UNVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

Institute of Sound and Vibration Research

Enhancing Brain-computer Interfacing through Advanced Independent Component Analysis Techniques

by

SUOGANG WANG

Thesis submitted for the degree of Doctor of Philosophy

March 2009

Abstract

A Brain-computer interface (BCI) is a direct communication system between a brain and an external device in which messages or commands sent by an individual do not pass through the brain's normal output pathways but is detected through brain signals. Some severe motor impairments, such as Amyothrophic Lateral Sclerosis, head trauma, spinal injuries and other diseases may cause the patients to lose their muscle control and become unable to communicate with the outside environment. Currently no effective cure or treatment has yet been found for these diseases. Therefore using a BCI system to rebuild the communication pathway becomes a possible alternative solution. Among different types of BCIs, an electroencephalogram (EEG) based BCI is becoming a popular system due to EEG's fine temporal resolution, ease of use, portability and low set-up cost. However EEG's susceptibility to noise is a major issue to develop a robust BCI. Signal processing techniques such as coherent averaging, filtering, FFT and AR modelling, etc. are used to reduce the noise and extract components of interest. However these methods process the data on the observed mixture domain which mixes components of interest and noise. Such a limitation means that extracted EEG signals possibly still contain the noise residue or coarsely that the removed noise also contains part of EEG signals embedded.

Independent Component Analysis (ICA), a Blind Source Separation (BSS) technique, is able to extract relevant information within noisy signals and separate the fundamental sources into the independent components (ICs). The most common assumption of ICA method is that the source signals are unknown and statistically independent. Through this assumption, ICA is able to recover the source signals. Since the ICA concepts appeared in the fields of neural networks and signal processing in the 1980s, many ICA applications in telecommunications, biomedical data analysis, feature extraction, speech separation, time-series analysis and data

mining have been reported in the literature. In this thesis several ICA techniques are proposed to optimize two major issues for BCI applications: reducing the recording time needed in order to speed up the signal processing and reducing the number of recording channels whilst improving the final classification performance or at least with it remaining the same as the current performance. These will make BCI a more practical prospect for everyday use.

This thesis first defines BCI and the diverse BCI models based on different control patterns. After the general idea of ICA is introduced along with some modifications to ICA, several new ICA approaches are proposed. The practical work in this thesis starts with the preliminary analyses on the Southampton BCI pilot datasets starting with basic and then advanced signal processing techniques. The proposed ICA techniques are then presented using a multi-channel event related potential (ERP) based BCI. Next, the ICA algorithm is applied to a multi-channel spontaneous activity based BCI. The final ICA approach aims to examine the possibility of using ICA based on just one or a few channel recordings on an ERP based BCI.

The novel ICA approaches for BCI systems presented in this thesis show that ICA is able to accurately and repeatedly extract the relevant information buried within noisy signals and the signal quality is enhanced so that even a simple classifier can achieve good classification accuracy. In the ERP based BCI application, after multi-channel ICA the data just applied to eight averages/epochs can achieve 83.9% classification accuracy whilst the data by coherent averaging can reach only 32.3% accuracy. In the spontaneous activity based BCI, the use of the multi-channel ICA algorithm can effectively extract discriminatory information from two types of single-trial EEG data. The classification accuracy is improved by about 25%, on average, compared to the performance on the unpreprocessed data. The single channel ICA technique on the ERP based BCI produces much better results than results using the lowpass filter. Whereas the appropriate number of averages improves the signal to noise rate of P300 activities which helps to achieve a better classification. These advantages will lead to a reliable and practical BCI for use outside of the clinical laboratory.

Acknowledgements

It is a pleasure to thank the many people for their support and encouragement during the course of my research and thesis preparation.

I would primarily like to thank my supervisors Dr Christopher, J. James and Professor Maria Stokes with their enthusiasm, their inspiration and their great efforts to explain things clearly and simply. Throughout my thesis-writing period, they provided encouragement, sound advice, good teaching, and lots of good ideas. I would have been lost without them.

I would like to thank staff and my postgraduate colleagues at the Signal processing and control group in the ISVR for their help in various aspects of my work. In particular I thank Disha Gupta, Norma Castenada-Villa, Charmaine Demanuele and Aida Jimenez-Gonzalez for discussing about ICA techniques and exchanging ideas.

I would also like to thank staff from the ERP laboratory in the School of Psychology for their help on the EEG side of things; staff from School of Health Sciences for their help on understanding the cognitive concepts.

I gratefully acknowledge receipt of the Lord Rayleigh Scholarship to support my research and living expense.

I wish to thank my parents, Zhijun Wang and Congxian Suo for their great encouragement and support for me studying abroad. To them I dedicate this thesis.

Lastly, and most importantly, I wish to thank my wife Ying Yang who moved more than 5,000 miles from China to UK, stands beside me and encourages me constantly for every step of my work and in preparing this thesis. My thanks are also to my baby, Yangjia Wang for giving me happiness and joy.

Content

List of Figures

List of Tables

Notations and Abbreviations

Notations

$\boldsymbol{\delta}$	delta band EEG frequencies
$\boldsymbol{\theta}$	theta band EEG frequencies
$\boldsymbol{\alpha}$	alpha band EEG frequencies
$\boldsymbol{\beta}$	beta band EEG frequencies
$\boldsymbol{\mu}$	mu band EEG frequencies
x	input signal
$\bar{\mathbf{x}}_t$	signal average
y	output signal
N	signal length
P	IIR feedforward filter order
Q	IIR feedforward filter order
a_i	filter coefficient
b_i	filter coefficient
φ_i	parameter of AR model
θ_i	parameter of MR model
s	source signal
A	mixing matrix
$\mathbf{a}(i)$	column vector of the mixing matrix
W	unmixing matrix
$\mathbf{w}(i)$	column vector of the unmixing matrix
m	number of mixed signals
n	number of source
ŝ	estimated source signal
f	function
$\mathbf{n}(t)$	additive random noise
V	linear transformation

$\boldsymbol{E}$	covariance
$\mathbf{C}$	correlation matrix
$\mathbf{H}$	differential entropy
$\mathbf{J}$	Negentropy
τ	time delay
$\mathbf{C}_{\tau}^{\mathbf{x}}$	signal cross-covariance matrix
$\mathbf{C}_{\tau}^{\mathbf{s}}$	source cross-covariance matrix
$\mathbf{U}$	orthogonal matrix
$\mathbf{A_c}$	spatial constraint vector in a mixing matrix
$\mathbf{A_u}$	unconstrained vectors in a mixing matrix
$g(\mathbf{w})$	closeness constraint for temporally constrained ICA
$\varepsilon(\mathbf{w})$	closeness
ξ	closeness threshold
r	reference signal for temporally constrained ICA
$f(\mathbf{w})$	contrast function
$\mathbf{c}_l(t)$	reference vector for spectrally constrained ICA
υ_k	delay vector
υ	delay matrix
D	number of degrees of freedom of the underlying system
$\widehat{m}$	dimension of embedding matrix
f_s	sampling frequency
f_L	lowest frequency of interest
$\mathbf{u}_i$	ith independent component
$\widetilde{\mathrm{x}}(t)$	space-time vector
x_{ij}	bandpass filtered j^{th} sample of the i^{th} trial
R	average power in the reference interval
$\mathbf{W}^{-1}$	inverse unmixing matrix
*	convolution
F^{-1}	inverse Fourier Transform
$F\{f\}$	Fourier Transform of $f(t)$
$G\{g\}$	Fourier Transform of $g(t)$
$\mathbf{C}_{\tau}^{x}$	lagged covariance matrix stack
$\mathbf{P}(i)$	trial spectrum
$\mathbf{P}_{ic}(j)$	trial spectrum in the j^{th} IC source

$\mathbf{h}_d$	frequency bin
$\mathbf{f}_p(i)$	power feature in an individual channel
r^2	r^2 measure (the coefficient of determination)

Abbreviations

A/D	Analog-to-digital
ALS	Amyotrophic Lateral Sclerosis
ANN	Artificial neural network
APRA	Advanced Research Projects Agency
AR	AutoRegressive model
ARMA	Autoregressive-moving-average
BCI	Brain computer interface
BSS	Blind source separation
cICA	constrained ICA
DE	Dynamical embedding
DFT	Discrete Fourier transform
DS	Dynamical system
ECG	Electrocardiography
ECoG	Electrocorticogram
EEG	Electroencephalogram
EM	Electromagnetic
EOG	Electrooculography
EP	Evoked potential
ERD	Event-related desynchronization
ERP	Event-related potential
ERS	Event-related synchronization
FFT	Fast Fourier Transform
FIR	Finite impulse response
fMRI	Functional magnetic resonance imaging
IC	Independent component
ICA	Independent component analysis
IIR	Infinite impulse response
LDA	Linear Discriminant Analysis
MA	Moving average

MEG	Magnetoencephalography
MI	Motor Imagery
NIRS	Near-infrared spectroscopy
PCA	Principal component analysis
PRBS	Pseudorandom binary sequence
SC-ICA	Single-channel ICA
SCP	Slow cortical potentials
SNR	Signal to noise ratio
ST-ICA	Space-time ICA
SVD	Singular value decomposition
SVM	Support vector machine
TD	Temporal decorrelation
VEP	Visual evoked potential

CHAPTER 1

Introduction

1.1 Background

Many different disorders, such as Amyotrophic lateral sclerosis (ALS), brainstem stroke, brain or spinal cord injury and numerous other diseases can disrupt the neuromuscular channels through which the brain controls and communicates with its environment. These kinds of severe diseases may cause people to lose verbal and nonverbal communication. 'Locked-in' usually refers to a situation where individuals are conscious and aware, but unable to control their muscles so that they cannot present their needs, wishes, and emotions. In short, the healthy brain is locked into a paralysed body. However the current knowledge about these disorders is rather limited. There are no effective treatments which can provide a 'cure' or even a significant recovery.

In the absence of methods for repairing the damage caused by these diseases, there is an option that we believe can provide an individual with a new, non-muscular communication – a brain-computer interface (BCI) for conveying messages and commands to use some devices such as assistive applications, computers etc. This

type of direct-brain interface would increase an individual's independence and improve his/her quality of life.

In theory, many measurements through monitoring brain activity might serve as a BCI. At present, the electroencephalogram (EEG) and related measurements can be applied to form a possible non-muscular BCI, for the EEG can function in most environments and requires relatively simple and inexpensive equipment. Regardless of the technology used in BCI, some specific features must be extracted from the raw signals in order to form a meaningful BCI. The kinds of signal patterns include visually evoked potentials (VEP), slow cortical potentials (SCP), P300, μ rhythms and cortical neuron activity. All of these methods have been shown to be useful for implementing practical working BCI systems [1].

The traditional VEP based word speller allows the user to select a letter on a virtual keyboard by looking directly at it. The system determines the target by detecting the VEP amplitude difference of the response or different frequency of the response in the EEG recordings. SCP BCI system allows the user to move a cursor horizontally or vertically on a computer screen by generating negative or positive potential shifts during the procedure. This SCP BCI gives locked-in patients with motor disability a chance to communicate to the outside since SCP potential is easy to handle and apply. In P300 based BCI, especially the famous P300 word speller uses a characteristic called the 'oddball paradigm' which involves frequent and in-frequent stimuli and the positive potentials are evoked by rare events. This word speller is examined as a training-free and processing-fast system. The idea behind spontaneous rhythmic activity based BCI is that people can learn to modify the amplitude of a particular rhythmic band (say μ band) after training. The difference of amplitude can be compared and translated into cursor movement. This BCI requires fewer electrodes and simple classification. Unlike the above signal patterns used in BCIs, cortical neuron activity is a different signal pattern which is recorded from implanted electrodes located in the motor cortex in BCI system [2]. The neuronal activity is

much clearer and clean and can be recorded by implanted electrodes. Studies have found the capacity of people to learn to control neuronal firing rate and more rapid and accurate control can be achieved. But the argument is that whether implanted methods are safe and whether they can provide significant improvement compared to non-implanted methods.

