



**BINDING SERVICES**

Tel +44 (0)29 2087 4949

Fax +44 (0)29 20371921

e-mail [bindery@cardiff.ac.uk](mailto:bindery@cardiff.ac.uk)



# Signal Processing Algorithms for Brain Computer Interfacing

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

Leor Shoker

Supervisor: Dr. S. Sanei

Centre of Digital Signal Processing  
Cardiff University

2006

UMI Number: U584852

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U584852

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

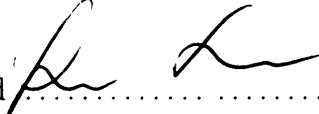
All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## Declaration


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

Signed  ..... (Candidate)

Date 25/9/06

## Statement 1

This thesis is the result on my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed  ..... (Candidate)

Date 25/9/06

## Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be available to outside organisations.

Signed  ..... (Candidate)

Date 25/9/06

## Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by the Graduate Development Committee.

Signed  ..... (Candidate)

Date 25/9/06

## Abstract

A brain computer interface (BCI) allows the user to communicate with a computer using only brain signals. In this way, the conventional neural pathways of peripheral nerves and muscles are bypassed, thereby enabling control of a computer by a person with no motor control. The brain signals, known as electroencephalographs (EEGs), are recorded by electrodes placed on the surface of the scalp.

A requirement for a successful BCI is that interfering artifacts are removed from the EEGs, so that thereby the important cognitive information is revealed. Two systems based on second order blind source separation (BSS) are therefore proposed. The first system, is based on developing a gradient based BSS algorithm, within which a constraint is incorporated such that the effect of eye blinking artifacts are mitigated from the constituent independent components (ICs). The second method is based on reconstructing the EEGs such that the effect of eye blinking artifacts are removed. The EEGs are separated using an unconstrained BSS algorithm, based on the principles of second order blind identification. Certain characteristics describing eye blinking artifacts are used to identify the related ICs. Then the remaining ICs are used to reconstruct the artifact free EEGs. Both methods yield significantly better results than standard techniques. The degree to which the artifacts are removed is shown and compared with standard methods, both subjectively and objectively.

The proposed BCI systems are based on extracting the sources related to finger movement and tracking the movement of the corresponding signal sources. The first proposed system explicitly localises the sources over successive temporal

windows of ICs using the least squares (LS) method and characterises the trajectories of the sources. A constrained BSS algorithm is then developed to separate the EEGs while mitigating the eye blinking artifacts. Another approach is based on inferring causal relationships between electrode signals. Directed transfer functions (DTFs) are also applied to short temporal windows of EEGs, from which a time-frequency map of causality is constructed. Additionally, the distribution of beta band power for the IC related to finger movement is combined with the DTF approach to form part of a robust classification system.

Finally, a new modality for BCI is introduced based on space-time-frequency masking. Here the sources are assumed to be disjoint in space, time and frequency. The method is based on multi-way analysis of the EEGs and extraction of components related to finger movements. The components are localised in space-time-frequency and compared with the original EEGs in order to quantify the motion of the extracted component.

# Acknowledgements

Firstly, I would like to gratefully acknowledge my supervisor Dr Saeid Sanei for his enthusiastic supervision, valuable input, and guidance throughout this work.

I would also like to thank my second supervisor Professor Jonathon Chambers for his helpful advice and suggestions.

Finally, I would like to thank my family for their support, encouragement, and education.



# Statement of Originality

Chapters 4 to 8 of this thesis comprise of original work to the author's best knowledge except where referenced and stated. The novelty of the thesis stems from the following contributions

## **Chapter 4: Artifact Removal using Constrained Blind Source Separation**

- Implementation and application of a gradient version of SOBI for the separation of EEG signals
- Constraining the gradient SOBI algorithm so that it automatically rejects eye blinking artifacts.

## **Chapter 5: Artifact Removal from the EEGs using Blind Source Separation and Classification**

- Combining a gradient SOBI BSS algorithm with a classification system using support vector machines for automatic identification and removal of eye blinking artifacts.

## **Chapter 6: Brain Computer Interfacing by Localisation of Finger Movement Sources**

- Development and implementation of a localisation algorithm by combining blind source separation with a non-linear least squares algorithm for

localisation.

- Constraining the gradient SOBI algorithm so that it automatically rejects eye blinking artifacts.

### **Chapter 7: Brain Computer Interfacing by Extracting Propagation Factors using Directed Transfer Function**

- Application of directed transfer functions together with scalp projections of the independent components for the classification of left and right finger movements.

### **Chapter 8: Brain Computer Interfacing by Space-Time-Frequency Analysis**

- Development and application of a space- time-frequency masking method for extraction of the components that are disjoint in space, time, and frequency. The algorithm combines short-time Fourier transforms across all the EEG channels, with  $k$ -means clustering.

# Publications

The original contribution of this thesis is partially supported by the following publications

M. A. Latif, J. A. Chambers, S. Sanei, L. Shoker, "Localization of Abnormal EEG Sources using Blind Source Separation Partially Constrained by the Locations of Known Sources" *IEEE Signal Processing Letters*, vol. 13 (3), 2005, pp. 117-120.

L. Shoker, S. Sanei, and A. Sumich, "Distinguishing Between Left and Right Finger Movement from EEG using SVM" *Proceedings of the IEEE Engineering in Medicine and Biology Annual Conference, Shanghai, China*, September 1-4, 2005.

L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from Electroencephalograms using a hybrid BSS-SVM algorithm", *IEEE Signal Processing Letters*, vol. 12 (10), 2005, pp. 721-724.

J. Corsini, L. Shoker, S. Sanei, and G. Alarcon, "Epileptic seizure prediction from scalp EEG incorporating BSS", *IEEE Transactions on Biomedical Engineering*, vol. 53(5), 2006, pp. 790-799.

L. Shoker, S. Sanei, W. Wang, and J. Chambers, "Removal of eye blinking artifact from EEG incorporating a new constrained BSS algorithm", *IEE Journal*

*of Medical and Biological Engineering and Computing*, vol. 43, 2005, pp. 290-295.

L. Shoker, S. Sanei, and J. Chambers, “A hybrid algorithm for the removal of eye blinking artifacts from electroencephalograms”, *Proceedings of the Statistical Signal Processing Workshop, SSP2005*, France, February 2005.

S. Sanei and L. Shoker, “Artefact Removal from EEGs Using a Hybrid BSS-SVM Algorithm”, Invited talk, *IEE Biomedical Signal Processing Workshop*, London, 2004.

L. Shoker, M. A. Latif, S. Sanei “A New Constrained BSS Algorithm for Separation of EEG Signals with Eye-Blinking Artifact”, *Proceedings of the IEEE Sensor Array And Multichannel Signal Processing Workshop, SAM2004*, Spain, 2004.

# List of Abbreviations

<b>AIC</b>	Akaike Information Criterion
<b>AMUSE</b>	Algorithm for Multiple Unknown Signal Extraction
<b>AP</b>	Action Potential
<b>AR</b>	Auto Regressive
<b>BP</b>	Bereitschafts Potential
<b>BCI</b>	Brain Computer Interface
<b>BSS</b>	Blind Source Separation
<b>BSS-SVM</b>	Hybrid Blind Source Separation Support Vector Machine
<b>CBSS</b>	Constrained Blind Source Separation
<b>CNS</b>	Central Nervous System
<b>CSP</b>	Common Spatial Patterns
<b>CT</b>	Computed Tomography
<b>CV</b>	Cross Validation
<b>DTF</b>	Directed Transfer Function
<b>ECG</b>	Electrocardiogram
<b>ECoG</b>	Electrocorticogram
<b>EEG</b>	Electroencephalograph
<b>EMG</b>	Electromyogram

<b>EOG</b>	Electrooculogram
<b>EP</b>	Evoked Potential
<b>ERD</b>	Event Related Desynchronisation
<b>ERS</b>	Event Related Synchronisation
<b>ERP</b>	Event Related Potential
<b>FFT</b>	Fast Fourier Transform
<b>GA</b>	Gradient Algorithm
<b>HOS</b>	Higher Order Statistics
<b>IC</b>	Independent Component
<b>ICA</b>	Independent Component Analysis
<b>InfoMax</b>	Information Maximisation
<b>ISTFT</b>	Inverse Short Time Frequency Transform
<b>KKT</b>	Karush Kuhn Tucker
<b>KL</b>	Kullback-Leibler Divergence/Distance
<b>LDA</b>	Linear Discriminant Analysis
<b>LFP</b>	Local Field Potential
<b>LHS</b>	Left Hand Side
<b>LS</b>	Least Squares
<b>MRI</b>	Magnetic Resonant Imaging
<b>MVAR</b>	Multivariate Auto Regressive
<b>OA</b>	Ocular Artifact
<b>OSH</b>	Optimal Separating Hyperplane
<b>PCA</b>	Principal Component Analysis
<b>PDF</b>	Probability Density Function
<b>PET</b>	Positron Emission Tomography
<b>RBF</b>	Radial Basis Function

<b>RHS</b>	Right Hand Side
<b>RP</b>	Readiness Potential
<b>SCP</b>	Slow Cortical Potential
<b>SDTF</b>	Short Time Directed Transfer Function
<b>SOBI</b>	Second Order Blind Identification
<b>SOS</b>	Second Order Statistics
<b>STF</b>	Space-Time-Frequency
<b>STFD</b>	Space-Time-Frequency Distribution
<b>STFT</b>	Short Time Frequency Transform
<b>SV</b>	Support Vector
<b>SVM</b>	Support Vector Machine
<b>WSS</b>	Wide Sense Stationary

# List of Nomenclatures

$t$	Discrete time index
$(\cdot)^T$	Transpose of matrix or vector
$(\cdot)^\dagger$	Pseudoinverse of matrix
$(\cdot)^{-T}$	Pseudoinverse and transpose of matrix
$\det(\mathbf{W})$	Determinant of matrix $\mathbf{W}$
$\ \mathbf{W}\ _F$	Frobenius norm of matrix $\mathbf{W}$
$ \cdot $	Absolute value
$p(x)$	Probability density function of variable $x$
$\mathbf{s}(t)$	Vector of sources at discrete time $t$
$\mathbf{x}(t)$	Vector of observed EEGs at discrete time $t$
$\mathbf{y}(t)$	Vector of estimated sources at discrete time $t$
$\mathbf{A}$	Mixing matrix
$\mathbf{W}$	Separation or unmixing matrix
$\mathbf{I}$	Identity matrix
$E\{\cdot\}$	Expectation
$\mathbf{R}_x(k)$	Estimated covariance matrix of $\mathbf{x}(t)$ at lag $k$
$[R]_{ij}$	The $ij^{th}$ element of matrix $\mathbf{R}$
$T_B$	Data length in samples



$\arg \min_{\mathbf{W}} f(\mathbf{W})$	The value of $\mathbf{W}$ that minimises the function $f(\mathbf{W})$
$\max(\cdot)$	maximum value
$\mathbf{P}$	Permutation matrix
$\mathbf{B}_w$	Whitening or sphering matrix
$\mathbf{0}$	Matrix of zeros
$\sigma_v$	Standard deviation of variable $v(t)$
$\text{diag}(\cdot)$	Diagonal of a matrix
$KL(p(x_1)  p(x_2))$	Kullback-Leibler distance between PDF of variables $x_1$ and $x_2$
$\alpha_i$	Lagrangian Multiplier used in SVM
$\varphi(\cdot)$	Nonlinear function used in SVM
$C$	Regularisation parameter used in SVM
$K(\cdot, \cdot)$	Kernel function used in SVM
$\mathbf{g}_i$	$i^{\text{th}}$ feature vector used in SVM

# Contents

<b>1</b>	<b>Brain Computer Interfacing; Introduction</b>	<b>23</b>
1.1	Aims and Objectives . . . . .	25
1.2	Thesis Outline . . . . .	26
<b>2</b>	<b>Overview of the Electroencephalogram</b>	<b>28</b>
2.1	Introduction . . . . .	28
2.2	Brain Imaging . . . . .	29
2.3	Anatomical Makeup of the Brain . . . . .	31
2.3.1	The Structure of Neurons . . . . .	32
2.3.2	Cortical Areas . . . . .	35
2.4	Acquisition Methods . . . . .	37
2.4.1	Electrode placement and Configuration . . . . .	38
2.5	EEG Signal Properties . . . . .	39
2.5.1	Rhythmicity . . . . .	39
2.5.2	Event Related Potential . . . . .	40
2.5.3	Event Related Desynchronisation / Synchronisation . . . . .	43
<b>3</b>	<b>State of the Art in Brain Computer Interfacing</b>	<b>45</b>
3.1	Artifact Rejection . . . . .	46
3.2	Blind Source Separation in Artifact Rejection . . . . .	48

3.2.1	Whitening . . . . .	52
3.2.2	Information Theoretic Based BSS . . . . .	53
3.2.3	Second Order Blind Identification . . . . .	54
3.3	Approaches to Brain Computer Interfacing . . . . .	58
3.3.1	Autoregressive Modelling . . . . .	58
3.3.2	Complexity Measure . . . . .	59
3.3.3	Time-Frequency Methods . . . . .	61
3.3.4	Common Spatial Patterns . . . . .	61
3.3.5	Blind Source Separation in BCI . . . . .	62
<b>4</b>	<b>Artifact Removal using Constrained Blind Source Separation</b>	<b>64</b>
4.1	Joint Diagonalization of Correlation Matrices . . . . .	65
4.2	Constrained Learning . . . . .	66
4.3	Experiments . . . . .	71
4.3.1	Simulated Source Signals . . . . .	71
4.3.2	Removing The Effect of Eye Blinking From Real EEG Data	75
4.3.3	Removing The Effect Of ECG From EEG Real Data . . . .	78
4.4	Conclusions . . . . .	80
<b>5</b>	<b>Artifact Removal from the EEGs using Blind Source Separation and Classification</b>	<b>88</b>
5.1	Introduction . . . . .	88
5.2	Methods . . . . .	89
5.2.1	Feature Extraction . . . . .	89
5.2.2	Classification . . . . .	92
5.3	Experiments . . . . .	96
5.3.1	Dataset for analysis . . . . .	96
5.3.2	Testing the Features . . . . .	96

5.4	Conclusions . . . . .	100
<b>6</b>	<b>Brain Computer Interfacing by Localisation of Finger Movement</b>	
	<b>Sources . . . . .</b>	<b>104</b>
6.1	Methods . . . . .	105
6.1.1	BSS and the Permutation Problem . . . . .	105
6.1.2	Localisation Algorithm . . . . .	107
6.1.3	Motion Characterisation . . . . .	111
6.2	Experiments . . . . .	113
6.2.1	Data Collection . . . . .	113
6.2.2	Results . . . . .	113
6.3	Conclusions . . . . .	116
<b>7</b>	<b>Brain Computer Interfacing by Extracting Propagation Factors</b>	
	<b>using Directed Transfer Function . . . . .</b>	<b>120</b>
7.1	Introduction . . . . .	120
7.2	Methods . . . . .	124
7.2.1	Feature Extraction . . . . .	124
7.3	Experiments . . . . .	129
7.3.1	Data Collection . . . . .	129
7.3.2	Testing the Features . . . . .	130
7.4	Conclusions . . . . .	134
<b>8</b>	<b>Brain Computer Interfacing by Space-Time-Frequency Analysis</b>	<b>142</b>
8.1	Introduction . . . . .	142
8.2	Methods . . . . .	144
8.2.1	Space-Time-Frequency Analysis . . . . .	144
8.2.2	Clustering . . . . .	146

8.2.3	Reconstruction . . . . .	148
8.2.4	Motion Characterisation . . . . .	148
8.3	Experiments . . . . .	149
8.3.1	Testing the Algorithm . . . . .	149
8.4	Conclusions . . . . .	151
<b>9</b>	<b>Conclusions and Future Work</b>	<b>154</b>
9.1	Conclusions . . . . .	154
9.2	Future work . . . . .	158

# List of Figures

2.1	A cross-sectional view of the brain. The main areas are highlighted and labelled [1]. . . . .	31
2.2	Structure of a neuron [1]. The neuron is made up from a cell body, dendrites, axon, and synapse. . . . .	33
2.3	Ion exchange between the inside of a neuron cell and extracellular fluid. . . . .	34
2.4	The cortical areas of one hemisphere and their associated functions [2]. . . . .	37
2.5	A block diagram of the EEG acquisition system. . . . .	37
2.6	A diagram of the international 10-20 system for electrode placement. The diagram shows a 21 electrode system [3]. . . . .	39
2.7	An example of the P100 evoked potential. The labels on the y-axis are the number of trials that have been averaged. . . . .	41
3.1	An example of EEG that is contaminated by eye blinking artifact	47
3.2	The ensemble average of autocorrelation at $r_x(\tau)$ , $\tau = 3$ , for 30 real EEGs (solid line). The dotted line shows the standard deviation in the estimate of the ensemble average of $r_x(\tau)$ at each consecutive window. . . . .	51

4.1	The solution space for an optimisation problem is shown in the area within the solid line and a solution which satisfies a constrained optimisation problem, i.e. $\mathcal{W}$ is shown in the hatched area. . . . .	67
4.2	Original speech like source signals. The third signal represents the artifact. . . . .	72
4.3	Artificially mixed signals. . . . .	73
4.4	Estimated sources with the artifact minimised. . . . .	73
4.5	Convergence performance of the proposed algorithm (solid line) compared with that of an unconstrained gradient SOBI [4] algorithm (dotted line). . . . .	74
4.6	A selection of five EEG channels from a 16 channel EEG recording. The EEGs on these channels are corrupted by the ocular artifact between samples 600 to 900. . . . .	76
4.7	The vertical EOG signal measured from the right eye. . . . .	77
4.8	A selection of five independent components (ICs) derived from the EEG primarily corrupted by ocular artifact. The ICs represent the EEG with the EOG artifact removed. . . . .	78
4.9	The EEGs heavily contaminated by eye blinking artifact. . . . .	79
4.10	The ICs of the EEGs in Fig. 4.9 using the proposed CBSS algorithm. . . . .	80
4.11	The ICs of the EEGs in Fig. 4.9 using an unconstrained BSS algorithm. . . . .	81
4.12	The EEGs in Fig. 4.9 separated using PCA. . . . .	82

4.13	The reprojected ICs to the scalp electrodes after application of the proposed CBSS algorithm. Each scalp plot represents the projection strength of one IC in Fig. 4.10, designated by the label beneath the scalp plot. The colour represents projection strength of the IC onto each electrode and is normalised to unity across all electrodes (arbitrary units). The frontal electrodes are located towards the top of the scalp plot. In comparison with Fig. 4.14 the projection strengths of the CBSS ICs are not smeared over the frontal electrodes. . . . .	84
4.14	The reprojected ICs to the scalp after application of the unconstrained BSS algorithm. Each scalp plot represents the projection strength of one IC in Fig. 4.11, designated by the label beneath the scalp plot. The colour represents projection strength of the IC onto each electrode and is normalised to unity across all electrodes (arbitrary units). The frontal electrodes are located towards the top of the scalp plot. Point ‘A’ in the figure highlights the limitation of the unconstrained BSS algorithm, in that the projections are smeared over a number of electrodes. A similar pattern can be seen in a number of other electrodes. . . . .	85
4.15	A selection of five channels from the EEG recording. There is an obvious ECG artifact present in the first and fourth channels of the figure. . . . .	86
4.16	A selection of five independent components after the CBSS algorithm has removed the ECG. . . . .	87
4.17	The measured ECG reference signal. . . . .	87
5.1	Block diagram of the BSS and SVM system. . . . .	89



5.2	(a) Three features plotted against each other forming the feature space. (b) The optimum separating hyperplane for the feature space in (a). . . . .	93
5.3	The feature space for a nonseparable case. The circled points are the support vectors calculated by minimising (5.5). The slack parameter $\gamma_i$ enables the use of Lagrangian theory since it can now account for the overlapping features. . . . .	95
5.4	(a) A plot of the two largest principal components of the feature space. There are 200 feature vectors, 100 from normal EEG (+) and 100 from EEG containing eye blinks ( $\circ$ ). (b) A histogram plot showing the output of the classifier pre $sgn(\cdot)$ using the linear kernel.	97
5.5	The (a) classification rate and (b) number of support vectors required for various parameter values of the RBF kernel. . . . .	102
5.6	A selection of 8 electrodes from a 16 electrode EEG recording. The OAs are clear in (a) between samples 400 to 600, 900 to 1400, and 1700 to 1900. They are more prominent over the frontal electrodes (FP1, FP2 etc.). (b) The same segment of EEGs after the eye blinking artifacts are removed using the proposed BSS-SVM algorithm. . . . .	103
6.1	A block diagram of the localisation based BCI system. . . . .	105
6.2	Part of the scalp including three electrodes, and the location of the source to be identified (assuming the head is homogenous) . . . .	107
6.3	The relationship between the conductivity and the distance from the source to the electrode. When the source is close to the scalp the conductivity is large compared to sources that are deep within the brain as they have to travel through the skull. . . . .	109

6.4	The relationship between the cross correlation and the distance from the source. Sources located closest to the electrode have larger cross correlation value than those located farther away. Cross correlation approximately reflects the conductance of various regions in the head (as in Fig. 6.3). . . . .	110
6.5	The results of the localisation algorithm for synthetic stationary and moving sources. The squares ( $\square$ ) are the known sensor locations, the circles ( $\circ$ ) are the known source locations, ( $\times$ ) are the locations calculated directly from the actual distances, and asterisks ( $*$ ) are the source locations calculated from the cross correlation value in (6.4). The numbers close to the estimated locations of the moving source identify the order of the consecutive localisation windows, i.e. 1 being the first processing window, and so on. . . .	117
6.6	The results of the localisation algorithm for real EEGs containing left and right finger movement. The squares ( $\square$ ) are the known sensor locations and are labelled with the corresponding electrode name. Each $+$ and $\circ$ represent the source location of one processing window (for one IC) for left and right finger movement respectively. The lines between the $+$ and $\circ$ represent the trajectory of one estimated source between two consecutive localisation windows. For clarity the source trajectories of left finger movement are plotted using a solid line while right finger movement trajectories are plotted using a dash-dot line. The source locations are calculated assuming that they are proportional to the cross correlation value as in (6.4) . . . . .	118

6.7	The histogram plot for the motion characterisation algorithm collapsed across all trials. The number of times that an angle between two consecutive processing blocks for left and right finger movement trials are represented by black and white bar plots, respectively. This shows that the angle between motion vectors does not reveal any significant distinctive features between left and right finger movements. . . . .	119
7.1	An example of the spectral coherence for the EEGs one second prior to finger movement. . . . .	122
7.2	A block diagram of the proposed BCI system based on extracting the propagation of EEG sources and the location of beta activity.	124
7.3	A block diagram showing the localisation of beta band activity feature. The EEGs are processed in blocks of $T_B$ samples. The outputs of this feature are the reprojected beta band power values of the IC corresponding to finger movement i.e. the IC with the largest reprojected to the electrodes located over the motor cortex.	125
7.4	Block diagram of the system for classification of the propagation features. $\mathbf{X}$ is an EEG block. . . . .	126
7.5	The recording procedure for the BCI data. Each trial lasted for a total of 9 seconds. Between each trial there was an interval in which the user was able to blink or get comfortable. . . . .	130

- 7.6 SDTF calculated for left finger movements. Frequency is plotted on the  $y$  axis and time (seconds) along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds. The direction of flow is read from the electrode denoted above the column to the electrode denoted by the label on the left of the rows. Electrode FC6 (column) is the most active because of the high value of STDF (red), suggesting that the source of finger movement is located close the that electrode. . . . . 135
- 7.7 SDTF calculated for right finger movements. Frequency is plotted on the  $y$  axis and time (seconds) along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds. The direction of flow is read from the electrode denoted above the column to the electrode denoted by the label on the left of the rows. In this case electrodes CP5 and CP3 (columns) are the most active at the time of finger movement, suggesting that the source is located close to those electrodes. . . . . 136
- 7.8 The SDTF calculated for electrodes located over the left hemisphere for left finger movement trials averaged over the alpha band. The  $y$  axis shows the SDTF value, where a value of 1 indicates maximal causal relation between the electrode denoted above the column to the electrode denoted to the left of the row. Time (seconds) is shown along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds, which is indicated by the dotted line. . . 137

7.9	The SDTF calculated for electrodes located over the right hemisphere for left finger movement trials averaged over the alpha band. The axis are as in Fig. 7.8 Lower values of SDTF are observed near the time of finger movement compared with the left (ipsilateral) hemisphere (Fig. 7.8) . . . . .	138
7.10	The SDTF calculated for electrodes located over the left hemisphere for right finger movement trials averaged over the alpha band. The axis are as in Fig. 7.10. For right finger movement the SDTF is lower in the left hemisphere compared with that of the right hemisphere (Fig. 7.11) . . . . .	139
7.11	The SDTF calculated for electrodes located over the right hemisphere for right finger movement trials averaged over the alpha band. The axis are as in Fig. 7.8 . . . . .	140
7.12	A histogram plot showing the distribution of classifier output values using the RBF kernel tested on 100 trials, 50 from each class. The outputs for each class are clearly separated, which generally indicates that features are significant for detection of left and right finger movements. . . . .	141
8.1	Block diagram of the space-time-frequency extraction algorithm. . . . .	145
8.2	The cluster centres for the extracted atoms for 45 left and 45 right finger movement trials. . . . .	151
8.3	The time-frequency representation of the extracted atoms for a left finger trial. . . . .	152

# Chapter 1

## Brain Computer Interfacing; Introduction

In the world of computing there is an emphasis on the interface between human and the core program. The ergonomics of the interface between human and computer is known as Human Computer Interaction (HCI). The field of HCI has received a wealth of interest amongst researchers and industry which has led to many innovative methods of interacting with a computer e.g. via voice recognition, gesture recognition, and handwriting recognition. Underlying many advances in HCI is the assumption that the user has 'normal' motor control and coordination. Therefore, a person who suffers from a muscular disorder or a quadriplegic, who has no physical control, cannot reap the benefits that a computer brings.

A system with a direct communication path between the brain and the computer has long been a fantasy confined to the imagination of science fiction writers. However, with the advances in computing power and signal processing technology this dream is becoming more realisable. Direct brain-computer interaction,

or brain-computer interface (BCI), has inspired many interdisciplinary research associations between biomedical and engineering faculties. A BCI is defined as a system which allows the user to interact with a computer using brain signals alone.

There are a number of modalities in measuring the cognitive function of the brain, such as functional magnetic resonance imaging (fMRI), Magnetoencephalography (MEG), and Electroencephalography (EEG). The most viable modality for the design of a BCI is EEG, since it is a relatively low cost option for monitoring the brain function. EEGs have a very high temporal resolution since they are only limited by the acquisition speed of the analogue to digital converters. The current temporal resolution is  $200\mu s$ . However, the main disadvantage of EEGs is its poor spatial resolution, since they are limited to the number of electrodes that can be placed over the scalp. Embedded within the EEGs are cortical processes for many tasks, most of them unknown. The main aim of a BCI is to identify and exploit distinguishable states from the EEGs.

In general a BCI system comprises of three major components; an input, a translation algorithm, and an output [5]. The input to the BCI is, generally, an extractable feature of the brain's normal function. Traditionally, these features may be the activity in certain frequency bands over time over a certain region of the brain such as the motor cortex, which is involved with voluntary movements, or time domain features such as P300, or the action potentials of individual neurons. The translation algorithm converts the inputs (brain waves) to the outputs (commands) i.e. it decodes the features generated from the brain waves into output commands. In general, if the features are well separated in the feature space then linear classifiers are used, however if the feature space is not linearly separable non-linear methods for classification, such as those based on artificial neural networks (ANNs), will be favorable. BCIs can be divided into two main

categories; invasive and non-invasive. The former uses intracranial electrodes or subdural electrodes implanted, respectively, deep inside or on the surface of the brain, whereas the latter uses surface electrodes placed over the scalp.

The first attempt in the design of a BCI system was by Vidal [6] in 1973. In his attempt for BCI he used a parsimonious representation of the EEGs, namely using principal component analysis [6], for a variety of experimental setups. Current BCIs use one of a number of extractable EEG components, such as rhythmicities [7] in the data, such as slow cortical potentials (SCP) [8], or evoked potentials (EPs) [9]. EPs such as P300 are time-locked events which are, generally, extracted by averaging many trials of the same event.

## 1.1 Aims and Objectives

The main interfering signals in EEGs are eye blinking artifacts. These signals generally have detrimental effects on BCIs in that they may increase misclassification errors and thus lead to lower command throughput. Therefore, a requirement for a successful BCI is to remove such artifacts from the EEG signals. The effect that eye blinking artifacts have on EEGs and their signal characteristics will be investigated. The signal characteristics of the eye blinking artifacts will determine the *a priori* information that can be extracted from such signals and thereby facilitate their removal from the EEGs. The aim of the artifact removal algorithm is to condition the EEGs such that the BCI algorithm is presented with uncontaminated EEG signals.

Once suitable artifact correction algorithms are developed a survey of current BCI systems can be made. During planned finger movement the prefrontal region of the motor cortex is activated, then during execution of the movement the posterior region of the motor cortex is activated [10]. This suggests that the source



related to finger movement follows a particular path. Another explanation is that the two regions of the brain communicate in order to coordinate the movement. Therefore, for the BCI system the aim is to develop a BCI that identifies this characteristic through electrodes placed on the surface of the scalp and determines whether the left or right finger has been moved.

## 1.2 Thesis Outline

Chapter 2 gives the reader a general overview of the electroencephalogram (EEG) and other brain imaging techniques. The purpose of this chapter is to introduce the reader to the terminology and concepts of brain imaging techniques and become familiar with the basics in neuroscience.

In Chapter 3 a survey is carried out on the state of the art techniques used to process EEGs for the application to BCI. The background work will be used as a framework for the design of the BCI system and artifact rejection algorithms.

In Chapters 4 and 5 two types of preprocessing algorithm are developed for the removal of eye blinking artifacts. Both are based around blind source separation. In Chapter 4 a constrained blind source separation algorithm is developed to separate the EEGs into constituent independent components and at the same time mitigate the effect of eye blinking artifact. The algorithm uses the concepts derived from nonlinear programming, and these are used to solve the constrained optimisation problem. The preprocessing algorithm proposed in Chapter 5 uses a hybrid approach where blind source separation is coupled with classification in order to reconstruct the artifact free observations.

Chapters 6, 7, and 8 propose three approaches to the design of a brain computer interface. The method in Chapter 6 is based on localisation of the ERP sources related to finger movement. The algorithm provides a solution to the

geometrical LS, where the geometric distances are derived from ICs of a BSS algorithm. Though this method doesn't yield particularly acceptable results it gives an insight to the problems in localising the ERPs such as non-homogeneity of the brain regions. Chapter 7 uses the spectral correlations between electrodes to determine the direction of cortical flow during finger movement. The direction of cortical flow and the scalp projections of the ERP sources are used in the classification of left and right finger movements. The BCI system proposed in Chapter 8 is based on the extension of time-frequency masking to include the spatial dimensions.

Finally in Chapter 9, conclusions are presented with suggestions for further work.

# Chapter 2

## Overview of the Electroencephalogram

### 2.1 Introduction

The electroencephalogram (EEG) was discovered by Hans Berger in 1926. The EEG is commonly defined as the electrical activity recorded from the surface of the head, which is called the scalp as opposed to invasive methods which enable measurement of the EEG from inside the brain. The most common type of electrodes used to record the EEGs are the metal disc electrodes which are applied to the scalp with a saline gel. The saline gel acts as an electrolyte medium between the scalp and the electrode. Another type of EEG is electrocorticogram (ECoG), which uses subdural electrodes. Subdural electrodes are inserted into the scalp and measure the electrical activity from the dura, a membrane covering the brain. In intracellular recordings the individual neurons' activation is measured using an electrode inserted into the cell. Where as extracellular recordings are made using electrodes places within the brain tissue sufficiently close to the neurons.

It reflects the activation of local neurons, and hence is also known as local field potential (LFP). For the purpose of this research only surface electrodes will be used.

## 2.2 Brain Imaging

There are a variety of electro-physiological measurement technologies used to measure phenomena from different organs. These include the electro-oculogram (EOG) which is the measurement of the electrical activity generated by the muscles around the eye, the Electromyogram (EMG) which reflects the measurement of muscle activity, and the Electro-cardiogram (ECG) which shows the heart's activity.

Brain imaging techniques can be divided into those that reflect the anatomical information and those that highlight the functional regions within the brain. Additionally, imaging techniques that reflect brain functionality suffer from a trade off between spatial and temporal resolution. For example functional imaging techniques with high spatial resolution tend to have a low temporal resolution. An example of an anatomical imaging modality is computed tomography (CT) in which a series of X-ray beams strike the body at different angles and the cross-sectional image is reconstructed by the computer. The spatial resolution in CT is very high but the temporal resolution is very low, on the order of 10 seconds per scan for brain. The brightness of the image is proportional to the density of the tissue. The CT images only reflect anatomical information and are used in conjunction with other imaging modalities to highlight any pathological information. The main disadvantage of CT is that it exposes the subject to radiation, which limits the number of scans that can be performed on one person, since exposure to ionising radiation may cause malignancy.

Magnetic resonance imaging (MRI) is a safer option to CT. It measures changes in electromagnetic activity within a large magnetic field. Pulses of radio waves build up a two/three dimensional image of the brain. MRI provides high resolution (approximately 1mm) two/three dimensional images of the anatomical make up of the brain, but no functional information. Functional magnetic resonance imaging (fMRI), however, reflects the reaction of the oxygen molecules against a controlled magnetic field and hence highlights the functional properties of the brain. fMRI has a high spatial resolution (between 2mm to 4mm) but low temporal resolution (approximately 4 seconds) and the equipment is very expensive.

Magnetoencephalography (MEG) refers to measurement of magnetic fields produced by the electrical activity in the brain. More specifically, MEG measures magnetic fields that are generated as a result of ionic currents flowing through the dendrites (see next section). The magnetic fields generated by the brain are on the order of 100 to 1000 fT (femto Tesla, femto =  $10^{-15}$ ) and therefore require specialised (and expensive) equipment to detect such small magnetic fields. Typically, MEG is acquired from 300 recording electrodes (Superconducting Quantum Interference Devices - SQUIDS) placed around the scalp in a radio frequency and magnetically shielded room. Unlike EEGs, MEGs are not distorted by the body and hence make it ideal for localisation studies. The most common use for MEG is localising sources in the primary auditory, somatosensory, and motor areas.

Positron emission tomography (PET) measures the metabolic rate in the blood. Before recording PET, the subject is administered a short lived radioactive substance, known as radiotracer, which emits a positron as it decays. When the positron strikes the electron of oxygen molecule, two photons are emitted in equal and opposite directions, which is then detected by the scanning device. More active regions in the brain having larger blood flow would elicit a larger

response on the scanning device. PET reflects the metabolic processes and therefore typically combined with CT or MRI to show the structural and metabolic information simultaneously. PET typically has a lower spatial resolution than fMRI and suffers from a poor temporal resolution of about 15-20 seconds.

The above mentioned imaging technologies show physiological or pathological properties. Although EEGs are mainly used to determine functional properties of the brain, by using advanced signal processing techniques other properties can also be investigated.

## 2.3 Anatomical Makeup of the Brain

The brain can be divided into six major anatomical areas; cerebral hemisphere, thalamus, midbrain, cerebellum/pons, and medulla oblongata (medulla), as shown in Figure 2.1. The medulla contains neurons passing through it into the rest of

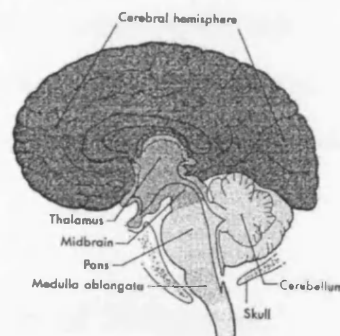


Figure 2.1: A cross-sectional view of the brain. The main areas are highlighted and labelled [1].

the brain. In general the neuronal information that ascends through the medulla is associated with sensory inputs to the brain, such as touch, smell, etc. On the contrary, the information passing through from the medulla to the rest of the

body are associated with voluntary movements. The medulla is also responsible for the regulation of respiratory organs and heart. The cerebellum is considered as part of the motor system and contains the source of facial nerve endings. The midbrain is considered as the link between the brainstem and the forebrain [1]. It contains neural pathways connected to visual, auditory and motor systems. The midbrain is also responsible for the transmission of pain. The thalamus can be seen as the central junction for all of the sensory inputs, with the exception of smell. Its function is to distribute the information to the relevant parts of the cerebral hemisphere. The cerebral hemisphere is the largest part of the brain. It is divided into three areas; cerebral cortex, subcortical white matter, and basal ganglia. On the surface of cerebral cortex there are peaks and valleys of 0.5cm deep, which give the well known look of the brain. The peaks and valleys are respectively known as gyri and sulci. Certain areas of the cerebral cortex are associated with sensory input, such as vision, touch etc, whereas others are associated with voluntary movement and cognitive thought.

