEE Trans. Sig. Process.*, vol. 54, no. 3, pp. 1028 – 1040, 2006.

[159] A. Belouchrani and M. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. Sig. Process.*, vol. 46, no. 11, pp. 2888–2897, 1998.

[160] C. Févotte and C. Doncarli, "Two contributions to blind source separation using time-frequency distributions," *IEEE Sig. Process. Lett.*, vol. 11, no. 3, pp. 386–389, 2004.

[161] Y. Luo, W. Wang, J. A. Chambers, S. Lambotharan, and I. Proudler, "Exploitation of source nonstationarity in underdetermined blind source separation with advanced clustering techniques," *IEEE Trans. Sig. Process.*, vol. 54, no. 6, pp. 2198–2212, 2006.