However, the task of developing a BCI is extremely difficult, since a number of large problems concerning EEG recordings must be overcome. Moreover, artifacts (such as movement artifacts, eyeblinks, etc) change the raw EEG and render the recording virtually unusable. In general, signal processing techniques such as coherent averaging, filtering, FFT and AR modelling, etc. are used to reduce the noise and extract components of interest. However these methods can only deal with the data processing on the observed mixture domain which mixes components of interest and noise. Such a limitation dictates that extracted EEG signals possibly still contain the noise residue or the removed noise contains embedded part of EEG signals. Blind Source Separation (BSS) techniques such as Independent Component Analysis (ICA) extract the relevant information buried within noisy signals and allow the separation of measured signals into their fundamental underlying (independent) components. The ultimate aim in applying ICA is to accurately and repeatedly extract the relevant information buried within noisy signals, so that the BCI system is more reliable and robust – especially for use outside of the clinical laboratory. This body of work will examine the use of existing and newly developed BSS techniques in conjunction with other relevant signal processing methods.

1.2 Overview of the thesis

The rest of this thesis is divided into seven chapters. Chapter 2 provides an introduction to the human brain, the EEG, the technique of BCI and signal processing

techniques for BCI, giving an overview of the discovery of the EEG, recording techniques, the definition and structure of BCI and the goal of this thesis.

Chapter 3 introduces the concept of ICA and ICA techniques. Following a brief historical overview of its development, it presents the generic technique of ICA and the fundamental assumptions based on which the algorithm works. The different possible algorithms in the ICA literature are discussed, as well as the existing ICA applications in biomedical signal processing field.

The rest of the chapters are about proposed ICA techniques on specific BCI applications. These ICA techniques are able to improve the performance of BCI applications and help develop more practical BCI systems based on the use of automated independent source selection methods. Chapter 4 presents preliminary analyses for a Southampton BCI pilot study. It is the first study performed within this BCI research group in which it followed the paradigms in the literature and examined the possibility of using evoked potentials and spontaneous activity within the Southampton BCI program. Chapter 5 reviews the P300 potential detection problem in a P300 word speller application. Several ICA techniques based on spatial constraint and templates are then proposed to enhance the performance. Chapter 6 presents another important BCI application which uses spontaneous rhythmic activity as the control feature. An ICA method based on time structure together with temporal constraints is proposed to deal with the problem of inefficiently detecting power changes generated by spontaneous rhythmic activities. The final performances are compared and discussed. Essentially, all the above ICA techniques are built upon multi-channel data recordings. But in the real world not so many circumstances allow us to use multiple channels, for example, multi-channel recording devices are not available or only a few (or even only one) channel of all recordings from the multi-channel data is of good quality. Chapter 7 examines the possibility of using ICA which is only based on one or a few channel recordings on the above same P300 word

speller application. The results are assessed and compared with the performance achieved in the multi-channel version.

The thesis then concludes with a discussion on the proposed ICA techniques and the improvement in their applications. Also areas which could benefit from future work are discussed.

1.3 Publications

During the period of my doctoral study, the following papers and abstracts have been accepted for publication.

Journal Papers

S. Wang, and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis", *Computational Intelligence and Neuroscience*, 2007(ID41468), 9pp, 2007.

Conference Papers

C. J. James and S. Wang, "Single Channel ICA on P300 based BCI", *Proceedings of the IET Medical Signal and Information Processing Conference MEDSIP 2008*, Italy, 14-16 July, CD-ROM, 2008

M. Davies, C. J. James and S. Wang, "Space-Time ICA and EM Brain Signals", *Proceeding of 7th International Conference, ICA 2007*, 577-584, 2007.

S. Wang, and C. J. James, "On the independent component analysis of evoked potentials through single or few recording channels", *29th International Conference of IEEE Engineering in Medicine and Biology Society (EMBC2007)*, Lyon, France, 23-26 August 2007, 5433-5436, 2007.

C. J. James and S. Wang, "Blind source separation in single-channel EEG analysis: An application to BCI", *Proceedings 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 31 August - 2 September 2006, CD-ROM, 2006,

S. Wang, and C. J. James, "Enhancing evoked responses for BCI through advanced ICA techniques", *Proceedings of the IET Medical Signal and Information Processing Conference MEDSIP 2006*, Glasgow, Scotland, 17 - 19 July 2006, CD-ROM, 2006.

Conference Abstracts

S. Wang, C. J. James and M. Stokes, "The Southampton BCI Research Programme", *Abstract book of Satellite symposium of IEEE EMBS 27th Annual Conference*, Beijing, China, 2005.

S. Wang, and C. J. James, "Preprocessing the P300 word speller with ICA for Brain-Computer Interfacing", *Abstract book of ICA Research Network Workshop on Applied Blind Source Separation and Independent Component Analysis*, Southampton, 2005.

CHAPTER 2

Electroencephalography and Brain-computer Interfacing

2.1 Introduction

An electroencephalogram (EEG) is a recording of the brain's electrical activity. Nowadays, EEG recording has become a routine clinical procedure as well as a useful tool to investigate many disorders such as epilepsy in particular. A BCI system is a direct communication pathway between the brain and an external device. In BCI applications, the EEG has shown the most potential to record the signal input, mainly due to its fine temporal resolution, ease of use, portability and low set-up costs.

This chapter first gives a brief introduction of brain anatomy and an overview of the history and origin of the EEG. The constitution of the EEG is introduced and discussed next along with the technology's susceptibility to noise. Next the chapter introduces the BCI concept in general, including the definition, history, structure and present day techniques. Finally the chapter gives a short summary of BCI systems in use today and introduces the goal of the BCI studies in this thesis.

2.2 Brain Anatomy

The brain is the most complex part of the human body. It controls thought, memory, emotion, touch, motor skills, vision, respirations, temperature, hunger, and every process that regulates our body. The brain can be divided into the cerebrum, brainstem, and cerebellum [3] (Figure 2.1):

2.2.1 Brainstem

The brainstem is located at the bottom of the brain and connects the cerebrum to the spinal cord. Functions of this area include: movement of the eyes and mouth, relaying sensory messages (hot, pain, loud, etc.), hunger, respirations, consciousness, cardiac function, body temperature, involuntary muscle movements, sneezing, coughing, vomiting, and swallowing.

2.2.2 Cerebellum

The cerebellum is located at the lower back of the head and is connected to the brain stem. It is the second largest structure of the brain and is made up of two hemispheres. The cerebellum controls complex motor functions such as walking, balance, posture, and general motor coordination.

2.2.3 Cerebrum

The cerebrum is the largest part of the brain and is composed of the right and left hemispheres. Functions of the cerebrum include: initiation of movement, coordination of movement, temperature, touch, vision, hearing, judgment, reasoning, problem solving, emotions, and learning. More specifically four lobes make up the cerebrum, the frontal, temporal, parietal, and occipital.

1) Frontal lobe: the frontal lobe is the largest section of the brain located in the front of the head. It controls attention, behavior, abstract thinking, problem solving, creative thought, emotion, intellect, initiative, judgment, coordinated movements, muscle movements, smell, physical reactions, and personality.

2) Parietal lobe: the parietal lobe is the middle part of the brain, the parietal lobe helps a person to identify objects and understand spatial relationships between the person and objects around him/her. The parietal lobe is also involved in interpreting pain and touch in the body.

3) Occipital lobe: the occipital lobe is the back part of the brain and controls vision.

4) Temporal lobes: the temporal lobes are sited on both sides of the brain, these temporal lobes are involved in auditory and visual memories, language, some hearing and speech, language, plus some behavior.

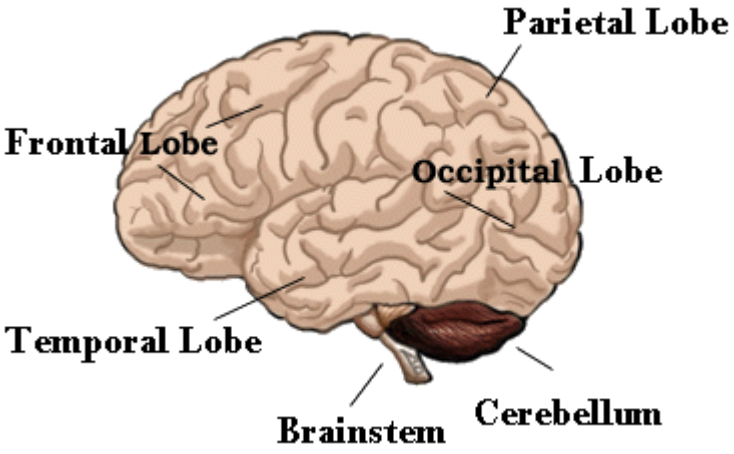

Figure 2.1: *The three main components of the brain -- the cerebrum, the cerebellum, and the brainstem. The cerebrum is the largest and most developmentally advanced part of the human brain. It includes the frontal, temporal, parietal, and occipital lobes* [3].

2.3 The electroencephalogram

2.3.1 EEG history and origin

EEG is a recording, through electrodes attached to the scalp, of electrical activity produced by the brain. The first EEG recording of the human brain was made by the German psychiatrist Hans Berger in 1924 (Figure 2.2), and his publication of this recording appeared in the Archives of Psychiatry, 1929 [4]. Nowadays, EEG plays one of the most important roles to evaluate neurological disorders in the clinic. Most commonly it is used to show the type and location of the activity in the brain during a seizure in epilepsy [5], [6]. It is also used to investigate those having problems associated with brain functions in the laboratory [7], [8].

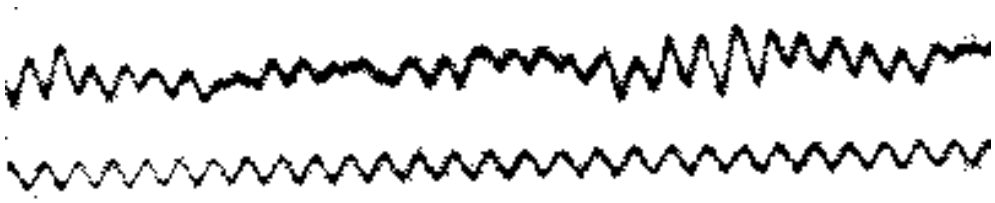

Figure 2.2: *The first EEG recording by Hans Berger* [4].

The EEG represents the averaged activity of many neurons (Figure 2.3 shows a simple schematic example of neurons [9]) in the brain. However the origin of the neuronal activity is not completely understood, it is generally accepted that the nerve cells communicate with each other by producing and sending action potentials through neighbouring axons. To let action potentials pass across the chemical junction interfaces ─ synapses between axons, a chemical substance called a neurotransmitter is released to bridge synapses and activity receptors which send a flow of ions into or out of the dendrite. This results in compensatory currents in the extracellular space. After passing through layers of fat, bone and cerebrospinal fluid these extracellular currents are summed and contribute to the generation of EEG potentials [10].