In order to understand how the electrical currents are generated within the brain, it is important to understand the most basic unit of the make up of the brain, the neuron.

### **2.3.1 The Structure of Neurons**

The central nervous system (CNS) consists of 100 billion cells for which the two most common types are Neurons and Glia cells [1]. Neurons, also known as nerve cells, are the fundamental building blocks of the brain. A neuron transmits, receives and processes information from other neurons or tissue by changing its biochemical properties. All neurons produce the same electrical signal [11]. Neurons are made up from three major parts as in Figure 2.2;

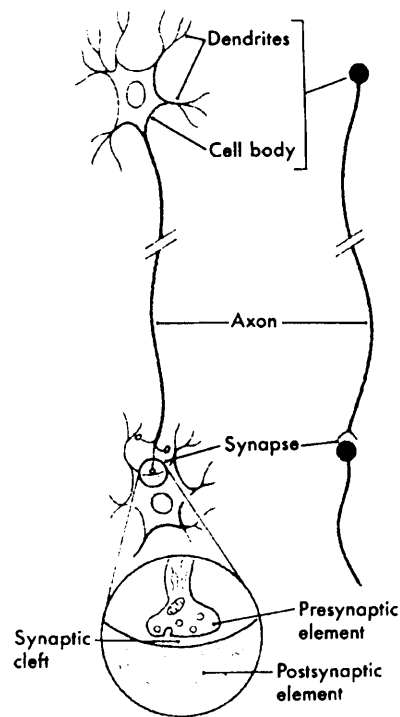


Figure 2.2: Structure of a neuron [1]. The neuron is made up from a cell body, dendrites, axon, and synapse.

## Dendrites

Dendrites are tree-like structures which extend from the central part of neuron, the cell body (see next section). They receive signals from synapses, which are junctions between the tips of the dendrites and cell bodies.

## Cell body (Soma)

The cell body, or soma, is the central part of neuron. It is similar to a regular cell with the exception that it can modify its biochemical processes to communicate with other neurons. It contains deoxyribonucleic acid (DNA) in its nucleus, ribosomes in its cytoplasm (the plasma between the nucleus and the cell wall) for



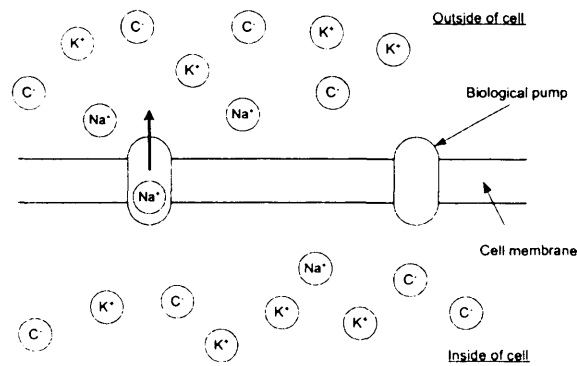


Figure 2.3: Ion exchange between the inside of a neuron cell and extracellular fluid.

building proteins. The cell body draws its energy from mitochondria within the cytoplasm.

### Axon

An axon is a long trunk like structure, typically greater in length compared to dendrites, which carries the information from the cell body to the dendrites. Its cytoplasm contains a large number of microtubules and neurofilaments which are tiny tube like structures that carry metabolic information between the cell body and dendrites. The speed at which the axon can transmit information is directly proportional to its thickness. Most human axons are  $30\mu m$  thick, which is very thin compared to that of other animals such as squid.

As mentioned earlier a neuron communicates by changing its biochemical properties. The axon's membrane is permeable to positively charged potassium ( $K^+$ ) and negatively charged chloride ( $Cl^-$ ) ions but is impermeable to sodium ( $Na^+$ ). Since an unequal amount of  $Cl^-$  is distributed across the cell wall a voltage gradient forms at the junction between the intracellular plasma and the

outside of the cell (see Fig. 2.3). The inside of an inactive cell (intracellular) rests at  $-70\text{mV}$  relative to the outside of the neural cell (extracellular), this is known as the membrane potential or resting potential. The resting potential is maintained with biological pumps on the membrane which regulate the amount of  $\text{Na}^+$  inside the axon.

When a soma transmits an electrical pulse the intracellular concentration of sodium within the axon decreases causing the membrane voltage to increase to approximately  $30\text{mV}$ . This is known as an action potential (AP). Long axons, such as peripheral nerve axons, are insulated with a myelin sheath with gaps at regular intervals, where the axon is not insulated. These gaps are known as Ranvier node. The myelin sheath prevents the axon from producing AP, i.e. it creates an ionic barrier. Only regions of the axons which are not insulated can transmit action potentials and the distance between Ranvier nodes is sufficiently small such that the action potential can propagate along the axon. This overcomes the relative slowness of information transfer since the AP propagates in discrete steps from one Ranvier node to the next.

When an AP reaches the end of the axon it is at a part called a synapse. A synapse is the junction between the dendrites, muscle or glands etc. It contains a chemical called neurotransmitter which either excites or inhibits the recipient cell. The action depends on the type of cell it is connected to (the postsynaptic cell).

### **2.3.2 Cortical Areas**

The cerebral cortex is divided into six neuronal layers which are parallel with the scalp, and consists of a large number of pyramidal neurons. The dendrites of the pyramidal neurons are perpendicular to the scalp and hence have the greatest

effect on the EEG. Most of the neurons in the cerebral cortex are arranged in this way and hence the observed EEG proportional to summation of the potentials radiating from these dendrites. Additionally, the thalamus and brainstem regulate the rhythm of the pyramidal cells by acting as a biological pace maker [12]. The neuronal rhythms can also be affected by respiration, heart rate, or any events which affect the oxygen and glucose level in the cells. The EEG can only detect large populations of neurons that are in near synchrony.

The cerebral cortex is divided into two halves or hemispheres. In general, the left hemisphere controls the right half of the body whereas the right hemisphere controls the left half of the body. Each hemisphere is divided into various functional areas as shown in Figure 2.4. The premotor cortex is associated with the preparation of voluntary movements, where as the primary motor cortex is associated with voluntary movement. If the cells in one part of the motor area are electrically stimulated the subject will move one side of the body, depending on which section of the motor cortex is stimulated. In an experiment by Stewart *et al.* [13], 12 subjects were asked to listen to a familiar piece of music while the brain activity was being monitored using fMRI. The aim of this experiment was to monitor the cortical associations between piano players and non piano players. The experiments show that the motor cortex areas associated with the fingers and thumbs would be active in the piano players but not in the non piano players. The sensory area receives sensory information from the rest of the body, such as taste, smell, touch etc. The frontal lobe acts as a short term memory storage, where imagery and higher cognitive function take place.

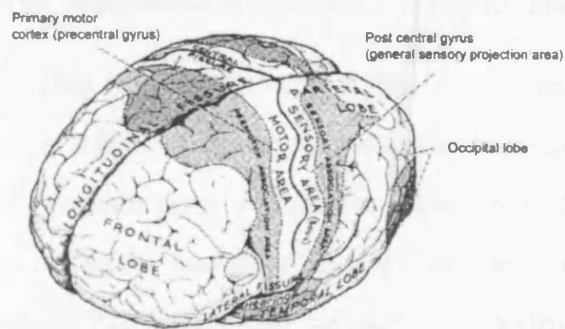


Figure 2.4: The cortical areas of one hemisphere and their associated functions [2].

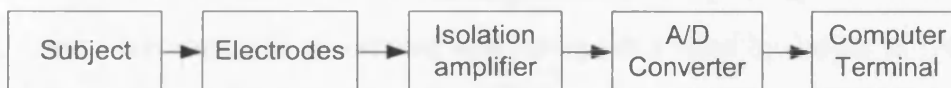


Figure 2.5: A block diagram of the EEG acquisition system.

## 2.4 Acquisition Methods

The signals recorded using scalp EEG electrodes are on the order of microvolts and are digitised before being processed by the clinician. A block diagram for a typical EEG recording is shown in Figure 2.5. The electrodes are silver-silver/chloride disc electrodes and are applied to the scalp individually or using a pre configured electrode cap. The isolation amplifier protects the subject from the large currents from the computer terminal and amplifies the signal before being digitised. The computer terminal stores and displays the EEG on the screen for the clinician. Typically the EEGs are bandpass filtered between 0.1Hz and 100Hz. Additionally, a 50Hz notch filter is applied to remove the effects of mains frequency.

### 2.4.1 Electrode placement and Configuration

During the EEG recording it is important that the connection impedance for each electrode is kept below  $5K\Omega$ . In order to keep the impedance low, the skin is lightly abraded so that a thin layer of the skin is removed. Additionally an ionic conductive paste is applied to the electrodes to facilitate current flow from the brain to the electrodes. Electrodes can be applied individually to the scalp or by using an electrode cap. An electrode cap is an elastic cap with the electrodes attached at predefined locations. If an electrode cap is used, a blunt siring is used to inject the saline gel into the electrode and abrade the skin.

In most cases the electrodes are arranged on the scalp using the International 10-20 system for electrode placement which was introduced by Jasper in 1958 [14]. The 10-20 system is used for uniformity of the recordings between EEG recordings and across different patients. The name 10-20 system arises from the distances between electrodes, as they are spaced at regular intervals defined by 10% or 20% of the distance between the nasion and inion. The nasion is the bone between the eyes and the inion is the prominent bump located over the occipital cortex (to the rear of the head). A 10-20 system for 21 electrodes is shown in Figure 2.6. The letters in the 10-20 system refer to the region of the brain that the electrode is closest; frontal lobe (F), parietal lobe (P), temporal (T), central (C) and occipital (O). Odd label numbers are located on the left hemisphere and even numbers are located on the right hemisphere, the label 'z' is located on the midline. In some EEG recordings, such as localisation studies, up to 256 electrodes are used and therefore researchers modify the 10-20 system to accommodate the larger number of electrodes.

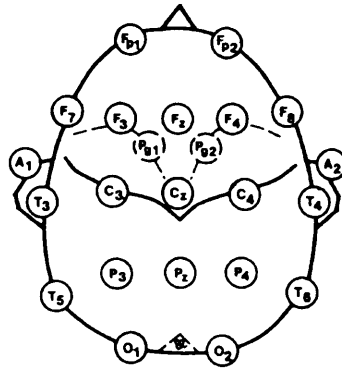


Figure 2.6: A diagram of the international 10-20 system for electrode placement. The diagram shows a 21 electrode system [3].

## 2.5 EEG Signal Properties

This section explains various ways that EEG can be grouped based on properties of the signals. The most obvious descriptor of EEG is its morphology for example, rhythmicity, spindles, or spikes. The frequency content, amplitude, periodicity, and distribution are all examples of EEG descriptors.

### 2.5.1 Rhythmicity

The frequency spectrum of the EEG can be divided into four main frequency components known as delta, theta, alpha, beta, and gamma.

**Delta (0Hz - 4Hz)** Dominant in infants and are active in the central cerebellum and frontal lobes. They appear in adults during deep sleep and when brain damage has occurred.

**Theta (4-8)** Appears over the frontal, temporal, and parietal regions in drowsy or emotionally stressed normal adults. Diffuse theta commonly signifies abnormal activity.

**Alpha (8-13)** The most prominent rhythm in normal adults observed over the occipital and parietal regions. Alpha occurs when the person is relaxed and comfortable. The non-dominant hemisphere (right for right handed persons) typically has a larger amplitude in the alpha band. They are more prominent when the eyes are closed and when the person is in a relaxed state.

**Beta (>13)** Is mainly observed over the anterior region of the brain and has a low amplitude compared with the amplitude in the other frequency ranges. It can be divided into three categories based on its distribution over the scalp; Frontal beta is blocked by movement, Widespread is unreactive, and Posterior reacts to opening eyes

The frequency band above 30Hz is also known as gamma band.

## 2.5.2 Event Related Potential

Event related potentials (ERPs) are potential changes that are in response to a stimulus or event. They are used to evaluate a variety of cognitive processes and neurological disorders, by analysing the spatial and temporal distribution of the ERP. In general, ERPs are time locked to the stimulus and therefore the most common method for extracting the ERP is to ensemble average over a number of trials, see Fig. 2.7 for an example of the P100 ERP. Trials that contain artifacts are manually excluded from the ensemble. This method for extraction assumes that the background EEG is a zero mean ergodic process and therefore will cancel out revealing the ERP. The fundamental flaw in this method is that it ignores the event related activity that is not well represented by the aforementioned assumption [15].

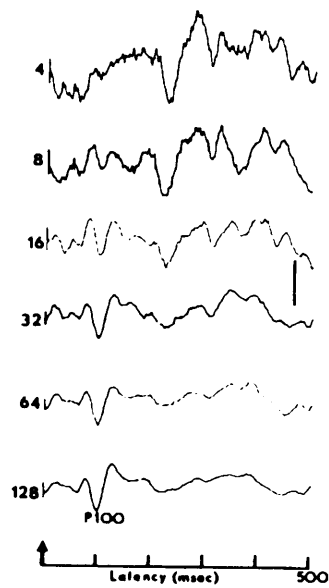


Figure 2.7: An example of the P100 evoked potential. The labels on the y-axis are the number of trials that have been averaged.

## Evoked Potential

Evoked potentials (EP) are time locked responses that evoke a response from the sensory organs, such as visual and auditory systems, as opposed to those evoked responses from other neural structures. The EPs are direct results of reorganisation of the phases in neural structures related to the sensory organs. They have a number of distinguishable features which are used by clinicians for diagnosing schizophrenia and other neurological and mental disorders [16]. The most obvious identifiers of EPs are the latency and the sign of the EP. Typically, the notation for EPs is that a P or N, indicating positive or negative potential, precedes a number which signifies the latency of the EP. As an example, the EP in Fig. 2.7 is a P100 meaning that it is a positive potential with a latency of 100 ms. The spatial distribution of the EP is directly related to the stimuli presented, for



example an inverting checkerboard pattern elicits a response in the visual cortex, while an auditory click elicits a response in the auditory cortex. This type of EP is termed visually evoked potential (VEP). An experiment by Sutton *et al.* [17], proved that there is a larger response when the subject is uncertain to what the stimulus will be. An experimental paradigm where there was an uncertainty to whether the stimuli would be a flashing light or an auditory click after a random interval proved that, in comparison with certainty in the stimuli, when the stimuli is unknown *a priori* the EP is more pronounced and larger in magnitude. This uncertainty is known as the ‘oddball’ paradigm.

A number of BCI researchers have used EPs as the control signals in their development of BCIs. For example the P300 was used to implement a spelling device [18] by randomly flashing rows of a matrix containing letters of the alphabet. In another application P300 was used to manoeuvre the user through virtual worlds [9].

### **Bereitschafts Potential**

Bereitschafts potential (BP), also known as readiness potential (RP) or slow cortical potential (SCP), is a slow moving wave (0.1-0.5Hz) which precedes imagined or real movement by up to one second [19]. Planning and organisation of movement involves a number of brain structural components such as the primary motor area, premotor area, secondary somatosensory area, basal ganglia, supplementary motor area (SMA), thalamic nuclei, and the cerebellum [20]. It is believed that the communication and spatial organisation of these structures are the reason for the SCP [21]. One can learn to control the amplitude of one’s SCP [22]. BCIs based on regulation of SCP exploit the adaptive nature of the brain by using biofeedback. In biofeedback the brain is placed in the adaptive loop of the BCI via a, typically visual, feedback. A typical experimental setup of biofeedback

places the subject in front of a computer monitor with a cursor at the centre of the screen. The user is then asked to make the cursor move up. The user is not instructed on how to achieve this goal, however they are 'rewarded' for making the cursor move up or down. After a number of experimental sessions the user is be able to successfully move the cursor on demand. In one experiment the user thought about playing in the park in order to move the cursor upwards [23]. Essentially the brain was placed in the feedback loop and therefore adapted in order to achieve the goal of moving the cursor up or down.

### 2.5.3 Event Related Desynchronisation / Synchronisation

Another example of an ERP is the Event Related Desynchronisation/Synchronisation (ERD/ERS)[24][3]. The ERD/ERS is the result of a decrease or increase in the synchrony of a neuronal population for ERD and ERS respectively. This results in a decrease/increase in power in a certain frequency band. This power variation reflects the local change in interneuron activity, and this in turn reflects the frequency spectrum of the ongoing EEG. A typical task that elicits an ERD is perceptual, judgemental or memory task, where the magnitude of the desynchronisation is proportional to the complexity of the task [3]. The behavior of EEG just before and during voluntary finger movement was first quantified in [25]. The experiments in [25] quantified the ERDs as follows:

1. Band pass filter the EEG between frequencies 1.6Hz - 30Hz.
2. Average a 2s window beginning 4s prior to movement, call this value R.
3. Average consecutive 250ms windows, call them A(t).
4. Then  $ERD(t) = \frac{R-A(t)}{R} 100$ .

It was observed that the ERD occurs on average, 1.7 seconds before the finger movement [20]. This means that the brain knows which finger will move well before the action has taken place. The ERD is most prominent over the contralateral motor cortex in the upper alpha band (11-12Hz). ERS can be observed 0.5 to 1 second after the finger movement over the ipsilateral hemisphere. ERD/ERS are traditionally extracted by averaging 10-20 trials of the same event, such as pressing a micro switch with the index finger. This method of extraction assumes that the ERP has a fixed time delay and the background EEG behaves like additive uncorrelated noise [26]. However, this simple assumption does not hold since, in general, ERD/ERS is time locked but not phase locked. Therefore, a method based on frequency analysis may be more effective in extracting the ERD/ERS. There are many methods for the quantification of ERD/ERS e.g. the band power method [27] and those based on the Hilbert transform [28].

# Chapter 3

## State of the Art in Brain

## Computer Interfacing

The field of BCI has made considerable advances since Vidal's [6] first attempt in the early 1970s. There are a number of BCI systems with a variety of signal processing techniques to extract and classify the signals of interest e.g. those associated with left or right finger movements. Majority of these techniques can be classified into two major approaches: *pattern recognition* and *operant conditioning*. The former relies more on the signal processing algorithms to extract a natural phenomena of the brain's activity, such as ERS/ERD or P300, whereas the latter relies more on the BCI user training the brain to adjust a phenomena of normal brain activity, such as increasing the power in a particular frequency band. A common problem for all BCI systems is that the EEGs are usually corrupted by artifacts. In this section the requirements for a successful BCI system are identified and the methods applied in BCI are outlined.

### 3.1 Artifact Rejection

Ocular artifacts (OAs), also called electro-oculograms (EOGs), are the main source of interference for EEG signals. These pose a significant problem to clinicians and neurologists because of the large amount of data that can be lost due to their presence. The majority of BCI researchers exclude the trials that contain eye blinks [29].

OAs refer to the potential difference that is generated when the eye moves in its socket or when a blink occurs. OAs propagate to other recording electrodes and impose themselves over the existing EEG. They can be measured by placing electrodes around the eyes. Horizontal eye movement can be measured by placing electrodes on either side of the eyes, whereas vertical movement and blinks can be measured by the electrodes placed above and below the eyes.

The interfering eye blinks generate a signal that is on the order of ten times larger than cortical signals. Eye blinks can last between 200 to 400ms. The eyeball can be considered as a dipole rotating in a socket, since the cornea remains at 0.4 to 1mV positive with respect to the retina. Rotations of the eyeball in saccadic eye movements cause large external field variations that can contaminate EEG readings [30]. Due to the magnitude of the blinking artifacts and the high resistance of the scalp, OAs can contaminate majority of the electrode signals, even those in the occipital area. An example of EOG artifact is shown in Fig 3.1.

It is possible to ask patients to fixate on a point, which will reduce the number of eye movements, but involuntary movements, such as eye blinks, are just as troublesome. Asking the patients to suppress eye blinks will distract them from the clinician's instructions and proves to be impossible, for example, when examining children. Closing the eyes results in increased involuntary eye movements. Eye blinks may be in response to a cognitive task, therefore simply rejecting the

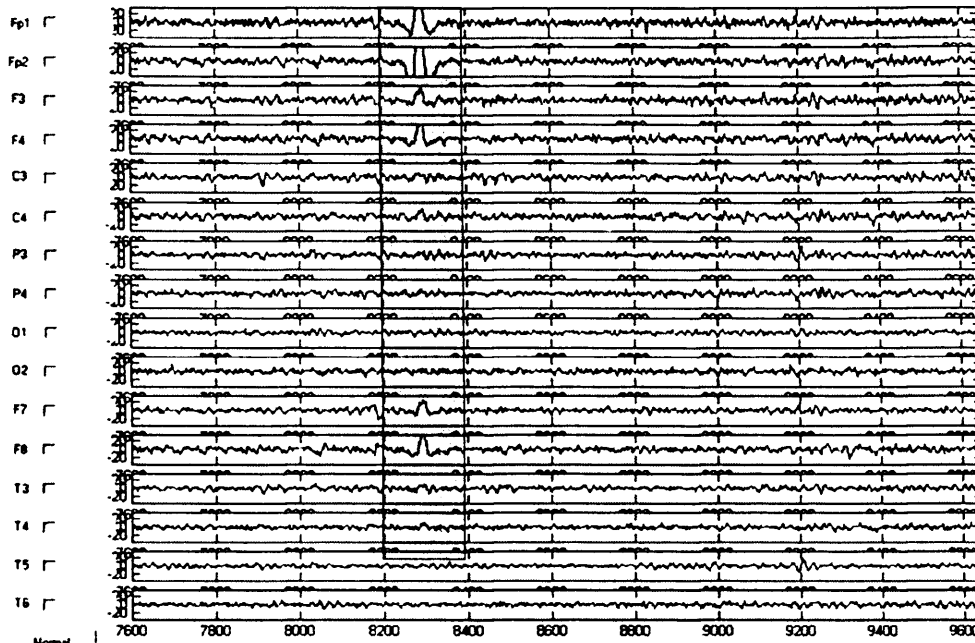


Figure 3.1: An example of EEG that is contaminated by eye blinking artifact

data segment will result in the loss of important information.

The main reason why EOGs cannot be simply removed using the conventional frequency domain filtering techniques is because of the spectral overlap between EOG and the underlying EEGs. Numerous methods have been employed for removal of OAs that exploit the use of regression analysis which is incorporated into popular EEG monitoring software, such as Neuroscan. Part of the EOG is subtracted from the EEG such that  $Corrected\ EEG = Raw\ EEG - \gamma EOG$ , where the EOG is measured at the mastoids which removes the need for a horizontal and vertical EOG measurement [31]. The parameter  $\gamma$  has been determined in numerous ways, such as the ratio between EEG and EOG. In [32]  $\gamma$  was determined by the maximum covariance between EOG and EEG. However, due to volume conduction, OAs contain some EEG information which will inevitably be subtracted using these techniques.

Adaptive filters have been implemented for the removal of EOG artifacts [33]. In these methods the vertical and horizontal EOGs were measured and used as reference inputs to the adaptive algorithm. In another approach [34] a method has been proposed which does not require a reference input for removing the EOG artifact. Their adaptive algorithm estimated the EOG by predictive filtering techniques. The mentioned adaptive filtering techniques show promising results, however, they operate on one EEG channel at a time, which can be computationally expensive, especially when a large number of electrodes are used. Moreover, these techniques do not consider all the information within the EEG signals, hence their use in artifact rejection is not efficient. Another method for removing blinking artifacts from EEG was proposed by [35] using principal component analysis (PCA). It finds orthogonal directions of greatest variance in the EEG signals. PCA is based on explicit spectral matrix factorization of the EEG signals, therefore the application of PCA is generally superior to the traditional aforementioned regression technique. The main drawback of PCA lies in the fact that neurobiological signals are not believed to be orthogonal, hence OAs will not always be effectively removed [36].

## **3.2 Blind Source Separation in Artifact Rejection**

One area which has sparked interest in the biomedical field is the use of Independent Component Analysis (ICA) in blind source separation (BSS). ICA is a method of estimating the sources given that only the mixtures are available. This is achieved by making as few assumptions as possible about the original sources. One common assumption is that the source signals within

$\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$  are statistically independent and zero mean, where  $t$  denotes the discrete time sample and  $(\cdot)^T$  is vector transpose. This is a stronger claim than uncorrelatedness because it assumes that the joint probability density of the sources can be factorised into the product of marginal densities [37] i.e

$$p(\mathbf{s}) = \prod_i p_i(s_i). \quad (3.1)$$

The mixtures can be modelled by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \quad (3.2)$$

where  $\mathbf{A}$  is the  $M \times N$  full column rank mixing matrix,  $N$  is the number of sources,  $M$  ( $M > N$ ) is the number of linear mixtures,  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$  observed at the electrodes and  $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_M(t)]^T$  is the additive zero mean sensor noise. It is assumed that the sensor noise is temporally uncorrelated i.e.  $E\{\mathbf{v}(t)\mathbf{v}^T(t-k)\} = \mathbf{0} \forall k \neq 0$ , and uncorrelated with the sensor data  $E\{\mathbf{v}(t)(\mathbf{A}\mathbf{s}(t))^T\} = \mathbf{0}$ . The output of the ICA system (i.e. the estimated original sources) is given by

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \quad (3.3)$$

where  $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_N(t)]^T$  is the vector of the estimated sources,  $\mathbf{W}$  is the  $N \times M$  separation matrix. Equation (3.3) implies that  $\mathbf{A} = \mathbf{W}^\dagger$ , where  $^\dagger$  is the pseudoinverse (henceforth,  $\mathbf{W}$  is assumed to be square, i.e.  $M = N$ ). However this is not generally true, since there is a scale and permutation ambiguity in most BSS algorithms. A more accurate definition for the sources is

$$\mathbf{y}(t) = \hat{\mathbf{s}}(t) = \mathbf{P}\mathbf{D}\mathbf{W}\mathbf{x}(t) \quad (3.4)$$

where  $\mathbf{P}$  is a permutation matrix of size  $N \times N$  with each row having only one column equal to one where the permutation has occurred, and  $\mathbf{D}$  is a diagonal scaling matrix.



Some BSS algorithms require that the data is stationary or at least wide-sense stationary (WSS). EEGs can be considered stationary stochastic processes when data is not corrupted by artifacts for up to 25 seconds [38]. In the general case EEGs can be considered, at the least, WSS processes for shorter intervals. A WSS processes is described to have the following properties [37];

1. The time average of the data  $\mu(t)$  is time invariant.
2. The autocorrelation function is independent of time, in that  $r_x(\tau) = E\{x(t)x(t+\tau)\}$  for all values of  $t$ . This property is demonstrated for EEG signals in Fig. 3.2. One can see that the auto correlation function of EEGs do not vary significantly with time.
3. The variance,  $\sigma^2 = r_x(0) = E\{x(t)x(t)\}$  is finite, which holds true for EEGs.

Most BSS algorithms estimate the probability distribution of the data, and are typically applied to blocks of data. Therefore, the stationarity requirement of the sources is not as great as the stationarity of the mixing system [37], i.e. the mixing matrix must not change with time. Within the context of EEGs, this requirement is met when processing short blocks of EEGs. However, location of the sources cannot be guaranteed between two consecutive nonoverlapping blocks. The goal of an ICA algorithm is to estimate sources that are statistically independent. There are three prominent approaches to attaining this goal:

1. Assuming that the source signals are stationary and non Gaussian, they can be reconstructed by measuring the statistical distance between the joint distribution and the product of marginal densities. One example of this measure is the Kullback-Leibler divergence (KL). However, the drawbacks of this method is that it is sensitive to noise and the measure is also sensitive to the estimation of the probability densities, i.e. a poor estimation of

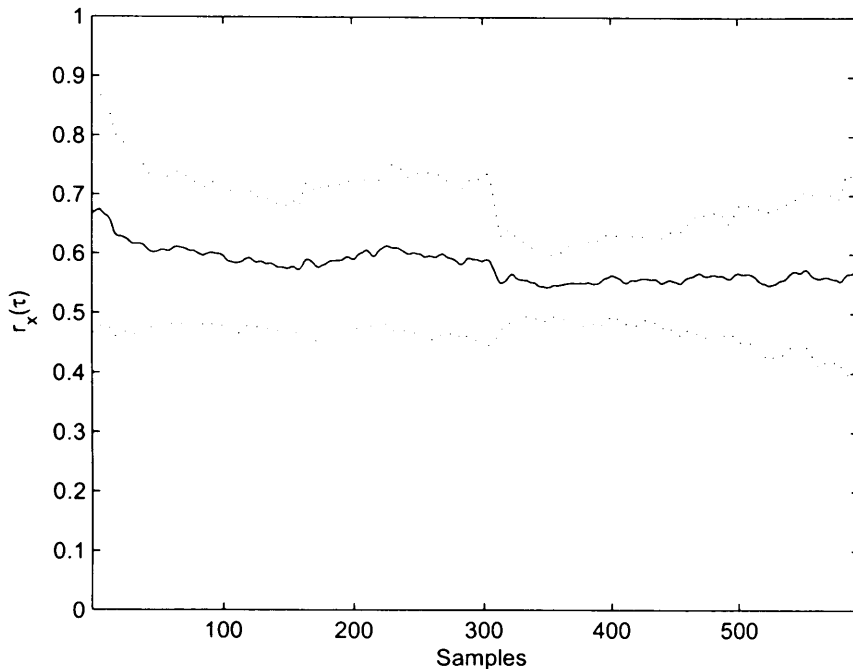


Figure 3.2: The ensemble average of autocorrelation at  $r_x(\tau)$ ,  $\tau = 3$ , for 30 real EEGs (solid line). The dotted line shows the standard deviation in the estimate of the ensemble average of  $r_x(\tau)$  at each consecutive window.

probability densities, for example in nonstationary source signals, results in poor performance of the separation algorithm.

2. Simultaneous diagonalization of several covariance matrices for multiple blocks of data, i.e. time varying covariance. If the sources are non-stationary and mutually independent then the sum of the off diagonal elements in the covariance matrices will be close to zero. By applying the decorrelation algorithm the unmixing matrix can be found. The advantage of this method is that it only uses second order statistics to account for non-stationarities in the signal, hence making it computationally attractive.

3. Minimising the mutual temporal correlation over a number of time lags.

The main advantage of this method over the second method is that it is able to separate source signals that contain white noise. This and the second method use second order statistics only, hence, both have low computational cost.

### 3.2.1 Whitening

A common preprocessing step in ICA is known as *whitening* or *sphering* the data. Whitening, which is also the basis of PCA, is defined as a linear coordinate transformation of an arbitrary multivariate density function into a spherical one [39] [40]. This decorrelates the data. Decorrelation is a weaker condition compared to independence. Two variables  $y_1$  and  $y_2$  are said to be uncorrelated if  $E\{y_1 y_2\} - E\{y_1\}E\{y_2\} = 0$  whereas for independence  $E\{h(y_1)h(y_2)\} - E\{h(y_1)\}E\{h(y_2)\} = 0$  where  $h(\cdot)$  is an arbitrary function. The transformation matrix is found by decomposing the covariance matrix of the observed data,  $\mathbf{R}_x = E\{\mathbf{x}(t)\mathbf{x}^T(t)\}$ , into its eigenvalues and corresponding eigenvectors. The observed data  $\mathbf{x}(t)$  is whitened by the following transformation

$$\mathbf{B}_w = \mathbf{U}\mathbf{Q}^{-1/2}\mathbf{U}^T \quad (3.5)$$

Where  $\mathbf{U}$  is a matrix of orthonormal eigenvectors of  $\mathbf{R}_x$ ,  $\mathbf{Q}$  is the diagonal matrix where the diagonal elements are the eigenvalues of  $\mathbf{R}_x$  and  $\mathbf{B}_w$  is the whitening transformation matrix. This means that the covariance matrix of the whitened observed data  $\mathbf{R}_z = E\{\mathbf{z}(t)\mathbf{z}^T(t)\} = \mathbf{I}$  after whitening, where  $\mathbf{z}(t) = [z_1(t), \dots, z_M(t)]^T$ , is the whitened observed data at discrete time  $t$ , and  $\mathbf{z}(t) = \mathbf{B}_w\mathbf{x}(t)$ .

As mentioned above a number of authors have attempted to use PCA to separate OAs [41] from EEGs based on the assumption that OAs are algebraically

orthogonal to neurobiological signals. The principal components are ranked by the value of the eigenvalues, in that, a principal component containing most of the information has a large eigenvalue.

This helps the ICA algorithm by reducing the solution to finding an orthogonal mixing matrix. This means that the joint distribution of the data will only need to be rotated in  $N(N - 1)/2$  degrees of freedom, as opposed to unwhitened data, which has  $N^2$  degrees of freedom [37]. Pre-processing by whitening the data considerably reduces the complexity of the problem. With this assumption in mind, many algorithms are designed so that the estimated sources meet this criterion, albeit approximately in practice.

### 3.2.2 Information Theoretic Based BSS

One semi-autonomous algorithm for the removal of artifacts from EEGs is based on the Information Maximization (Infomax) theorem [42]. The Infomax algorithm is based on the principles of maximum likelihood (ML) estimation. Using the assumption given in (3.1) and substituting the estimated sources  $\mathbf{y} = \mathbf{W}\mathbf{x}$  for the actual sources in (3.1), for a given number of observations,  $M$ , the log-likelihood expression can be formulated as [43] [37]

$$\frac{1}{T_B} \log L(\mathbf{W}) = E\left\{\sum_{i=1}^M \log p_i(\mathbf{w}_i^T \mathbf{x})\right\} + \log |\det \mathbf{W}| \quad (3.6)$$

the incremental update rule is given by

$$\Delta \mathbf{W} \propto \frac{1}{T_B} \frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} = [\mathbf{W}^{-T} - \mathbf{f}(\mathbf{y})\mathbf{x}^T] \quad (3.7)$$

where  $\mathbf{f}(\mathbf{y}) = [f(y_1), f(y_2), \dots, f(y_N)]^T$  is a nonlinear function applied to each element of the estimated output  $\mathbf{y}$ ,  $(\cdot)^{-T}$  is the pseudoinverse and transpose, and  $T_B$  is the data length. The update equation is given by,

$$\mathbf{W}(t + 1) = \mathbf{W}(t) + \eta \Delta \mathbf{W}, \quad (3.8)$$

where  $\eta$  is the learning rate and is typically set to a low value ( $\eta < 1$ ).

In [36] the form of  $f(y_i)$  is proportional to the cumulative density function of the sources. Where as in [42] the nonlinear function has been defined as  $f(y) = (1 + e^{-y})^{-1}$ . However, in most cases the nonlinearity is restricted to a sigmoidal function which limits the separation to super-Gaussian sources.

In order to avoid the matrix inversion and to speed up the convergence the gradient equation in (3.7) is modified by post-multiplying the term  $\mathbf{W}^T \mathbf{W}$  [44]. The natural gradient becomes

$$\frac{\partial L(\mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = [\mathbf{I} - \mathbf{f}(\mathbf{y})\mathbf{y}^T] \mathbf{W} \quad (3.9)$$

Jung *et al.* [43] proposed a system to segregate the EEG signals using the Infomax algorithm, and then the effect of eye blinking artifact was removed via manually inspecting the ICs for flagged components that contained eye blinking artifacts. The separated signals are then recombined to reconstruct the artifact free EEG.

### 3.2.3 Second Order Blind Identification

Second Order Blind Identification (SOBI) [45] has been used by a number of researchers for the separation of EEG signals [46] [47] and particularly for the separation of EEGs contaminated by EOGs [48]. It is the ability of recovering correlated sources which has attracted researchers to use SOBI for artifact removal. SOBI is particularly popular in experiments investigating the frontal lobe activity, such as short term memory tasks. Eye blinking artifacts severely corrupt the signals from electrodes located at these sites. Moreover, when movement related potentials are of interest, OAs interfere with the parietal electrodes, which are associated with movement preparation and hence these ERPs can be misclassified as OAs by the majority of BSS algorithms. The main reason why SOBI performs so well on temporally correlated data is because it exploits the

temporal information in order to estimate the separated sources. On the contrary, BSS algorithms based on an information theoretic approach exploit the PDF of the sources and hence the time structure is completely neglected i.e information theoretic BSS algorithms are insensitive to the order of the data points.