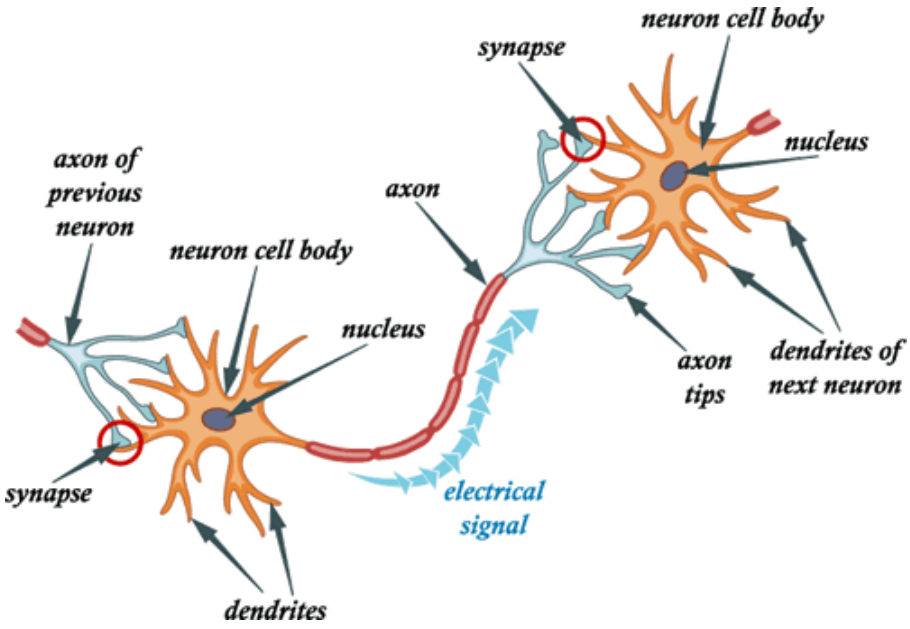

Figure 2.3: *A simple schematic of neurons* [9].

2.3.2 Types of electrodes and the 10/20 system

The recording of these activities is obtained by placing electrodes on the scalp, usually after preparing the scalp area by light abrasion and application of a conductive gel to reduce impedance [11]. To record good EEG signals, one of the keys is to select the appropriate type of electrodes for the measurement. Electrodes that make the best contact with a subject's scalp and contain materials (conductive gel) that most readily conduct EEG signals provide the best EEG recordings. Generally there are four types of EEG electrodes available to the related applications (Figure 2.4):

1) **Reusable disks:** These electrodes can be placed close to the scalp, even in a region with hair. A small amount of conductive gel is needed to be applied under each disk. Disks normally are made either from tin, silver, and gold. Since these electrodes can be repeatedly used for years, the overall cost is low.

2) **EEG Caps with disks:** EEG caps are used to hold the electrodes in place to the subject's head. They are available with different numbers and types of electrodes.

The conductive gel is injected under each disk through a hole in the back of the disk. This EEG accessory offers a quick and easy way to place multi-electrodes in precise locations.

3) **Adhesive Gel Electrodes:** These are the disposable silver/silver chloride electrodes for recording from regions of the scalp without hair. Since the size of these electrodes is large, they cannot be placed close enough to the scalp or in regions with hair.

4) **Subdermal Needles:** These electrodes normally are single-use needles that are placed under the skin. However due to the potential risk of a needle stick injury (for example, skin infection), in some situations, permission is needed before needle electrodes are used.

The standard method for scalp electrode localization is the International 10/20 electrode system [12]. The “10” and “20” represent the actual distances between neighbouring electrodes are either 10% or 20% of the total front-back or right-left distance of the skull. The positions are determined by the following two reference points: *nasion*, which is the point between the forehead and the nose, level with the eyes; and *inion*, which is the bony prominence at the base of the skull on the midline at the back of the head. From these points, the skull perimeters are measured in the transverse and median planes [13]. The electrode positions are showed in Figure 2.5. Each location uses a letter to identify the lobe and a number to identify the hemisphere location. The letters F, T, C, P and O stand for Frontal, Temporal, Central, Parietal and Occipital respectively. A "z" refers to an electrode placed on the midline. Even numbers refer to electrode positions on the right hemisphere, whereas odd numbers refer to those on the left hemisphere.

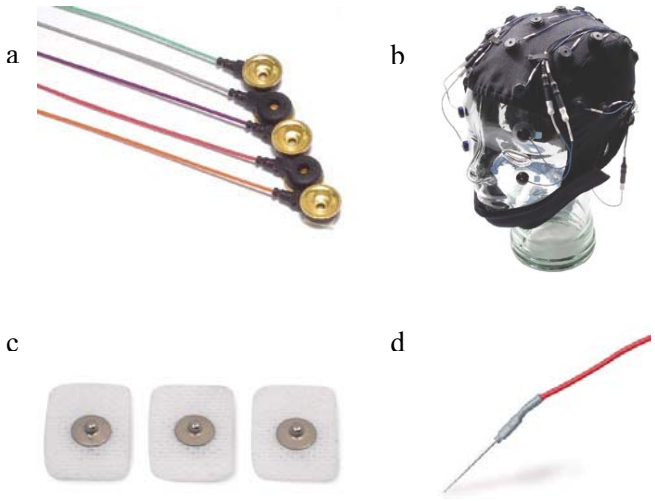

Figure 2.4 *Four types of EEG electrodes: a) Reusable disks; b) EEG Caps with disks; c) Adhesive Gel Electrodes; d) Subdermal Needles.*

For the recording, each electrode and a system reference electrode are connected to a differential amplifier as two inputs. This amplifier amplifies the voltage between the measurement and the reference electrodes. In an analog EEG system, the signal is plotted on paper for investigation by specialists. Nowadays most EEG systems are able to digitize the amplified analog signal through an analog-to-digital (A/D) converter. The typical analog-to-digital sampling rate is within the range of 240Hz to 512Hz in clinical scalp EEG. The digital signal then can be displayed on a computer screen or stored electronically for later use. The amplitude of a normal adult EEG is about 10~100 μV when measured on the scalp.

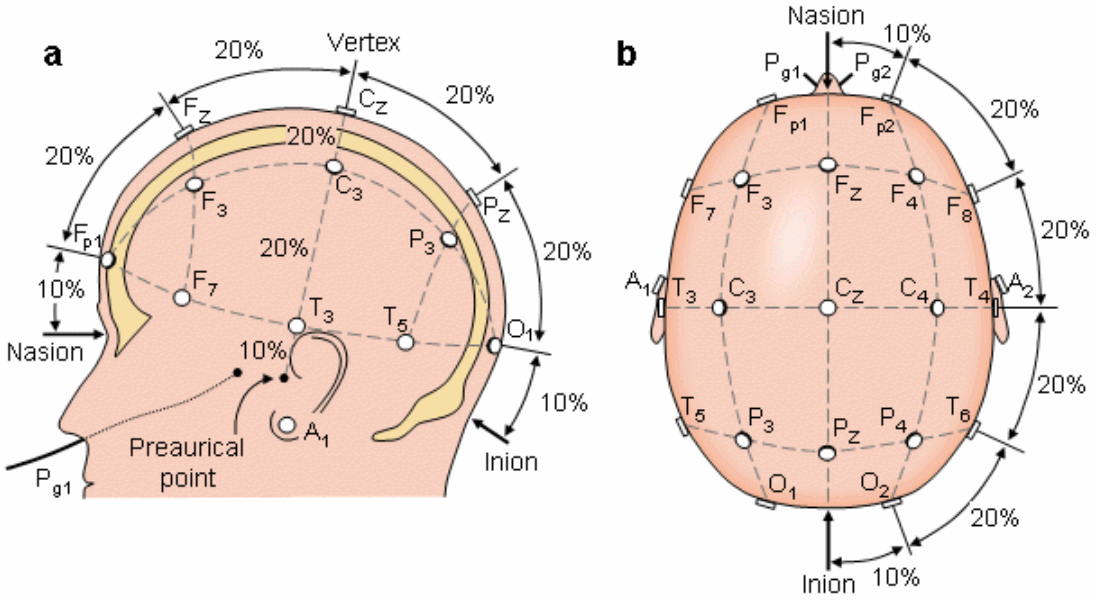

Figure 2.5: *The International 10-20 system seen from left (a) and above the head (b). A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital, T=temporal.*[12]

2.3.3 Rhythmic activity

Historically EEG is often divided into four major types of continuous rhythmic sinusoidal waves (α, β, δ and θ) based on a series of frequency ranges. There is no precise agreement on the frequency ranges for each type:

1) $\boldsymbol{\delta}$ (delta) is in the frequency range: < 4 Hz. The shape is observed as the highest in amplitude and the slowest waves. It is often seen frontally in adults during deep sleep and also seen posteriorly in babies. It may occur focally with subcortical lesions and in general distribution with diffuse lesions and certain encephalopathies.

2) $\boldsymbol{\theta}$ (theta) is in the frequency range: 4 Hz ─ 8 Hz. It is associated with drowsiness, childhood, adolescence and young adulthood. This EEG frequency can sometimes be produced by hyperventilation. $\boldsymbol{\theta}$ waves can be seen during hypnagogic states such as trances, hypnosis,

deep day dreams, lucid dreaming and light sleep and the preconscious state just upon waking, and just before falling asleep.

3) α (alpha) is in the frequency range: 8 Hz ─ 12 Hz. It is characteristic of a relaxed, alert state of consciousness and is present by the age of two. α rhythms can be best detected with the eyes closed whilst it attenuates with drowsiness and open eyes. It can be seen on both sides in the posterior area and higher amplitude on the dominant side. An α-like normal variant called μ rhythm is sometimes seen over the motor cortex (central scalp). μ rhythm attenuates with movement of limbs, or mental imagery of movement.

4) β (beta) is in the frequency range: 13 Hz ─ 30 Hz. It is seen in low amplitude with multiple and varying frequencies symmetrically on both sides in the frontal area. It is often associated with active, busy or anxious thinking and active concentration. β rhythm with a dominant set of frequencies is associated with various pathologies and drug effects.

2.3.4 Artifacts and other effects

Although the EEG records brain activity, it also records electrical activity arising from sites other than the brain. Recorded activity that does not originate in the brain is referred to as an artefact. Artifacts can be divided into physiologic and extraphysiologic artifacts. While physiologic artifacts are generated by the subject, they arise from sources other than the brain (i.e. the body, e.g. muscles, heart etc). Extraphysiologic artifacts arise from outside the body (i.e. equipment, environment) [14].

The major physiologic artifacts, eye movement and blinks, produce electrical potentials and electo-magnetic fields that are often much larger than those deriving from brain sources. They can cause big errors in peak measurement or source localization. Attempts to control this artifact by instructing subjects to fixate their gaze on a point or not to blink are often ineffective, particularly if the subject is psychotic or cognitively impaired [15]. Other artifacts, such as muscle activity, line noise, body movements etc can also generate potentials that may even mimic cerebral activity [16].

Many methods have been proposed to remove artifacts (such as eye movement and blinks) from EEG recordings including: trial rejection and regression based methods [17], [18]. However, simply rejecting contaminated EEG epochs results in a considerable loss of collected information. Regression is performed in the time or frequency domain on EEG recordings to derive parameters characterizing the appearance and spread of electrooculography (EOG) artifacts in the EEG channels. But EOG records also contain brain signals, so the regression method inevitably involves subtracting a portion of the relevant EEG signal from each recording as well, which is not desirable.

2.3.5 Further EEG application

Over time, people have also speculated that the EEG could have further applications, that it could be used to decipher thoughts, or intent, so that a person could communicate with others or control devices directly by means of brain activity, without using the normal channels of peripheral nerves and muscles. This is where the concept of brain-computer interfacing (BCI) first appears [19].

2.4 Brain computer Interfacing

2.4.1 Severe physical disability through brain damage

Amyotrophic Lateral Sclerosis (ALS) is a rare progressive and ultimately fatal neurological disease in which the degeneration of specific motor neurons causes muscles to stop receiving movement command [20] [21] [22]. People with ALS will ultimately lose the ability to control their muscles, to communicate and eventually to breathe. However, ALS commonly does not affect a person's ability to think, their intelligence or memory. Therefore a healthy brain is "locked in" a paralysed body [23]. The cause of ALS is still unknown, and there are no effective treatments which can provide a 'cure' or even a significant recovery. Brainstem stroke, brain or spinal cord injury and numerous other diseases produce similar severe disability where neuromuscular channels between a brain and its environment are broken. Therefore, maintaining the basic communication and control capabilities so that these individuals can express their desires to caregivers or even operate word processing programs or neuroprostheses has become a very important clinical issue.

2.4.2 The definition and a brief history of BCI

BCI is primarily a communication system in which messages or commands that an individual sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles, but rather pass through a computer based system.

Early BCI research was started in the 1970's. Professor Jacques J. Vidal first introduced the idea of direct brain computer communication in 1973 [19]. In his BCI Laboratory at the University of California Los Angeles, a successful project demonstrated that a computer-generated visual stimulation is able to evoke people to

produce a certain response which could provide a communication channel between the subject and a computer. In that decade, several scientists and organizations also tended to develop similar communication systems driven by the recorded brain activity. For example, the Advanced Research Projects Agency (ARPA) in the Defence Department of the USA planned to develop an interaction application between a human brain and a computer for use in military applications [24]. This project was proposed to improve the performance of military personnel especially in tasks involving heavy mental loads. The research produced valuable insights, but made minimal progress toward its goals due to limitations in the techniques used.

Over the past decades, BCI research has grown rapidly and become a very popular research topic around the world. Several leading BCI research groups, such as the Berlin BCI group [25], University of Tübingen [26], Germany; Wadsworth Center, US [27] and Graz BCI research, Austria [28] have already made contributions in this area. For the most part, developing a possible augmentative communication option for people with severe motor disabilities has become the most important objective in BCI research.

2.4.3 BCI structure

Like any communication system, a BCI system contains several units: BCI input, feature extraction, translation algorithm and BCI output. Figure 2.6 shows these elements and their principal interaction [29], each component is discussed in more detail next.

(1) BCI input

In theory, any brain activity has the potential to be used as the basis of a BCI. The most common one is the recording of electrical activity. Other technologies such as

magnetoencephalography (MEG) [30], near-infrared spectroscopy (NIRS) [31], and functional magnetic resonance imaging (fMRI) [32] are used as well but MEG and fMRI equipment are currently too large and expensive, for the practical application. NIRS is small and affordable, but it is based on hemodynamic responses [33], and thus the time constants involved are relatively long. As a result, most BCI research groups are focused on bioelectrical brain signals recorded by EEG and electrocorticogram (ECoG) [34] [35]. The electrocorticogram (ECoG) is the recording of brain activity by using epidural or subdural electrode arrays from the cerebral cortex. Because a craniotomy is required for implantation of the electrodes, ECoG is an invasive procedure. That is to say that it is not possible to use ECoG for recording outside of clinical laboratory. Since EEG is recorded from multiple electrodes placed on the scalp, which normally does not involve an invasive procedure. EEG has become a popular measure of brain activity and is used for the most part as the BCI input. The raw EEG sequences come out of an A/D converter and are digitized at a sampling rate of several hundreds Hz per recording channel as input data. The detailed EEG acquisition method has been introduced in the early EEG section of this chapter.

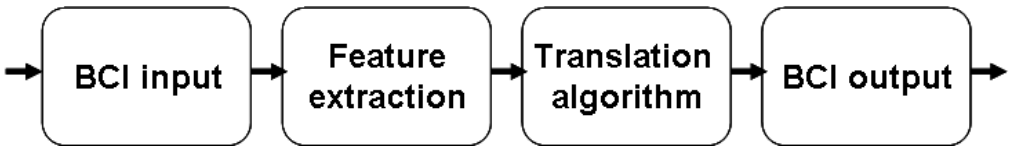

Figure 2.6: *The basic design and operation of a BCI system.*

(2) Feature extraction

The BCI system receives commands sent by the user in the form of EEG patterns. Feature extraction procedures, such as spatial filtering, voltage amplitude measurements and spectral analysis etc., extract the signal features that encode the user's messages or commands. BCI can use signal features that are in the time domain (such as evoked potential amplitudes or neuronal firing rates) or the frequency domain (such as μ or β rhythm amplitudes). A BCI may also use both time-domain and frequency-domain signal features together to improve performance. The knowledge of elements such as the location, size, and function of the cortical area generating a rhythm or an evoked potential can indicate how the signal should be recorded, how users might best learn to control its amplitude, and how to recognize and eliminate the effects of physiologic artifacts.

(3) Translation algorithm

Before the next step – the translation algorithm – begins, several important factors much be taken care of. First, when a new user is first introduced to a BCI system the algorithm must adapt to that user's signal features. For example, if the signal feature is the amplitude of rhythmic activity, the algorithm should adjust to the user's specific frequency band as well as the power spectra; if the feature is a particular waveform in time, it adjusts to the user's own feature characteristics. Since this is a basic requirement, all BCI processing should reach this level of adaptation. It assumes that the user's performance remains stable and never changes again. However, EEG and other electrophysiological signals display variations linked to time of day, hormonal levels, immediate environment, recent events, fatigue, illness, and other factors. Thus, effective and advanced BCIs need to deal with this higher level of adaptation through periodic online adjustments to reduce the impact of such spontaneous variations. A good algorithm will adjust to these variations so as to match as closely as possible the

user's current range of signal feature values to the available range of device command values. This is even more critical in subjects with neuron degenerative disorders, where BCI paradigms may even have to be changed over time.

In order to find out what the user wants, a BCI system has to classify the processed data after the previous operations. This means that the system does not attempt to understand the user's intentions, but it compares the data symbolizing a segment to representatives of a limited number of classes and selects a class which is best fitted to the classification criteria. This translation process might use linear classification methods or nonlinear classification methods. All the methods change independent variables (signal features) into dependent variables (e.g. the word in the P300 word speller which is going to introduce in Section 2.5, the direction in a control panel. etc).

a. Linear classification methods

The goal of a linear classifier is to group real vectors into classes by making a classification decision based on the value of the linear combination of the features. Normally a linear classifier has a function that maps a high-dimensional input into a two-dimensional space (two-class problem) by using a hyperplane: all points on one side of the hyperplane are classified as the first class, while the others are classified as the second class. The most popular linear methods, for example, include: simple threshold method [36], linear discriminant analysis (LDA) [37] and linear support vector machine (SVM) [38].

A simple threshold method finds a suitable threshold from a set of fully labeled data called a training data set, then maps all unlabeled values in a testing set which are above the threshold to the first class and all other values to the second class. LDA maximizes the ratio of between-class variance to the within-class variance and finds a linear transformation ("discriminant function") which can transforms values into their

own classes. An SVM constructs a separating hyperplane in a high dimensional space. In order to achieve maximum separation, a margin between the data belonging to two classes is maximum so that the distance from the hyperplane to the nearest data point is maximized. Such a hyperplane is known as the maximum-margin hyperplane and can separate more complex data element into certain classes.

A linear classifier is often used in situations where the speed of classification is an issue, since it is often the fastest classifier, also, linear classifiers often work very well when the number of dimensions in the input space is large.

b. Nonlinear classification methods

If a classification problem exists and it cannot be solved linearly. The classification may be solved by using nonlinear methods. One of popular nonlinear methods, the artificial neural network (ANN) often called a "neural network" (NN) [39], is a mathematical model or computational model based on biological neural networks. The ANN works like a “black box” which takes the input vector and generates an output vector. The processing between the input and output is typically associated with an adjustable set of weights which are computed during the training phase. The weighted sum of the input will be transformed into an output value telling which of classes the input vector belongs to.

Another well-considered nonlinear method, the nonlinear SVM classifier, applies a kernel function to maximum margin hyperplanes [40]. Since linear SVM is not adequate for cases when complex relationships exist between input parameters and the class of a pattern, the SVM model can be fitted with nonlinear functions to provide efficient classifiers for hard-to-separate classes of patterns. The nonlinear SVM is formally similar to a linear SVM, except for a non-linear kernel function. An advantage of SVMs is that whilst ANNs can suffer from multiple local minima, the

solution to an SVM is global and unique. The computational complexity of SVMs does not depend on the dimensionality of the input space [41].

Compared to linear methods, nonlinear methods can achieve more accurate classification performance in complex classification problems, however several drawbacks in practice are that the computation complexity of nonlinear methods is expensive and normally the speed of training is slow.

(4) BCI output

Most BCI output devices are a computer screen. The output is the selection of targets, letters, or icons representing the commands. In addition to being the intended product of BCI operation, some output, such as cursor movement toward the item prior to its selection, acts as feedback that the brain can use to maintain and improve the accuracy and speed of communication. Some studies are exploring BCI control of a neuroprosthesis or orthosis that provides hand closure to people with cervical spinal cord injuries [42] [43] [44]. In such specific BCI application, the output device is the user's own hand. A few studies have also used monkeys with implanted electrodes allowing the animals to control a robotic arm [45] [46].

2.4.4 BCI categorisation

BCI systems fall into one of two classes: dependent and independent BCI systems. A dependent BCI minimally requires a partially intact neuron muscular communication pathway. For example, one dependent BCI could be a matrix of letters that are flashing on a computer screen one at a time at different frequencies, and the user chooses a specific letter by fixating upon at it. In this case, the brain's way of communication in this example is EEG, but the flashing letter at different frequency activates extraocular muscles and the cranial nerves to generate the EEG signal. The

EEG activity with a certain frequency is detected to match a flashing letter with a different frequency as a way of the letter selection. Because such a dependent BCI select a target of interest by simply staring at it and it does not involve higher brain cognitive process. Therefore a dependent BCI system does not give the brain a new communication channel that is independent of conventional channels [47].

An independent BCI works with signals that are independent of muscle activity. Moreover the generation of the EEG signal relies mainly on the user's intent. For example, one independent BCI presents the user with a matrix of letters that flash one at a time, and the user selects a specific letter not by gazing but by "thinking" which produces certain electrical patterns, e.g. a P300 evoked potential (EP) [48] (this will be explained in more detail in Section 2.5.3). Therefore independent BCIs need a higher level brain cognitive process. Because of this higher cognitive process, independent BCIs provide the brain with wholly new output pathways, for people with the most severe neuromuscular disabilities, who may lack all normal output channels, independent BCIs are likely to be the most useful.

2.5 Current BCI techniques in the literature

Present-day BCIs can be divided into five groups based on the electrophysiological signal patterns they use rather than the terminology of dependence and independence explained above. The types of patterns that have been employed for this purpose include: 1) visual evoked potentials (which form the only dependent BCI in literature), 2) slow cortical potentials, 3) P300 evoked potentials, 4) spontaneous rhythmic activity and 5) cortical neuron activity, which form independent BCIs.

2.5.1 Visual evoked potential

The visual evoked potential (VEP) is the electrical response of the brain's primary visual cortex to a visual stimulus [49]. Figure 2.7 shows an example of VEP activities for use in VEP based BCI. Jacques Vidal first used the term 'brain-computer interface', in his work, he developed a system that satisfied the above definition of a dependent BCI [19]. The work used the VEP recorded from the scalp over the visual cortex to determine the direction of eye gazes and then to determine the direction in which the user wished to move a cursor.

Sutter introduced a new term – 'Brain response interface' [50] and developed a system based on VEP patterns. The scheme uses the VEPs produced by brief visual stimuli and recorded from the scalp over the visual cortex. The user faces a screen displaying 64 symbols in an 8×8 matrix and looks at the symbol he or she wants to select. Subgroups of these 64 symbols undergo a red/green flashing alternation or a red/green check pattern alternation 40–70 times/s. Each symbol is included in several subgroups, and the entire set of subgroups is presented several times. Each subgroup's VEP amplitude about 100 ms after the stimulus is computed and compared to a VEP template already established for the user. Then the system determines with high accuracy the symbol that the user is looking at. Normal volunteers can use such a word processing program at 10-12 words/min.