In SOBI it is assumed that the covariance of the observed vector,  $\mathbf{x}(t)$  is structured as follows [45]:

$$\mathbf{R}_x(0) = E\{\mathbf{x}(t)\mathbf{x}^T(t)\} = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^T + \sigma_n^2\mathbf{I} \quad (3.10)$$

$$\mathbf{R}_x(\tau) = E\{\mathbf{x}(t+\tau)\mathbf{x}^T(t)\} = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^T \quad (3.11)$$

where  $\mathbf{I}$  is the identity matrix,  $\sigma_n^2$  is the variance of the noise and

$$\mathbf{R}_s(\tau) = E\{\mathbf{s}(t+\tau)\mathbf{s}^T(t)\} \quad (3.12)$$

is the covariance of the source signals at time lag  $\tau$ . In practice the covariance matrix is estimated from a finite length of data given by,

$$\mathbf{R}_s(\tau) = \frac{1}{T_B - \tau} \sum_{t=1}^{T_B} \mathbf{s}(t+\tau)\mathbf{s}^T(t) \quad (3.13)$$

assuming that the sources have zero mean. Further references to covariance matrices are assumed to be calculated in this fashion. Since the sources are mutually uncorrelated, the covariance matrix of the source signals,  $\mathbf{R}_s(0)$  will be the identity matrix, assuming that the sources have unit variance. Hence, the covariance matrix of the mixtures at  $\tau = 0$  will be  $\mathbf{R}_x(0) = \mathbf{A}\mathbf{A}^T + \sigma_n^2\mathbf{I}$ . The eye blinking artifact introduces a non-stationary source into the EEGs, which will hinder the separation performance of any BSS algorithm which assumes that the sources are stationary. Since SOBI can only derive optimal separation parameters when the sources are stationary, a common preprocessing step when the sources are not stationary is symmetrising the correlation matrices, given by

$$\tilde{\mathbf{R}}(\tau) = [\mathbf{R}(\tau) + \mathbf{R}^T(\tau)]/2 \quad (3.14)$$

where  $\bar{\mathbf{R}}(\tau)$  is the symmetrised correlation matrix for a given time lag  $\tau$ . This has the effect of extinguishing any non-stationarity in the covariance matrices, which is indicated by the off diagonal elements of  $\mathbf{R}(\tau)$  having the property,  $[R]_{ij}(\tau) \neq [R]_{ji}(\tau), \forall \tau > 0$ .

The first step in SOBI is to apply the whitening transform using (3.5) such that  $\mathbf{R}_z(0) = \mathbf{B}_w \mathbf{R}_x(0) \mathbf{B}_w^T = \mathbf{B}_w \mathbf{A} \mathbf{A}^T \mathbf{B}_w^T = \mathbf{I}$ , assuming that the sources are mutually uncorrelated and variances normalised to unity.  $\mathbf{R}_z(0)$  is the covariance matrix of the whitened data. This implies that

$$\mathbf{U} = \mathbf{B}_w \mathbf{A} \quad (3.15)$$

is a  $N \times N$  unitary matrix. Thereby, the mixing matrix can be determined by

$$\mathbf{A} = \mathbf{B}_w^{-1} \mathbf{U}, \quad (3.16)$$

subject to scaling and permutation ambiguities.

Following these aims SOBI algorithm attempts to find a unitary factor  $\mathbf{U}$  that will diagonalize a set of covariance matrices. Using the properties defined in (3.11) and (3.16) the following is obtained

$$\mathbf{R}_z(\tau) = \mathbf{B}_w \mathbf{R}_x(\tau) \mathbf{B}_w^T = \mathbf{U} \mathbf{R}_s(\tau) \mathbf{U}^T, \quad (3.17)$$

Where  $\mathbf{R}_z(\tau)$  is the covariance matrix for the whitened observations. Since the observations are whitened, i.e. all cross terms are discarded, then  $\mathbf{U}$  can be any unitary matrix.

The covariance matrix of the estimated sources can be recovered by

$$\mathbf{R}_y = \mathbf{V}^T \mathbf{R}_z(\tau) \mathbf{V} \quad (3.18)$$

where  $\mathbf{R}_y$  is a diagonal matrix in which the diagonal terms contain the autocorrelation values for the estimated sources at time lag  $\tau$ , i.e.

$$[\mathbf{R}_y]_{ii} \neq [\mathbf{R}_y]_{jj} \quad \forall 1 \leq i \neq j \leq N. \quad (3.19)$$

The matrix  $\mathbf{V}$  is a unitary factor and since any unitary factor is a diagonalizer of  $\mathbf{R}_z(\tau)$  then  $\mathbf{V} \equiv \mathbf{U}$ . More formally the diagonalization of a matrix is defined as

$$\text{off}(\mathbf{M}) \triangleq \sum_{1 \leq i \neq j \leq n} |M_{ij}|^2 \quad (3.20)$$

Provided that the source covariance matrix,  $\mathbf{R}_s(\tau)$ , satisfies the spectral theorem for normal matrices, namely  $\mathbf{M}^T \mathbf{M} = \mathbf{M} \mathbf{M}^T$ , then it is said to be unitarily diagonalizable, i.e. there exists a diagonalizing matrix  $\mathbf{V}$ .

This property led to the popular time-structure based BSS algorithm known as Algorithm for Multiple Unknown Signals Extraction (AMUSE) [49] [50]. Many applications have been found in the processing of EEG signals, particularly for the removal of artifacts from the EEGs [51]. However, the main drawback of this method is the requirement set by (3.19) which means that for successful diagonalization of  $\mathbf{R}_z(\tau)$ ,  $\tau \neq 0$ , the eigenvalues of  $\mathbf{R}_s(\tau)$  must be distinct. This in general, cannot be checked *a priori*. Furthermore, as the spectral shapes of the sources become more similar, the estimate of  $\mathbf{U}$  becomes poorer. With this in mind and the fact that the distribution of EEGs are yet unknown the AMUSE algorithm is not an ideal candidate for optimal separation of EEG sources.

One solution to this shortcoming of AMUSE, proposed by [45], is to jointly diagonalize multiple covariance matrices at different time lags, i.e.  $\{\mathbf{R}_z(\tau_k), k = 1, \dots, K\}$  (where  $K$  is the maximum time lag) whitened covariance matrices, given by,

$$\mathbf{R}_y(\tau_k) = \mathbf{V}^T \mathbf{R}_z(\tau_k) \mathbf{V} \quad \forall k = 1, \dots, K \quad (3.21)$$

under the condition that

$$[R_y(\tau_k)]_{ii} \neq [R_y(\tau_k)]_{jj} \quad \forall 1 \leq i \neq j \leq N \text{ and } \forall 1 \leq k \leq K \quad (3.22)$$

The unitary diagonalizer in (3.21) is found by computing the product of Givens rotations [45]. Since  $\mathbf{V}$  is an approximate joint diagonalizer the criterion defined



in (3.20) applied to (3.21) will not, generally, be equal to 0. This is because of the estimation errors in the covariance matrices as they are estimated from a finite set of data.

### 3.3 Approaches to Brain Computer Interfacing

A brain computer interface (BCI) is a system which allows the user to interact with a computer using brain signals only. BCIs can be divided into two main categories; invasive and non-invasive. The former uses intracranial electrodes or subdural implanted deep inside or on the surface of the brain, whereas the latter uses surface electrodes placed over the scalp. Non-invasive BCIs will be considered and further will give references to BCIs imply the non-invasive type. Current BCIs use one of a number of extractable EEG signals, such as rhythmicities [7] in the data or a particular component, such as slow cortical potentials (SCP) [8], or evoked potentials (EPs) [9]. EPs such as P300 are time-locked events which are, generally, extracted by averaging many trials of the same event.

In this section the focus will be on the signal processing techniques used for BCI.

#### 3.3.1 Autoregressive Modelling

A number of researchers have modelled the EEGs using auto regressive (AR) modelling on the basis that the EEGs are temporally correlated within a reasonably short term. In all cases they assume that each EEG sample can be calculated from its previous samples as

$$x(t) = \sum_{m=1}^p a_m x(t-m) + e(t) \quad (3.23)$$

where  $x(t)$  is the EEG sample at discrete time  $t$ ,  $a_m$  is the AR coefficient for an all-pole filter,  $e(t)$  is the residual error, which is considered as additive noise, and  $p$  is the model order. The parameters are determined by the Akaike information criterion (AIC) [52] give by,

$$AIC(p) = \ln(\sigma_v^2) + \frac{2p}{T_B} \quad (3.24)$$

where  $\sigma_v^2$  is the variance of the residual error and  $T_B$  is the data length. Typically the AR coefficients are used directly as features to classify the EEGs into differentiating states with an application to BCI. However the main drawback of this method is that the AR coefficients are highly interdependent, which results in poor generalisations performance as the number of AR coefficients change. One solution to this limitation in AR modelling was proposed by Curran *et al.* [53], in which the reflection coefficients were considered instead of the prediction coefficients. Curran *et al.* [53] managed to classify left and right motor imagery with an accuracy of 71% using a nonlinear classification method.

### 3.3.2 Complexity Measure

Complexity is defined as the ‘randomness’ of a given dataset. This has been used to detect ERD/ERS on the basis that EEGs over the motor cortex become more complex during ERD and less complex during ERS. Roberts *et al.* [54] compared several measures of complexity and assessed their suitability for identifying finger movements. One of the successful complexity measures was the embedding-space decomposition which describes the EEG as a sequence of repeatedly windowed versions of the same data. It is described as follows

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{T-(i-1)})^T \quad (3.25)$$

where  $\bar{T}$  is the number of windows, and

$$\mathbf{x}_i = (x_i, x_{i+J}, \dots, x_{i+(l-1)J})^T \quad (3.26)$$

where  $J$  is the jump interval. The parameter  $l$  is the embedding dimension and according to [55] must satisfy  $l \geq 2U + 1$ , where  $U$  is the dimensionality of the attractor<sup>1</sup>. If  $l$  is chosen large enough such that a phase portrait of the attractor is obtained with redundancy, then the amount of redundancies can be exploited to describe the complexity of the EEGs, as in [54]. Firstly, the eigenvalues of the embedding matrix in (3.25) are calculated, denoted by  $\lambda_i$   $i = 1, \dots, l$ . The entropy of the eigen spectrum of the embedding matrix in (3.25) is given by,

$$H = - \sum_{i=1}^{\bar{T}} \lambda_i \log \lambda_i \quad (3.27)$$

where  $\lambda_i$  are the eigenvalue above the noise floor. The number of states is given by,  $\Omega = 2^H$ . Later Roberts *et al.* [54] extended the method to provide a measure of spatio-temporal complexity by reconstructing the embedding matrix in order to accommodate the spatial dimension, as follows:

$$\mathbf{X}_{tot} = (\mathbf{X}_1, \dots, \mathbf{X}_N) \quad (3.28)$$

where  $\mathbf{X}_i$  are constructed from (3.25) for  $N$  EEG channels. The results showed that there was spatio-temporal localisation of complexity at the point of finger movement over the contralateral motor cortex. The advantage of this proposed method is it takes into account all of the information available (spatial and temporal) to detect an event. However the method has yet to be followed by a classification algorithm to evaluate its performance.

---

<sup>1</sup>An attractor is a point in the phase space to which the trajectory of embedding matrix follows.

### 3.3.3 Time-Frequency Methods

The time-frequency analysis is commonly used to visualise the changes in the frequency spectrum over time. In particular, short term Fourier transform (STFT) [56] is used to highlight the time varying frequency characteristics of the EEGs. The STFT algorithm applies the fast Fourier transform to short overlapping windows of the data. This yields a fixed resolution time-frequency representation for one EEG channel. This information was used by He *et al.* [57] and Qin *et al.* [58]. They extracted the region of the time-frequency representation (TFR) that was related to finger movement (namely the alpha band) and reconstructed the time domain signal by applying time frequency masks [59] to the TFR of the EEGs. The use of spatial and temporal information is becoming more popular in BCI for extracting movement related potentials (MRPs) since the brain signal source related to finger movement is localised over the contralateral motor cortex within a specific frequency band.

### 3.3.4 Common Spatial Patterns

A popular feature for BCI is to extract the characteristics in the common spatial patterns (CSP) [60] [61], which are parsimonious representations of the EEGs. Since left and right finger movements are localised within the contralateral motor cortex then the patterns in the distribution of potential over the scalp can identify which finger is being moved. CSP is based on decorrelating the sum of covariance matrices,  $\mathbf{R}_{tot} = \mathbf{R}_{left} + \mathbf{R}_{right}$ , of the EEGs for each of the classes, in the same way as PCA (or whitening), to yield a whitening matrix  $\mathbf{B}_{tot}$ . Then the class covariance matrices,  $\mathbf{R}_{left}$  and  $\mathbf{R}_{right}$  are transformed by the whitening matrix  $\mathbf{B}_{tot}$ , yielding  $\mathbf{S}_{left}$  and  $\mathbf{S}_{right}$ . Next, an eigenvalue decomposition of  $\mathbf{S}_{left}$  and  $\mathbf{S}_{right}$  yields a common eigen vector matrix  $\mathbf{U}$ . Finally, the projections onto the

electrodes are given by  $\mathbf{P} = \mathbf{U}^T \mathbf{B}_{tot}$ . The columns of  $\mathbf{P}$  are used to determine the projection strengths onto the scalp electrodes and hence to discriminate between each of the classes. Ramoser *et al.* [60] and Müller-Gerking *et al.* [62] achieved a 92% classification rate using CSP with a linear discriminant (LDA) classifier. In summary, CSPs find a common describing factor to both classes as a way to maximally discriminate between the left and right finger movements. The technique was further extended to multi-class BCI by [63] to discriminate between finger/foot movements.

### 3.3.5 Blind Source Separation in BCI

BSS has been applied to BCI by a number of researchers [64] [65] [46]. The EEG source signal related to finger movement is embedded within the background EEGs. In most cases the head is considered a homogeneous region and the sources isotropic. Therefore, it is generally assumed that EEG source related to finger movement will be mixed with the background EEGs.

Understandably, BSS is an attractive candidate for processing the EEGs since there is an array of observation sensors, unknown number of sources (presumed to be less than the number of sensors provided that the number of recording electrodes is high), and an unknown mixing channel. The motivation for using BSS in BCI is that it is assumed that the source related to finger movement or imagination constitutes one of the independent sources. Kamousi *et al.* [64] used the FastICA algorithm [66] to separate the EEGs and then used a dipole fitting algorithm [57] to localise the ICs. The results therein demonstrated that the ipsilateral ERS at finger movement would be correctly localised for their test datasets. Serby *et al.* [65] used a higher order statistics based BSS [67] method for extracting the P300 component for their BCI. In order to determine which

component contained the P300 they compared the ICs with a P300 extracted offline [65].

The use of SOBI [45] algorithm (see Section 3.2.3) for the detection of ERPs has been validated in [46]. Wang *et al.* [68] used SOBI to discriminate between left and right hand median nerve stimulation from two channels of EEGs. Essentially the projection strengths of the SOBI components were used to determine the median nerve stimulation source. The SOBI components were presented to a back propagation neural network (BPNN) and achieved classification rates of up to 83%. Tang *et al.* [69] separated the EEGs using a hybridised algorithm which extracts spatial and temporal information from the SOBI components to localise the ERPs. The SOBI components that were correlated with a previously extracted ERP were flagged and formed a subset. Additionally, ICs that projected onto the motor cortex were flagged and formed another subset. Finally only ICs that were in both IC subsets were in both subsets were fitted to the equivalent current dipoles. This method thereby, neglects the redundant ICs and only localises the ICs that are significant to detecting the median nerve stimulation.

## Chapter 4

# Artifact Removal using Constrained Blind Source Separation

The challenge is to separate the signals into their independent constituent sources while automatically removing the artifact and retaining any diagnostic information about the brain disorder. A pre-determined reference is incorporated into the minimisation algorithm hence yielding an automated artifact rejection system. The significance of the algorithm is also due to its performance in the case of an undetermined number of sources.

EEGs are said to be instantaneous mixtures since the potentials are due to emission from the volume regions of electromagnetic dipoles and the bandwidth of the signal (and accordingly the required sampling frequency) is very low (Bandwidth  $< 50\text{Hz}$ ). This in turn means that the signals measured at the electrodes are received with a negligible delay i.e. in much less than one sample interval (linearly mixed), hence an instantaneous type of ICA is used for separation of

EEG signals. Although the number of signal sources within the brain is yet unknown, an initial assumption is that, for a reasonably large number of electrodes, the number of sources,  $N$ , is less than the number of electrodes  $M$  i.e. an over-determined system has been considered.

## 4.1 Joint Diagonalization of Correlation Matrices

The standard SOBI algorithm (see Section 3.2.3) is based on finding a rotation matrix such that several matrices are diagonalized. Presented here is a gradient based algorithm for joint diagonalization of multiple correlation matrices. The separation matrix can be found by minimising a cost function  $J(\mathbf{W})$ , which provides a measure of independence of the estimated sources. Therefore the goal of the diagonalization algorithm is to find a  $\mathbf{W}$  that diagonalizes the output covariance matrix  $\mathbf{R}_Y(k)$  diagonal,  $k \in \{1, 2, \dots, K\}$ , where  $K$  is the maximum time lag. Hence minimising  $J(\mathbf{W})$  will ensure that the estimated sources are as independent as possible. The covariance matrix  $\mathbf{R}_Y(k)$  to be diagonalized, is given by

$$\mathbf{R}_Y(k) = \mathbf{W}[\mathbf{R}_X(k) - \mathbf{R}_V(k)]\mathbf{W}^T \quad (4.1)$$

where in practice  $\mathbf{R}_X(k)$  is the estimate of the time lagged covariance matrix of the signal mixtures and  $\mathbf{R}_V(k)$  is the estimate of the covariance matrix of the sensor noise. Since it is assumed that the noise is spatially uncorrelated,  $\mathbf{R}_V(0)$  will be a diagonal matrix and  $\mathbf{R}_V(k) = \mathbf{0}$  for  $k \neq 0$  [70].

$$\mathbf{R}_X(k) = \mathbf{A}\mathbf{R}_S(k)\mathbf{A}^T + \mathbf{R}_V \quad (4.2)$$



where  $\mathbf{R}_S(k)$  is a diagonal covariance matrix of the independent source signals. Following [70], the least squares (LS) estimate of  $\mathbf{W}$  is

$$\mathbf{W}_{opt} = \arg \min_{\mathbf{W}} \sum_{k=1}^K \|J_M(\mathbf{W}, k)\|_F^2 \quad (4.3)$$

where  $\mathbf{W}_{opt}$  is the optimum separation matrix,  $\|\cdot\|_F^2$  is the squared Frobenius norm, and  $J_M(\mathbf{W}, k)$  is the error to be minimised between the covariance of the source signals  $\mathbf{R}_S(k)$  and the estimated sources  $\mathbf{R}_Y(k)$ . This criteria can be replaced by a suitable cost function based upon minimising the off-diagonal elements for multiple lagged covariance matrices, as

$$\mathbf{W}_{opt} = \arg \min_{\mathbf{W}} \sum_{k=1}^K \|\mathbf{R}_Y(k) - \text{diag}(\mathbf{R}_Y(k))\|_F^2 \quad (4.4)$$

where  $\text{diag}(\cdot)$  is an operator which zeros the off-diagonal elements of a matrix.

## 4.2 Constrained Learning

Minimising the cost function in (4.4) alone is not enough to remove the EOG from the underlying EEG, as there is no constraint to minimise the effect of the EOG. This is very important in places where there is an undetermined number of sources such that the output independent components (ICs) may not represent the actual sources. In this case, minimisation of the cost function should be subject to an equality constraint as

$$\min J_M(\mathbf{W}) \quad \text{s.t.} \quad \mathbf{d}(\mathbf{W}) = \mathbf{0} \quad (4.5)$$

where  $\mathbf{d}(\mathbf{W}) = [d_1(\mathbf{W}), d_2(\mathbf{W}), \dots, d_r(\mathbf{W})]^T$ ,  $r \in \mathbb{N}$  is the constraint term, and  $r \geq 1$  indicates that there is more than one constraint. The problem of constrained optimisation is shown in Fig. 4.1.

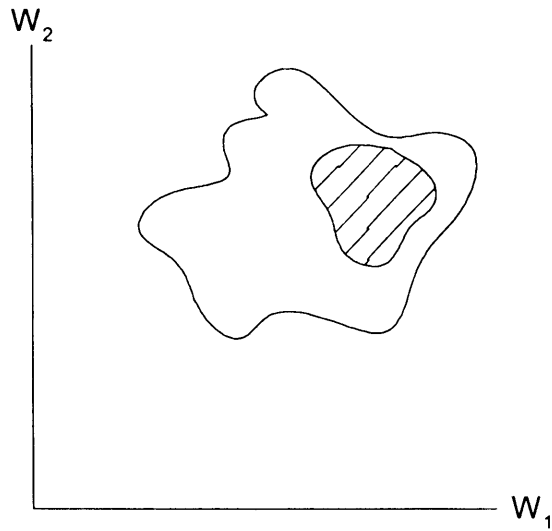


Figure 4.1: The solution space for an optimisation problem is shown in the area within the solid line and a solution which satisfies a constrained optimisation problem, i.e.  $\mathcal{W}$  is shown in the hatched area.

Penalty function based approach is a method for solving the constrained optimisation problem. It converts a constrained problem (4.5) into an unconstrained one by introducing a penalty parameter to the constraint term (right hand side of (4.5)) and incorporating it into the cost function. The penalty parameter asserts a penalty when the constraint term is in violation of the objective function. In general, penalty functions can be classified into two classes; interior and exterior functions. For equality constraints, as defined in (4.5), exterior penalty functions are best suited and interior penalty functions are generally used for inequality constraints such as  $\mathbf{d}(\mathbf{W}) > \mathbf{0}$ . The exterior penalty function is defined as in [71] and in matrix form in [72] as follows; Define a closed subset  $\mathcal{W} \in \mathbb{R}^M$  and a sequence of continuous scalar valued functions  $G_q(\mathbf{W})$  with  $q \in \mathbb{N}$ .  $G_q(\mathbf{W})$  is a

sequence of exterior penalty functions if the following conditions are true

$$G_q(\mathbf{W}) = 0, \quad \forall \mathbf{W} \in \mathcal{W} \quad (4.6)$$

$$0 < G_q(\mathbf{W}) < G_{q+1}(\mathbf{W}) \quad \forall \mathbf{W} \notin \mathcal{W} \quad (4.7)$$

$$G_{q+1}(\mathbf{W}) \rightarrow \infty \quad \text{as } q \rightarrow \infty \quad (4.8)$$

Let  $\mathbf{c}(\mathbf{d}(\mathbf{W})) = [c_1(d_1(\mathbf{W})), c_2(d_2(\mathbf{W})), \dots, c_r(d_r(\mathbf{W}))]^T$  be a positive penalty function vector corresponding to the constraints  $\mathbf{d}(\mathbf{W})$  and is weighted by  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_r]^T$ . The constrained problem is then converted into the minimisation of the following unconstrained problem

$$\min(J_M(\mathbf{W}) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{d}(\mathbf{W}))) \quad (4.9)$$

An example of function  $c_i$  is given by,

$$c_i(\mathbf{W}) \triangleq \xi_i \|\mathbf{d}(\mathbf{W})\| \quad (4.10)$$

where  $\xi_i$  is a weighting factor such that  $\xi_{i+1} > \xi_i > 0$ . This ensures that the adaptation will converge monotonically [72]. It is important to note that minimisation of the cost function in (4.4) will not guarantee that the conditions in (4.5) will be met. Only when the conditions in (4.9) are satisfied the unmixing parameter will be equivalent [71]. Incorporation of the Lagrangian multipliers is another approach for solving constrained optimisation problems. It works by finding the optimal solution where the gradients of the constraint terms coincide with those of the cost function [72] and the Lagrangian multipliers are defined at this point. For nonlinear optimisation Lagrangian multiplier may only locally minimise the objective function. These local minimisers of the cost function may not be global minimisers, especially in nonlinear optimisation which is often encountered in real world situations. Therefore, in general, penalty functions globally minimise

nonlinear optimisation problems. For further proof of the convergence properties of penalty function approach applied to BSS read [72] and [4].

In removal of eye blinking artifact, for example, the constraint is the cross correlation between the EOG electrode and the EEG electrodes. The constrained optimisation problem (4.5) can be converted into an unconstrained one as in (4.9) by using the penalty function method or by using Lagrangian multipliers. Define a new cost function,  $J_T(\mathbf{W})$ , such that

$$\begin{aligned} \mathbf{W}_{opt} &= \arg \min_{\mathbf{W}} J_T(\mathbf{W}) \\ &= \arg \min_{\mathbf{W}} \sum_{k=1}^K (J_M(\mathbf{W}, k) + \Lambda J_G(\mathbf{W}, k)) \end{aligned} \quad (4.11)$$

where  $\Lambda = \{\Lambda_{ii}\}$  ( $i = 1, \dots, N$ ) penalty coefficient which is governed by the cross-correlation between the EOG and estimated EEG source signals, defined by

$$\{\Lambda_{ii}\} = P E\{y_i(t)g(t)\} \quad (4.12)$$

where  $P \in \mathbb{R}^+$  is an adjustable constant,  $g(t)$  is the reference artifact signal and  $y_i(t)$  is the  $i^{th}$  IC. This applies a penalty at each iteration on each of the estimated sources that is proportional to its cross correlation with the artifact source. Therefore a component that is uncorrelated with the artifact will have no penalty applied to it. The constraint term is given by

$$J_G(\mathbf{W}, k) = \frac{1}{2} \sum_k \|\text{diag}(\mathbf{R}_{yg}(k))\|^2 \quad (4.13)$$

where  $\mathbf{R}_{yg}(k) = E\{\mathbf{y}(t)\bar{\mathbf{g}}^T(t+k)\}$  is the cross correlation between ICs and artifact signal,  $\bar{\mathbf{g}}(t) = [g_1(t), \dots, g_M(t)]^T$  is a vector with the reference signal copied  $M$  times. A gradient algorithm (GA) [73] was used to find the  $\mathbf{W}$  that minimises  $J_M(\mathbf{W}, k)$  as well as the constraint term  $J_G(\mathbf{W}, k)$ . The general GA update equation is

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \Delta\mathbf{W}(t) \quad (4.14)$$

where  $\Delta \mathbf{W}(t)$  is the incremental update of  $\mathbf{W}(t)$  given by [74]

$$\Delta \mathbf{W}(t) = -\mu(t) \frac{\partial J_T(\mathbf{W})}{\partial \mathbf{W}} \quad (4.15)$$

The adaptive learning rate  $\mu(t)$ , as used in [72], is dependent on the spread of the data and on the gradient of the total cost function. It is given by

$$\mu(t) = \mu_0 \left( \frac{1}{\sum_{k=1}^K \|\mathbf{R}_X(k)\|_F^2} + \frac{2}{\zeta + \|\Delta J_T(\mathbf{W})\|^2} \right) \quad (4.16)$$

where  $\mu_0$  is a positive constant typically  $\mu_0 < 1$  and  $\zeta$  is a regularisation parameter, which prevents the learning rate from being too large when the gradient becomes small. The typical value of the parameter  $\zeta$  is 0.05 and  $\Delta J_T = J_T(\mathbf{W}(t-1)) - J_T(\mathbf{W}(t))$ . Finding the gradient of (4.11) yields

$$\begin{aligned} \frac{\partial J_T(\mathbf{W})}{\partial \mathbf{W}} = & 4 \sum_{k=1}^K [\mathbf{R}_Y(k) - \text{diag}(\mathbf{R}_Y(k))] \mathbf{W} [\mathbf{R}_X(k) - \mathbf{R}_V(k)] \\ & + \Lambda \sum_k \text{diag}(\mathbf{W} \mathbf{R}_{XG}(k)) \text{diag}(\mathbf{R}_{XG}(k)) \end{aligned} \quad (4.17)$$

where  $\mathbf{R}_{XG}(k) = E\{\mathbf{x}(t)\bar{\mathbf{g}}^T(t+k)\}$  is the cross correlation between mixtures and artifact signal.

When the noise of the system is unknown its covariance can be estimated in the following fashion

$$\hat{\mathbf{R}}_V(t+1) = \eta \hat{\mathbf{R}}_V(t) + (1-\eta) \Delta \hat{\mathbf{R}}_V(t) \quad (4.18)$$

where  $\Delta \hat{\mathbf{R}}_V(t) = \hat{\mathbf{R}}_X(k) - \mathbf{W}^{-1} \hat{\mathbf{R}}_Y(k) (\mathbf{W}^T)^{-1}$ ,  $\hat{\mathbf{R}}_X(k)$  and  $\hat{\mathbf{R}}_Y(k)$  are respectively moving window estimates of the observation and output covariance matrices and  $\eta \in (0, 1)$ . The adaptation stops when the error falls below an acceptable level i.e. when  $\|\mathbf{W}(t-1) - \mathbf{W}(t)\|^2 \simeq 0$ .

In the following section the algorithm is examined using a set of simulated signals, a set of EEG contaminated by eye blinking artifact, and a set of EEG contaminated by ECG.

## 4.3 Experiments

In this section the CBSS algorithm is applied to both simulated signals and real EEG data and the results are analysed. The performance of the proposed algorithm will be evaluated in terms of the convergence speed and the ability to remove the artifact from each of the components.

### 4.3.1 Simulated Source Signals

In the first experiment the algorithm is presented with a synthetic set of signals affected by a simulated artifact. The synthetic signals were two speech like signals of 5000 samples length sampled at 12kHz. The artifact was generated using a sampled sinc function as shown in Fig. 4.2. Here it is assumed that the artifact signal can be easily extracted from the mixtures. The source signals and artifact are artificially mixed using an  $M \times N$  matrix ( $M = N$ ).  $\mathbf{W}$  is initialized to  $\mathbf{I}$  and the other parameters are set as follows;  $P = 0.01$ ,  $\mu_0 = 0.1$  and  $\eta = 0.1$ . Since the original sources are available, the mean square error (MSE),

$$\varepsilon^2 = E\|\mathbf{y} - \mathbf{s}\|^2 = \frac{1}{N} \sum_{i=1}^N E\{|y_i(t) - s_i(t)|^2\}, \quad (4.19)$$

is used to evaluate the resemblance between the estimated and the original sources. The performance of the algorithm is measured by finding the waveform similarity in  $dB$  defined by  $\varepsilon_{dB}^2 = 10 \log_{10}(1 - \varepsilon^2)$ . It is assumed that the signals are zero mean and unit variance. The mixed signals and the estimated sources are shown respectively in Fig. 4.3 and Fig. 4.4. By inspection of the estimated sources it is possible to see that the artifact has been removed from the signals of interest. The algorithm was tested using 10 data sets of synthetic signals mixed with the same mixing matrix. The waveform similarity index is compared with the SOBI [45] [4] algorithm ( $\mathbf{\Lambda} = \mathbf{0}$ ) and the proposed algorithm

$(\Lambda \simeq P \text{diag}(\mathbf{R}_{GY}))$  for each data set.

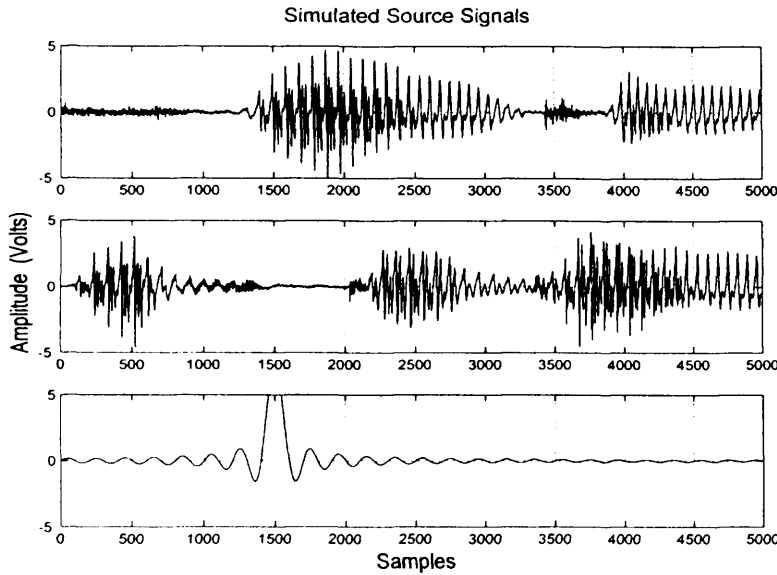


Figure 4.2: Original speech like source signals. The third signal represents the artifact.

The waveform similarity for the proposed algorithm and SOBI algorithm are  $\varepsilon_{dB}^2 = -0.27dB$  (*s.d.*  $0.02dB$ ) and  $-0.53dB$  (*s.d.*  $0.01dB$ ) respectively, which is a difference of  $0.26dB$ . This shows that sources estimated by the proposed CBSS algorithm significantly improves the quality of the signals compared to the unconstrained BSS algorithm. The performance of the algorithm was further examined by comparing the cross-correlation between the estimated sources and the artifact. Table 4.1 shows the performance improvement over SOBI algorithm. The goal of the algorithm is to minimise the effect of the artifact by minimising the cross-correlation between the estimated sources and the artifact. The artifact component may not be completely eliminated since the number of iterations in (4.15) is finite, which means that  $\mathbf{W}$  will approach  $\mathbf{W}_{opt}$  as  $t \rightarrow \infty$ . The convergence performance, shown in Fig. 4.5, is comparable with that in [4].

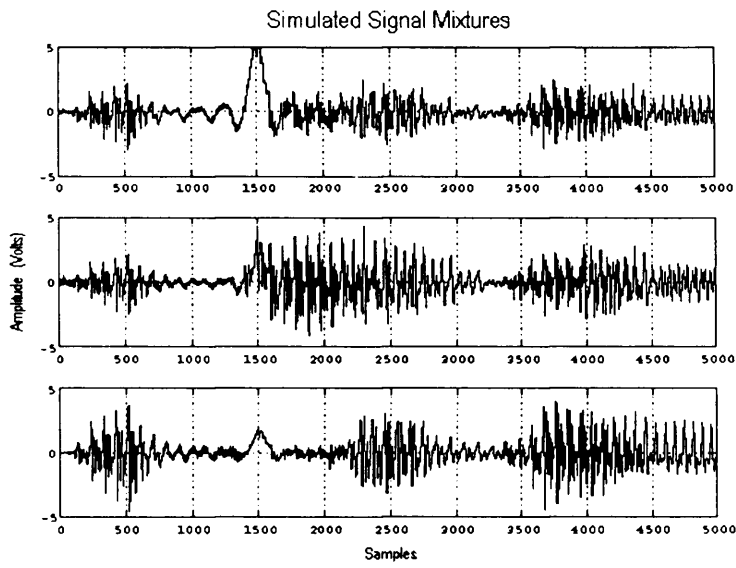


Figure 4.3: Artificially mixed signals.

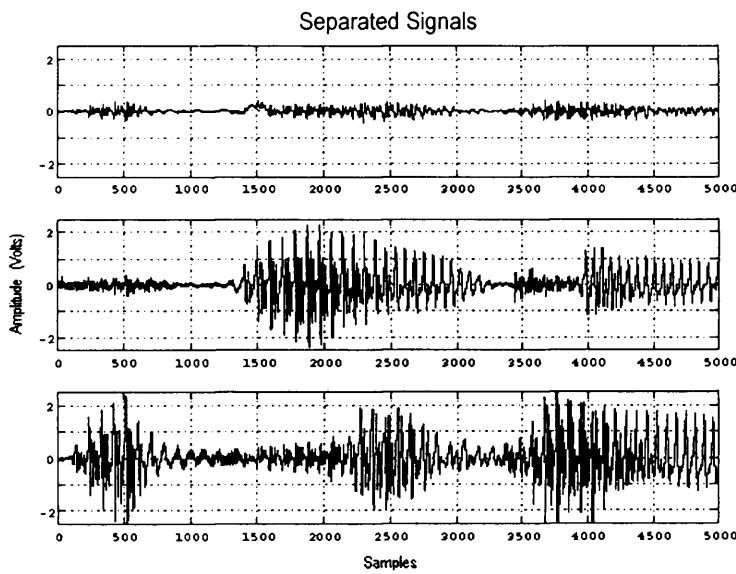


Figure 4.4: Estimated sources with the artifact minimised.



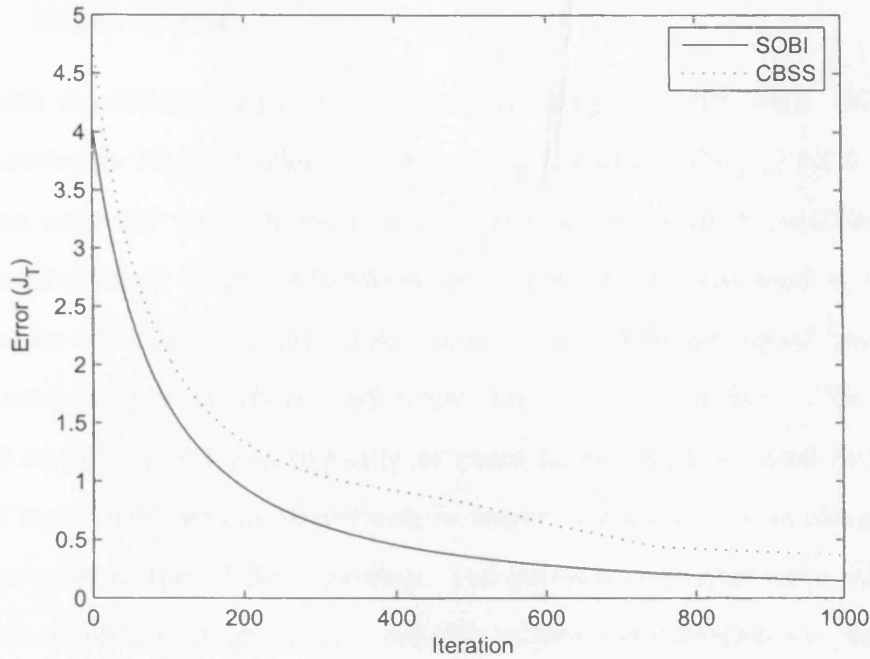


Figure 4.5: Convergence performance of the proposed algorithm (solid line) compared with that of an unconstrained gradient SOBI [4] algorithm (dotted line).

Table 4.1: Performance of the CBSS algorithm is based on the measurement of cross-correlation; the average cross-correlation between the artifact and the mixtures is compared with the cross-correlation between the artifact and the estimated sources. In this experiment the artifact is a sampled sinc signal. The results are also compared with SOBI algorithm

Average correlation between synthetic artifact and estimated sources by		
mixtures	SOBI	CBSS
0.82 (s.d. 0.3)	0.19 (s.d. 0.01)	0.09 (s.d. 0.01)

### 4.3.2 Removing The Effect of Eye Blinking From Real EEG Data

The CBSS algorithm is further examined by using real EEG data. The signals were recorded in King's College London using a Cadwell Easy II EEG amplifier. EEG was collected from 16 electrodes placed on the scalp at locations defined by the conventional 10-20 electrode system. The earlobe was used as a common reference for all of the channels. The ocular artifact reference signal was obtained from electrodes placed above and below the left or right eye. The data was sampled at 200Hz and was digitally lowpass filtered with a cutoff frequency of 40Hz. Twenty data sets of 10 seconds in length containing eye blinking artifacts were presented to the CBSS algorithm. Ten datasets were used since this was the minimum number that provided a reliable measure of performance. Each of the data sets was standardised to have unit variance and zero mean. A threshold is applied to the artifact so that any details concerning other brain signals presented in the EOG will not contribute to the penalty term. The artifact signal then becomes

$$\begin{cases} g(t) & \text{if } g(t) \geq \varsigma \quad \forall t \\ 0 & \text{if } g(t) < \varsigma \quad \forall t \end{cases} \quad (4.20)$$

The parameter  $\varsigma$  was empirically found to be 0.2 for normalised signals. The performance was evaluated by finding the cross-correlation between the artifact and each of the mixtures and comparing them with the cross-correlation between the artifact and the estimated sources. EEG sensor data and the artifact reference are shown in Fig. 4.6 and Fig. 4.7 respectively. The resulting separated sources are shown in Fig. 4.8.

From Table 4.2 its possible to see that by applying the constrained algorithm

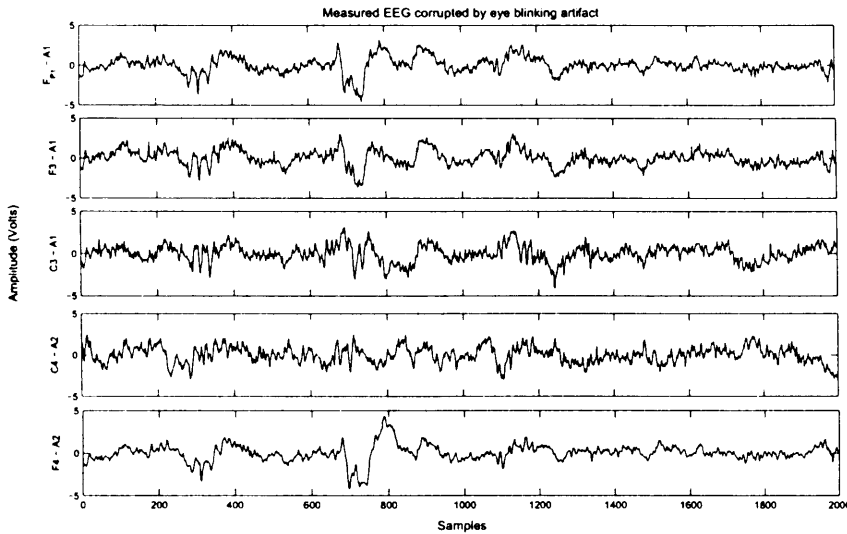


Figure 4.6: A selection of five EEG channels from a 16 channel EEG recording. The EEGs on these channels are corrupted by the ocular artifact between samples 600 to 900.

the cross-correlation between the estimated sources and the artifact has been considerably reduced. The penalty term  $\Lambda$  is adjusted in proportion to the cross correlation between the artifact and the estimated sources, i.e.  $E\{g(t)y_i(t)\}$ . Therefore, the higher the cross-correlation between the estimated source and the artifact, the larger the penalty on that component.

Another dataset presented to the algorithm is shown in Fig. 4.9. Eye blinks occur bilaterally, therefore in this experiment  $F_{P1}$  electrode was used as a reference signal, which can be seen in the first row of Fig. 4.9. The artifact corrected signals by the proposed CBSS algorithm are shown in Fig. 4.10 and the same EEGs separated by an unconstrained BSS algorithm and PCA are shown respectively in Fig. 4.11 and Fig. 4.12. Visual inspection of the ICs shows that

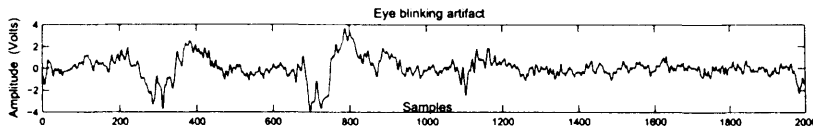


Figure 4.7: The vertical EOG signal measured from the right eye.

Table 4.2: Performance of the CBSS algorithm is based on the measurement of cross-correlation between the EEG and the EOG artifact.

Average correlation between artifact and		
mixtures	estimated sources by SOBI	estimated sources by CBSS
0.75 (s.d. 0.02)	0.23 (s.d. 0.02)	0.16 (s.d. 0.01)

the effect of eye blinking artifact is not entirely removed by unconstrained BSS algorithm or PCA, i.e. the estimated sources resemble the reference signal. Separation performance of PCA was slightly better than the unconstrained algorithm since the eye blinking artifact is clearly defined in PC10 of Fig. 4.12, however the remaining components still contain remnants of the eye blinking artifact. On the contrary, the signals separated by the CBSS algorithm do not resemble the reference signal. Furthermore, in order to validate the ICs extracted using the CBSS and unconstrained BSS algorithms, each IC was reprojected to all electrodes by

$$\mathbf{Y}_i^{rep} = (\mathbf{W})^{-1} \tilde{\mathbf{Y}}_i \quad (4.21)$$

where  $\tilde{\mathbf{Y}}_i$  is a  $N \times T_B$  matrix of zeros with only the  $i^{th}$  row being the  $i^{th}$  IC, with  $T_B$  defined by the length of the signal.  $\mathbf{Y}_i^{rep}$  are the projection strengths of the IC to each of the electrodes. The projection results for the proposed CBSS algorithm and unconstrained BSS are shown in Fig. 4.13 and Fig. 4.14 respectively. The projection strengths of the ICs are similar for both constrained and unconstrained algorithms. However the ICs extracted by CBSS contain fewer projections to the

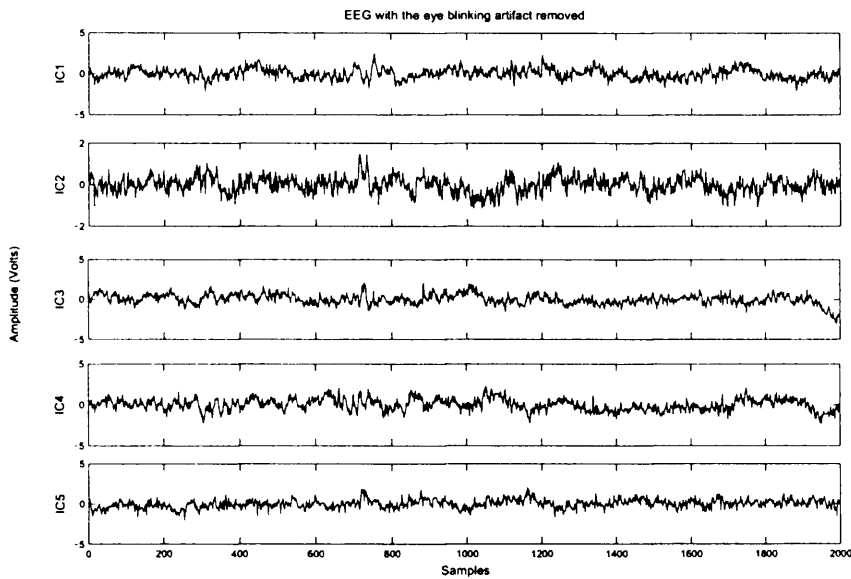


Figure 4.8: A selection of five independent components (ICs) derived from the EEG primarily corrupted by ocular artifact. The ICs represent the EEG with the EOG artifact removed.

frontal electrodes as can be seen by comparing Fig. 4.13 and Fig. 4.14. The projection strengths of the ICs using the unconstrained algorithm are similar to those extracted from CBSS algorithm. Tang *et al.* [46] demonstrated the validity of the extracted SOBI component and therefore similar projection patterns of the ICs suggests the validity of the components extracted by CBSS algorithm.

### 4.3.3 Removing The Effect Of ECG From EEG Real Data

The proposed system was also tested on EEG signals contaminated by ECG, and the performance was examined. The ECG was measured using Eindhoven's Triangle for the electrode configuration [75]. The ECG data were acquired by the Cadwell Easy II amplifier and sampled at 200Hz. In this experiment appropriate

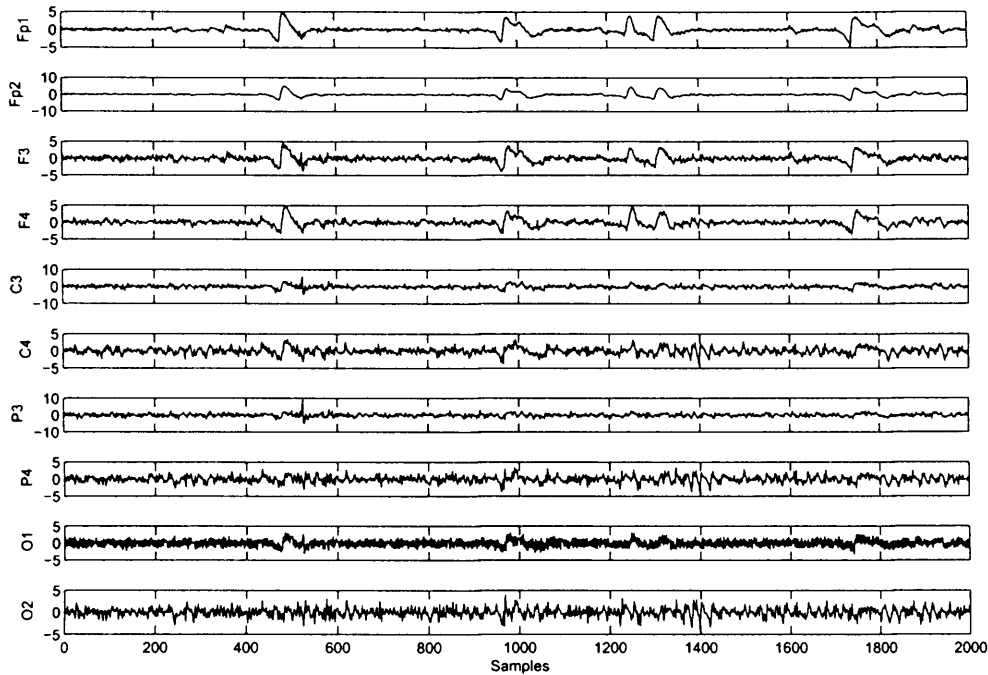


Figure 4.9: The EEGs heavily contaminated by eye blinking artifact.

values for  $\mu_0 = 0.01$ ,  $\eta = 0.1$  and  $P = 0.01$  were found empirically. The performance of the system in terms of the cross correlation between the artifact and the estimated output is illustrated in Table 4.3. An 8 second long segment of contaminated EEG is shown in Fig. 4.15 and the measured ECG reference signal is shown in Fig. 4.17. The EEG after removing the artifact is shown in Fig. 4.16.

Based on 20 sets of EEGs it was found that the average correlation for the proposed CBSS algorithm was 0.16, with standard deviation 0.01. As the distribution of the estimator was not known but variance was known ( $0.01^2$ ), Chebychev's inequality was used given by,  $Prob\{|0.16 - R| < \epsilon\} \geq 1 - 0.01^2/\epsilon^2$ , where  $R$  is the true value of the average cross correlation. This shows that the estimate of the mean correlation value was within  $\epsilon$  of the true value of  $R$  with a probability

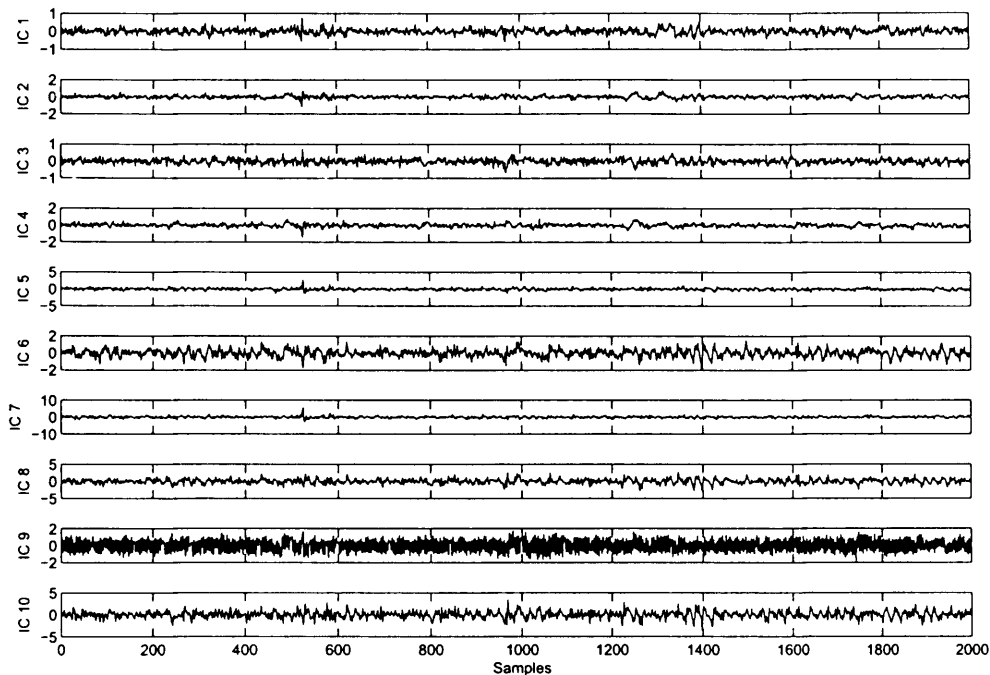


Figure 4.10: The ICs of the EEGs in Fig. 4.9 using the proposed CBSS algorithm.

of 90%.

From Table 4.3 it is possible to see that the CBSS algorithm has successfully separated the mixtures and its decorrelation performance in the undetermined case of EEG is, on average, better than that of SOBI algorithm [45]. The extent to which the artifact has been removed can also be verified by visual inspection of the output (Fig. 4.16).

## 4.4 Conclusions

As a requirement for preprocessing of the EEGs, a constrained BSS system for removing the eye blinking artifact has been developed by introducing nonlinear

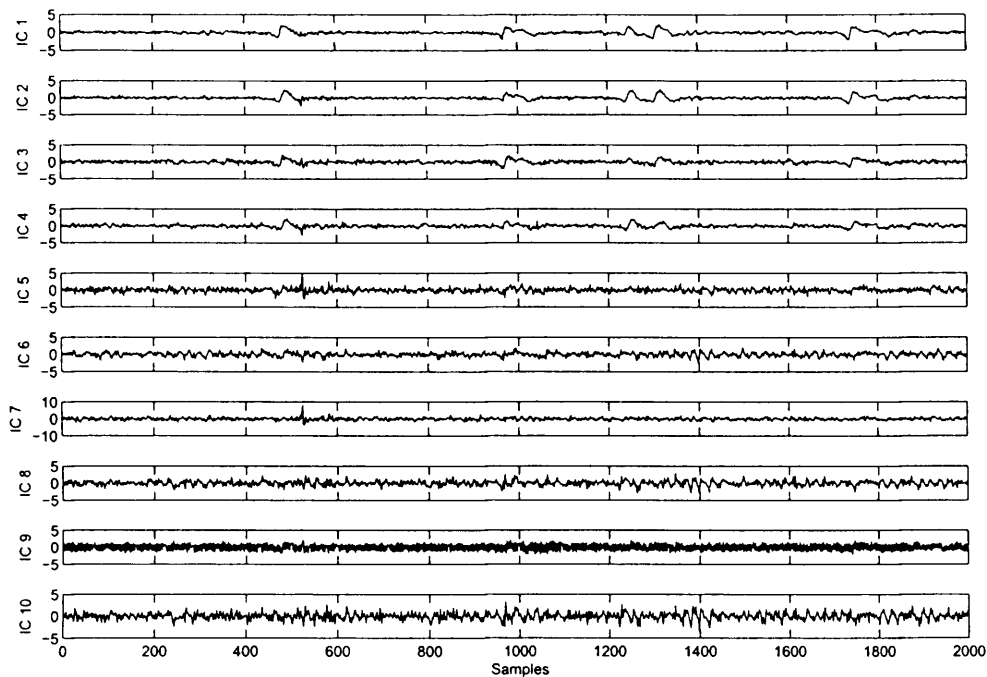


Figure 4.11: The ICs of the EEGs in Fig. 4.9 using an unconstrained BSS algorithm.

penalty functions. The penalty terms incorporate the constraints into the main objective function, thereby converting the constrained problem into an unconstrained problem.

The algorithm was first tested using synthetic signals corrupted by a known artifact. Synthetic signals were used to highlight the efficacy of the constraint in mitigating the artifact source. The effect of the constraint was objectively quantified by using the waveform similarity index (4.19). The effect of the undesired (interfering) signal was highly reduced and the desired components were extracted, as highlighted by the results. The results of the first experiment show that the quality of the separated signals has been improved and the convergence



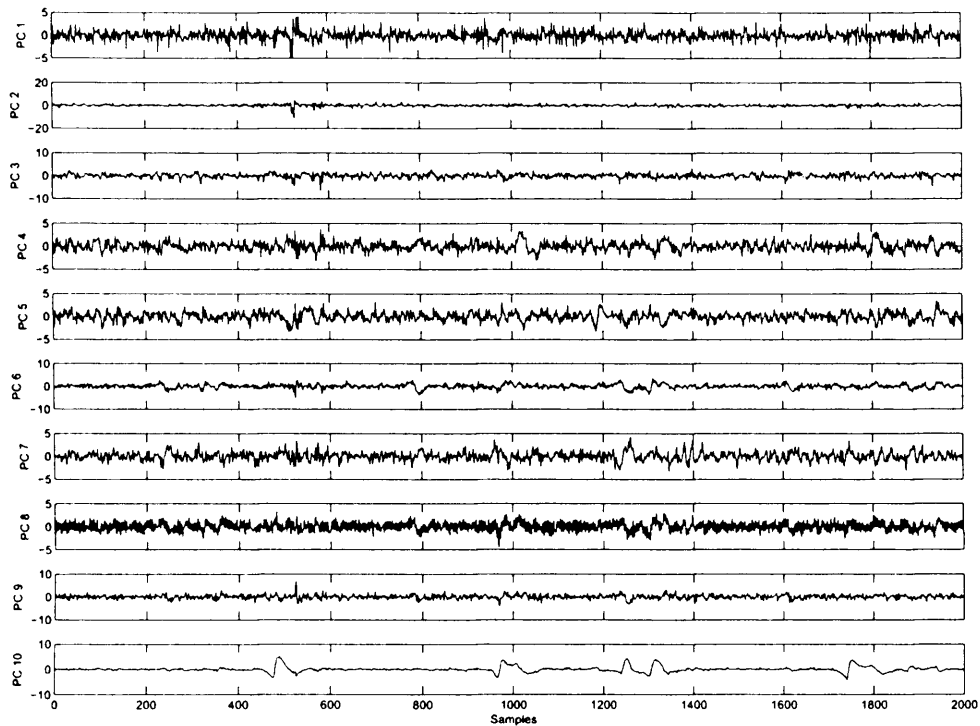


Figure 4.12: The EEGs in Fig. 4.9 separated using PCA.

performance was comparable to that of Joho *et al.* [4] [76].

The results of experiments with real EEG signals show that the solution space has been found which meets the criteria defined by (4.11), albeit approximately. The main advantage of the proposed CBSS method is that it performs the separation and mitigation of the interfering signal, which can be utilised by the clinician without further identification of artifacts related to eye blinks.

The result of the algorithm may be extended to removal of other interferences such as electroglottograms (EGG) from EEGs. As for the case of online EEG processing the permutation ambiguity of BSS must be resolved [77].

Table 4.3: Performance of system based on cross-correlation between EEG and ECG.

Average correlation between the artifact and		
mixtures	estimated sources by SOBI	estimated sources by CBSS
0.76 (s.d. 0.23)	0.21 (s.d. 0.01)	0.17 (s.d. 0.02)

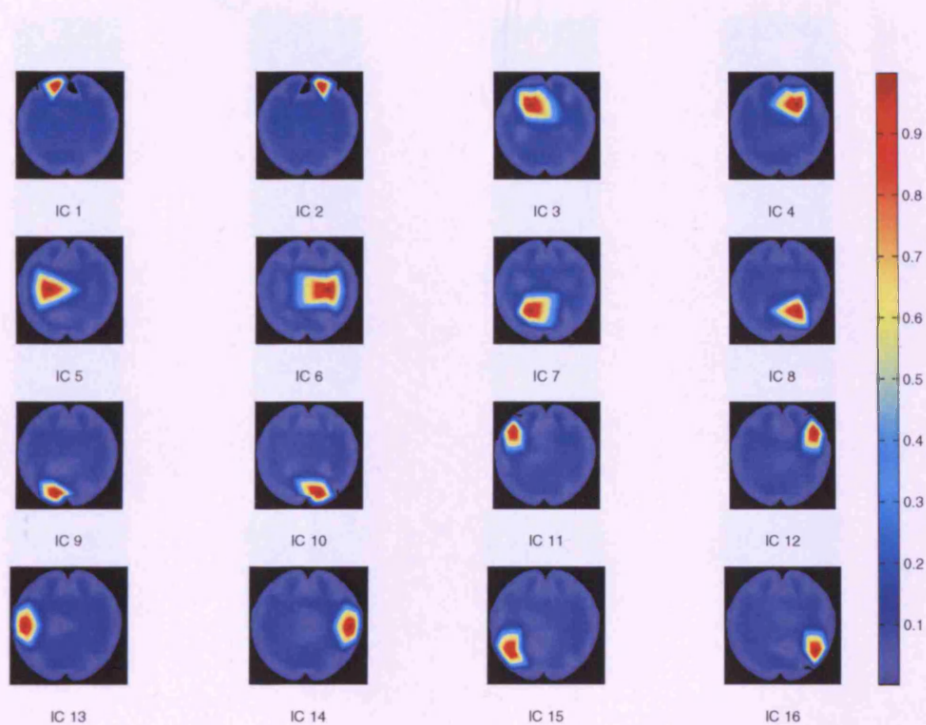


Figure 4.13: The reprojected ICs to the scalp electrodes after application of the proposed CBSS algorithm. Each scalp plot represents the projection strength of one IC in Fig. 4.10, designated by the label beneath the scalp plot. The colour represents projection strength of the IC onto each electrode and is normalised to unity across all electrodes (arbitrary units). The frontal electrodes are located towards the top of the scalp plot. In comparison with Fig. 4.14 the projection strengths of the CBSS ICs are not smeared over the frontal electrodes.

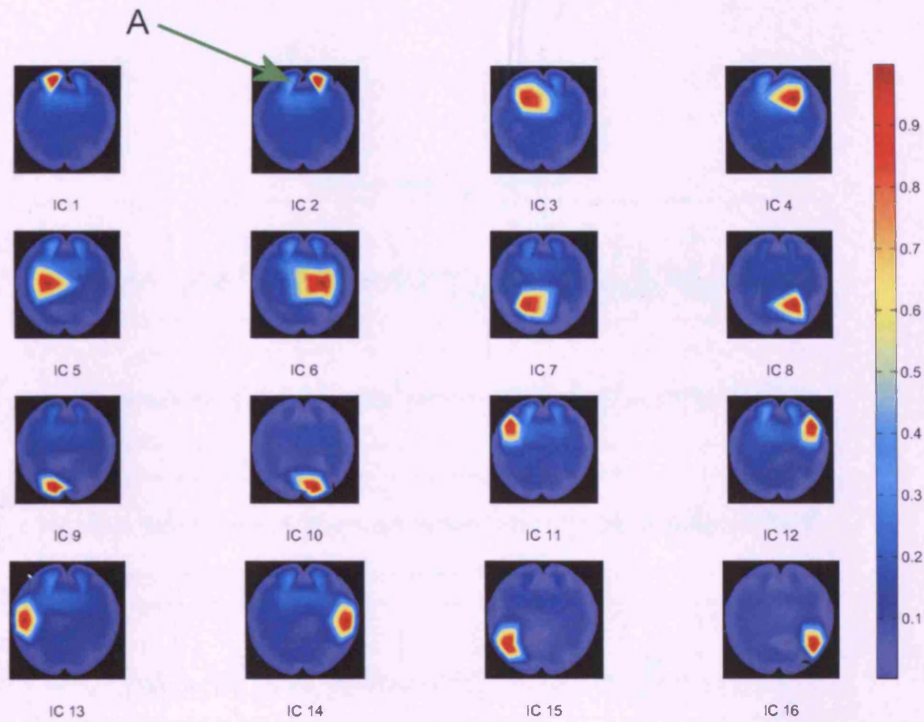


Figure 4.14: The reprojected ICs to the scalp after application of the unconstrained BSS algorithm. Each scalp plot represents the projection strength of one IC in Fig. 4.11, designated by the label beneath the scalp plot. The colour represents projection strength of the IC onto each electrode and is normalised to unity across all electrodes (arbitrary units). The frontal electrodes are located towards the top of the scalp plot. Point ‘A’ in the figure highlights the limitation of the unconstrained BSS algorithm, in that the projections are smeared over a number of electrodes. A similar pattern can be seen in a number of other electrodes.

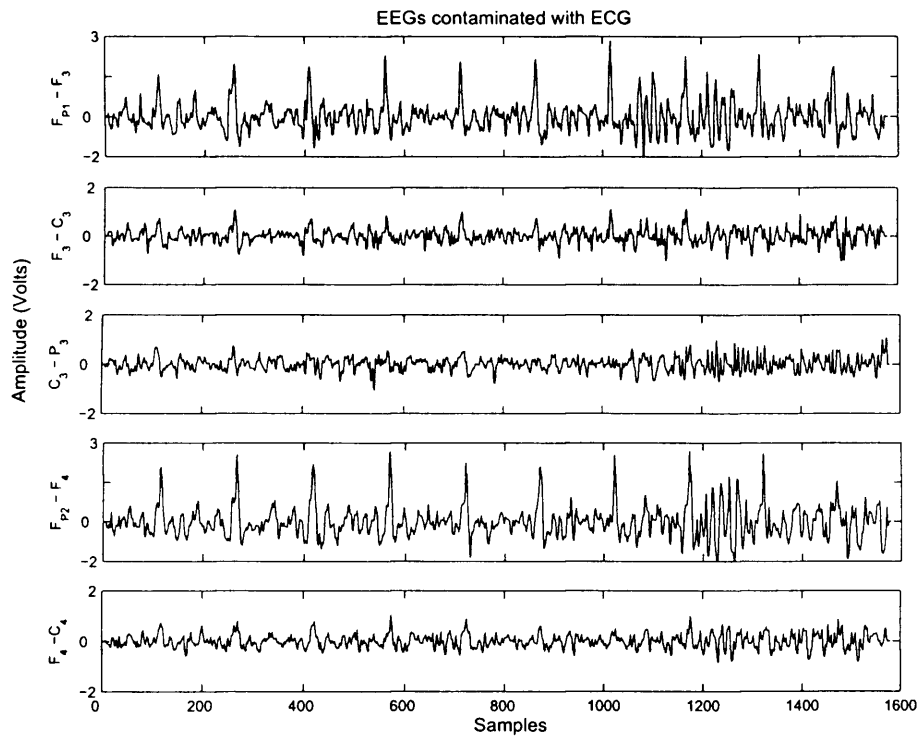


Figure 4.15: A selection of five channels from the EEG recording. There is an obvious ECG artifact present in the first and fourth channels of the figure.

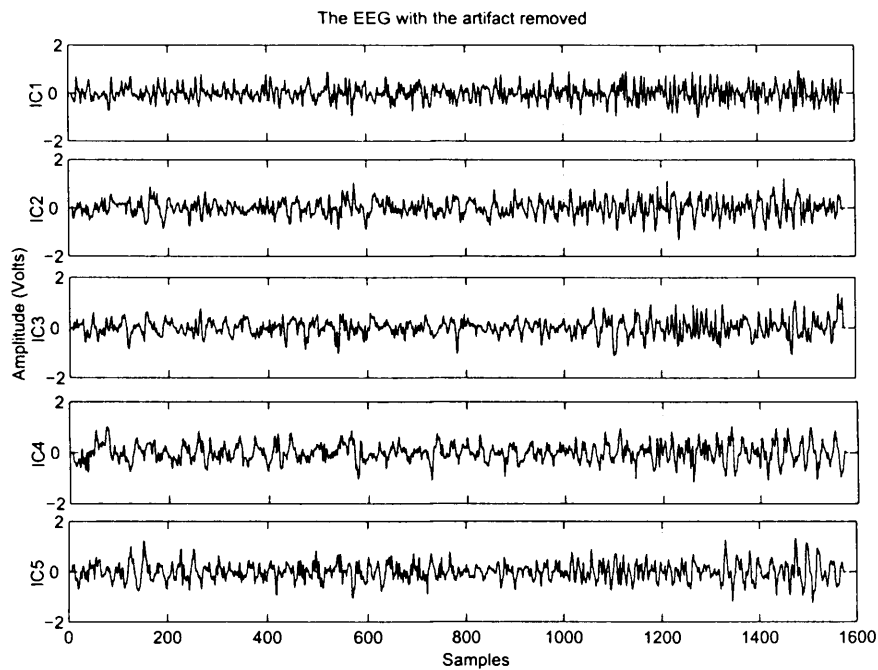


Figure 4.16: A selection of five independent components after the CBSS algorithm has removed the ECG.

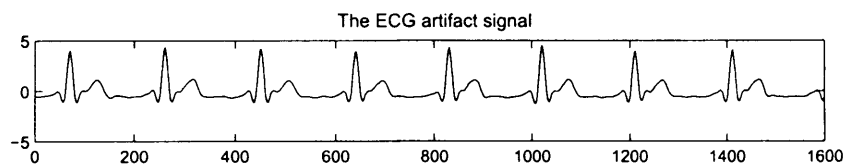


Figure 4.17: The measured ECG reference signal.

## Chapter 5

# Artifact Removal from the EEGs using Blind Source Separation and Classification

### 5.1 Introduction

An automated method for removing OAs from the EEGs has been described by Joyce *et al.* [48]. Joyce *et al.* used a BSS algorithm based on second order statistics, to separate the EEG and measured EOG into statistically independent sources. The separation is then performed a second time on the raw EEGs but with a selection of EOG channels inverted. The ICs which have been found after inversion are compared with the ICs of the previous separation and those which inverted are removed. In addition, ICs that are above a threshold of correlation with the measured reference are removed, as are the ICs with high power in the low frequencies. The main drawback of this method is that it is restricted to having the reference EOG channels, which may not be available if one would like

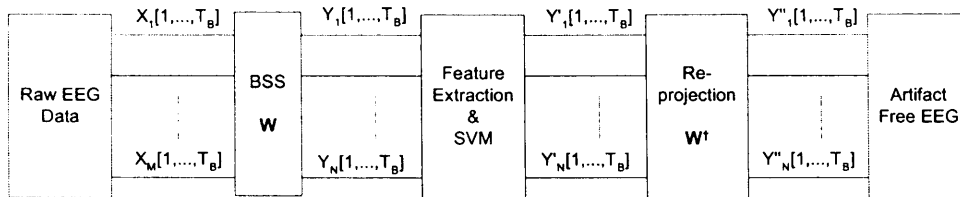


Figure 5.1: Block diagram of the BSS and SVM system.

to process sets of previously recorded data.

A block diagram of the proposed system is shown in Fig. 5.1. In the first stage of the block diagram the EEGs are acquired and stored, these are denoted in the diagram as  $X_{1, \dots, M}(1, \dots, T_B)$  where  $M$  is the number of electrodes and  $T_B$  is the data block length. Similarly, the intermediate signals within the figure are denoted as  $Y_{1, \dots, N}$ ,  $Y'_{1, \dots, N}$ ,  $Y''_{1, \dots, N}$ , where  $N$  is the number of the estimated sources. In the second stage of Fig. 5.1, the EEGs are separated into statistically independent sources using BSS. The SOBI algorithm was used for separation due to its excellent separation of OAs from the background EEGs [48]. Then, features are extracted from the independent sources and used to establish whether the source contains eye blinking artifact. Finally, the sources which are not identified as artifact are used to reconstruct the artifact free EEGs through re-projection.

## 5.2 Methods

### 5.2.1 Feature Extraction

Artifacts such as eye blinks and ECGs have certain waveform shapes, statistical, and temporal characteristics. Therefore, these characteristics are present in the ICs related to eye blinking artifacts. The four most effective features found which efficiently discriminate the artifact signal from the normal EEG are as follows:



## Feature 1

A large ratio between the peak amplitude and the variance of a signal suggests that there is an out of range amplitude in the data. This is a typical identifier for the eye blink because it causes a large deflection on the EEG trace. The equation describing this feature is given by

$$f_1 = \frac{\max(|\mathbf{u}_n|)}{\sigma_u^2} \quad n = 1, \dots, N \quad (5.1)$$

where  $\mathbf{u}_n$  is one of the  $N$  ICs,  $\max(\cdot)$  is a scalar valued function that returns the maximum element in a vector,  $\sigma_u$  is the standard deviation of  $\mathbf{u}_n$  and  $|\cdot|$  is the absolute value applied element-wise in (5.1). The normal EEG activity is tightly distributed about its mean value, therefore a low ratio is expected for it in contrast to ICs containing eye blink sources for which a high value is expected.

## Feature 2

This feature corresponds to a third order statistic of the data. The normalised skewness for each IC is given by

$$f_2 = \left| \frac{E\{u_n^3(i)\}}{\sigma_u^3} \right| \quad n = 1, \dots, N \quad (5.2)$$

for zero mean data. An EEG containing eye blinks typically has a positive or negative skewness since the eye blinking artifact increases locally the asymmetry of the signal segment. Therefore ICs corresponding to the source of eye blinking artifact will also have a positive or negative skew. The absolute value of the skewness is used because both negative and positive skew are associated with eye blinking artifacts. The significance of this feature in the overall classification is high since the eye blink signal has larger skewness than that of normal EEGs, which are approximately symmetrically distributed.