In [51], Middendorf *et al*. described another method for using VEPs in order to determine gaze direction. Several virtual buttons appear on a screen and flash at different frequencies. The user looks at a button and the system determines the frequency of the photic driving response over the visual cortex. When this frequency matches that of a button, the system concludes that the user wants to select it. In [52], Xiaorong Gao *et al*. applied a similar method in an environmental controller for a motion disabled user. Multiple targets are placed on a visual panel and flicker with different frequencies. The system then detects the fundamental frequency of the VEPs

matching the same frequency of the flickering target. Results have shown that this system can distinguish between at least 48 targets and provide an information transfer rate up to 70bits/min. They also found that a lower stimulus frequency gave higher accuracy but a slower speed; with the stimulus frequency in the α range (8-12Hz), the average speed of selection was high.

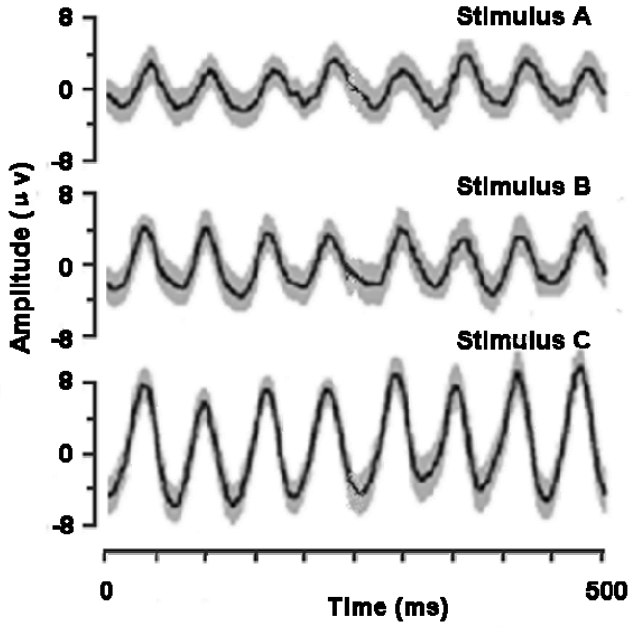

Figure 2.7: *Visual-evoked potentials (VEP) based BCI, it shows VEP activities for responding to different types of stimulation. The shades are the overlapped VEP signals and bold lines indicate the averaged VEPs.* [49]

Several efforts have been established to modify the model design to speed up the experiment at time for VEP based BCIs. For example in [53] the original framework includes Feature Extractor, Feature Translator and Control Interface. After investigation on the actual performance, the original three functional models were found incompatible. A new plan of optimization for the VEP based BCI framework design was proposed and tested in [54]. The proposed algorithms separate the simulator from the Control Interface. The Stimulator and the associated Stimulus Mechanism along with all the other components between the user and the Control

Interface can be treated as a conceptual BCI Transducer. The advantage of this functional delineation is that different BCI Transducer technologies can be connected to the same Control interface. It can represent a range of system configurations, including multi-modal designs, and supports any type of device.

In summary, The VEP is a naturally occurring response and more dependent on the stimulus presentation than subject attentiveness. VEP based BCI requires little or even no training for the subjects. The transfer rate of this dependent BCI is often faster than that of independent BCI.

2.5.2 Slow cortical potentials

Negative or positive potential shifts in the EEG lasting over 0.5-10.0s over the cortex are called slow cortical potentials (SCPs). SCP activity has been applied to control movement of an object on a computer screen and to choose a letter by using a series of two-choice selections in a word speller [55]. Figure 2.8 shows SCP activity for use in SCP based BCI. Basically, EEG is recorded from electrodes over the vertex area. SCPs are extracted by appropriate filtering and the different level of amplitude of SCPs is used to control a cursor to choose a target either at the top or at the bottom. In SCP based BCI studies [56] [57] [58], the investigators have shown that after training, people are able to learn to control SCPs and thereby control movement of an object and word speller on the computer. It also has been tested in locked-in patients and has proved able to supply basic communication capability. After sufficient practice, normally several months, the system can achieve high classification accuracy (70%~80%). However the process of this SCP based BCI is slow, requiring more than an hour for patients to write 100 characters with the cursor, while training often took many months. For this reason, the range of possible application of this method is quite limited.

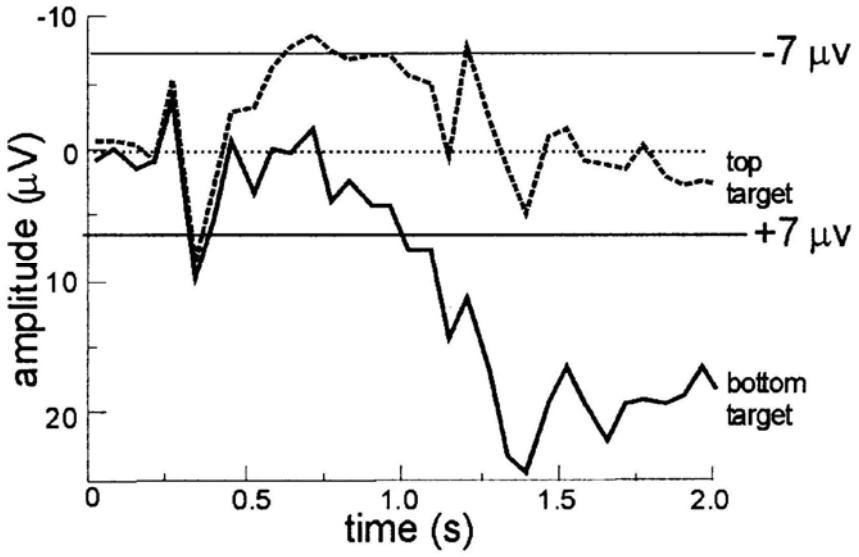

Figure 2.8: Slow cortical potential (*SCP) based BCI. Users learn to control SCPs to move a cursor toward a target by generating more positive SCPs (at the bottom) or more negative SCPs (top) on a computer screen* [57].

2.5.3 P300 evoked potentials

An event-related potential (ERP) or (evoked response (EP)) is any electrophysiological response to an internal or external "event" [59]. This event may be a sensory stimulus (such as a visual flash or an auditory sound), a mental event (such as recognition of a specified target stimulus), or the omission of a stimulus (such as an increased time gap between stimuli). More simply, it is any measured brain response that is directly the result of a thought or perception.

The well-known ERP response known as the P300 EP is a late positive wave that occurs over the parietal cortex at about 300 milliseconds after the onset of a meaningful stimuli and the principal frequency is below 10 Hz. The P300 response is able to occur regardless of the stimulus presented: visual, tactile, auditory, etc, because of this general invariance in regard to stimulus type, this ERP has been widely used in cognitive tasks. It forms a well known 'oddball paradigm' which involves frequent and infrequent external stimuli, the latter is able to elicit P300 EP

around 300-450 ms. A word speller based on this paradigm is one of the most important BCI applications to date, to detect real-time P300 waveforms and translate them into letters (and then words), this idea was introduced in [48]. Figure 2.9 illustrates the modality of the P300 for use in this type BCI. The advantages of the P300 are that the response occurs regardless of the stimulus presented: visual, tactile, auditory, etc and it requires no initial user training. Because of the general invariance, this ERP has become a popular research topic in BCI studies. A study has explored the relationship between speller matrix size and target detection accuracy [60]. The results show the speller matrix size has no significant effect on P300 latency but has an effect on P300 amplitude, and that a larger matrix produced larger differences in amplitude values which might decrease the difficulty of recognition [61]. Other studies also examined the application of advanced EEG preprocessing methods (e.g. Independent Component Analysis) together with newer classification methods (e.g. SVM) in these ERP based BCIs [62] [63] [64]. The results demonstrated that with appropriate methods, it is possible to increase the detection accuracy and simultaneously reduce the processing time.

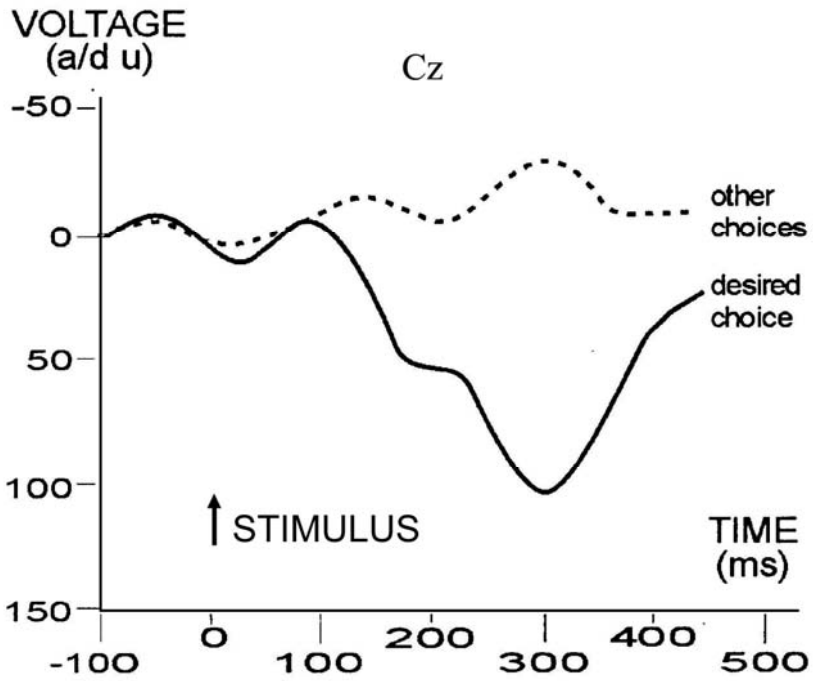

Figure 2.9: *P300 based BCI in which a matrix of possible choices is presented on a screen and these choices flash in succession. Only the choice desired by the user evokes a large P300 potential* [48].

2.5.4 Spontaneous rhythmic activity

In BCI studies, an emerging popular function is that which allows users to control the amplitude of their μ (8 – 12Hz) or β (18 – 22 Hz) brain rhythmic activity over the sensorimotor cortices. This is caused by Motor Imagery (MI) [65], (i.e. imagining hand or foot movement). For MI, the users are instructed to imagine a specific motor action without any related motor output. The imagination of different movements is followed by the different power of the EEG or an effect known as event-related (de)/synchronization (ERD/ERS) [66] on the sensorimotor cortex. When an ERD is present, it is relatively detectable and can be used as a feedback signal to control specially designed electrical devices, for instance, to control the movement of a cursor on a computer screen or to drive/ steer a wheelchair. In addition, humans can learn to

modify the amplitude of the μ rhythm after training. The analysis suggests that μ rhythms could be good signal features for EEG-based communication [67]. The classification can be very simple - just compare power in the frequency range of interest. Of more importance is that the accuracy is relatively high, at up to 95% [68]. Figure 2.10 shows the μ rhythmic activity in a μ rhythm based BCI. Because this type of BCI requires training and a proper training session could determine the quality of the experiments, a study has explored and examined the ways of designing efficient training sessions [69]. This study showed that with longer and better training, more accurate performance can be achieved. Some other studies carried out modifications to improve the classification. These include the use of parameters derived by autoregressive (AR) frequency analysis [70], and use of alternative spatial filters (e.g. PCA, ICA) [71] [72].

Many signal processing techniques have been developed and used in these BCI studies, such as AR modelling [73], and common spatial patterns (CSP) [74]. These methods tend to find a spatial filter to maximally improve the signal to noise ratio (SNR). For example CSP derives weights that are applied to each channel in order to emphasise or de-emphasise activity with a specific focus (or focii). Each set of weights can be derived and referred to as a spatial filter because when these are applied to multi-channel EEG they have the effect of selectively attenuating activity disparate spatial regions. Each spatial filter can then be attributed to one or more neurophysiological source or artefact. In the case of an artefact, a spatial filter which de-emphasises, say, ocular activity, will strongly attenuate the EEG amplitude around the frontal region whilst leaving other areas relatively unattenuated. Other spatial filters would work on other sources in a similar manner. In order to reach an optimal performance, some additional processing methods are required as pre-processing steps before the application of, for example, band-pass filtering, common average reference and/or manual artifact rejection. A combination of pre-processing methods can improve the performance, but also results in a less flexible and robust BCI system.