### Feature 3

As the third feature, the correlation between the IC and an independent dataset containing eye blinking artifact from six electrodes including the frontal electrodes close to the eyes ( $F_{P1}$ ,  $F_{P2}$ ,  $F_3$ ,  $F_4$ ) and the electrodes on the occipital lobe ( $O_1$ ,  $O_2$ ) is used. The reference dataset i.e. the EEG from the aforementioned electrodes, is distinct from the training and test datasets. This will make the classification more robust by introducing a measure of the spatial location of the eye blinking artifact. The mean of the maximum value of cross-correlation between each of the electrode locations and the IC is used as:

$$f_3 = \frac{1}{6} \sum_{i=1}^6 \max_{\tau} (|E\{x_i^0(t)u_n(t+\tau)\}|) \quad n = 1, \dots, N \quad (5.3)$$

where  $u_n(t)$  is the  $n^{th}$  independent component and  $x_i^0(t)$  are eye blinking reference signals, where  $i$  indexes each of the aforementioned electrode locations. The value of this feature will be larger for ICs containing eye blinking artifact, since they will have a larger correlation for a particular value of  $\tau$  in contrast to ICs containing normal EEG activity, the maximum is empirically chosen  $\tau \approx \sqrt{T_B}$ . This ensures that the cross correlation estimate will be accurate.

### Feature 4

The fourth feature is the statistical distance between the probability density function (PDF) of an IC and the PDF of a reference IC known to contain OA. The OA reference IC is taken from a dataset which is also distinct from the training and test datasets. Here it is assumed that the PDF of the IC containing the artifact is identical to that of the reference signal containing the artifact. To measure the statistical distance between the two PDFs the Kullback-Leibler (KL)

distance is used, given by

$$\begin{aligned} f_4 &= KL(P(u_n)||P(x_{ref})) \\ &= \int_{\text{inf}}^{\text{sup}} p(u_n) \ln \frac{p(u_n)}{p(x_{ref})} du_n \quad n = 1, \dots, N \end{aligned} \quad (5.4)$$

where  $P(u_n)$  and  $P(x_{ref})$  are the PDFs of one of the  $N$  ICs and a previously measured artifact, respectively. When the IC contains OAs the KL distance between its PDF and the PDF of the reference IC will be approximately zero, whereas the distance to the PDF of a normal EEG signal will be larger. Since the KL distance is related to the mutual information it reflects effectively the information shared between the IC and the reference signal.

### 5.2.2 Classification

While the classification stage is an important part of a BCI system, the choice of classifier is not as important as the choice of features. In EEG recording a number of classification methodologies have been used. For example in artifact classification neural networks (NN) have been used together with decision trees [78], a Bayesian network in [79] and even basic thresholding techniques [80]. An SVM is used for the classification method due to its generalisation and established empirical performance [81]. The standard SVM algorithm is a supervised learning algorithm for the classification of two classes. This means that the SVM is trained using a feature set in which target values are known *a priori*, thereby the testing feature set is classified based on these training values. Modifications of the traditional versions of SVM algorithms have been applied to allow semi-supervised classification (see [82] and [83] for more details). The goal of an SVM is to find a separating hyperplane for a given feature set that is optimal in the sense that the closest feature vector is furthest from the separating hyperplane. This is called the optimal separating hyperplane (OSH) [81]; see Fig. 5.2(a).

There are generally two regimes for finding the OSH; by maximising the margin between the two classes (Maximal Margin Classifier) with the assumption that there is no overlap between the classes in the feature space, and the soft margin classifier with the assumption that there is a degree of overlap between the classes. More often than not the feature space for real data is best suited to a soft margin classifier, therefore a soft margin classifier will be treated in this section.

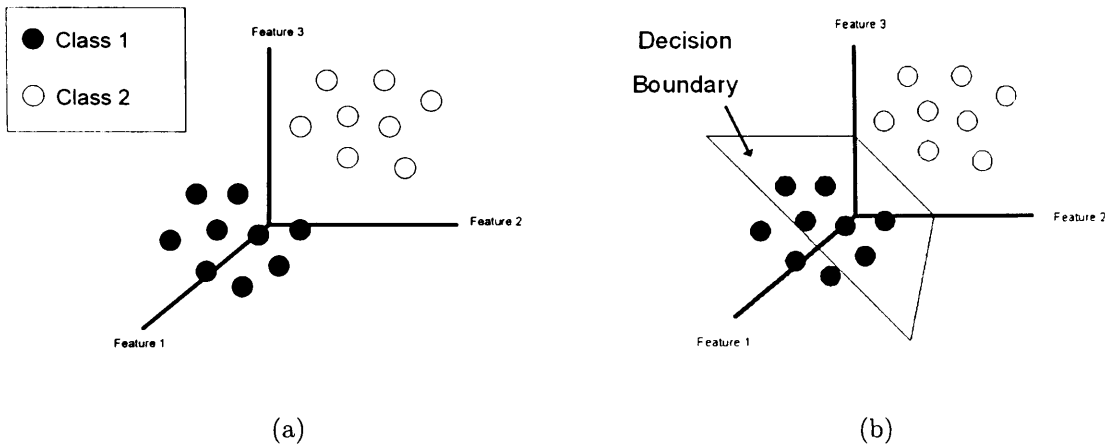


Figure 5.2: (a) Three features plotted against each other forming the feature space. (b) The optimum separating hyperplane for the feature space in (a).

The OSH is found by solving the following constrained optimisation problem:

$$\begin{aligned} \min_{\mathbf{z}, b, \gamma_{i=1, \dots, l}} & \left( \frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^l \gamma_i \right) \\ \text{s.t. } & q_i (\mathbf{z}^T \mathbf{g}_i - b) + \gamma_i \geq 0, \quad \gamma_i \geq 0 \quad i = 1, \dots, l \end{aligned} \quad (5.5)$$

where,  $l$  is the number of training vectors and  $q_i \in \{\pm 1\}$  are the output targets,  $\|\mathbf{z}\|^2 = \frac{1}{2} \mathbf{z}^T \mathbf{z}$  is the squared Euclidean norm. The parameter  $\mathbf{z}$  determines the orientation of the separating hyperplane,  $\gamma_i$  is the  $i^{\text{th}}$  positive slack parameter and places an upper bound on the number of training errors [84],  $\mathbf{g}_i$  is a vector containing the features  $\mathbf{g}_i = [f_1(i) \ f_2(i) \ f_3(i) \ f_4(i)]^T$ .

The constrained optimisation equation in (5.5) is solved using an extension of Lagrangian theory to solve inequality statements, known as Karush-Kuhn-Tucker (KKT) theory [85]. Forming the Lagrangian primal according to KKT from (5.5) yields,

$$L_P(\mathbf{z}, b, \gamma, \alpha, \mathbf{r}) = \frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^l \gamma_i - \sum_{i=1}^l \alpha_i [q_i(\mathbf{z} \cdot \mathbf{g}_i + b) - 1 + \gamma_i] - \sum_{i=1}^l r_i \gamma_i \quad (5.6)$$

where  $r_i$  are the Lagrangian multipliers introduced in to keep  $\gamma_i$  positive. The non negative parameter  $C$  is the (misclassification) penalty term, and can be considered as the regularisation parameter and is selected by the user. It places an upper bound on the Lagrangian multipliers, such that  $0 \leq \alpha_i \leq C$ . A larger  $C$  is equivalent to assigning a higher penalty to the training errors. The parameter  $C$  is set to a value which yields the lowest cross-validation (CV) test error. Support vectors (SVs) are the points from the dataset that fall closest to the separating hyperplane. Any vector  $\mathbf{g}_i$  that corresponds to a non-zero  $\alpha_i$  is a SV of the optimal hyperplane. It is desirable to have the number of SVs small to have a more compact and parsimonious classifier. The diagram in Fig. 5.3 illustrates the soft margin classifier for a two dimensional case, but the principle is the same for an  $n$ -dimensional feature space.

Minimising with respect to each of the parameters on left hand side (LHS) of (5.6) and substituting back into (5.6) yields,

$$L_D(\mathbf{z}, b, \gamma, \alpha, \mathbf{r}) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l q_i q_j \alpha_i \alpha_j \mathbf{g}_i \mathbf{g}_j \quad (5.7)$$

The OSH (generally nonlinear) is then computed as a decision surface of the form

$$f(\mathbf{g}) = \text{sgn} \left( \sum_{i=1}^{L_s} q_i \alpha_i K(\mathbf{g}_i^s, \mathbf{g}) + b \right). \quad (5.8)$$

where  $\text{sgn}(\cdot) \in \{\pm 1\}$ ,  $\mathbf{g}_i^s$  are SVs,  $K(\mathbf{g}_i^s, \mathbf{g})$  is the nonlinear kernel function (if  $K(\mathbf{g}_i^s, \mathbf{g}) = \mathbf{g}^T \mathbf{g}_i^s$  the SVM is linear), and  $L_s$  is the number of support vectors.

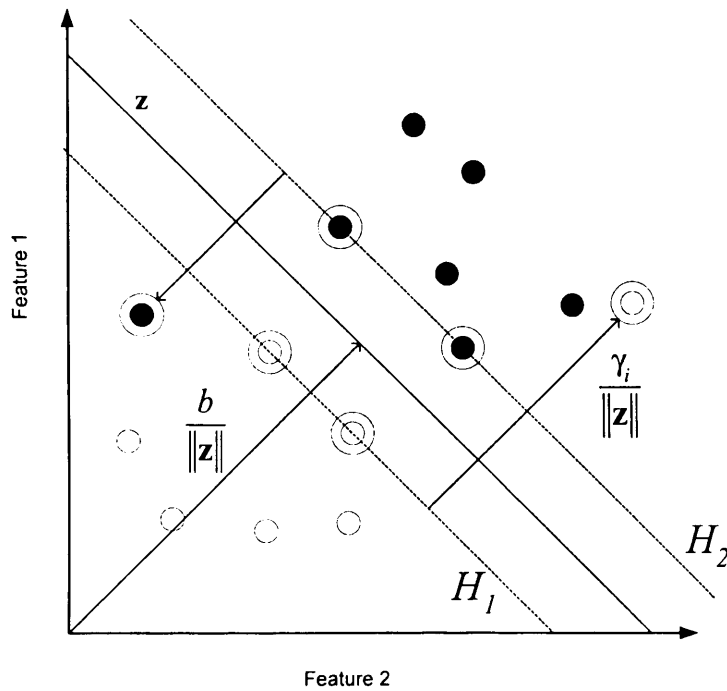


Figure 5.3: The feature space for a nonseparable case. The circled points are the support vectors calculated by minimising (5.5). The slack parameter  $\gamma_i$  enables the use of Lagrangian theory since it can now account for the overlapping features.

A kernel for a nonlinear SVM projects the samples to a feature space of higher dimension via a nonlinear mapping function.

Using Mercer's theorem [86], the high computational cost required in projecting samples into the high-dimensional feature space can be replaced by a simpler kernel function satisfying the condition  $K(\mathbf{g}_i, \mathbf{g}) = \varphi(\mathbf{g}_i) \cdot \varphi(\mathbf{g})$ . Where  $\varphi(\mathbf{g}_i)$  is nonlinear function which satisfies Mercer's condition [84].

Among nonlinear kernels the radial based function (RBF) defined as  $K(\mathbf{g}_i, \mathbf{g}) = \exp(-|\mathbf{g} - \mathbf{g}_i|^2 / (2\rho))$ , where the adjustable parameter  $\rho$  governs the variance of the function is more popular. Another kernel is the  $n^{th}$  order polynomial defined



as  $K(\mathbf{g}_i, \mathbf{g}) = (\mathbf{g}^T \mathbf{g}_i + 1)^n$ .

The main advantage of SVM is that unlike neural networks, it does not suffer from local minima since it has a convex convergence and therefore the local minima is also the global minima. Furthermore, with the correct choice of regularisation parameters it is robust against outliers. This makes the use of SVM ideal for classifying EEGs since the signals may be corrupted by noise and artifact, which may transpire in outlying features.

## 5.3 Experiments

### 5.3.1 Dataset for analysis

The data were provided by King's College Hospital, London U.K. and are available online [87]. The data represent a wide range of patients and therefore gives a comprehensive set of data for the evaluation of the proposed method. The scalp EEG was obtained using Silver/Silver-Chloride electrodes placed at locations defined by the 10-20 system. The data were acquired using a Beekeeper Telefactor EEG amplifier, sampled at 200Hz and bandpass filtered with cutoff frequencies of 0.3Hz and 70Hz. The independent components are obtained by applying BSS to blocks of data, 10 seconds in length. It is assumed that the number of sources is the same as the number of electrodes (i.e.  $N=M$ ). Then, the features are extracted from each of the ICs. The classifier was trained using the ICs from different patients.

### 5.3.2 Testing the Features

The features were tested using 200 ICs; 100 ICs containing eye blinks and 100 free of artifact. The classifier [88] was tested using a variety of kernels. For

each kernel the average error values were estimated with 4-fold cross validation i.e. using 75% of the data as training examples and 25% for testing with no overlapping. The cross-validation was performed 10 times, each time the data were randomly rearranged in order to yield a better estimate of the error. To find the value of parameter  $C$  the average CV test error is evaluated for a range of values for  $C$ . The optimum value of  $C$  was found to be 64 in the case of the linear and cubic polynomial. For the RBF kernel the parameters  $C$  and  $\rho$  were adjusted and found the optimal values for the RBF kernel as  $C = 72$  and  $\rho = 7$ . The CV error results are shown in Table 8.1.

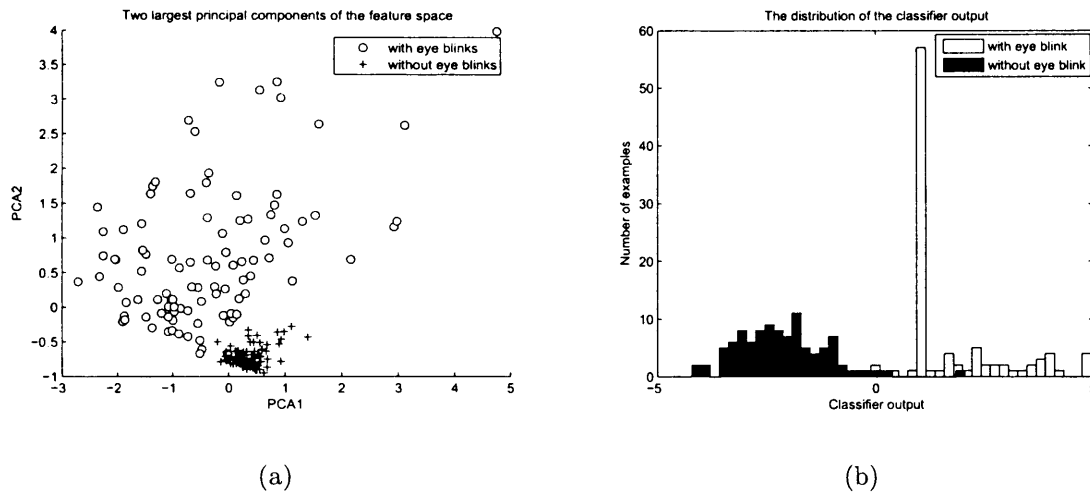


Figure 5.4: (a) A plot of the two largest principal components of the feature space. There are 200 feature vectors, 100 from normal EEG (+) and 100 from EEG containing eye blinks (o). (b) A histogram plot showing the output of the classifier pre  $sgn(\cdot)$  using the linear kernel.

Illustration of the distribution of the feature space becomes difficult when the dimension of the features is greater than 3. In order to understand the distribution of the feature space, one can use an RBF kernel with varying  $C$  and  $\rho$  to give



Table 5.1: The performance of the classifier based on the average number of correctly classified points. Three kernels are compared in the classification.

Kernel	Average classification rate (%) (s.d.)		
	Overall	Normal	Eye Blinks
Gaussian RBF	98.50 (1.00)	98.26 (1.17)	99.03 (1.35)
Cubic Polynomial	94.50 (1.92)	91.15 (2.31)	97.91 (2.04)
Linear	99.00 (1.15)	99.24 (1.11)	99.21 (0.97)

further insight into the optimum shape of the separating hyperplane. One would expect the number of SVs to decrease as  $\rho$  decreases. A linear kernel corresponds  $\rho \rightarrow \infty$ . Fig. 5.5(a) and Fig. 5.5(b) show the hyperparameter space for the RBF kernel. With the kernel width parameter  $\rho$  being finite, and regularisation parameter kept constant the classifier yields its highest classification rate with the lowest number of SVs. Therefore the feature space can be considered a linear one. The RBF kernel can be considered as both a linear and non-linear kernel depending on the parameter values,  $C$  and  $\rho$ , that are chosen.

In the case of cubic polynomial and linear kernels the number of support vectors found were 18% and 3.3% respectively of the training dataset size. The results in Table 8.1 show that, with the exception of linear kernel, the classifier had lower classification rates when classifying normal EEG. This may be due to non ocular related artifacts present in the EEG such as spikes, which produce similar feature values to that of the true eye blinks.

The training error was found by using the training data to test the SVM. The training error was found to be 2% (av) and the test error was 3% (av). This avoids any overfitting since the training error is close to the training error.

The classifier was further evaluated by plotting the distribution of the classifier output for 200 test points. It is calculated by applying the classification function in (5.8) without the  $\text{sgn}(\cdot)$  function. The result from the training data using the linear kernel is shown in Fig. 5.4(b). The ICs containing eye blinks are clustered around and above +1 and the ICs containing normal EEG activity around and below -1. There is minimal overlap between the classifier outputs, indicating that the proposed features are sufficiently significant to the detection of eye blinking artifacts for the test datasets.

For the dataset tested there is only 0.5% difference in the overall classification rate between the linear kernel and the RBF kernel. The cubic polynomial had the lowest overall classification rate. The largest difference in classification performance was between the RBF and cubic polynomial kernel when classifying normal EEGs, there was a difference of 7.1%. The reason for the close overall classification rates is mainly due to the separability of the feature space. Since the linear kernel requires fewer SVs in calculating the OSH and due to its computational simplicity, the linear kernel will be used to classify eye blinks in the following experiments. In order to test the significance of proposed features their eigenvalues were evaluated as 2.97, 0.68, 0.66 and 0.25; this testifies that the proposed features are significant to the detection of eye blinks in EEGs. A plot of the two largest principal components is shown in Fig. 5.4(a). From Fig. 5.4(a) it can be verified that the multidimensional feature space was linearly nonseparable, in the sense that there was an overlap between the features extracted from ICs containing eye blinking artifacts and those related to normal EEGs.

The BSS-SVM algorithm was applied to 10 real EEG datasets, each were 7 minutes long. The performance of the algorithm can be seen by comparing the EEG data obtained at the electrodes (see Fig. 5.6(a)) and the same segment of data after being processed by the proposed algorithm (see Fig. 5.6(b)). The

significance of the results was subjectively justified by a clinician at King’s College Hospital. The proposed algorithm was compared to EEGs reconstructed by manual artifact rejection (i.e. manually identifying and cancelling the artifact) by calculating the cross correlation between the BSS-SVM reconstructed EEGs and the manually reconstructed EEGs. The average value of cross correlation between the reconstructed EEGs is 0.92 (s.d. 0.02). In a number of trials the effect of ECG has been automatically detected and removed, whereas the complete removal has not been achieved with the method based on the manual selection. This had a detrimental effect on the cross-correlation measure since the BSS-SVM output will be less correlated with the manually reconstructed outputs, but has a positive effect on the output since there is less artifact present in the output.

As a second criterion for measuring the performance of the overall system a segment of EEG,  $x_{seg}$ , and the reconstructed EEG,  $\tilde{x}_{seg}$ , that do not contain any artifact were selected, and measured the waveform similarity,

$$\varepsilon_{dB} = 10 \log \left( \frac{1}{M} \sum_{i=1}^M (|1 - E\{(x_{i,seg}[n] - \tilde{x}_{i,seg}[n])\}|) \right).$$

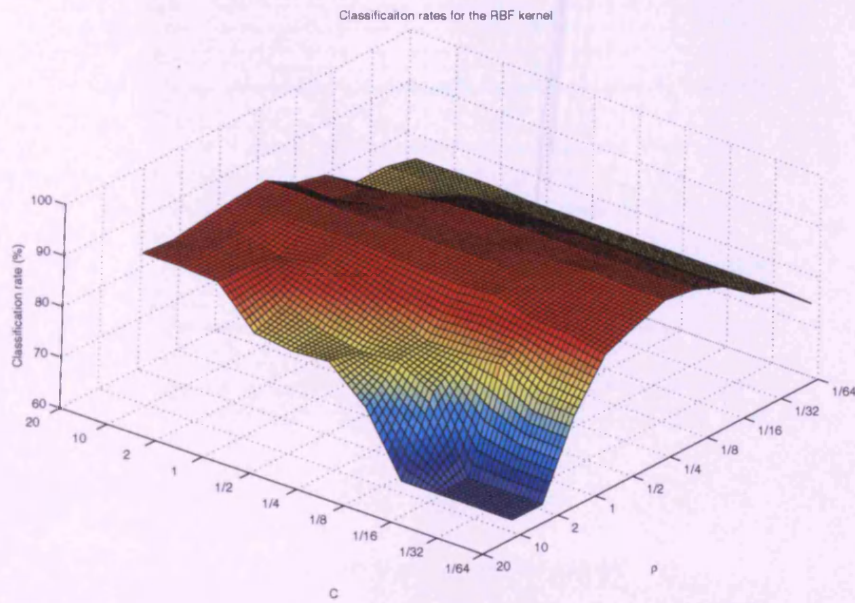
When the value of  $\varepsilon_{dB}$  is zero, the original and reconstructed waveforms are identical. From ten sets of EEGs the average waveform similarity was  $\varepsilon_{dB} = -0.009dB$  (standard deviation  $10^{-4}dB$ ). These results suggest that the observations have been faithfully reconstructed both in terms of subjective visual inspection and objective performance metrics.

## 5.4 Conclusions

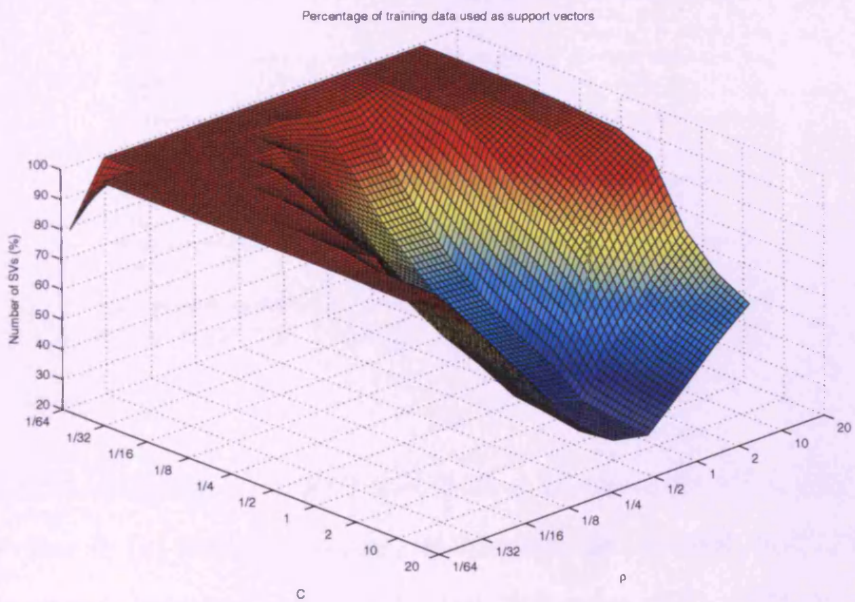
A robust method for removing ocular artifacts from EEGs by fusing BSS and SVM methods is presented in this chapter. The results show that the proposed algorithm identifies and removes the effect of eye blinking artifacts. A second

order methods was used to separate the sources which are spatially and temporally uncorrelated. The main advantage of using second order methods is that it requires fewer samples than the HOS methods, which lends itself to a lower computational complexity and hence shorter processing times. The efficacy of the SOBI algorithm in separation of OAs has been demonstrated in [48] and was exploited in this algorithm to extract features from the ICs. A second order method for source separation was used since, unlike higher order methods, it exploits the time structure of the EEGs. The EEGs are separated using the time lagged SOBI algorithm and the identified artifacts are autonomously cancelled, then the EEG is reconstructed from the remaining ICs.

Four features were identified as effective descriptors of eye blinking components. The selection of features were based on statistical measurements such as KL distance, cross correlation, power ratio, and skewness. The experiments herein demonstrate that for the test dataset the eye blinking sources are effectively classified by using the introduced features especially when the linear kernel is used for the SVM. It was demonstrated that the feature space is linearly separable by fixing the RBF kernel width parameter  $\rho$ , adjusting the slack parameter  $C$ , examining the number of support vectors found, and the corresponding classification rate. Based on the experimental data the BSS-SVM algorithm consistently removes the effect of eye blinking artifacts from the EEGs. When removing the artifacts from long data sets, manual removal of artifacts becomes infeasible and therefore automated techniques are required.

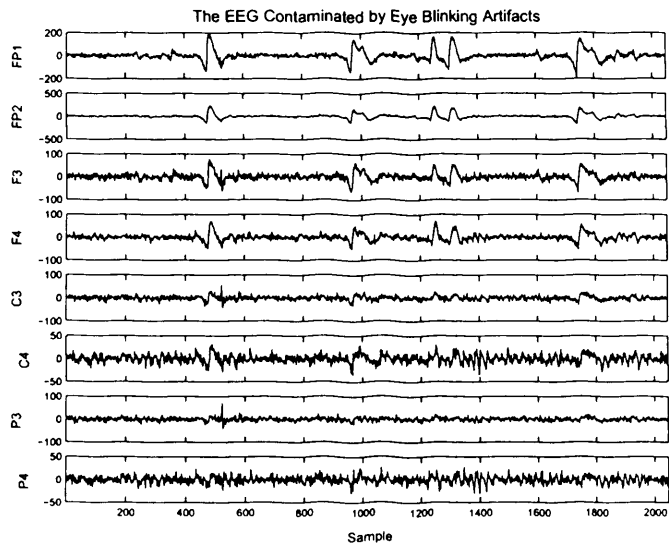


(a)

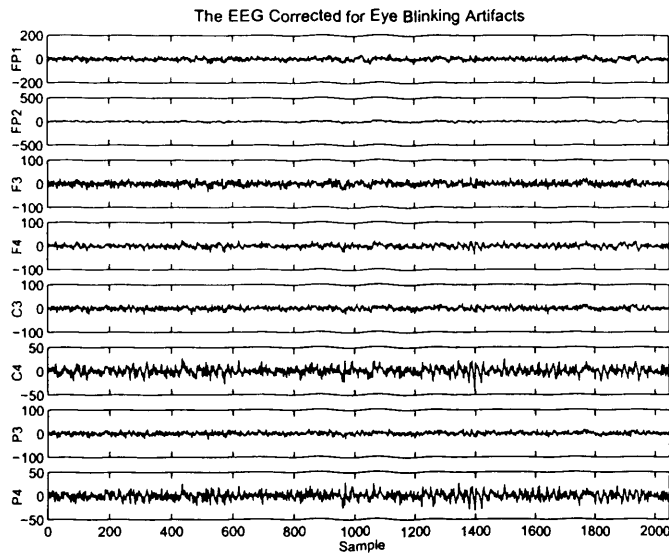


(b)

Figure 5.5: The (a) classification rate and (b) number of support vectors required for various parameter values of the RBF kernel.



(a)



(b)

Figure 5.6: A selection of 8 electrodes from a 16 electrode EEG recording. The OAs are clear in (a) between samples 400 to 600, 900 to 1400, and 1700 to 1900. They are more prominent over the frontal electrodes (FP1, FP2 etc.). (b) The same segment of EEGs after the eye blinking artifacts are removed using the proposed BSS-SVM algorithm.

## Chapter 6

# Brain Computer Interfacing by Localisation of Finger Movement Sources

In this chapter the aim is to localise the source of finger movement and track the motion of source location. Brain imaging and localisation techniques have been applied in order to localise brain abnormalities [89]. Source localisation techniques based on beamforming and direction of arrival (DOA), such as MUSIC and ESPRIT, perform well in noisy conditions, however they do not exploit the nature of the EEGs in calculating the location of the source, i.e. these methods do not use any prior knowledge in order to calculate the location of the sources.

A block diagram of the proposed localisation based BCI system is shown in 6.1. Firstly, a window of EEGs,  $\mathbf{X}(t, t + 1, \dots, t + L)$  is separated into its statistically independent components using the CBSS algorithm explained in chapter 4, where  $L$  is the window length. Next, the ICs are reordered and corrected for the sign and scale ambiguity of the BSS algorithm, as explained in the next section. Then each

source is localised (see Section 6.1.2) and the motion with respect to the previous processing window is characterised, as in Section 6.1.3. Finally the location of finger movement is identified and classified.

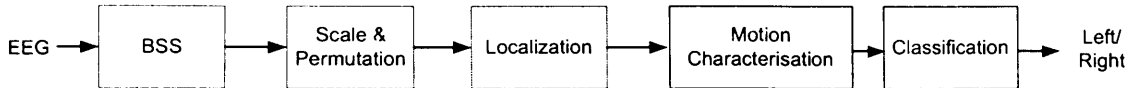


Figure 6.1: A block diagram of the localisation based BCI system.

## 6.1 Methods

### 6.1.1 BSS and the Permutation Problem

The first step in the BCI system is to separate the EEGs into their independent components. However, one of the ambiguities of BSS algorithms is that there is a permutation in the output. This means that the order of the ICs may change from one processing block to another. This problem is mitigated by following the overlapping window approach, as used in [90]. The constrained BSS algorithm proposed in Chapter 4 was used to separate the sources into independent components while removing any artifacts relating to eye blinks. The BSS and localisation algorithm are applied to a fixed length window  $L$  with overlap  $O$  samples ( $O < L$ ). Assume that  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$  represents the scalp EEG recording at time  $t$ , where  $M$  is the number of sensor signals and  $\mathbf{X}(t) = [\mathbf{x}(t), \mathbf{x}(t+1), \dots, \mathbf{x}(T)]^T$  is the entire scalp recording where  $T$  is the total recording length in samples. Therefore two consecutive overlapping windows would be represented by  $\mathbf{X}_1 = [\mathbf{x}(t), \mathbf{x}(t+1), \dots, \mathbf{x}(t+L)]$  and  $\mathbf{X}_2 = [\mathbf{x}(t+L-O), \mathbf{x}(t+1+L-O), \dots, \mathbf{x}(t+2L-O)]$  for  $0 < t < T - 2L + O$ .



The BSS algorithm is applied to  $\mathbf{X}_1$  and  $\mathbf{X}_2$  which gives the estimated sources for each of the windows  $\hat{\mathbf{S}}_1 = [\hat{\mathbf{s}}(1), \hat{\mathbf{s}}(2), \dots, \hat{\mathbf{s}}(L)]$ , where  $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), \hat{s}_2(t), \dots, \hat{s}_N(t)]^T$  are the sources estimated from  $\mathbf{X}_1$  and  $\hat{\mathbf{S}}_2 = [\hat{\mathbf{s}}^T(1), \hat{\mathbf{s}}^T(2), \dots, \hat{\mathbf{s}}^T(L)]^T$  where  $\hat{\mathbf{s}}^T(t) = [\hat{s}_1(t), \hat{s}_2(t), \dots, \hat{s}_N(t)]^T$ . The estimated sources may be permuted with respect to each other within the overlapping region. Therefore the overlapping region can be represented by  $\hat{\mathbf{S}}_2(t)_{t=1, \dots, O} = \mathbf{P} \cdot \mathbf{D} \cdot \hat{\mathbf{S}}_1(t)_{t=L-O, \dots, L}$ , where  $\mathbf{P}$  is an  $N \times N$  permutation matrix and  $\mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_N\}$  is the scaling matrix.

In order to measure the similarity between each of the estimated sources in the overlapping region the cross-correlation coefficient was used and is given by,

$$\rho_{s_i s_j} = \frac{r_{s_i s_j}}{\sigma(s_i)\sigma(s_j)}, \quad (6.1)$$

where  $r_{s_i s_j}$  is the cross-correlation between  $\hat{s}_i(t)$  and  $\hat{s}_j(t)$  (from hereon denoted by  $s_i$  and  $s_j$ , respectively) and is given by  $E\{s_i(t)s_j(t)\}$ , with  $t$  being within the overlapping region,  $\sigma(s_i)$  and  $\sigma(s_j)$  are the variances of  $s_i(t)$  and  $s_j(t)$  respectively. The cross-correlation coefficient satisfies  $-1 \leq \rho_{s_i s_j} \leq 1$ . When  $\rho_{s_i s_j} = 1$   $s_i$  and  $s_j$  are perfectly correlated, i.e.  $s_i = s_j$ . If  $\rho_{s_i s_j} = -1$   $s_i$  and  $s_j$  are anti-correlated, i.e.  $s_i = -s_j$ . When  $\rho_{s_i s_j} = 0$  there is no correlation between  $s_i$  and  $s_j$ . BSS yields statistically independent sources, and since independence is a stronger criteria than uncorrelatedness it is expected that the output of only one source in the overlapping region will be perfectly correlated. The permutation matrix can be found by  $\mathbf{P}_{ij} = \rho_{s_i s_j}$ .

Another ambiguity of BSS is the scaling and sign change of the ICs between each window. After performing BSS the variances of the output sources are normalised to unity and therefore matching sources from each window,  $\check{s}_i(t)$  and  $\check{s}_j(t)$  respectively, both have equal variance. Since the signals only share an overlap of  $O$  samples, the energy within the overlap segment of these signals will

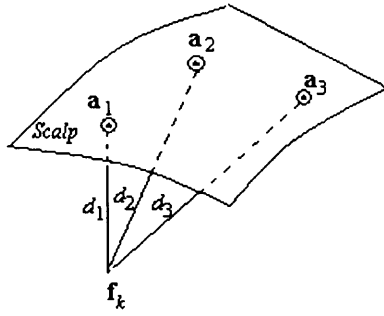


Figure 6.2: Part of the scalp including three electrodes, and the location of the source to be identified (assuming the head is homogenous)

generally be different and therefore can be used to solve the amplitude ambiguity. In particular

$$\hat{s}_i(t + L - O) = \text{sgn}(\check{\rho}_{ij}) \frac{\sigma_i}{\sigma_j} s_j(t), \quad \text{for } t = 1, \dots, L - O, \quad (6.2)$$

where  $\check{\rho}_{ij}$  is calculated for  $\hat{s}_i(t)$  and  $\hat{s}_j(t)$  within the overlap segment, and  $\hat{\sigma}_i$  and  $\hat{\sigma}_j$  are the variances of  $\hat{s}_i(t)$  and  $\hat{s}_j(t)$  respectively, within the overlap segment.

### 6.1.2 Localisation Algorithm

For the localisation algorithm the sources are considered to be isotropic signals within a homogenous and isotropic medium [91]. This means that the mixing media only mixes and attenuates the sources, i.e. there is no delay in the conduction from the source to electrodes. The attenuation corresponds to the distance between the electrode and the sources, as shown in Fig. 6.2. From Fig. 6.2 the distance between the source and electrode is given by,

$$d_j = \sum_{j=1}^3 \|\mathbf{f}_k - \mathbf{a}_j\|_2 \quad (6.3)$$

where  $\mathbf{f}_k \in \mathbb{R}^3$  is the location of the  $k^{\text{th}}$  estimated source and  $\mathbf{a}_j \in \mathbb{R}^3$  is the coordinates of the electrode. The parameter  $d_j$  is proportional to the distance between

the electrode and estimated source. In reality the head is a non-homogeneous region, therefore a more accurate estimation of the source position requires information about the non-homogeneity of the media between the source and the sensors [91]. However, since the aim is to characterize relative motion, an accurate location is not required and therefore the assumption of a homogenous medium is sufficient for this application. The relationship for conductivity between the source and scalp electrodes is shown in Fig. 6.3. The conductivity in Fig. 6.3 shows a simple representation of the non-homogeneity in the head, which would need to be accounted for when determining the distances between the source and the electrodes. The thickness of the skull varies between sexes and geographic origin [92]. Accounting for various skull thicknesses is beyond the scope of this research and is left for a later date. The skull has the lowest conductance and hence has the largest attenuation factor for the sources located beneath the skull, which constitutes most of the sources of interest. The distances to the electrodes can be represented as the inverse square of the cross-correlation between the source and the electrode. The relationship between the distance and cross-correlation is shown in Fig. 6.4.

In order to determine the value of  $d_j$  the three electrodes which contribute most to a source are found by estimating the correlation between each IC and all of the electrodes. The largest contributors are defined as the three electrodes that are most correlated with the IC to be localised. Then the distance parameter is given by

$$d_j = \frac{H}{E\{y_l(t)x_i(t)\}^2} \quad i = 1, 2, 3 \quad (6.4)$$

where  $y_l(t)$  is the estimated source from the BSS algorithm and  $x_{i=1,2,3}(t)$  are the three largest contributors to source  $y_l(t)$ ,  $H$  is a scaling factor which is empirically adjusted to accommodate any bias in the cross-correlation model of the distance

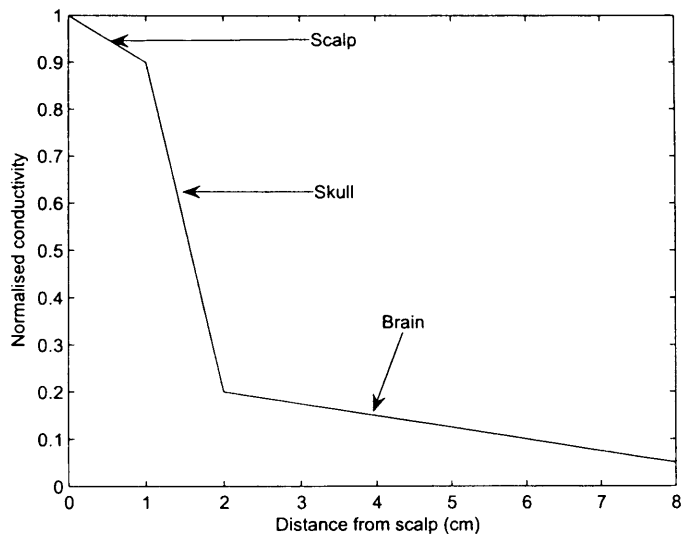


Figure 6.3: The relationship between the conductivity and the distance from the source to the electrode. When the source is close to the scalp the conductivity is large compared to sources that are deep within the brain as they have to travel through the skull.

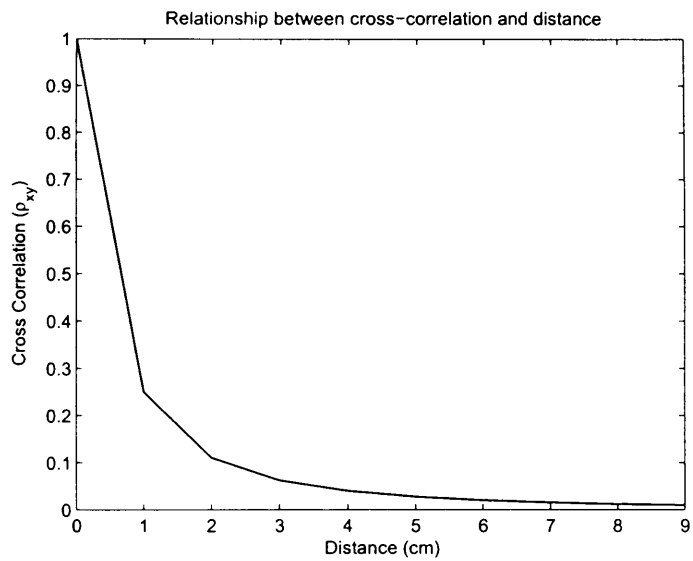


Figure 6.4: The relationship between the cross correlation and the distance from the source. Sources located closest to the electrode have larger cross correlation value than those located farther away. Cross correlation approximately reflects the conductance of various regions in the head (as in Fig. 6.3).

measure. The three dimensional coordinates of the estimated sources can be found [93] by solving the LS problem, formulated from in (6.3), given by,

$$\arg \min S(\mathbf{f}_k) \quad (6.5)$$

where

$$S(\mathbf{f}_k) = \sum_{j=1}^3 \{ \|\mathbf{f}_k - \mathbf{a}_j\|_2 - d_j \}^2. \quad (6.6)$$

The iterative solution to (6.5) is given by,

$$\mathbf{f}_k(n) = \mathbf{f}_k(n-1) + \mu \nabla S(\mathbf{f}_k(n)) \quad (6.7)$$

where  $\mu$  is the iterative learning rate and  $\nabla S(\mathbf{f}_k(n))$  is the gradient of (6.5), which is defined by

$$\nabla S(\mathbf{f}_k(n)) = 2 \sum_{j=1}^3 (\|\mathbf{f}_k(n) - \mathbf{a}_j\| - d_j) (\mathbf{f}_k(n) - \mathbf{a}_j). \quad (6.8)$$

The location defined at the minima of (6.8) yields the optimum location in the LS sense.

### 6.1.3 Motion Characterisation

The motion characterisation algorithm is based on the difference in the estimated source location between successive signal segments. The motion vector is calculated by

$$\mathbf{v}_k(t) = \mathbf{f}_k(t-1) - \mathbf{f}_k(t). \quad (6.9)$$

It is assumed that any movement in the location of finger movement sources is localised within the contralateral hemisphere as discussed by Pfurtscheller *et al.* [26] and demonstrated with LFP studies [94]. The number of motion vector trajectories is currently equivalent to the number of estimated sources (via BSS), i.e.  $k = 1, \dots, N$ . The number of sources related to finger movement, denoted as

$N_{mov}$ , is typically lower than the number of active sources in normal EEGs (typically  $N_{mov} \ll N$ ). Features that are extracted from sources that are unrelated to finger movement will introduce redundancy, which in turn will reduce the performance of the classifier. Therefore it is important to reduce the number of sources that are unrelated to finger movements. The physiological characteristics during finger movements were demonstrated by Pfurtscheller *et al.* [26]. They showed that the sources related to finger movements are localised to within the contralateral motor cortex. Therefore any motion vectors representing unexpected jumps from one side of the head to the other are considered as noise and removed. Large movements are detected by measuring the variance for each of the axis in  $\mathbf{f}_k(t)$ . If the variance is above a threshold then that component is rejected. Finally the angle between each motion vector is calculated to characterize the motion given by,

$$\phi(t) = 90 - 180 \arccos \left( \frac{\mathbf{v}_k(t)^T \mathbf{v}_k(t+1)}{\|\mathbf{v}_k(t)\| \|\mathbf{v}_k(t+1)\|} \right) \pi^{-1}. \quad (6.10)$$

where  $\phi(t)$  is the angle between motion vectors and  $\mathbf{v}_k(t+1)$  is the motion vector for the next processing window. Therefore, it is expected that for a source with a circular trajectory, the angle between motion vectors remain constant, either negative or positive depending on the direction (clockwise or counter-clockwise). The main reason for including this step rather than using the motion vectors or actual locations is to reduce the number of features. This method enables classification of a series of scalar values, representing angles between motion vectors, rather than a series of vectors representing motion values. This method effectively reduces the number of features from  $3Q$ , where  $Q$  is the number of processing windows for calculating the location of a source, down to  $Q$ . This also reduces the number of features required for training the classifier.

## 6.2 Experiments

### 6.2.1 Data Collection

The data was provided by King's College Hospital. The EEG was collected using 64 electrodes using a Synamp acquisition system from Neuroscan. The electrodes were placed using the extended 10-20 system referenced to linked mastoids. During acquisition the electrode impedance was kept below  $5k\Omega$ . The signal was sampled at  $2kHz$  and lowpass filtered with a cutoff frequency of  $200Hz$ . An able bodied subject was seated with arms resting on a table. The subject was asked to press either a left or right microswitch on a button box at approximately 5-7 second intervals. The test dataset consisted of 50 left and 50 right button presses with left or right hand respectively.

### 6.2.2 Results

#### Synthetic Data

In the first experiment the performance of the localisation algorithm when the distance between source and electrode was estimated by cross correlation, as in (6.4) was assessed. Four sources were generated by bandpass filtering white noise, such that there was minimal overlap between the spectra of each of the sources. One of the sources was a moving source with a trajectory shown in Fig. 6.5 by the circles and the others were static. This means that between consecutive processing blocks the geometric location of the source was moved (others kept constant) and then the localisation algorithm was applied. This was repeated for six consecutive processing blocks (as shown in Fig. 6.5). The sources were artificially mixed where the elements of the mixing matrix were proportional to the distance to the electrode. The scaling factor  $H$  was set to 0.75.



The results of the initial experiment are shown in Fig. 6.5. When the distances between the source and the electrodes were used instead of the cross correlation estimate in (6.4), the LS algorithm localises the sources accurately indicated by  $\times$  in Fig. 6.5. When the cross correlation estimate (6.4) is substituted for the known distance there was a constant bias in the calculated location of each source. This bias in the location was different for each of the sources. The bias in the locations of the estimated sources may have been introduced because the distance between the source and electrodes was based on the an estimate of cross correlation (6.4). The cross correlation is estimated from a finite data length, thereby introducing a source of error. Of importance, was that the moving source (labelled in Fig. (6.5)) was tracked and clearly in motion relative to the geometrically stationary sources.

The purpose of this experiment was to clarify that the relative motion of the source between consecutive localisation windows could be identified and not to exactly localise each source. Since the relative motion of the sources are used the exact location of the source is not required.

### **EEG Data with finger movement**

The aim of the next experiment was to assess the performance of the localisation algorithm with real EEGs. The window length,  $L$ , was set to 1000 samples and the overlap,  $O$ , was set to 95% of the window length. The centre of the first processing window (500<sup>th</sup> sample) was aligned such that it was it coincided with 500ms prior to the finger movement. It was suggested by Pfurtscheller *et al.* [26] that ERD occurs up to one second prior to finger movement. Therefore, an interval of 500ms prior to movement would increase the probability that the ERD is captured within the processing window. Furthermore, in order to reduce the processing time, the number of electrodes was limited to 21 over the sensory

motor cortex (FC, C, and CP electrodes).

Each consecutive EEG processing block was localised and plotted in Fig. 6.6. The general locations of the sources occurred in the contralateral hemisphere as described in [95] and were located within the motor cortex area. From Fig. 6.6 the localisation points occurred in the regions below the electrodes located over the sensory motor cortex, as opposed to the frontal or occipital regions of the brain. The contralateral locality of sources did not occur in all trials, i.e. the location of left finger trials occurred in the ipsilateral hemisphere. This may be due to the assumption that the conducting medium is homogeneous. Additionally the instantaneous cerebral blood flow changes the conduction characteristics of the brain as a function of time and mental task. Therefore the conduction characteristics may change between two consecutive processing windows. This may explain the large changes in location between two consecutive processing blocks thereby leading to unsatisfactory performance in the localisation algorithm.

It is difficult to interpret the individual source trajectories from Fig. 6.6, therefore the directionality of the sources were characterised by calculating the angles between the motion vectors as in (6.10). The histogram plot of the angle between consecutive motion vectors for 100 trials is presented in Fig. 6.7. It was expected that the distribution of angles between motion vectors would be distinct such that the sources that are related to left finger movement, for example, would be mainly positive angle values and those related to right finger movement would have negative angle values. This would indicate that the source of finger movement moves in distinct patterns for left and right finger movements. However, as shown in Fig. 6.7 the distribution of angles between successive motion vectors for left and right finger movements were almost completely overlapped. Assuming that the location of the sources are correct in relation to the previous processing window, the angle between successive motion vectors cannot be used

to discriminate between left and right finger movements.

### 6.3 Conclusions

In this chapter a BCI system based on BSS and localisation of moving sources was presented. The algorithm was tested on synthetic and real data. The synthetic data sets were modelled by instantaneously mixed isotropic sources with one source moving within a homogeneous medium. In this case the mixing matrix was proportional to the cross correlation between the estimated sources and electrodes, which was demonstrated in Fig. 6.5. The results indicated that this model was able to localise and track a moving source, albeit with a bias.

For real EEG experiments however, the localisation algorithm was not effective enough in localising the moving ERP sources related to finger movement. The main causes of this were due to the assumption that the head is a homogeneous medium and the number of sources was high. The high number of sources indicates that there are redundant ICs that are not related to the source of finger movement. This in turn leads to poor estimates in the locality of the sources in comparison with the synthetic datasets, where the number of sources was known *a priori*.

Visual inspection of the histogram plot for angles between successive positions (Fig. 6.7) for each class shows that by using this feature one cannot discriminate between left and right finger movements. In addition, a robust method for identifying the components related to ERPs, such as those methods proposed in [69] is required.

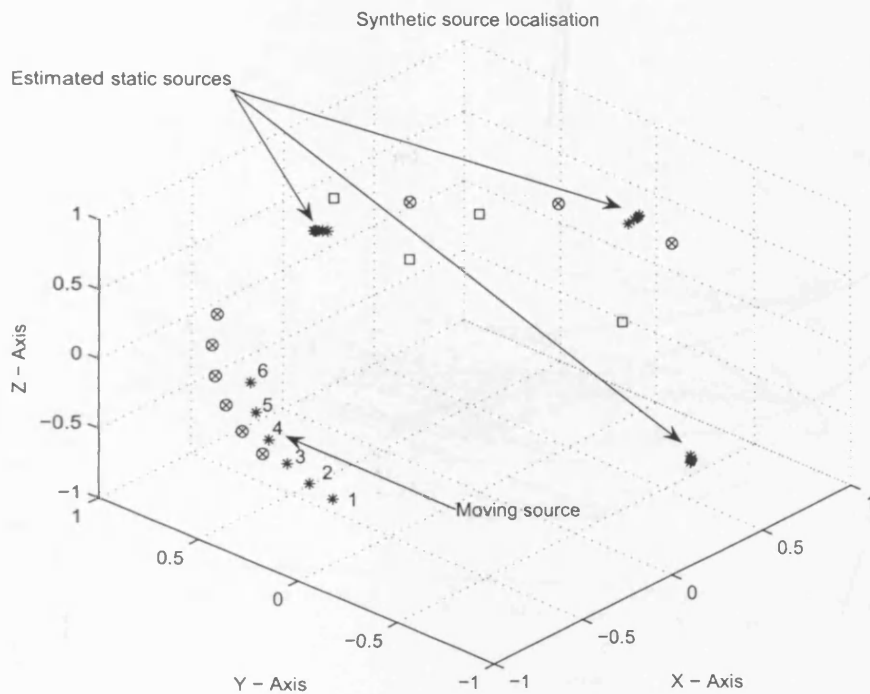


Figure 6.5: The results of the localisation algorithm for synthetic stationary and moving sources. The squares ( $\square$ ) are the known sensor locations, the circles ( $\circ$ ) are the known source locations, ( $\times$ ) are the locations calculated directly from the actual distances, and asterisks ( $*$ ) are the source locations calculated from the cross correlation value in (6.4). The numbers close to the estimated locations of the moving source identify the order of the consecutive localisation windows, i.e. 1 being the first processing window, and so on.

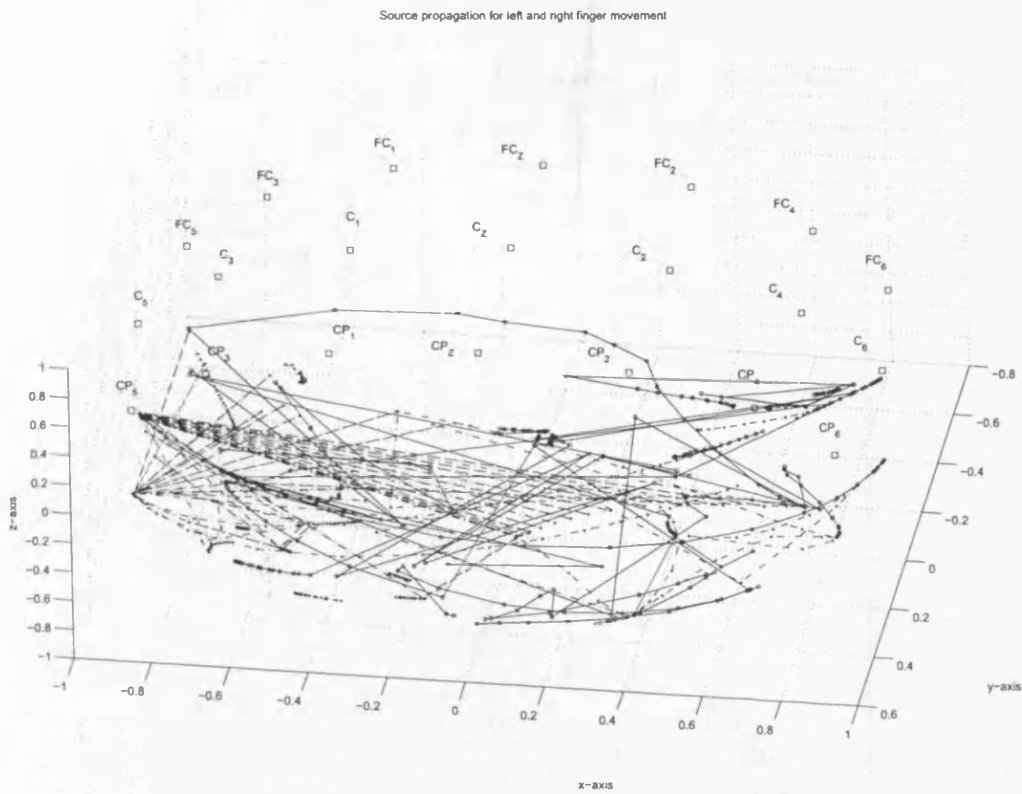


Figure 6.6: The results of the localisation algorithm for real EEGs containing left and right finger movement. The squares ( $\square$ ) are the known sensor locations and are labelled with the corresponding electrode name. Each + and  $\circ$  represent the source location of one processing window (for one IC) for left and right finger movement respectively. The lines between the + and  $\circ$  represent the trajectory of one estimated source between two consecutive localisation windows. For clarity the source trajectories of left finger movement are plotted using a solid line while right finger movement trajectories are plotted using a dash-dot line. The source locations are calculated assuming that they are proportional to the cross correlation value as in (6.4)

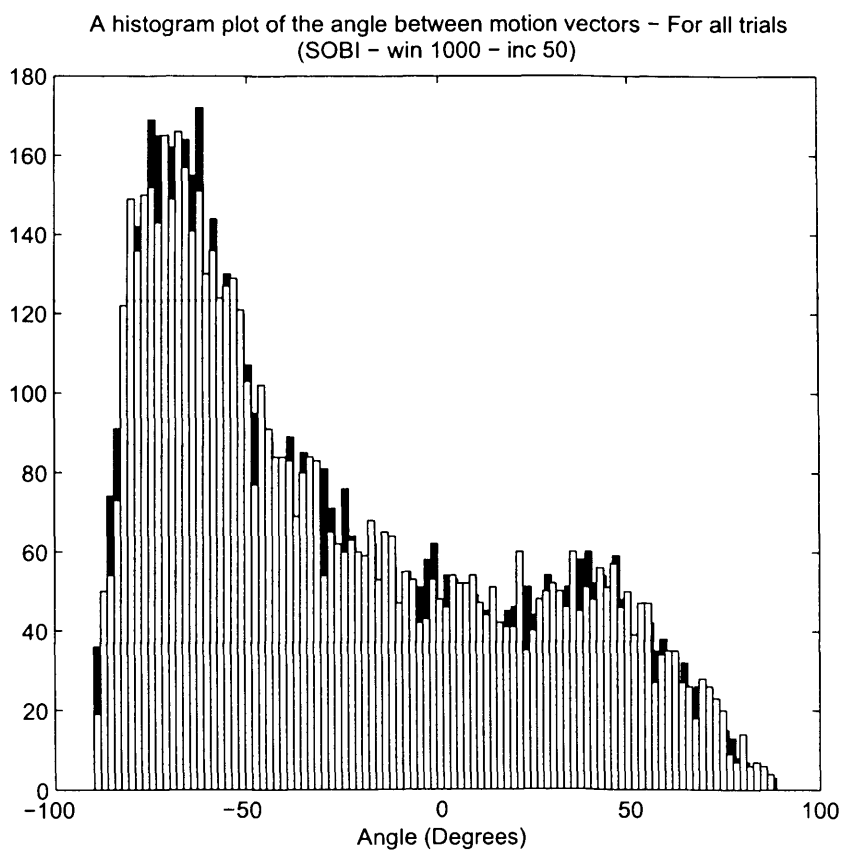


Figure 6.7: The histogram plot for the motion characterisation algorithm collapsed across all trials. The number of times that an angle between two consecutive processing blocks for left and right finger movement trials are represented by black and white bar plots, respectively. This shows that the angle between motion vectors does not reveal any significant distinctive features between left and right finger movements.

# Chapter 7

## Brain Computer Interfacing by Extracting Propagation Factors using Directed Transfer Function

### 7.1 Introduction

In this chapter the efficacy of features derived from the propagations of EEG sources during voluntary finger movements is evaluated. There are many methods used in BCI research for the quantification of EEG signal states, some mentioned in Section 3.3. This section outlines and exploits the interaction, or ‘cross-talk’, between EEG signals. In order to bring to light the relationship between the underlying processes during cognition, the frequency content and the phase relations between each of the electrodes are commonly used to determine the connectivity during performance of specific tasks. It has been well documented that different regions within the brain communicate with each other during mental tasks. The various cortexes that are involved in a particular mental task may be considered

as statistically independent processes [96]. However, the interaction between the various cortexes may not be captured with the statistical independence assumption. Rodriguez *et al.* [97] demonstrated that within the framework of facial recognition tasks there are transient periods of synchrony between various spatial locations within the brain. This manifests itself as a synchronisation of the phases in the electrodes located near to the corresponding cortex. It is not possible to directly infer causal relationships by visually inspecting the EEGs, and hence one must use advanced signal processing techniques in order to infer such causal relationships. One time series is said to be causal to another if the information contained in that time series enables the prediction of the other time series.

A common method for determining the synchrony in EEG activity is the spectral coherence [98], which is given by,

$$Coh_{ij}^2(f) = \frac{|C_{ij}(f)|^2}{C_{ii}(f)C_{jj}(f)} \quad (7.1)$$

where  $C_{ij}(f) = X_i(f)X_j^*(f)$  is the Fourier transform of the cross correlation coefficients between channel  $i$  and channel  $j$  of the EEG. An example of the cross-spectral coherence for an EEG signal 1 second prior to finger movement is shown in Fig. 7.1. Typically, a feature is created by averaging the coherence over a certain frequency band of interest, as  $Coh_{ave} = \frac{1}{f_2-f_1} \sum_{f_1}^{f_2} Coh_{ij}^2(f)$  [99], where  $f_1$  and  $f_2$  are the lower and upper frequency bounds. As can be seen from Fig. 7.1 coherence provides no information on the directionality of the coupling between two recording sites. Furthermore it is only capable of detecting zero time lag synchronisation and nonzero but fixed time lag synchronisation, which may occur when there is a significant delay between the two neuronal population sites [100].

Another method for extracting directionality from the EEGs is Granger causal-



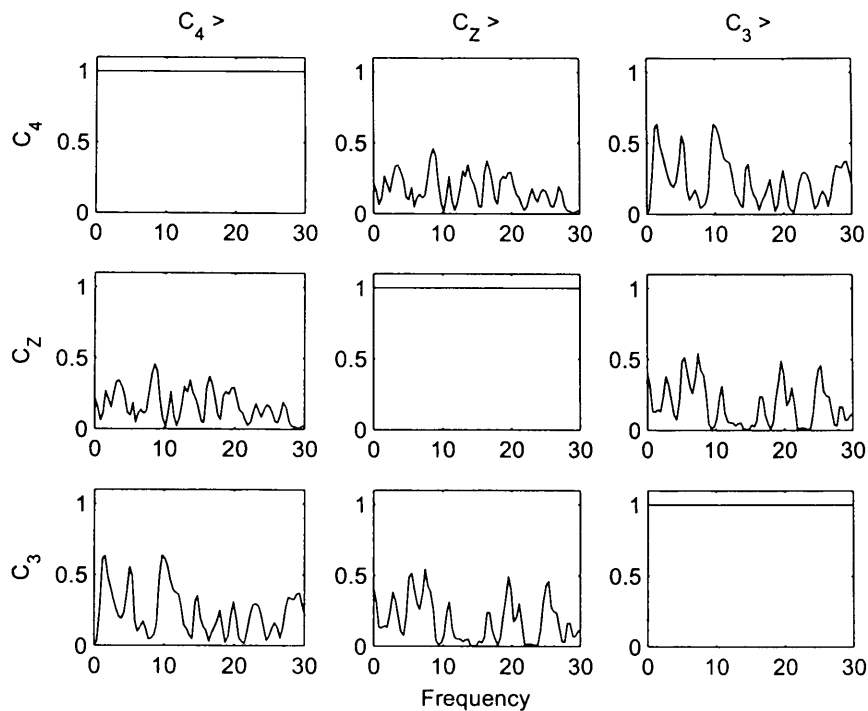


Figure 7.1: An example of the spectral coherence for the EEGs one second prior to finger movement.

ity [101] (also referred to as Wiener-Granger causality). It is based on bivariate auto regressive estimates of the data in order to infer causal relationships between the two time series. When determining directionality from more than two electrodes, Granger causality is typically calculated from pair-wise combinations of electrodes. In an experiment by Benasconi *et al.* [102] a trained cat performed a visiomotor behavioral task guided by events on the video screen. Intracortical local field potentials were recorded from various locations of the cat's visual system, while performing the task. The Granger causality was calculated for pairwise combinations of electrodes. They concluded that the Granger causality measure yielded satisfactory results but needed to be interpreted with caution be-

cause it often resulted misleading results, e.g. causal relations that were known to be untrue.

The directed transfer function (DTF) [103] has been developed in order to provide a method for detecting the direction of the coupling when applied to multivariate data. It was demonstrated in [104] that the DTF is a multivariate extension of the Granger causality measure. Granger causality can be applied to multivariate data by processing pair-wise combinations of the data. However, this approach often results in misleading results, i.e. pair-wise Granger infers causal relations that are untrue [102] [104] e.g. in experiments to highlight alpha rhythm dynamics while the eyes are closed [105], it was expected that the source of activity should be focused on the posterior region of the brain. DTF produced results inline with clinical explanations in that sources were located at  $P_4$ ,  $O_2$ , and  $O_z$  electrodes. In contrast, pair-wise Granger causality yielded causal relationships that were not precisely representative of clinical explanations, in that the source of activity was smeared over the majority of posterior electrodes. Another disadvantage of processing pair-wise Granger is that for each pair-wise combination the multivariate autoregressive (MVAR) parameters must be calculated, which accordingly increases the processing time. The advantage of DTF over spectral coherence is that it can determine the directionality in the coupling when the frequency spectra of the two brain regions have overlapping spectra. DTF has been adopted by many researchers for determining the directionality in the coupling [106] [105].

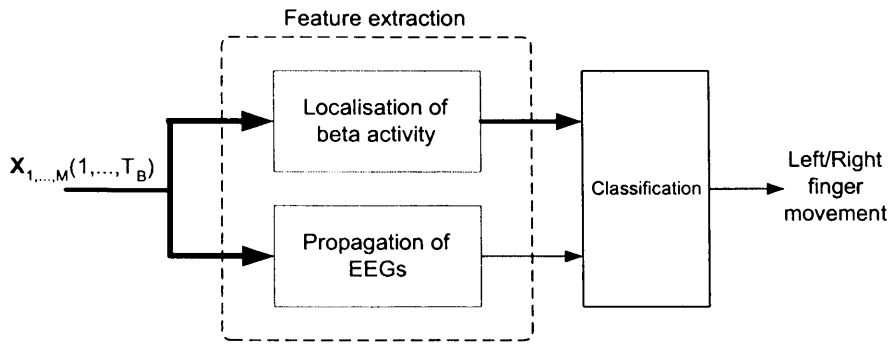


Figure 7.2: A block diagram of the proposed BCI system based on extracting the propagation of EEG sources and the location of beta activity.

## 7.2 Methods

### 7.2.1 Feature Extraction

A block diagram for the proposed classification system is shown in Fig. 7.2. The EEGs, denoted by  $\mathbf{X}_{1,\dots,M}(1, \dots, T_B)$  are processed in blocks of length  $T_B$ . The EEGs are then classified into left or right finger movement, based on the extracted features. The proposed features describing beta band activity and propagation of EEG activity are as follows:

#### Feature 1 - Localisation of beta band activity

It is well established that after imagined and real finger movement there is an increase in the amplitude of the beta band, which is known as event related synchronisation (ERS), over the contralateral side of the sensory motor cortex to the moved or imagined finger [95]. A block diagram for extraction of this feature is shown in Fig. 7.3. The inputs to this system are the EEGs, denoted as  $X_{1,\dots,M}(1, \dots, T_B)$ , where  $M$  is the number of electrodes and  $T_B$  is the data block length. The first stage in extracting this feature is to bandpass filter the

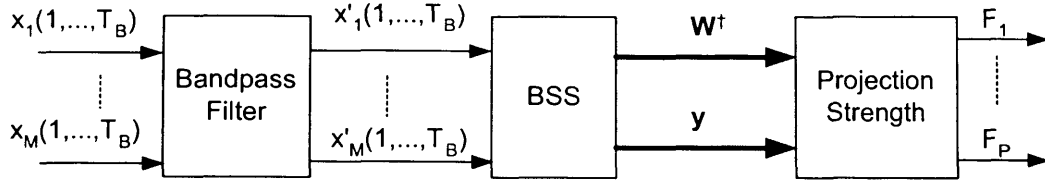


Figure 7.3: A block diagram showing the localisation of beta band activity feature. The EEGs are processed in blocks of  $T_B$  samples. The outputs of this feature are the reprojected beta band power values of the IC corresponding to finger movement i.e. the IC with the largest reprojection to the electrodes located over the motor cortex.

EEGs, with cutoff frequencies of  $14Hz$  and  $20Hz$ , denoted as  $X'_{1,\dots,M}(1, \dots, T_B)$ . Then the filtered EEGs are separated into statistically independent sources, using the proposed constrained BSS algorithm (as in Section 4), denoted by  $y$  and the projection strength of the IC onto the scalp electrodes is defined by  $\mathbf{W}^\dagger$ . When separating the EEGs it is assumed that the number of sources is equivalent to the number of electrodes ( $N = M$ ) and therefore the unmixing matrix will be square. The parameter  $F_{1,\dots,P}$  is the extracted feature, where  $P$  is the total number of electrodes over the motor cortex. A gradient implementation of SOBI was used to separate the sources due to its proven separation performance in ERP extraction [69] [46].

The next stage is to find the ICs whose projections are over the sensory motor cortex, which correspond to electrodes  $T_3$ ,  $T_4$ ,  $C_3$ ,  $C_4$  and  $C_z$ . The ICs projected onto other electrodes, i.e. those that are not located over cortical regions associated with finger movement, are considered to be unrelated to finger movement or noise. The columns of the inverse of the unmixing matrix  $\mathbf{W}^{-1}$  reflect the projection strength of the ICs onto the electrodes. One may use the

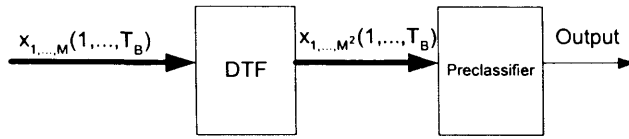


Figure 7.4: Block diagram of the system for classification of the propagation features.  $\mathbf{X}$  is an EEG block.

rows of  $\mathbf{W}^{-1}$  that correspond to the electrodes over the sensory motor cortex as a feature. For example, the projection strength for all ICs onto one electrode would be in the  $i^{th}$  row of  $\mathbf{W}^{-1}$ . The scale ambiguity of BSS means that using the inverse of unmixing matrix, directly, will yield an inaccurate representation of the projection strengths when comparing two different ICs. However, the scale information is preserved when combining  $\mathbf{W}^{-1}$  with the IC  $y_i(t)$ .

Therefore the ICs are reprojected to the electrodes, given by

$$\hat{\mathbf{x}}_i(t) = \check{\mathbf{w}}_i y_i(t) \quad (7.2)$$

where  $\check{\mathbf{w}}_i$  is the  $i^{th}$  column of  $\mathbf{W}^{-1}$ ,  $y_i(t)$  is the  $i^{th}$  IC, and  $\hat{\mathbf{x}}_i(t)$  is the reprojection of the  $i^{th}$  IC to all electrodes. Then the power within beta band (14Hz-20Hz) is calculated for each of the electrodes located over the motor cortex and the IC corresponding to the largest beta band power values over the motor cortex is used as the feature.

## Feature 2 - Propagation of the EEGs

The second feature is the flow of cortical information between electrodes, within the  $8Hz - 13Hz$  range. Ginter *et al.* [107] demonstrated that there is a directed flow of information or ‘cross-talk’ between the sensors around the sensory motor area before finger movement. A block diagram for this feature is shown in Fig. 7.4. The flow of information between EEG electrodes in the al-

pha band is used to extract a feature value by using a preclassifier. The method used to calculate the information flow, initially developed by Kaminski *et al.* [103], is called directed transfer function (DTF). The DTF is based on fitting the EEGs to a multivariate autoregressive (MVAR) model. Assume that the process  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$  is sufficiently described by

$$\mathbf{x}(t) + \mathbf{L}(1)\mathbf{x}(t-1) + \dots + \mathbf{L}(m)\mathbf{x}(t-m) = \mathbf{v}(t) \quad (7.3)$$

where  $\mathbf{x}(t)$  is the  $M$  channel EEG and  $t$  denotes the time index,  $\mathbf{L}(i)$  is an  $M$ -by- $M$  matrix containing the model order coefficients and  $\mathbf{v}(t)$  is a vector containing the residual prediction error and can be considered as zero mean noise [108], with a covariance matrix  $\mathbf{\Omega}$ . Post multiplying (7.3) by  $\mathbf{x}^T(t-k)$  and taking expectations, where  $k = 1, 2, \dots, m$  and  $m$  is the model order which is chosen using Akaike AIC criterion [52], yields the following Yule-Walker equation [108]

$$\mathbf{R}(-k) + \mathbf{L}(1)\mathbf{R}(-k+1) + \dots + \mathbf{L}(m)\mathbf{R}(-k+m) = \mathbf{0} \quad (7.4)$$

where  $\mathbf{R}(q) = E\{\mathbf{x}(t)\mathbf{x}^T(t+q)\}$  is the covariance matrix of  $\mathbf{x}(t)$ , and the noise vector is  $\mathbf{0}$  because  $\mathbf{x}(t)$  is spatially and temporally uncorrelated with the noise vector  $\mathbf{v}(t)$ , i.e.  $E\{\mathbf{v}(t)\mathbf{x}^T(t+q)\} = \mathbf{0} \quad \forall q$ . There are a number of methods for calculating the MVAR coefficients such as Levinson-Wiggins-Robinson (LWR) algorithm [109].

In order to yield an accurate representation of the EEGs, long stationary data are required for estimating the correlation matrices and MVAR coefficients. EEGs can be considered stationary over a short interval and since the sampling rate is typically 250Hz there are not enough data points to yield an accurate estimate of the correlation matrix. Ding *et al.* [108] proposed an algorithm that reduces redundancy in the data given multiple realisations of the same process. This is usually the case for ERP data when there are a number of trials of the same

action. The inter trial ensemble average of the estimated cross correlation can be used to reduce any redundant information and thereby increasing the efficiency of the estimate, given by

$$\tilde{\mathbf{R}}(q) = \frac{1}{N_r} \sum_{r=1}^{N_r} \mathbf{R}_r(q) \quad (7.5)$$

where  $\mathbf{R}_r(q)$  is the cross correlation matrix for trial  $r = 1, \dots, N_r$ , and  $N_r$  is the total number of trials. Provided that the number of trials is large, the data used to calculate the cross correlation matrix can be as short as  $m + 1$  samples.