Moreover the application of additional processing methods brings with it the problem of increased computation time.

Several pilot studies have investigated the possibility of using certain brain activities during performing different types of non-motor imagery tasks such as the imagery of spatial navigation (around a favourite place), auditory (of familiar music) to drive BCI systems [75]. Stronger responses should be observed over the temporal cortical area for the spatial task while over the parietal cortical area for the auditory task. As the outcome from pilot studies, these non-motor imagery tasks achieved better classification results than those using motor imagery [76]. The non-motor imagery tasks can be an alternative option for the subjects who are not suitable to perform the motor imagery. For example, some paralyzed subjects especially since birth may not be able to access motor imagery tasks.

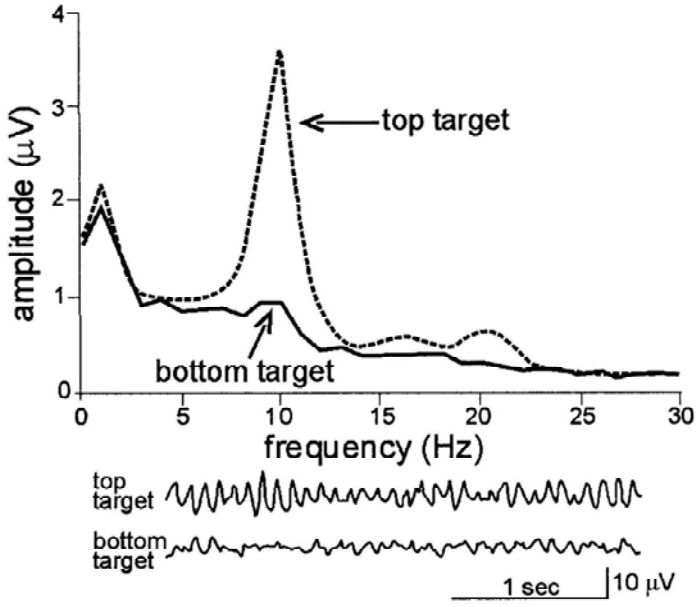

Figure 2.10: *μ rhythm activity based BCI. Users control the amplitude of an 8–12 Hz μ rhythm to move a cursor to a target may at the top of the screen or at the bottom. Frequency spectra for top and bottom targets show that control is clearly focused in the μ-rhythm frequency band* [58].

2.5.5 Cortical neuron Activity

Unlike the patterns described in previous sections, the cortical neuron activity benefits from implanted electrodes. These devices are very small and normally placed as an electrode array [77] (for example, BrainGate™ is a brain implant system developed by the bio-tech company Cyberkinetics in 2003 in which individual electrode products are 1 mm long, several μm wide in 10×10 grid). The electrode array is implanted about 5 mm deep into the cortex around the motor areas. The ensemble activity of multiple neurons are then detected and translated into motor commands and typically, a computer is needed to translate these commands into useful outputs.

A few studies [78] [79] [80] have shown the application of action potentials of single neurons from animal subjects – monkeys in particular. The monkeys have shown the capacity to control neuronal firing/spiking rate which is detected by the implanted electrodes to move a cursor on the computer screen or a robotic arm. For example, monkeys could learn to control the discharge of single neurons in the motor cortex. The classification processor compares different patterns and translates them into commands. The results show that the firing rates of a set of cortical neurons can reveal the direction and nature of movement [81].

The use of implanted electrodes on human subjects appeared in the latest studies [82]. Matthew Nagle, a 25-year-old quadriplegic was linked to the computer by BrainGate™ which was implanted into the motor cortex of his brain in June 2004. During the experiments he imagined limb motions to modulate the neural firing. By detecting the certain neural response evoked by imagined actions (imagined hand or distal arm actions), the firing patterns are transformed into a two-dimensional output signal displayed as a cursor position on a screen. From the results Nagle used this BCI to turn on lights, change TV channel, read Email and even draw something with a painting programme, all by moving the cursor through his cognitive actions. The

results also show that in a specially designed letter input system, the transfer rate is possible to reach up to 6.5 bits/s, or about 15 words/min.

2.6 Non-implanted vs implanted methods

Implanted methods require electrodes which have to be implanted into the cortex during surgery, and then brain tissue grows around the electrodes to secure them against any movement [83]. However it takes about three/four months to stabilize the signals following implantation. Because of the lack of knowledge to access accurately the location of neurons into the motor cortex area, a number of electrodes will fail to acquire the brain activity. Therefore careful selection of active electrodes with certain neural evoked responses postsurgery becomes important. Apart from these, the performance results have shown that a simple signal processing method (e.g. a threshold method) is enough to achieve good performance [84]. It is believed by some that the use of implants will increase the clinical viability of BCIs in humans.

However others believe that is not necessary to literally tap into the brain [85]. Non-implanted methods are based on traditional brain surface electrodes. Unlike the implanted electrodes and wires exposed at the scalp which carry risks of infection and other complications, surface electrodes are more convenient to place and to vary the position and number of electrodes. However for scalp electrodes the acquired data contain many kinds of noise, or artifacts, which are both electrical and biological in nature (as discussed previously in Section 2.3.4). In scalp recording the major challenge is to extract meaningful brain activity from the background signals in the presence of this noise.

2.7 Signal processing techniques used in BCI in the literature

The main task of a BCI is to recognize patterns by interpreting sequences of signals. In a multi-channel BCI, 64 channels or more may be used, then the system has to deal with up to 10000 values per second at a suitable sampling rate. Moreover the raw data may be contaminated by all sorts of artifacts. Therefore it could be very difficult to find useful signals without applying advanced signal processing to the data. The signal processing could reduce the artifacts and improve the signal SNR ratio so that the extracted signal is clear and easily detected by the translation algorithm. The major signal processing methods in the literature include: coherent averaging, filtering, the Fast Fourier Transform (FFT) and blind source separation (BSS).

2.7.1 Coherent averaging

One of the standard ways to minimize the noise and enhance the signal quality in clinical data is through the process of coherent averaging [86]. In the averaging process, it is important to ensure that the time locking of the signal for the averaging process is accurate. When this requirement is met, the signal of interest will be averaged over many epochs. The noise is assumed to be random and different to the signals and so during the process the noise will be averaged out and reduced in amplitude.

Normally ERPs benefit from signal averaging to enhance their SNR [87]. A stimulus or other synchronizing event defines the time epoch of interest within the signal. The event is repeated, and a time-locked signal average (or coherent average) is calculated across trial epochs for each time point of the epoch. Coherent averaging reduces the variance of the noise, while preserving the amplitude of signals that are synchronous with the beginning of the stimulus. If $\mathbf{x}_{j(t)}$ is the electrical potential (voltage) or magnetic field strength at some electrode or sensor location at time t and trial j, the signal average $\bar{\mathbf{x}}_t$ is defined as:

$$\overline{\mathbf{x}}_t = \frac{1}{J}\sum_{j=1}^{J} \mathbf{x}_{j(t)} \tag{2.1}$$

where J is the number of signal trials for the coherent averaging. With coherent averaging, the SNR ratio of a noisy signal can be improved. It is widely accepted that under the conditions of (a) noise stationary, (b) physiological invariability, and (c) no correlation between signal and noise, through coherent averaging the SNR can be increased by a factor of $\sqrt{J}$ [86]. In reality, although these conditions cannot be said to be entirely met, nevertheless coherent averaging results in a useful increase in SNR and furthermore it is simple and easy to apply. With the key assumption of coherent averaging that the signal needs to be invariant across trials. However in the real world the ERP brain activity is not precisely time locked. For example P300 for the same subject could experience a varying time delay over the duration of a trial. Therefore in order to extract the true amplitude of the signal, the data acquisition needs to be long enough such that a good number of trial data is captured enabling a good average to be obtained. In practice, this could slow down the speed of processing or even cause habituation [88] which is a decrease in response to a stimulus after too many repeated presentations. For example, during the habituation the signal amplitude might decrease and the peak latency might increase. Therefore the more averaging that is applied will not necessarily achieve a better extraction result.

In the BCI literature, some work has already used coherent averaging to extract ERP activities [89] [90]. However, in order to achieve good performance studies sometime also require a method, for example filtering, to preprocess the data

.

2.7.2 Filtering

One of the most common signal processing approaches in the time or space domain is to enhance the input signal through a method called filtering. Filtering generally

consists of some transformation of a number of surrounding samples around the current sample of the input or output signal. Filtering generally works by accepting an input signal, blocking prespecified frequency components, and passing the original signal minus unwanted frequency components to the output.

The above filtering functions can be implemented by various types of methods. In the literature the most popular type of techniques are: finite impulse response (FIR) filtering method, infinite impulse response (IIR) method.

(1) FIR filter

A filter that has an impulse response which reaches zero in a finite number of sample intervals is called FIR filter. An N^{th} order FIR filter has a response to an impulse that is N+1 samples in duration. It can be described by the following difference equation which defines how the input signal is related to the output signal.

$$\mathbf{y}[n] = b_0 \mathbf{x}[n] + b_1 \mathbf{x}[n-1] + \cdots + b_N \mathbf{x}[n-N] \tag{2.2}$$

where $\boldsymbol{x}[n]$ is the input signal, $\mathbf{y}[n]$ is the output signal and b_i are the filter coefficients. N is known as the filter order or the filter length. An N^{th}-order filter has (N + 1) terms on the right-hand side. The goal of filter design is to select the filter's length and coefficients such that it achieves the desired filtering functions.

(2) IIR filter

An IIR filter is another type of digital signal filter, in which the output is the weighted sum of the current and past samples of input. The definition of IIR filter is given by the following difference equation.

$$\mathbf{y}[n] = b_0\mathbf{x}[n] + b_1\mathbf{x}[n-1] + \cdots + b_N\mathbf{x}[n-P] - a_1\mathbf{y}[n-1] - a_2\mathbf{y}[n-2] - \cdots - a_Q\mathbf{y}[n-Q] \quad (2.3)$$

where P is the feedforward filter order; b_i are the feedforward filter coefficients; Q is the feedback filter order; a_i are the feedback filter coefficients; $\mathbf{x}[n]$ is the input signal and $\mathbf{y}[n]$ is the output signal. The order of an IIR filter is the largest of P and Q.

Compared to the FIR filter, the IIR filter can achieve a given filtering characteristic using less memory and calculations than a similar FIR filter. However the IIR filter is more susceptible to problems of finite-length arithmetic so that a direct consequence of feedback occurs when the output isn't computed perfectly and is fed back, the imperfections can become compounded.

The overall advantage of using filtering is its simplicity that unwanted frequency components can be easily removed. However this method fails when the neurological phenomenon of interest overlap or lie in the same frequency band [91]. In BCI applications, some noise or artifacts have a frequency range that overlaps the EEG signal. As a result, a simple filtering approach cannot remove, for example, EMG or EOG artifacts without removing any unique neurological phenomenon. More specifically, since EOG artifacts generally consist of low frequency components, using a high-pass filter will remove most of the artifacts. However, for BCI systems that depend on low frequency neurological phenomena (such as ERP activities), these methods are not desirable, since these neurological phenomena may lie in the same

frequency range as that of the EOG artifacts. In the case of removing EMG artifacts from EEG signals, filtering specific frequency bands of the EEG can be used to reduce the EMG activity. Since artifacts generated by EMG activity generally consist of high-frequency components (most EMG exists in a frequency range between 20 and 200 Hz), using a low-pass filter may remove most of these artifacts. But they cannot be effective for BCI systems that use a neurological phenomenon with high-frequency content (such as μ, β rhythms).