Once the MVAR coefficients are found to satisfy the model in (7.3) it is transformed into the frequency domain

$$\mathbf{L}_f(f)\mathbf{x}(f) = \mathbf{v}(f) \quad (7.6)$$

where

$$\mathbf{L}_f(f) = \sum_{j=1}^m \mathbf{L}(j)e^{-i2\pi fj/N_f} \quad f = 0, 1, \dots, N_f - 1 \quad (7.7)$$

with  $\mathbf{L}(0) = \mathbf{I}$  and  $N_f$  is the window length of the FFT. Rearranging (7.6) yields

$$\mathbf{x}(f) = \mathbf{L}_f^{-1}(f)\mathbf{v}(f) = \mathbf{H}(f)\mathbf{v}(f) \quad (7.8)$$

where  $\mathbf{H}(f)$  is the transfer matrix of the system [108] [103] [104]. The DTF or causal relationship between channel  $i$  to  $j$  can be defined directly from the transform coefficients [104] given by

$$\Theta_{ij}^2(f) = |H_{ij}(f)|^2 \quad (7.9)$$

where  $|\cdot|$  is the absolute norm. Electrode  $i$  is causal to  $j$  at frequency  $f$  if  $\Theta_{ij}^2(f) > 0$ . The DTF is calculated for a window size of 500 samples (500 ms). A time-varying DTF is generated by calculating the DTF over short windows to yield the short time DTF (SDTF). A window size of 500 samples was chosen for calculating the short time DTF in order to capture at least four cycles of the

lowest frequency of interest and thereby improving the frequency estimate. The window is then shifted by 50 samples (90% overlap) and the DTF is calculated for the new window. The same model order was used for all of the windows and trials.

The final stage in the extraction for this feature is to calculate the mean amplitude of the DTF in the alpha frequency band and over time for each electrode. The feature vector is made up from the average flow of information between electrodes  $i$  to  $j$  in the alpha frequency band. Then the feature vector is passed through a preclassifier, from which a value is obtained as the feature. The reason for the preclassification stage is to reduce the number of features in the main classification stage.

## 7.3 Experiments

### 7.3.1 Data Collection

The data was provided by King's College Hospital. The EEG was collected using 64 electrodes using Neuroscan Synamp<sup>2</sup>. The electrodes were placed using the extended 10-20 system referenced to linked mastoids using an electrode cap for placement. The signal was sampled at  $1kHz$  and lowpass filtered with a cutoff frequency of  $200Hz$ . The experimental paradigm was as shown in Fig. 7.5. The blank screen period at the beginning of each trial allows the user to blink and get comfortable. This region of the trial was rejected from the datasets since it doesn't contain any useful information. At  $t = 2.5s$  either an 'E' or 'I' is presented at the centre of the screen informing the user that the action should be done explicitly or imagined, respectively. At  $t = 3.5$  seconds the user was presented with either an 'L' for left hand, 'R' for right hand, or 'N' for no response. At  $t = 4.5$  seconds



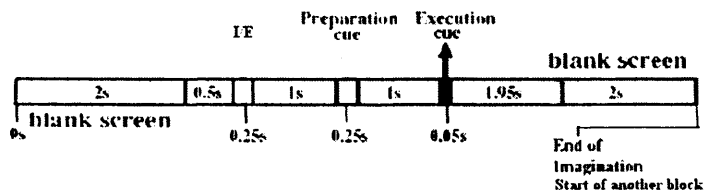


Figure 7.5: The recording procedure for the BCI data. Each trial lasted for a total of 9 seconds. Between each trial there was an interval in which the user was able to blink or get comfortable.

an 'X' is presented on the screen at which point the user is to initiate the required task dictated by the previous two cues. The proceeding 2 seconds contain the actions being performed.

Each session consisted of two blocks with a five minute break in between blocks. Each block consisted of 45 trials for each combination of the above commands, i.e. 45 explicit left finger movements, 45 imagined right finger movements, etc. The combination of commands was generated randomly for each block.

### 7.3.2 Testing the Features

In this study the features were tested using 180 trials in total; 90 for left finger movement and 90 for right finger movement. An additional 50 trials consisted of 25 left and 25 right finger movements.

When using all of the 64 electrodes to calculate the DTF, the algorithm was very slow and was very difficult to visualise. This is because there would be  $64^2$  time-frequency plots for the SDTF corresponding to every combination of the inter-electrode flow. A solution to this problem was proposed by [107] on the basis that there is negligible cortical flow between the hemispheres. Therefore, the electrodes on each hemisphere will be processed separately.

An example of the SDTF ensemble averaged over all trials for left and right finger movement is shown in Fig. 7.6 and Fig. 7.7 respectively. The SDTF over the right hemisphere for left finger movements (Fig. 7.6) shows that during finger movement ( $t = 3$ ) the FC6 electrode is active, which is indicated by the red areas in the column of FC6 electrode. This suggests that the source related to left finger movement is located in the vicinity of that electrode. Furthermore, between  $t = 2$  and  $t = 3$  seconds (before finger movement) there is a reduction in the SDTF which can be seen by the blue areas between those two time points in the column of electrodes C6 and CP6. For further clarification of the characteristics of the SDTF during left finger movement it was averaged over the alpha band and shown in Fig. 7.9. In Fig. 7.9 a value of one (on the y-axis) indicated that there is a causal relationship in the alpha band between the electrode denoted above the column to the electrode denoted to the left of the row. The dotted line indicates the point of finger movement. Comparing with the left hemisphere (ipsilateral hemisphere) shown in Fig. 7.8, the SDTF of the left hemisphere is attenuated in comparison with that of the right hemisphere at  $t = 3$ .

The STDF calculated over the left hemisphere for right finger movement is shown in Fig. 7.6. In this case the active electrodes are CP5 and CP3 indicated by the red areas in the figure. At the time of finger movement ( $t = 3$ ) CP3 is causal to FC3 and FC5 indicated by the red area at that time index. Another interesting observation is that while CP3 is active CP5 is inactive between  $t = 2$  and  $t = 3$  (blue areas prior to finger movement). This suggest that there is movement in the causality from electrode CP5 to CP3.

Comparing the left hemisphere (Fig. 7.10) alpha band STDF with that of the right hemisphere (Fig. 7.11) one sees that the SDTF is more attenuated at the point of finger movement ( $t = 3$ ). This is expected since experiments by Pfurtscheller *et al.* [27] using a similar experimental paradigm showed that

the contralateral hemisphere to finger movement is more desynchronised during finger movement compared with the ipsilateral hemisphere. Furthermore, attenuation of the SDTF at the locations identified as sources during finger movement are expected since during desynchronisation the EEGs become more chaotic or ‘complex’, therefore are not linearly interrelated thereby yielding a lower SDTF value. It is difficult to compare and contrast the SDTF plots for left and right finger movement by visual inspection, however the features derived from SDTF are separable on a higher dimensional feature-space.

The preclassifier was trained using the 50 trial set and the SVs were saved. The preclassifier was the discriminating function of the SVM defined in (5.8), but without the  $sgn(\cdot)$  function. For the main classification task an SVM was used with the feature vector given by

$$\mathbf{g}_i = [F_1(i), \dots, F_P(i), F_{P+1}(i)]^T \quad (7.10)$$

where  $F_{1,\dots,P}(i)$  are the beta band power values for the  $P$  electrodes located over the motor cortex, and  $F_{P+1}(i)$  is the output of the preclassifier.

In order to test the overall classification rate 4-fold cross-validation (CV) with no overlap was used, i.e. using 75% of the data for training and 25% for testing. The CV was performed 10 times, in each time the data was chosen at random from the trial pool. The classifier was used with three kernels, linear, RBF and cubic polynomial, for which the error is shown in Table 8.1. The optimum value for parameter  $C$  was found by calculating the CV classification rate for a range of values i.e.  $C$  in the range of 0.1 (non overlapping classes) to 1000 (overlapping classes). For the RBF kernel the optimal parameter values,  $C$  and  $\rho$ , were found by varying each parameter in turn and using the combination that yields the highest classification rate. The optimum values for  $C$  and  $\rho$  for RBF kernel were respectively, 10 and 0.9. The model order,  $m$ , for the propagation feature was

found to be 8 by applying the AIC algorithm [52] and was kept constant across all trials.

The average number of SVs calculated when using the RBF kernel was 54.9% of the training examples. When using the linear kernel the average number of SVs found was 12.4% and for cubic polynomial it was 16.5%. The training error was found by using the training data as test data. The training error was found to be 5% (ave.) and the test error was 7% (ave.). Since they are close together this gives an indication that the overfitting of the separating hyperplane has been avoided.

From the classification results in Table 8.1 it is possible to see that the RBF kernel performs best in terms of overall, left and right finger movement classification. This testifies that the feature space is not linearly separable because the classification rate increases when applying the RBF kernel compared with the linear kernel. Given that the optimum kernel width parameter of the RBF kernel was greater than 0 supports the use of nonlinearity in the OSH. An interesting pattern emerges from the results of Table 8.1, in all cases right hand finger movement yields a higher classification rate than the left. Pfurtscheller *et al.* [20] observed that that the contralateral hemisphere to the dominant hand yielded a larger amplitude in the alpha band than the contralateral hemisphere of the less dominant hand. The subject was right handed and therefore may explain the higher classification rates for right finger trials.

The classifier was further evaluated by plotting the distribution of the classifier output. It was calculated by using (5.8) before  $sgn(\cdot)$  operation. The results are shown in Fig. 7.12. The trials which refer to right finger movement are clustered around  $-1$  and those corresponding to left finger movement are clustered around  $+1$ . There is a small amount of overlap indicating that these features are significant for detecting left and right finger movements for the test datasets.

Table 7.1: The performance of the classifier based on the average number of correctly classified points. Three kernels are compared in the classification.

Kernel	Average classification rate (%) (s.d.)		
	Overall	Right	Left
Gaus. RBF	83.50 (1.0)	81.16 (1.2)	78.23 (1.5)
Cubic Poly.	73.30 (1.4)	72.15 (1.0)	70.36 (1.0)
Linear	67.01 (1.3)	69.34 (1.4)	63.41 (1.0)

## 7.4 Conclusions

A robust method was proposed for distinguishing between left and right finger movements from scalp EEGs using the features corresponding to the beta rhythms and propagation of those sources.

SOBI was shown to be a useful tool for highlighting the ERPs related to finger movement. The projection strength of the ICs onto the electrodes over the motor cortex shows that for left finger movements the ICs project onto the motor cortex in the right hemisphere, and vice versa for right finger movements. The direction of information flow within the brain during finger movement, was used as another feature. In order to detect and quantify the direction of flow the DTF was applied using short sliding windows so that the time varying characteristics in information flow can be detected.

The experiments herein demonstrate that for the test dataset left and right finger movements are correctly classified by using the introduced features. A higher classification rate is achieved when the RBF kernel is used for SVM.

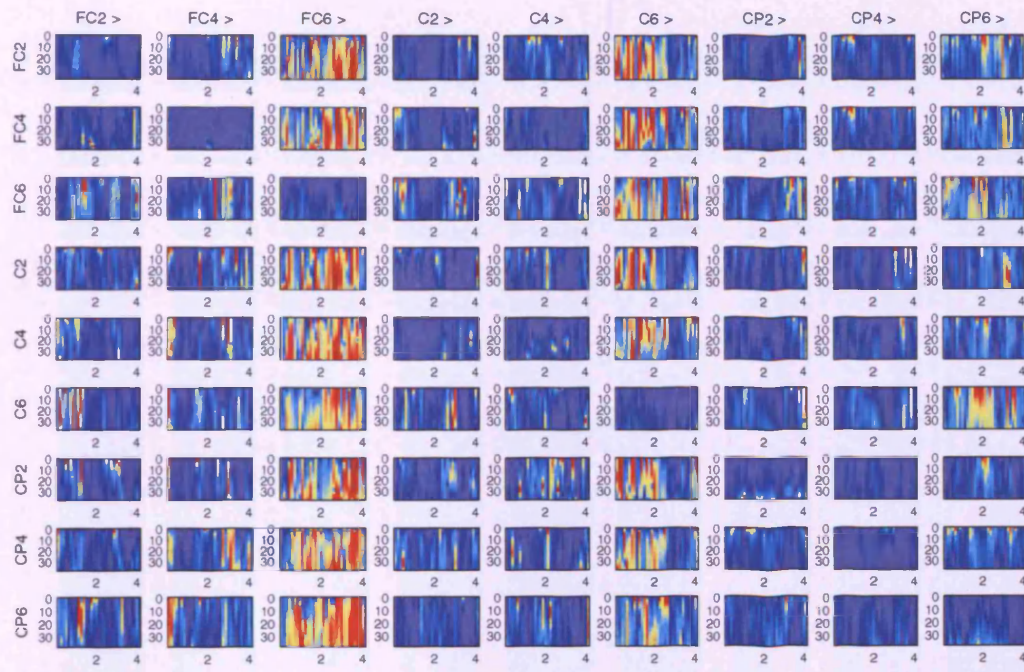


Figure 7.6: SDTF calculated for left finger movements. Frequency is plotted on the  $y$  axis and time (seconds) along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds. The direction of flow is read from the electrode denoted above the column to the electrode denoted by the label on the left of the rows. Electrode FC6 (column) is the most active because of the high value of STDF (red), suggesting that the source of finger movement is located close the that electrode.

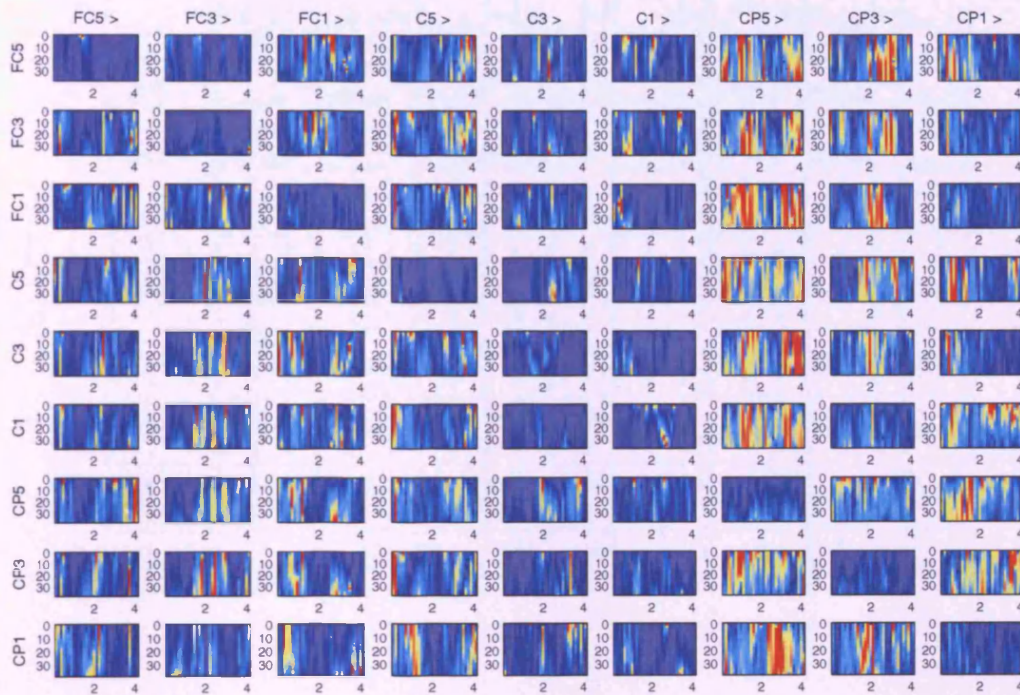


Figure 7.7: SDTF calculated for right finger movements. Frequency is plotted on the  $y$  axis and time (seconds) along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds. The direction of flow is read from the electrode denoted above the column to the electrode denoted by the label on the left of the rows. In this case electrodes CP5 and CP3 (columns) are the most active at the time of finger movement, suggesting that the source is located close to those electrodes.

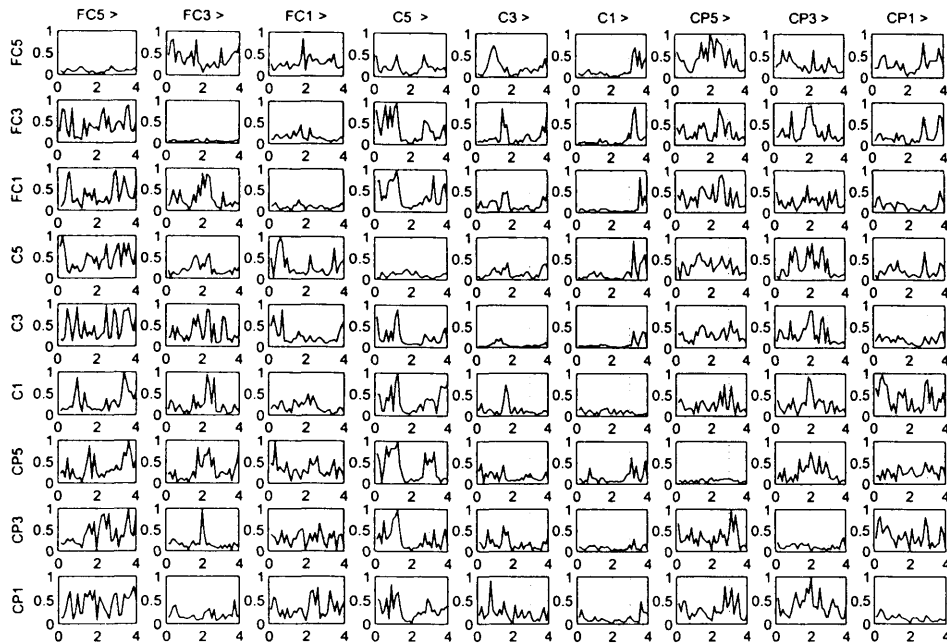


Figure 7.8: The SDF calculated for electrodes located over the left hemisphere for left finger movement trials averaged over the alpha band. The  $y$  axis shows the SDF value, where a value of 1 indicates maximal causal relation between the electrode denoted above the column to the electrode denoted to the left of the row. Time (seconds) is shown along the  $x$  axis. The movement of the left finger occurs at  $t = 3$  seconds, which is indicated by the dotted line.



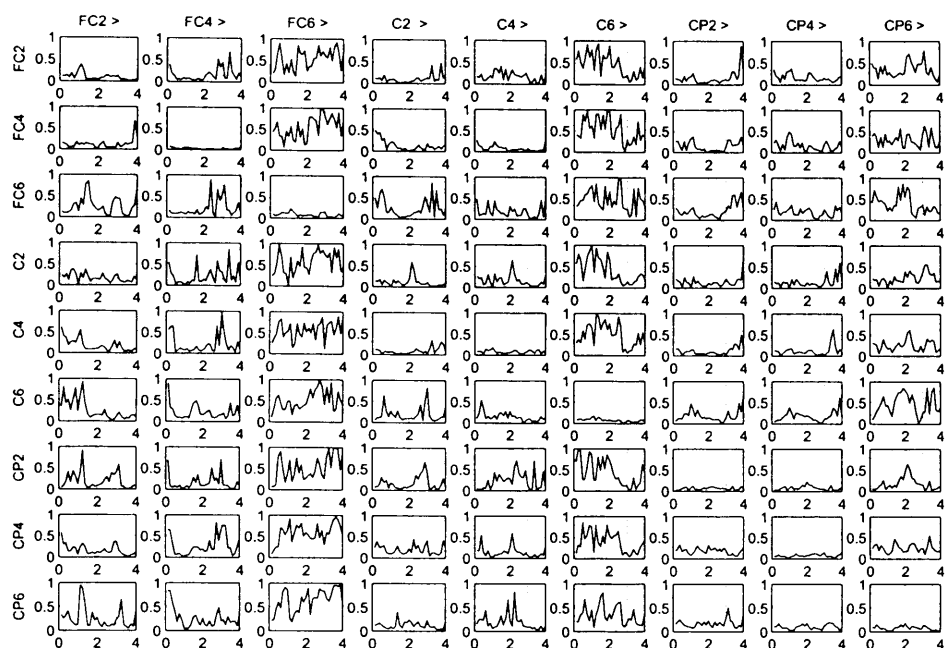


Figure 7.9: The SDTF calculated for electrodes located over the right hemisphere for left finger movement trials averaged over the alpha band. The axis are as in Fig. 7.8 Lower values of SDTF are observed near the time of finger movement compared with the left (ipsilateral) hemisphere (Fig. 7.8)

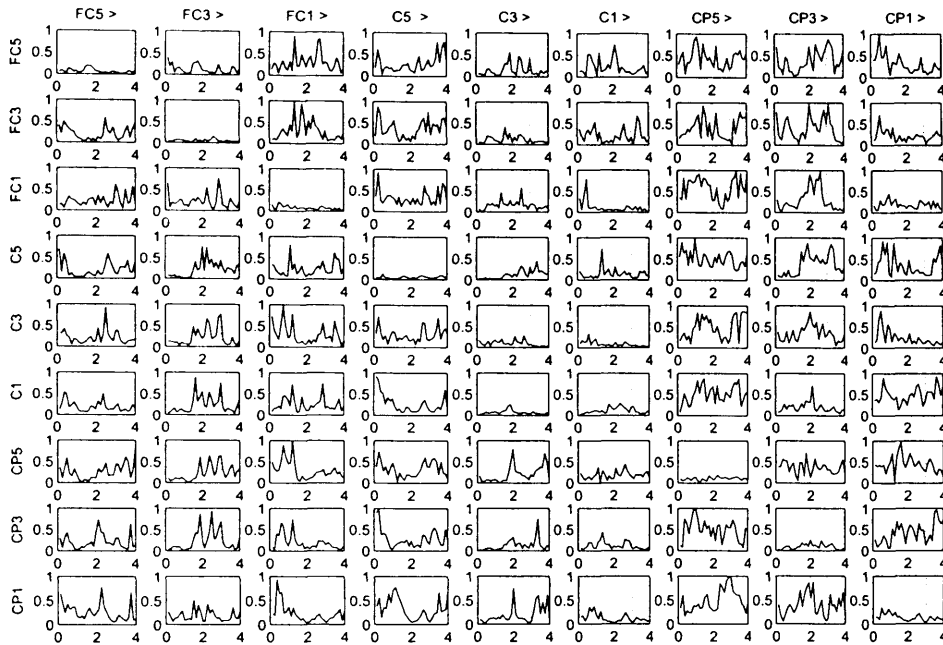


Figure 7.10: The SDTF calculated for electrodes located over the left hemisphere for right finger movement trials averaged over the alpha band. The axis are as in Fig. 7.10. For right finger movement the SDTF is lower in the left hemisphere compared with that of the right hemisphere (Fig. 7.11)

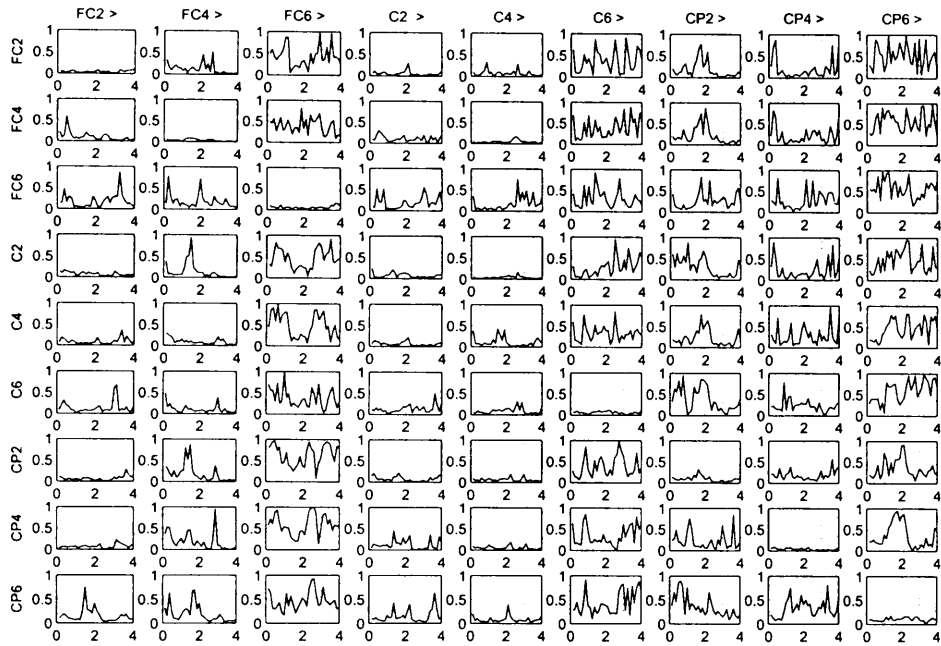


Figure 7.11: The SDTF calculated for electrodes located over the right hemisphere for right finger movement trials averaged over the alpha band. The axis are as in Fig. 7.8

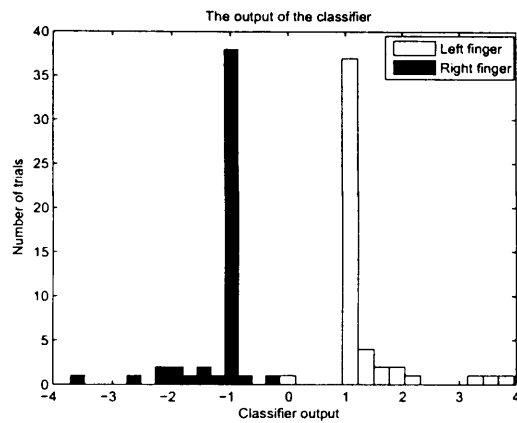


Figure 7.12: A histogram plot showing the distribution of classifier output values using the RBF kernel tested on 100 trials, 50 from each class. The outputs for each class are clearly separated, which generally indicates that features are significant for detection of left and right finger movements.

# Chapter 8

## Brain Computer Interfacing by Space-Time-Frequency Analysis

### 8.1 Introduction

This section demonstrates that in addition to the time and frequency information of the EEG signals, the spatial and directional information provide crucial indicators of intended left or right finger movement.

EEG is the result of a summation of a large number of synaptic potentials within the cortex [3]. These neurons tune into an oscillatory rhythm at a time interval, and location determined by the task at hand, e.g. finger movements causes a synchronisation of oscillatory rhythm in the beta band at approximately one second post movement located over the contralateral motor cortex. In visual processing tasks, a measurable response may be larger over the visual cortex. Therefore, the geometrical location provides important information about brain function. It is still debatable whether ERPs are a result of large amplitude bursts from a number of neurons, or the result of phase resetting in rhythmic activity

[110]. The latter view is generally supported by a growing number of clinicians and researchers.

Typically, in ERP analysis, time-frequency information is used to detect and extract relevant information. The geometric locations of the electrodes for the respective time frequency plots are also taken into account by the clinician, for example ERPs in the visual cortex are assumed to be related to visual processing. ERPs have the following properties:

1. ERPs are active within a certain region of the brain.
2. ERPs are active within a certain duration.
3. Typically active within a narrow frequency band.

These properties show that the geometric location of the ERP is an important factor in decomposing and understanding ERPs. Miwakeichi *et al.* [111] used Parallel Factor analysis (PARAFAC) to decompose the EEGs into space-time-frequency components. Traditional techniques such as PCA or ICA commonly analyse data in two dimensions<sup>1</sup>, space-time. Time-frequency representation over the space of electrodes effectively increases the dimensionality of data from two dimensions (space-time) to three dimensions (space-time-frequency) since the positions of the electrodes are taken into account. PARAFAC was used by Nazarpour *et al.* [112] to highlight the effect of finger movements on the EEGs by localising the alpha activity during finger movements.

Kaminski *et al.* in [104] demonstrated that there is a causal relationship between geometrically neighboring channels of the EEGs. Further works in [107] [113] and in the previous chapter showed that this can be used to distinguish between left and right finger movements from the EEG.

---

<sup>1</sup>The term ‘dimension’ is used in the context of PARAFAC to describe a modality, such as time, space, or frequency (i.e. not the number of recording channels).

## 8.2 Methods

In this section the EEGs are separated by assuming that the neuronal sources are disjoint in space, time, and frequency. A block diagram of the proposed system is shown in Fig. 8.1. In the first section the EEGs are converted into the time-frequency (TF) domain, then the TF representation of each electrode is arranged into a matrix where each element represents the x-y coordinates of the electrode. In the next block a space-time-frequency mask is created and the components within the mask are clustered. The cluster centres are one of the features used by the classifier. The other significant feature is the directionality of the moving reconstructed source signal, which is deduced from its cross correlation with the raw EEGs.

Section 8.2.1 explains the method for extracting the space-time-frequency distribution (STFD) from the EEGs. Then, Section 8.2.2 describes the clustering technique for extracting the atoms from the space-time-frequency distributions. Section 8.2.3 explains reconstruction of the signals from the clustered STFD. Section 8.2.4 describes the motion characterisation algorithm which forms one of the features used in the classification algorithm described in Section 5.2.2.

### 8.2.1 Space-Time-Frequency Analysis

The time-frequency distribution (TFD) of each electrode is constructed using short-term Fourier Transform (STFT) defined by

$$F_{s_j}(t, f) = \sum_{\tau} s_j(\tau)w(\tau - t)e^{-i\omega\tau} \quad (8.1)$$

where  $w(\cdot)$  is a window function and  $s_j(t)$  is the  $j^{th}$  electrode signal,  $i = \sqrt{-1}$ , and  $\omega = 2\pi f/f_s$ . The time-frequency plot for each electrode is arranged into a

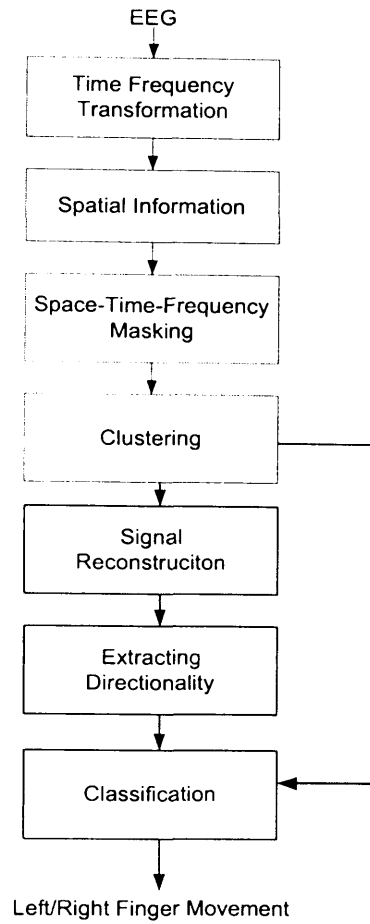


Figure 8.1: Block diagram of the space-time-frequency extraction algorithm.

four dimensional matrix such as

$$P(x, y, t, f) \triangleq F_{s_j} \quad j = 1, \dots, N \quad (8.2)$$

where  $x$  and  $y$  are the spatial coordinations of the electrodes, assuming that the scalp is represented by only two dimensions,  $t$  is the time index and  $f$  is the frequency index. The first two dimensions,  $x$  and  $y$  are sufficiently large so that all the electrodes can be arranged as they are defined by the 10-20 electrode placement system. For example,  $11 \times 11$  matrix is sufficient for a 64 electrode



EEG. The electrode Cz would be located at coordinates (6,6). The parameter  $N$  is the number of electrodes.

A space-time-frequency mask is constructed from the STFT plot based on the following criterion,

$$M(x, y, t, f) = \begin{cases} 1, & 20 \log(P(x, y, t, f)) > u \\ 0, & \text{otherwise} \end{cases} \quad (8.3)$$

where  $u$  is a threshold, which is empirically chosen to be  $0.25 \max(F_{s_j})$ .

## 8.2.2 Clustering

In order to extract the regions of activity from the STFT, the atoms in space-time-frequency must be identified and isolated from the background EEG. K-mean clustering algorithm [114] was used to identify and separate the active regions under the STFT mask  $M(x, y, t, f)$ .

The goal of k-means clustering algorithm is to find the cluster centres or means of a dataset given that the number of cluster centres is known *a priori*. The first step in clustering is to initialise the cluster centres randomly. Then cluster memberships are recomputed based the distance between each datapoint and the computed cluster centre as follows,

$$m_i = \min_k \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 \quad i = 1, \dots, n \quad (8.4)$$

where  $n$  is the number of datapoints,  $m_i \in [1, 2, \dots, k]$  is the membership vector. A datapoint is assigned a membership to the nearest  $k^{th}$  cluster centre. Then the cluster centres are recalculated for each cluster as,

$$\boldsymbol{\mu}_k = \frac{1}{k} \sum_{i \in m_k} \mathbf{x}_i \quad k = 1, \dots, \text{No. of clusters} \quad (8.5)$$

where  $k$  is the number of datapoints belonging to the  $k^{th}$  cluster. The process of updating (8.4) and (8.5) is repeated until the change in the value of  $\mu_k$  falls below a threshold.

Since, the number of clusters  $k$  is unknown, an estimate of the optimum number of clusters is required. The Gap statistic method proposed by [115] was used for this purpose. The within cluster compactness is given by

$$D_r = \sum_{i,j \in C_r} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (8.6)$$

where  $r \in \{1, \dots, K\}$ ,  $K$  is the number of clusters,  $C_r$  are the datapoints within cluster  $r$ , and

$$h_K = \sum_{r=1}^K \frac{1}{2N_r} D_r \quad (8.7)$$

is a compactness measure for  $K$  clusters and  $N_r$  is the number of datapoints within cluster  $C_r$ . Traditionally the optimal number of clusters is chosen by finding  $\max_{K_{opt}}(h_K - h_{K-1})$ , known as the L-Curve method. However the problem with this method is that the difference between  $(h_K - h_{K-1})$  is not normalised, which may give an incorrect estimate of the optimal number of clusters. The solution to this problem was proposed by the authors in [115] by comparing the clusters to a null reference dataset  $\mathbf{r}_b$   $\{b = 1, \dots, B\}$ , where  $B$  is the number of reference datasets. The number of clusters at which the clustered reference dataset is least similar to the clustered observations, is defined as the optimum  $K$ ,  $K_{opt}$ . The reference dataset is formed by scaling a uniformly distributed random dataset to the range of the principal components of the dataset. Then the reference dataset is clustered and  $h_{Kb}$  is evaluated, where  $b = 1, \dots, B$ . The Gap statistic is computed as

$$Gap(K) = \frac{1}{B} \sum_{b=1}^B \log h_{Kb} - \log h_K. \quad (8.8)$$

Next define

$$s_K = \sqrt{1 + \frac{1}{B}} \sigma_K \quad (8.9)$$

where  $\sigma_K$  is the standard deviation of  $\{\log h_{Kb}\}_{b=1,\dots,B}$ . The optimum number of clusters,  $K_{opt}$  is then defined as the smallest  $K$  which satisfies  $Gap(K) \geq Gap(K+1) - s_{K+1}$ . Essentially,  $K_{opt}$  is defined as the number of clusters which yields the largest difference between the clustered reference dataset and clustered observations.

### 8.2.3 Reconstruction

Each of the atoms are reconstructed by choosing the data points from the mask that belong to each cluster. Let  $M_c$  denote the mask with one cluster selected. The STFD for the cluster is given by

$$P_c = M_c \cdot P \quad c = 1, \dots, K_{opt} \quad (8.10)$$

where the space-time-frequency indices have been omitted,  $(\cdot)$  is the elementwise multiplication operator. Next the time series signal is reconstructed by computing the inverse short-time Fourier transform (ISTFT) of  $P_c(x, y, t, f)$  defined as

$$A_k(t) = \frac{1}{p} \sum_{l=1}^p \sum_{\omega} \sum_{\tau} P_c(x, y, t, \omega) w(\tau - t) e^{i\omega\tau}. \quad (8.11)$$

Where  $A_k(t)$   $k = 1, \dots, K_{opt}$  is the reconstructed atom and  $p$  is the number of electrodes that fall within spatial coordinates of atom  $k$ .

### 8.2.4 Motion Characterisation

In this section motion of the sources are quantified in order to determine whether there is left or right finger movement. The extracted atom's,  $A_k(t)$ , cross correlation with the raw EEGs is used within an overlapping window of length  $L$ ,

and with an overlap  $O$ . The absolute maximum value for each window of cross correlation is used as the location of the atom, given by

$$\rho_k(t) = \max_j (|E\{A_k(t)s_j(t)\}|) \quad (8.12)$$

and the location (coordinates) is deduced by the index  $j$ . For example, if the largest cross correlation between EEGs and extracted atom were electrode  $C_z$  at  $t = 1$  then  $C_3$  at  $t = 2$ , the transition would be from coordinates  $\rho_k(1) = \{6, 6\}$  to  $\rho_k(2) = \{4, 6\}$ . Since the atom is disjoint in time, space, and frequency, there should be only one peak in the cross correlation function for each window. Finally the average direction is given by

$$\mathbf{d}_k = \frac{1}{T} \sum_t \rho_k(t) \quad (8.13)$$

where  $\mathbf{d}_k$  is the direction for atom  $k$ , and  $T$  is the number of cross correlation windows.

## 8.3 Experiments

### 8.3.1 Testing the Algorithm

The dataset described in Section 7.3.1 was used to test the algorithm. Before applying the STF algorithm the EEGs were corrected for eye blinking artifacts as described in Section 5. In this study the features are tested using 180 trials in total; 90 for left finger movement and 90 for right finger movement. A feature set was created as  $\mathbf{g}_i = [m_{x_1}(i), m_{y_1}(i), d_{x_1}(i), d_{y_1}(i), \dots, m_{x_{k_{opt}}}(i), m_{y_{k_{opt}}}(i), d_{x_{k_{opt}}}(i), d_{y_{k_{opt}}}(i)]^T$ , where  $m_{x_k}(i)$  and  $m_{y_k}(i)$  are the  $x$  and  $y$  components of the cluster centres,  $d_{x_k}(i)$  and  $d_{y_k}(i)$  are the  $x$  and  $y$  components of the directional vector. For classification the SVM is used since it suited the highly nonlinear nature of the feature set.

In order to test the overall classification rate 4-fold cross-validation (CV) was used with no overlap, i.e. using 75% of the data for training and 25% for testing. The CV was performed 10 times, with each time the data was chosen at random from the trial pool. The classifier was used with three kernels, linear, RBF, and cubic polynomial, for which the error is shown in Table 8.1. For the test dataset the value chosen for the parameter  $C$  was 64 and for the case of the RBF kernel the parameter  $\rho$  was set to 0.5. The parameter  $B$  was set to 20 reference datasets such that the standard deviation parameter (8.9) can be accurately estimated. The maximum number of clusters,  $K$ , was set to 6, since the number of active sources at time of finger movement is generally low. This was demonstrated in [112]. A Hanning window function was used for the STFT algorithm. The window length,  $L$ , for the motion characterisation algorithm was set to 1000 samples, and the overlap,  $O$ , was 900 samples.

The cluster centres in the spatial domain are shown in Fig. 8.2. From the figure it can be seen that the cluster centres for the left finger movement occur on the contralateral hemisphere very close to  $C_2$  electrode location, which is located over the motor cortex. For right finger movements the location of the cluster centre is at electrode  $C_3$ , which is associated with right finger movements as explained in [95]. Figure 8.3 shows the time-frequency representation for the atoms of the left finger movement trial. The alpha band is the most dominant feature in the time-frequency domain for movement related tasks. This is highlighted by the results in Fig. 8.3 which shows alpha band activity is present before finger movement, then desynchronisation in the motor cortex causes the alpha band power to drop below the threshold,  $u$ . Post movement, the alpha band power returns to normal. The desynchronisation separates the clusters, i.e. before movement and post movement, and therefore two clusters are generally identified.

The average number of support vectors calculated when using the RBF kernel

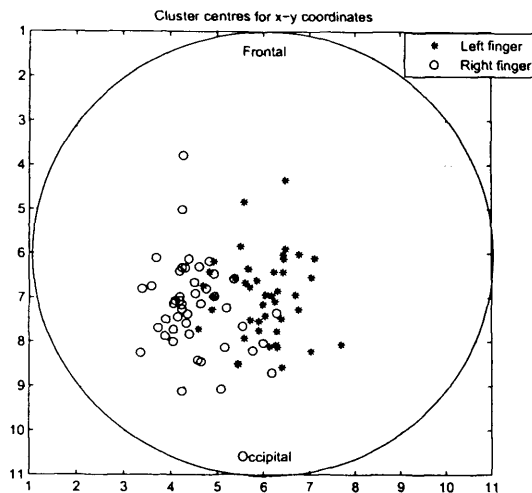


Figure 8.2: The cluster centres for the extracted atoms for 45 left and 45 right finger movement trials.

was 35.5% of the training examples. When using the linear kernel the average number of SVs found was 10.4% and for cubic polynomial it was 22.5%. The training error was found by using the training data as test data. The training error was found to be 7% (ave) and the test error was 9% (ave). Since the difference between two error values is small, this indicates that overfitting has been avoided.

## 8.4 Conclusions

A new method was presented based on space-time-frequency disjointness of the EEG sources for distinguishing between left and right finger movements from scalp EEGs using the features corresponding to the activity of alpha rhythms and directionality of the sources. The experiments herein demonstrated that for the test dataset the signals are correctly classified by using the introduced

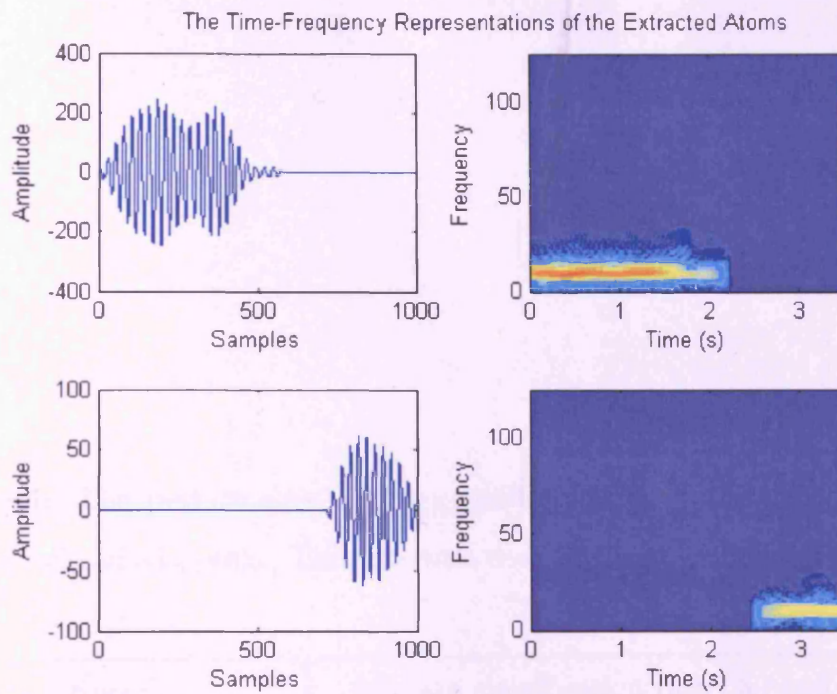


Figure 8.3: The time-frequency representation of the extracted atoms for a left finger trial.

features. Using k-mean clustering followed by the Gap statistic method enables accurate estimation of the number of disjoint factors, representing the brain's active sources. A higher classification rate is achieved when the RBF kernel is used for the SVM.

Table 8.1: The performance of the classifier based on the average number of correctly classified points. Three kernels used in the classification are compared.

Kernel	Average classification rate (%) (s.d.)		
	Overall	Right	Left
Gaus. RBF	75.50 (1.0)	74.16 (1.2)	72.43 (1.5)
Cubic Poly.	63.30 (1.4)	63.35 (1.0)	63.36 (1.0)
Linear	57.01 (1.3)	56.34 (1.4)	55.51 (1.0)



# Chapter 9

## Conclusions and Future Work

### 9.1 Conclusions

Over the last three decades BCI research has received increasing attention. As a result, BCI is becoming realisable because of the advances in signal processing techniques. In the past, communication between the neuroscience community and the signal processing community was very limited. However, with researchers from both areas realising the mutual requirement for each other, the two disciplines are, gradually, merging together. This means that the possibilities are endless in terms of diagnosing neuronal abnormalities and of course development of BCIs.

The main problem in BCI research is the destructive effects of artifact signals masking the signals of interest. One of the requirements for a BCI was to remove artifacts from the EEGs. The most common artifact signal in BCI, and for almost all clinical applications of EEG, is eye blinking artifact. Experimental results show that the majority of recording electrodes are affected during eye blinks.

Two methods have been developed to mitigate the effect of eye blinking arti-

fact. The first method described in Chapter 4 is based on extending the gradient implementation of the SOBI BSS algorithm [4] such that the artifact signal is mitigated from the recovered independent components. The assumptions of linear generation of EEGs and statistically independent neuronal sources are generally accepted by the majority of researchers in EEG analysis [46][47][69].

The proposed constrained BSS algorithm uses second order statistics only and therefore processing speeds are much faster than those based on higher order statistics [67]. Another advantage of using a second order statistic approach is that, in contrast to higher order statistics based BSS, it uses estimates of the covariance matrices at multiple time lags, hence is insensitive to certain noisy data or individual points out of the range of normal EEGs. The cost function was extended so that the calculated independent components were constrained in order not to be correlated with a reference signal. The validity of the constrained ICs was shown in Section 4.3.2 by reprojecting the ICs onto the scalp electrodes and comparing with those obtained from an unconstrained SOBI algorithm. The quality of the ICs was also subjectively quantified by visually inspecting the ICs and making comparisons with those from the unconstrained SOBI algorithm [4]. Furthermore, a visual comparison of the output of the CBSS algorithm with that of PCA [116] shows the improvement in separation performance of the CBSS algorithm over PCA. This algorithm was used in the preprocessing stage for detecting the predictability of epilepsy [90].

The second approach for removal of eye blinking artifact was based on combination of the SOBI algorithm with automatic classification of the ICs resulting from artifacts. The aim of the hybrid BSS-SVM method, described in Chapter 5, was to mitigate the eye blinking artifacts while preserving the EEGs. It was shown that the proposed features effectively characterised the ICs related to the eye blinking artifacts. The hybrid BSS-SVM algorithm is useful when long EEG

recordings are contaminated by eye blinking artifacts. The clinician would have to scroll through an entire EEG recording and manually remove the artifacts. Noting that EEG recordings can be on the order of hours makes this a tedious and time consuming process. In comparison with the manual artifact rejection algorithms [43] the proposed automated method can remove eye blinking artifacts from EEGs in a fraction of the time that manual techniques can. The hybrid BSS-SVM algorithm processes consecutive non-overlapping blocks of EEGs in turn, and since the ICs are reprojected to the electrodes, it is immune from any scale and permutation ambiguities of the BSS. Objective and subjective results in Section 5.3 show that the hybrid BSS-SVM algorithm effectively and consistently removes eye blinking artifacts from the EEGs. This method may be extended to remove other artifacts, such as ECG. Furthermore, unlike reconstructive artifact removal algorithms such as that proposed by Joyce *et al.* [48] the BSS-SVM algorithm does not require a reference channel to be recorded simultaneously with the EEGs.

Three BCI systems were proposed. They are based on the hypothesis that the EEG sources related to movement are geometrically nonstationary, and on the basis that the thalamus and various parts of the cortex communicate during finger movement. The evidence presented by Ginter *et al.* [107] suggests that there is a movement in the location of the source at the time of finger movement. The first system proposed in Chapter 6 was based on separating and localising consecutive overlapping blocks of EEGs such that any change in the location between consecutive windows would be interpreted as a moving source. From this system it was found that an accurate model was required for the conductivities of layers within the brain. This was demonstrated in Section 6.2.2 by a synthetic dataset in which the mixing medium was homogenous and signal sources were isotropic. On the contrary, when processing real EEGs the algorithm was unsuccessful in

localising the sources. This may be due to the assumption of homogeneity in the conducting medium. Furthermore, the conducting medium also changes with time as cerebral blood flow changes. This further degrades the accuracy of the estimated location of the source. Therefore an accurate model of the conductivities of the layers within the brain is required in order for this algorithm to be successful. This experiment gave an insight into localisation algorithms and the problems faced by this wide area of research.

The second BCI system described in Chapter 7 was based on extracting features from the causal relationship between electrodes and the projection strengths of the IC related to finger movements. The causal relationship feature gave highlighted the movement of a sources within the brain. It showed that the movement of the sources were generally from the fronto central to parietal electrodes as described by Ginter *et al.* [107]. Additionally, the location of the largest projection of the beta band activity highlighted the ERS observable within one second after finger movement. This characteristic of the time course of EEGs after finger movement was defined by Pfurtscheller *et al.* [20] and has been used to increase the classification performance of the proposed BCI. The EEGs were bandpass filtered within the beta band and separated by SOBI. Filtering the EEG signals effectively reduced the number of redundant active sources. The results of this experiment showed that the location of beta activity and propagation of EEG activity could be used to distinguish between left and right finger movements. Tang *et al.* [69] demonstrated that SOBI was a effective in separating the source of left and right finger peripheral nerve stimulation from the background EEGs. The results obtained by Tang *et al.* [69] corroborate the results obtained in Chapter 7 because the active brain regions are located on the contralateral hemisphere to side of finger movement.

The third experiment proposed in Chapter 8 was based on the assumption that

the EEGs are disjoint in space-time-frequency. This means that only one source is active within a given duration, within one region of the brain, and within a particular frequency band. The disjoint sources are known as atoms. The atoms were extracted by applying a mask to the space-time-frequency signatures. A desirable feature of this algorithm is that, unlike ICA, the sources can be extracted without assuming that they are statistically independent of each other. Therefore, fewer assumptions are made on the mixing model thereby allowing the sources to be interdependent which was evident from SDTF in Chapter 7. When combined with a clustering algorithm, the location of atoms can be deduced. The motion of the atom was quantified by cross correlation with the EEGs. Qin *et al.* [58] extracted the disjoint sources from just two electrodes located over the motor cortex. In the proposed algorithm the disjoint sources were extracted from 64 electrodes, yielding a more reliable estimate of the location of the atom. The experiments with EEGs during left and right finger movements show that the geometric location of the sources are localised to the contralateral hemisphere agrees with the second proposed BCI system proposed in Chapter 7.

## 9.2 Future work

Within the field of BCI there are endless possibilities for improvements and developments. The majority of current BCI systems operate on blocks of data (or per trial basis), meaning that the minimum time interval between user input commands is constrained by the time taken to sample and process one trial. Therefore, a BCI system that processes EEGs in real time would, in general, yield a higher command throughput. This in turn means that for the BCI systems based on BSS, online algorithms should be further investigated. Increasing the sampling rate by four or even eight times the typical 250Hz enables the extraction of the

dynamics of the finger movement from the EEGs. This may be considered in development of a convolutive BSS for separating the movement related sources.

In terms of the localisation approach in Chapter 6 a detailed model of the conductivities would be essential for the algorithm to perform to a satisfactory level. One solution may be to use the location of a known source as a reference point. The reference source will project to the electrodes and superimpose itself over the normal EEGs. The conduction strengths will be apparent from the amplitude of reference signal at each of the electrodes. The amplitude of the reference source at each electrode can then be used as *a priori* knowledge when localising the finger movement related EEGs. The reference signal may induced externally via an electrodes placed in close proximity with the centre of the brain.

The solution space of the estimated source location can be constrained by using *a priori* knowledge of the approximate location of the source. The propagation feature proposed in Chapter 7 gives a rough estimate of the source location and it's directionality. When localising successive overlapping windows of EEGs the SDTF may be calculated. It would be expected that the source would be travelling in the same direction defined by the SDTF estimates. Calculation of the location would then be a feedback system where the location is adjusted based on the SDTF estimates.

The system proposed in Chapter 8 was based on disjointness of the EEG sources in space-time-frequency. The space of electrodes were defined by a coarse  $11 \times 11$  matrix which was used to represent 64 electrodes according to the 10-20 system for electrode placement. One may investigate the effect of increasing the density of electrodes, this would presumably increase the accuracy of the geometric location of the atoms over the scalp. For the application of finger movement one may only need to increase the density of electrodes over the motor cortex area in order to gain an improvement. This is because the results in Fig.

8.2 show that the cluster centres are located around the motor cortex. One may also further investigate the effect of using other time-frequency transforms such as wavelet transform.

In terms of artifact rejection a more general algorithm for removal of multiple artifacts such as EOGs, EMGs, ECGs, and other artifacts, may be proposed. For example, to identify the ICs containing ECG with the BSS-SVM algorithm in Chapter 5, the ECG is almost periodic in nature. A template of one ECG pulse can be used in order to identify the corresponding ICs by finding the template ECG's cross correlation with the ICs. It would be expected that the IC containing ECG would be more correlated than ICs related to brain activity. EMG artifacts may be removed in a similar fashion, i.e. by finding physiological descriptors of the EMG. EMG typically has a broad frequency spectrum ( $20 - 400Hz$ ) that overlaps with normal EEGs. Therefore, the ratio between the EMG's lower frequency band (for example  $20 - 60Hz$ ) and higher frequency band ( $60 - 100Hz$ ) could be used to describe an IC containing EMG. It would be expected that the ratio between low and high frequency bands, denoted as  $\chi$ , would be  $\chi \approx 1$  for ICs containing EMG and  $\chi > 1$  for ICs unrelated to EMGs.

It was demonstrated that the constrained BSS algorithm in Chapter 4 could be extended to removal of ECG artifacts simply by using the ECG as a reference signal. It may be possible to extend the current constrained BSS algorithm so that it can accommodate multiple constraints related to the above mentioned artifacts simultaneously.

# Bibliography

- [1] D. Haines, *Fundamental Neuroscience, 1st Ed.* England, Churchill Livingstone, 1997.
- [2] A. B. McNaught and R. Callander, *Illustrated Physiology, 3rd Ed.* England, Churchill Livingstone, 1975.
- [3] E. Niedermeyer and F. L. D. Silva, *Electroencephalography, Basic Principles, Clinical Applications and Related Fields, 4th Ed.* Philadelphia, Williams & Wilkins, 1999.
- [4] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal processing," *Proc. IEEE Sensor Array And Multichannel signal processing workshop, Rosslyn VA, USA*, pp. 273–277, 2002.
- [5] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, and T. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2001.
- [6] J. J. Vidal, "Toward direct brain-computer communication," *Annual Review Biophysics*, vol. 2, pp. 157–180, 1973.



- [7] D. Mcfarland, W. Sarnacki, and J. Wolpaw, "Brain-computer interface (BCI) operation: optimizing information transfer rates," *Biological Psychology*, vol. 63, pp. 237–251, 2003.
- [8] N. Neumann and N. Birbaumer, "Predictors of successful self control during brain-computer communication," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 74, pp. 1117–1121, 2003.
- [9] J. Bayliss and D. Ballard, "A virtual reality testbed for brain-computer interface research," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 188–90, 2000.
- [10] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller, and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study," *Neuroimage*, vol. 10, pp. 658–665, 1999.
- [11] J. Nicholls, A. Martin, and B. Wallace, *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System, 3rd Ed.* Sunderland, MA, Sinauer Associates Inc., 1992.
- [12] B. Kolb and I. Whishaw, *Fundamentals of Human Neuropsychology, 5th Ed.* New York, W.H.Freeman, 2003.
- [13] L. Stewart, R. Henson, K. Kampe, V. Walsh, R. Turner, and U. Fritha, "Brain changes after learning to read and play music," *Neuroimage*, vol. 20, pp. 71–83, 2003.
- [14] H. Jasper, "Reports of committee on methods of clinical exam in EEG," *Electroencephalography & Clinical Neurophysiology*, vol. 10, pp. 370–375, 1958.

- [15] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, pp. 9–21, 2004.
- [16] K. A. Young, M. Smith, T. Rawls, D. B. Elliott, S. I. Russell, and P. B. Hicks, "N100 evoked potential latency variation and startle in schizophrenia," *Neuroreport*, vol. 12, no. 4, pp. 767–773, 2001.
- [17] S. Sutton and J. Z. M. Braren, "Evoked potential correlates of stimulus uncertainty," *Science*, vol. 150, pp. 1187–1188, 1965.
- [18] L. A. Farwell and E. Donchin, "Talking off the top of your head: towards a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, pp. 510–523, 1988.
- [19] C. H. M. Brunia and G. J. M. van Boxtel, "Wait and see," *International Journal of Psychophysiology*, vol. 43, pp. 59–75, 2001.
- [20] G. Pfurtscheller and A. Berghold, "Patterns of cortical activation during planning of voluntary movement," *Electroencephalography and Clinical Neurophysiology*, vol. 72, pp. 250–258, 1989.
- [21] C. Brunia and W. V. den Bosch, "Movement-related slow potentials I. a contrast between finger and foot movements in right handed subjects," *Electroencephalography and Clinical Neurophysiology*, vol. 57, pp. 515–527, 1984.
- [22] N. Birbaumer, A. Kbler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The thought translation device (TTD) for completely paralyzed patients,"

- IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 190–193, 2000.
- [23] U. Kraft, “Train your brain,” *Scientific American: Mind*, vol. 17, no. 1, pp. 58–63, 2006.
- [24] B. Graimann, J. Higgins, S. Levine, and G. Pfurtscheller, “Visualisation of significant ERD/ERS patterns in multichannel EEG and ECoG data,” *Clinical Neurophysiology*, vol. 113, pp. 43–47, 2002.
- [25] G. Pfurtscheller and A. Aranibar, “Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement,” *Electroencephalography and Clinical Neurophysiology*, vol. 46, no. 2, pp. 138–146, 1979.
- [26] G. Pfurtscheller and F. L. da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clinical Neurophysiology*, vol. 110, pp. 1842–1857, 1999.
- [27] G. Phurtscheller and A. Aranibar, “Event related cortical desynchronization detected by power measurements of the scalp EEG,” *Electroencephalography and Clinical Neurophysiology*, vol. 42, no. 6, pp. 817–826, 1977.
- [28] A. Burges and J. Gruzelier, “The reliability of event related desynchronisation: a generalisability study analysis,” *International Journal Psychophysiology*, vol. 23, pp. 163–169, 1996.
- [29] C. Guger, H. Ramoser, and G. Pfurtscheller, “Real-time EEG analysis with subject specific spatial patterns for a brain-computer interface (BCI),” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 447–456, 2000.

- [30] D. A. Overton and C. Shagass, "Distribution of eye movement and eyeblink potentials over the scalp," *Electroencephalography and Clinical Neurophysiology*, vol. 27, p. 546, 1969.
- [31] T. Elbert, W. Lutzenberger, B. Rockstroh, and N. Birbaumer, "Removal of ocular artifacts from the EEG - a biophysical approach to the EOG," *Electroencephalography and Clinical Neurophysiology*, vol. 60, pp. 455–463, 1985.
- [32] A. Schlogl and G. Pfurtscheller, "EOG and ECG minimization based on regression analysis," Institute for Biomedical Engineering, Tech. Rep., 1999.
- [33] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electroencephalogram by adaptive filtering," *Medical & Biological Engineering and Computing*, vol. 42, pp. 407–412, 2004.
- [34] P. Celka, B. Boashash, and P. Colditz, "Preprocessing and time-frequency analysis of newborn EEG seizures," *IEEE Engineering Medicine and Biology Magazine*, vol. 20, pp. 30–39, 2001.
- [35] P. Bewrg and M. Scherg, "Dipole modeling of eye activity and its application to the removal of eye artifacts from EEG and MEG," *Clinical Physiology and Physiological Measurements*, vol. 12(A), pp. 49–54, 1994.
- [36] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.
- [37] A. Hyvärinen, E. Oja, and J. Karhunen, *Independent Component Analysis, 1st Ed.* New York, Thomson Learning, Inc, 2001.

- [38] N. Kawabata, "Nonstationary analysis of the electroencephalogram," *IEEE Transactions On Biomedical Engineering*, vol. 20, pp. 444–452, 1973.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, 2nd Ed.* New York, John Wiley and Sons, Inc., 2001.
- [40] P. Comon, "Independent component analysis - a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.
- [41] O. G. Lins, T. W. Picton, P. Berg, and M. Scherg, "Ocular artifacts in EEG and event-related potentials, i: Scalp topography," *Brain Topography*, vol. 6, pp. 51–63, 1993.
- [42] T. Jung, S. Makeig, M. Mckeown, A. Bell, T. Lee, and T. J. Sejnowski, "Imaging brain dynamics using independent component analysis," *IEEE Procs.*, vol. 89, no. 7, pp. 1107–1122, July 2001.
- [43] T. P. Jung, C. Humphries, T. Lee, S. Makeig, M. McKeown, V. Iragui, and T. Sejnowski, "Extended ICA removes artifacts from electroencephalographic recordings," *Advances in Neural Information Processing Systems*, vol. 10, pp. 894–900, 1998.
- [44] S. Amari, T. P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.
- [45] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [46] A. Tang, M. Sutherland, and C. McKinney, "Validation of SOBI components from high-density EEG," *Neuroimage*, vol. 25, pp. 539–553, 2004.

- [47] A. C. Tang, B. A. Pearlmutter, N. A. Malaszenko, D. B. Phung, and B. C. Reeb, "Independent components of magnetoencephalography," *Neural Computing*, vol. 14, pp. 1827–1858, 2002.
- [48] C. A. Joyce, I. Gorodnitsky, and M. Kutas, "Automatic removal of eye movement and blink artifacts from EEG data using blind component separation," *Psychophysiology*, vol. 41, pp. 313–325, 2004.
- [49] L. Tong, V. Soon, R. Liu, and Y. Huang, "AMUSE: a new blind identification algorithm," in *Proc. ISCAS*, New Orleans, 1990.
- [50] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.
- [51] S. Romero, M. Mailanas, S. Clos, S. Gimenez, and M. Barbanoj, "Reduction of EEG artifacts by ICA in different sleep stages," *Proceedings of the 25<sup>th</sup> Annual International Conference of the IEEE EMBS Cancun, Mexico*, 2003.
- [52] H. Akaike, "A new look at statistical model order identification," *IEEE Trans Auto Cont*, vol. 19, pp. 716–723, 1974.
- [53] E. Curran, P. Sykacek, M. Stokes, S. J. Roberts, W. D. Penny, I. Johnsrude, and A. D. Owen, "Cognitive tasks for driving a brain computer interfacing system: A pilot study," *IEEE. Trans. Neural Systems and Rehabilitation Engineering*, vol. 12, no. 1, pp. 48–54, 2003.
- [54] S. Roberts, W. Penny, and I. Rezek, "Temporal and spatial complexity measures for EEG-based brain-computer interfacing," *Medical and Biological Engineering and Computing*, vol. 37, no. 1, pp. 93–99, 1998.

- [55] F. Takens, “Detecting strange attractors in turbulence,” in *Lectures Notes in Mathematics, Dynamical Systems and Turbulence, Warwick 1980*, D. A. Rand and L.-S. Young, Eds. Heidelberg, Berlin: Springer-Verlag, 1981, pp. 366–381.
- [56] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing, 1st Ed.* New-Jersey, Prentice-Hall, 1989.
- [57] B. He and T. Musha, “Equivalent dipole localization of spontaneous EEG alpha activity: Two moving dipole approach,” *Medical and Biological Engineering and Computing*, vol. 30, pp. 324–332, 1992.
- [58] L. Qin, B. Kamousi, Z. Liu, L. Ding, and B. He, “Classification of motor imagery tasks by means of time-frequency-spatial analysis for brain-computer interface applications,” *Proceedings of the 2nd International IEEE EMBS Conference on Neural Engineering, Arlington, Virginia, 2005*.
- [59] Özgür Yılmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1846, 2004.
- [60] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Transactions On Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [61] G. Dornhege, B. Blankertz, G. Curio, and K. Müller, “Combining features for BCI,” *Procs. Advances in Neural Information Processing Systems (NIPS 02)*, vol. 15, pp. 1115–1122, 2003.

- [62] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial eeg classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [63] G. Dornhege, B. Blankertz, G. Curio, and K. Müller, "Increase information transfer rates in BCI by CSP extension to multi-class," *presented at the Advances in Neural Information Processing Systems (NIPS 03) Vancouver, BC, Canada*, 2004.
- [64] 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 Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 2, pp. 166–171, 2005.
- [65] H. Serby, E. Yom-Tov, and G. F. Inbar, "An improved p300-based brain-computer interface," *IEEE Transactions on Neural Systems And Rehabilitation Engineering*, vol. 13, no. 1, pp. 89–98, 2005.
- [66] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [67] J. F. Cardoso, "Higher order contrasts for independent component analysis," *Neural Computing*, vol. 11, pp. 157–192, 1999.
- [68] Y. Wang, M. T. Sutherland, L. L. Sanfratello, and A. C. Tang, "Single-trial classification of erps using second-order blind identification (SOBI)," *Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai*, 2004.



- [69] A. C. Tang, M. T. Sutherland, and Y. Wang, "Contrasting single-trial ERPs between experimental manipulations: improving differentiability by blind source separation," *Neuroimage*, vol. 29, pp. 335–346, 2006.
- [70] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions on Speech Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.
- [71] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming Theory and Algorithms, 2nd Ed.* New York, Wiley, 1993.
- [72] W. Wang, S. Sanei, and J. A. Chambers, "Penalty function-based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1654–1669, 2005.
- [73] S. Haykin, *Adaptive Filter Theory, 4th Ed.* New Jersey, Prentice-Hall, 2002.
- [74] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications.* New York, John Wiley and Sons Ltd, 2002.
- [75] G. S. Wagner and H. J. L. Marriott, *Marriott's Practical Electrocardiography, 10th Ed.* Hagerstown MD, Lippincott Williams & Wilkins, 2001.
- [76] M. Joho, H. Mathis, and R. H. Lambert, "Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture," *Proc. International Conference on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, June 19–22*, pp. 81–86, 2000.

- [77] P. Smaragdis, "Information theoretic approaches to source separation," *Masters thesis, MIT Media Lab*, June 1997.
- [78] V. Schetinin and J. Schult, "The combined technique for detection of artifacts in clinical electroencephalograms of sleeping newborns," 2005. [Online]. Available: URL: <http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:cs/0504070> [Accessed: 7th June 2006]
- [79] Z. A. Keirn and J. I. Aunon, "Man-machine communications through brain-wave processing," *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, no. 1, pp. 55–57, 1990.
- [80] P. Sehatpour, S. Molholm, D. C. Javitt, and J. J. Foxe, "Spatiotemporal dynamics of human object recognition processing: An integrated high-density electrical mapping and functional imaging study of closure processes," *NeuroImage*, vol. 29, no. 2, pp. 605–618, 2006.
- [81] S. Gunn, "Support vector machines for classification and regression," Univ. of Southampton, URL: <http://www.ecs.soton.ac.uk/srg/publications/> [Accessed: 7th June 2006], Tech. Rep., 1998.
- [82] K. P. Bennett and A. Demiriz, "Semi-supervised support vector machines," *Proc. Neural Information Processing Systems, Denver*, 1998.
- [83] S. Sanei, "Texture segmentation using semi-supervised support vector machines," *IEEE Transactions on Neural Networks*, vol. 4, no. 2, pp. 131–142, 2004.
- [84] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 1–47, 1998.

- [85] N. Christianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines, 1st Ed.* New York, Cambridge University Press, 2000.
- [86] V. Vapnik, *The nature of statistical learning theory, 2nd Ed.* Berlin, Springer-Verlag, 1995.
- [87] Centre of DSP, 2006, Cardiff University website, URL: <http://www.engin.cf.ac.uk/research/group.asp?GroupNo=43> [Accessed: 15th May 2006].
- [88] S. Gunn, University of Southampton, URL: <http://www.isis.ecs.soton.ac.uk/resources/svminfo/> [Accessed: June 7th 2006], 1998.
- [89] S. Baillet, R. M. Leahy, M. Singh, D. W. Shattuck, and J. C. Mosher, "Supplementary motor area activation preceding voluntary finger movements as evidenced by magnetoencephalography and fMRI," *International Journal of Bioelectromagnetism*, vol. 3, no. 1, 2001.
- [90] J. Corsini, L. Shoker, S. Sanei, and G. Alarcon, "Epileptic seizure predictability from scalp EEG incorporating constrained blind source separation," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 790 – 799, 2006.
- [91] S. Sanei, M. Latif, and L. Shoker, "Detection and localization of epileptic seizures using ICA," *3rd IEEE Sensor Array And Multichannel signal processing workshop, Barcelona Spain*, 2004.
- [92] N. Lynnerup, J. G. Astrup, and B. Sejrsen, "Thickness of the human cranial dipole in relation to age, sex and general body build," *Head Face Medicine*, vol. 1, no. 13, pp. 1-7, 2005.

- [93] I. Coope, “Reliable computation of the points of intersection of  $n$  spheres in  $\mathbb{R}^n$ ,” *ANZIAM J.*, vol. 42, no. E, pp. C461–C477, 2000.
- [94] S. Ohara, A. Ikeda, T. Kunieda, S. Yazawa, K. Baba, T. Nagamine, W. Taki, N. Hashimoto, T. Mihara, and H. Shibasaki, “Movement-related change of electrocortigraphic activity in human supplementary motor area proper,” *Brain*, vol. 123, pp. 1203–1215, 2000.
- [95] G. Pfurtscheller and C. Neuper, “Synchronization of mu rhythm in the EEG over the cortical hand area in men,” *Neuroscience Letters*, vol. 174, pp. 93–96, 1994.
- [96] A. Delorme and S. Makeig, “EEG changes accompanying learned regulation of 12-hz EEG activity,” *IEEE Transactions On Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 133–136, 2003.
- [97] E. Rodriguez, N. George, J. Lachaux, J. Martinerie, B. Renault, and F. Varela, “Perception’s shadow: Long distance synchronization of human brain activity,” *Nature*, vol. 397, pp. 430–433, 1999.
- [98] G. Gerloff, J. Richard, J. Hadley, A. Schulman, M. Honda, and M. Hallett, “Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements,” *Brain*, vol. 121, pp. 1513–1531, 1998.
- [99] E. Gysels and P. Celka, “Phase synchronization for the recognition of mental tasks in a brain-computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 4, pp. 406–415, 2004.

- [100] A. Sharott, P. J. Magill, J. P. Bolam, and P. Brown, “Direcional analysis of coherent ocillatory field potentials in cerebral cortex and basal ganglia of the rat,” *Journal of Physiology*, vol. 562, no. 3, pp. 951–963, 2005.
- [101] C. W. J. Granger, “Investigating causal relations in econometric models and cross-spectral methods,” *Econometrica*, vol. 37, pp. 424–438, 1969.
- [102] C. Bernasconi and P. König, “On the directionality of cortical interactions studied by spectral analysis of electrophysiological recordings,” *Biological Cybernetics*, vol. 81, no. 3, pp. 199–210, 1999.
- [103] M. Kaminski and K. Blinowska, “A new method of the description of information flow in the brain structures,” *Biological Cybernetics*, vol. 65, pp. 203–210, 1991.
- [104] M. Kaminski, M. Ding, W. Truccolo, and S. Bressler, “Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance,” *Biological Cybernetics*, vol. 85, pp. 145–157, 2001.
- [105] R. Kuś, M. Kamiński, and K. J. Blinowska, “Determination of EEG activity propogation: Pair-wise versus multichannel estimate,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 9, pp. 1501–1510, 2004.
- [106] H. Jing and M. Takigawa, “Observtion of EEG coherence after repetitive transcranial magnetic stimulation,” *Clinical Neurophysiology*, vol. 111, pp. 1620–1631, 2000.
- [107] J. Ginter-Jr., K. Blinowska, M. Kaminski, and P. Durka, “Phase and amplitude analysis in time-frequency space - application to voluntary finger movement,” *Journal of Neuroscience Methods*, vol. 110, pp. 113–124, 2001.

- [108] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," *Biological Cybernetics*, vol. 83, pp. 35–45, 2000.
- [109] M. Morf, A. Vieira, D. Lee, and T. Kailath, "Recursive multichannel maximum entropy spectral estimation," *IEEE Transactions on Geoscience Electronics*, vol. 16, pp. 85–94, 1978.
- [110] S. Makeig, M. Westerfield, T. Jung, S. Enghoff, J. Townsend, E. Courchesne, and J. Sejnowski, "Dynamic brain sources of visual evoked responses," *IEEE Transactions on Neural Networks*, vol. 295, pp. 690–694, 2002.
- [111] F. Miwakeichi, E. Martínez-Montez, P. Valdés-Sosa, N. Nishiyama, H. Muzuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *Neuroimage*, vol. 22, pp. 1035–1045, 2004.
- [112] K. Nazarpour, S. Sanei, L. Shoker, and J. A. Chambers, "Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing," *Procs. European Signal Processing Conference (EUSIPCO), Florence, Italy*, 2006.
- [113] L. Shoker, S. Sanei, and A. Sumich, "Distinguishing between left and right finger movement from EEG using SVM," *Procs. IEEE Engineering in Medicine and Biology Society, Shanghai, China*, 2005.
- [114] A. Jain and K. Karu, "Learning texture discrimination masks," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 18, no. 2, pp. 195–205, 1996.

- [115] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a dataset via the gap statistic,” *Journal of the Royal Statistical Society, B*, vol. 63, pp. 411–423, 2001.
- [116] T. Jung, C. Humphries, M. Lee, V. Iragui, S. Makeig, and T. Sejnowski, “Removing electroencephalographic artifacts: Comparison between ICA and PCA,” *IEEE International Workshop on Neural Networks for Signal Processing*, pp. 63–72, 1998.