2.7.3 AR model

An autoregressive (AR) model is an IIR filter with some additional interpretation placed on it. It is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous inputs or outputs. Specifically a model which depends only on the previous outputs of the system is called an AR model, while a model which depends only on the inputs to the system is called a moving average model (MA), and a model based on both inputs and outputs is an autoregressive-moving-average model (ARMA).

The notation of P^{th} order AR model can be written:

$$\mathbf{y}[n] = \sum_{i=1}^{P} \varphi_i \mathbf{y}[n-i] \tag{2.4}$$

where φ_i are the parameters of the AR model and $\mathbf{y}[n]$ is the output signal.

The notation of Q^{th} order MA model can be written:

$$\mathbf{y}[n] = \sum_{j=0}^{Q} \theta_j \mathbf{x}[n-j] \tag{2.5}$$

where the θ_i are the parameters of the MA model and $\mathbf{x}[n]$ is the input signal.

The notation P+Q order ARMA refers to the model with P AR terms and Q MA terms. This model contains the AR and MA models:

$$\mathbf{y}[n] = \sum_{i=1}^{P} \varphi_i \mathbf{y}[n-i] + \sum_{j=0}^{Q} \theta_j \mathbf{x}[n-j] \tag{2.6}$$

AR models have been applied to EEG signals over many years now. The first applications of AR modeling on EEG are described in [92] [93], where the models were used to simulate artifacts. Other EEG applications include spectral analysis [94], segmentation [95] [96] and feature extraction [97]. In order to use AR modeling methods, a careful selected model order becomes a major issue. Normally the order selection can be achieved manually based on the prior knowledge of the application or automatically by techniques such as the Akaike Information Criterion [98] and Bayesian information criterion [99]. However, in general these automatic order selection criteria are of limited use since sometimes inappropriate model orders are estimated [100].

In BCI applications, AR modeling can be applied to remove EOG artifacts from EEG signals [101]. The process starts to calculate the AR parameters for EOG signals in a training data set so that the specific character of EOG for a particular subject can be caught. Those parameters then are used to separate the EOG artifacts in the subsequent data and the separated EOG signals are subtracted from the original EEG data to generate the non-contaminated EEG. One problem in using the AR model is that the EOG signal to be subtracted from the EEG signal is also contaminated with the EEG signal, so subtracting the EOG signal may also remove part of the EEG signal.

The reduction of EMG artifacts through AR modeling methods is more difficult since EMG has a more random presentation in time which means there is no reference to assist AR in calculating the appropriate parameters [102].

2.7.4 FFT

Sometimes, in order to analyse the data signals for their frequency content, data have to be converted from the time domain into the frequency domain. This is accomplished by the Fourier transform (FT) [103] [104]. The discrete Fourier transform (DFT) [105] is one of the specific forms of FT. An FFT allows the DFT to be obtained rapidly and efficiently [106] [107]. The definition of a DFT can be written as the following:

$$Z[k] = \sum_{n=0}^{N-1} z[n] e^{-\frac{2\pi i}{N} nk} \quad k = 0, \ \ldots, \ N\text{-}1 \tag{2.7}$$

where $z[n]$ are input signals, N is the length of data. In the case of N-sample signal, the total number of steps in the computation of the DFT is thus N^2. A FFT algorithm (for example Radix-2 algorithm [108]) requires the number of samples in the signal be a power of 2. The computing time for the radix-2 FFT is proportional to $N(\log_2 N)$. So for example a transform on 1024 points using the FFT runs about 100 times faster than using the DFT. In addition, the term of window is to generate a smoother and more reliable estimate of the distribution of power. This is achieved by averaging neighbouring periodogram intensities across frequencies.

The process of determining the amplitudes of frequency components of a signal is called the spectrum analysis [109]. The spectrum analysis separates the signal into its

specific sinusoidal and cosine waveforms. The resultant frequencies can be used to calculate signal power spectrum which shows a distribution of power values as a function of frequency. In the frequency domain, the power is the square of FFT's amplitude.

In EEG applications, the spectral analysis separates the relative contribution of the different frequencies in the signal. The analysis is able to reduce the noise which is achieved by selecting and discarding the noise's troublesome frequency component(s) from the signal [110] [111], then the result can be represented either as the power or as the energy within the particular frequency band. Moreover in BCI applications as the rhythmic activities (for example μ, β rhythms) are generally the components of interest, spectral analysis based on FFT is able to straightforward compute the power spectrum of components [112]. However this task becomes more difficult in the case that the frequencies of noise and components of interest are overlapped and fall in the same frequency band [113].

2.7.5 BSS

Multi-channel EEG signals represent a mixture of a large number of individual brain sources. This is due to two reasons. First, one single action potential cannot be recorded in isolation from the scalp, so any potential change contained in the EEG recording is the effect of thousands of neurons firing simultaneously. Second, and more importantly, the recordings from each electrode are influenced by multiple (perhaps independent) sources in the brain, e.g. activity in the auditory cortex not only affects the electrodes directly above this brain area, but all other recording sites as well. Moreover, severe contamination of EEG activity by EOG, EMG, heart and line noise lead to a more serious problem for EEG interpretation and analysis.

Source separation problems in signal processing are those in which several signals have been mixed together and the objective is to find out what the original signals were. BSS is a technique to separate a set of source signals from a set of mixed signals, without the aid of information about the source signals or the mixing process [114]. BSS relies on the assumption that, for example, the signals may be mutually statistically independent or decorrelated [115]. It thus separates a set of mixed signals into a set of underling source signals.

One of the key methods of BSS in recent years – Independent Component Analysis [116] – has the ability to extract the relevant information buried within noisy signals and allows the separation of measured signals into their fundamental underlying *independent* components (ICs). It has been widely applied to remove artifacts from EEG signals [117] [118] [119]. Studies in BCI applications have also shown that ICA is able to extract the signal based on morphology (in the time domain) [120] as well as the signal based on the spectrum (rhythmic activities) [121].

Techniques such as coherent averaging, filtering and AR modelling, etc. can only process the data on the observed mixture domain which mixes sources of interest along with aritfacts. Such a limitation ensures that the extracted EEG activities possibly still contain the noise residue or the removed noise contains embedded within it part of EEG signals [122]. Unlike those techniques, BSS has the ability to separate the mixture of components of interest, artifacts and other components into their fundamental sources. By this means, ICA can extract the components of interest and reduce the noise at the same time. It is then possible to apply further processing only to a limited number of those isolated signals. This certainly promises the ability to achieve a better extraction performance than other traditional methods [123] [124].

However it is worth noting that in the BCI literature ICA has been mainly used as a preprocessing step on the data [125] [126] – using ICA just to remove the artifacts, so not many studies focus on its applications beyond artefact removal. This thesis

explores the further functional ability of the ICA for the applications to BCI and extends BSS concepts by developing more advanced ICA techniques. The goal of applying BSS is to isolate the independent source signals so that the follow-on processes are able to achieve a much better classification performance even by using simple and easy techniques. Various methods for accomplishing this task will be discussed in more detail in Chapter 3.

2.8 General discussion on the use of BCI

In Summary, it has been shown that the EEG is a recording of the electrical activity in the brain, which is generally recorded at the scalp. It provides information pertinent to the diagnosis of a number of brain disorders and is now widely used in BCI. At present, the main aim in the development of BCI systems is to enable people with severe neuromuscular disabilities to drive computers directly by their brain activity rather than by physical means. An EEG based BCI system links the brain and a computer or other electrical devices together by using the ongoing EEG. By these means it helps the user to communicate with his/her environment and to provide them with additional assistance for a better quality of life.

The circadian rhythm [127] is an approximately 24-hour cycle in the biochemical, physiological or behavioural processes. More specifically, humans have a wake-sleep cycle which is associated with different levels of alertness (i.e., the state of paying close and continuous attention). Generally speaking, the alertness increases during the daytime and drops during the night. Moreover in term of quantitative measures, in the EEG recordings, the power in the frequency bands of brain activities can be observed to cycle based on this circadian rhythm.

In the BCI literature, there is a lack of studies on the impact on EEG based BCI diurnal variations since no studies have actually involved running a BCI system for 24 hours. However it is known that the key of a good BCI system, requires a high level of attention from users. That is to say, different levels of alertness in a wake-sleep cycle may affect the BCI performance. For example, because of the lack of alertness during the evening/night, users may even not be able to effectively concentrate their attention on the BCI tasks. Moreover if BCI runs over a long period and the brain activities used are related to rhythmic frequency band, since the spectral information varies over time then it may be that a re-calibration of the BCI classification criteria during the task will need to take place (for example, to update the threshold value which determines the classification decision point). From the BCI literature, since BCI experiments run over a relatively short period of time and the assumption is made that the alertness level and the spectral information remain the same during this time, therefore the performance of BCIs in this context should be stable.

BCI development, in order to succeed, requires a close interdisciplinary collaboration amongst the engineering, neuroscience, psychology, computer science and rehabilitation communities. The research needs to discern the nature of electrophysiological phenomena through psychological processes, seek appropriate BCI paradigms using various control patterns and develop processing methods to efficiently translate control patterns into final outputs. Further development of BCI technology will depend on basic neuroscientific research as well as applied signal processing techniques. However the essential neuroscientific and psychological foundations of the field are not sufficiently developed. As a result, most current BCI's have been developed primarily by engineers and other technically oriented groups with limited expertise in behavioural principles and methods. High-level intellectual and cognitive functional patterns are not able to benefit from the current BCI usage as yet.

In the BCI literature, five different EEG activities form five types of BCI paradigms. Current BCI systems are able to be used in basic environmental control, for example, a binary choice system for answering yes/no questions, and word spelling. They can even be used to operate assistance devices, for example, a wheelchair. Due to the different fundamental signals that are used as the control feature, and due to the fact that each are used in different applications, there is not a unique quantity that could be reliably used to measure the performance of BCI systems. However, a fair comparison in several aspects, such as: speed and accuracy, training time, actual trial length and experiment accessibility, can help researchers select the most suitable method for specific applications. The key measure of BCI performance depends on speed and accuracy. For example, VEP based BCIs could operate a word processing program at a high rate of 10~12 words/min with a maximum accuracy up to 90% [50] [51]. SCP based word spelling BCIs can write about 0.15~3 letters/min with an accuracy of 65~90% [55]. P300 based word spelling BCI can type up to 5 letters/min with about 70~90% accuracy [48]. Spontaneous activity based BCI performing a binary selection (for example, yes/no questions) can achieve 10-15 bits/min with a maximum accuracy up to 95% [68]. A cortical neuron activity based speller can reach 15 words/min [82].

However, another important measure to take into account is how long a training time is required. Basically VEPs and P300s are such natural brain activities that they therefore normally do not involve a training session. SCPs are, as their name suggests, very slow. SCP based BCIs require a long training time (several months in fact) so that users can learn to generate SCP activities. Spontaneous activity BCIs minimally need a few hours as a training session. Cortical neuron activity BCIs involve a specially designed training session so that users can learn how to control particular neuronal firing for specific applications.

The actual trial length of EEG recordings for BCI systems is also an important factor; that the longer such a segment is, the more time is needed to complete its processing. For example, VEP based BCIs detect signals of interest in the frequency domain and a trial lasting around 2.0 – 4.0s is able to provide sufficient spectral information for the processing. The duration of natural SCP potentials lasts 0.5 – 10.0s which indicates the actual trial length of such BCI systems needs to be long enough to capture the potentials. In theory P300 waveforms peak at 300ms, the length of these trials should thus be long enough to cover this P300 peak. The actual word speller paradigm [48] in P300 based BCI applies a faster stimulus representation so that 15 repetitions of a stimulus only needs 2.5 – 3.5s. Spontaneous activity based BCIs also process the signal in the frequency space. The actual trial length is around 4.0 – 6.0s. The trial length of cortical neuron activity based BCIs is short, around several tens to hundreds of millisecond.

Experimental accessibility can determine how easily users can access the BCI systems. Generally speaking, the preparation for VEP, SCP, P300 and spontaneous activity based BCIs are similar. They all use scalp EEG recordings, commonly using scalp electrodes following the 10/20 placement system, a computer program based instruction and stimulus presentation and output devices, such as a computer and wheelchair etc. Moreover the electrodes are easy to don/ remove. Cortical neuron activity based BCI requires extensive surgery to implant electrodes into the cortex. The whole process needs to be done by professionals and is very expensive and may have potential risks of brain infection and other such injury. Because of these reasons, cortical neuron activity based BCIs cannot be broadly applied at this stage.

From the above factors, one paradigm suitable for all application and all possible BCI users does NOT exist. Matching the user's need is essential so that BCIs can be broadly applied to assist users with different disabilities. From four existing BCI methodologies based on non-implanted recordings, if the text based communication is

the user's need, then VEP and P300 based BCIs appear the best choice based on a summary of their performance factors. If users want to control the computer cursor to select/click buttons on screen or even drive a wheelchair, then spontaneous activity based BCIs are most suitable. A BCI system that can combine various activities and can achieve different functions would provide more flexible and practical option for different users and tasks.

However from BCI literature, the signal processing for BCI is made more difficult by the fact that EEG recording is easily affected by artifacts. The BCI study in this thesis is undertaken with an aim to develop and examine advanced BCI techniques, mainly in the signal processing field. In traditional BCI applications, signal processing techniques such as coherent averaging, filtering, AR modelling and FFT were used to reduce the artifacts and improve signal SNR. However due to their functional limitation, those techniques do not extract the components of interest well. Therefore a number of additional processing steps have to be used to solve such problems.

A simple step is to apply a number of reinforcements in the data recording stage. The idea of applying the reinforcement ensures that enough temporal signal information gathering for the further processing, for example, for the use of the averaging and the majority vote technique. However, as a result, the experimental processing time has to be prolonged. Another step is to increase the number of recording electrodes so that the enough multi-channel signals are collected to find the best spatial electrode positions, which also will make the experiment time unnecessarily longer by electrode placement and processing very large data input.

As the BSS, especially ICA, separation technique has the unique ability to isolate the mixed signals into their underlying sources which can maximally extract the components of interest *and* reduce the artefacts, it has been widely applied to many

biomedical applications as well as to BCI systems. However in most study cases, ICA acts as a preprocessing step to clear the signals, then a number of further complex techniques take on the role to deal with feature extraction and final classification problems. This thesis proposes to apply and develop advanced ICA techniques so as to exert the advantages of ICA to this field. Therefore ICA plays the role of a combination of the signal preprocessing and feature extraction steps and at the same time simplifies the complexity of the classification process so that even a simple linear classifier can achieve a very good performance.

The proposed ICA techniques mainly work towards optimising two major issues: reducing the amount of reinforcement needed (i.e. reducing the recording time needed) and reducing the number of recording channels so that the total experimental time can be reduced. Moreover this work is in the context of increasing classification accuracy or at least with it remaining the same as the current final classification accuracy. The proposed techniques are demonstrated on two popular BCI systems in three aspects, multi-channel ICA algorithms for ERP based BCI; multi-channel ICA algorithms for spontaneous EEG based BCI; and, single channel ICA algorithms for ERP based BCI. As the time scale for this BCI study is limited, there is not a schedule to explore an application of single channel ICA on spontaneous EEG based BCI.

The next chapter introduces the standard and goes in some depth into the ICA concept, history, different types of implementations and applications. It also explores at a higher level, more advanced ICA techniques.

CHAPTER 3

Independent Component Analysis

3.1 Introduction

Blind source separation (BSS) refers to the problem of recovering signals from several observed linear blind mixtures. The attraction of the BSS model in signal processing is that only mutual statistical independence between the source signals is assumed and no *a priori* information: such as the characteristics of the source signals, the mixing matrix or the arrangement of the sensors is needed. Several simultaneously active signal sources at different spatial locations can then be separated by exploiting mutual independence of the sources. Nowadays, BSS methods such as Independent Component Analysis (ICA) are increasingly being used in biomedical signal processing and analysis.

This chapter first gives an introduction about the background to ICA, ICA theories followed by a detailed explanation of popular ICA algorithms which show the various architectures to achieve the goal of source separation. The chapter also introduces major processing steps to implement ICA algorithms. Moreover several advanced ICA algorithms with constraints and novel concepts of ICA applications on single channel or a few channels are highlighted and reviewed. Finally the chapter ends with

a summary of the existing ICA applications in the literature and a summary of this chapter.

3.2 ICA background

In the biomedical signal processing field, a major task is to extract information of interest from a set of observed measurements. The recorded biomedical signals, especially brain electromagnetic (EM) signals, contain a finite set of activities which are overlapped both in space and in time. These activities are generated by the body or are artifactual in nature. So basically the information of interest is seldom recorded in isolation but generally mixed with other background activities, for example artifacts from physiological and environmental origins, and the recorded data are nearly always contaminated by such 'background' noise. For this reason, the SNR of the desired signal is generally quite poor. From the point of view of the signal processing, it would be ideal to unmix and separate the sources of interest from a set of noisy biomedical signal measurements.

Within the above context, ICA, one type decomposition technique, has the ability to correct or remove signal contaminates. ICA is a statistical and computational technique for revealing hidden sources/components that underlie sets of random variables, measurements, or signals [128]. This technique attempts to unmix the observed signals into some number of underlying components and usually allow remixing those sources that would result in a clear version of the measured data.

An early ICA algorithm which made a linear source separation possible was proposed by Jeanny Herault & Christian Jutten in 1986 [129] and then the phrase 'independent component analysis' first appeared in their publication [130] in 1991, Further work to develop the ICA technique and for the first time explain it from the statistical point of view was written by Pierre Comon in his work [131]. There are a

large number of papers in the literature utilizing ICA for many applications. More details about the applications appear later in this chapter. A common application of the ICA separation problem is to solve the so called "cocktail party" problem where a number of people talking simultaneously are recorded in a party. The objective of this problem is to find out what the original individuals' voice signals were. Although this is a common example of ICA, it is actually a challenging problem due to the non-instantaneous mixing of the sources.

In order to use ICA algorithms successfully there are a few strong general assumptions that must be made about the sources themselves and the source mixing conditions before these can be applied to the measured data and any proper sense made of the results. A more common assumption of the ICA separation problem is made that the unknown underlying sources, which are independent of each other are linearly combined to form a mixed signal. When this independence assumption is correct, ICA can separate these mixed signals and returns the independent components (ICs).

For example, source signal vectors at time instant t $\mathbf{s}(t)=[s_1(t), s_2(t), \ldots, s_n(t)]^T$ are mixed and formed as vectors $\mathbf{x}(t)=[x_1(t), x_2(t), \ldots, x_m(t)]^T$ which are the observed (or mixed) signals, and the mixing matrix $\mathbf{A}$ with the dimension $m \times n$ describes the linear combination of sources $\boldsymbol{s}$:

$$\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t), \tag{3.1}$$

m stands for the number of mixed signals, n indicates the number of sources and both of them have the same number of samples. Generally, it is assumed that the number of sources is less than or equal to the number of measured data channels ($n \leq m$).

For example, we have four independent sources on the left in Figure 3.1, and these get mixed by a random mixing matrix **A**. The signals on the right in Figure 3.1 are the observed (mixed) signals.

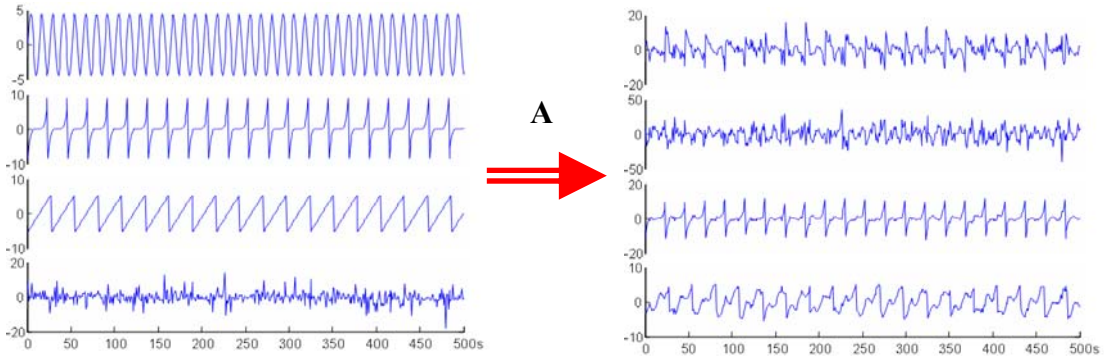

Figure 3.1: *Four independent sources are mixed by a random mixing matrix. The sources shown on the left and observed signals on the right.*

The task of ICA is to recover the original sources $\mathbf{s}(t)$ from the mixed signals $\mathbf{x}(t)$ through finding an unmixing matrix **W**, and obtain the independent component simply by:

$$\hat{\mathbf{s}}(t)=\mathbf{W}\mathbf{x}(t), \tag{3.2}$$

Where unmixing matrix **W** is equal to the inverse of the mixing matrix **A**, and $\hat{\mathbf{s}}(t)$ are the resulting estimates of the underlying sources $\mathbf{s}(t)$. Figure 3.2 shows the process to find the estimated sources by ICA from the observed signal which is mixed in Figure 3.1.

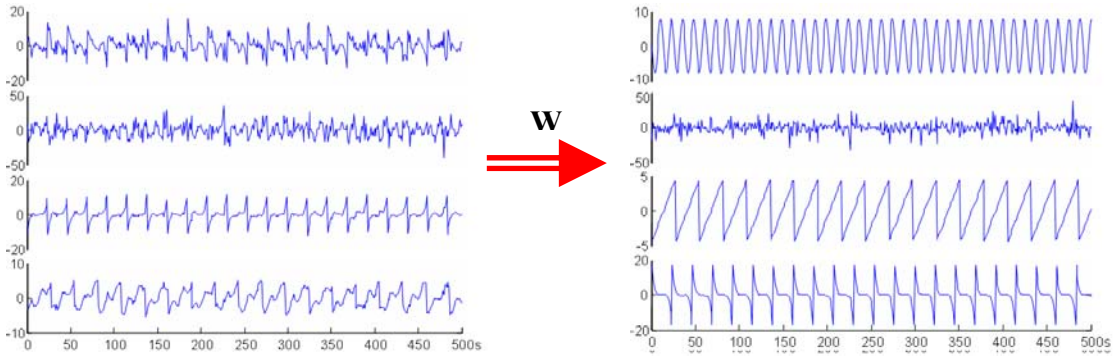

Figure 3.2: *The mixed signals (on the left) are recovered by ICA into the independent sources (right).*

3.2.1 ICA generative models

In actual fact equation 3.1 is a basic and simplified mixing model. A more general mathematical model which has no assumptions on the data signal (other than instantaneously mixing) can be summarized as the following,

$$\mathbf{x}(t)=\mathbf{f}\{\mathbf{s}(t)\} + \mathbf{n}(t) \tag{3.3}$$

where **f** indicates any unknown function and **n**(t) is the additive random noise vector. Now in equation 3.3 the ICA problem becomes, to obtain an unmixing matrix by inverting **f** and to map the observed signal to the source space without knowing information neither of **f**, **s** or **n**.

Based on assumptions of linear mixing, ICA can fall into two apparent models: linear ICA and nonlinear ICA.

3.2.2 Linear noisy ICA

The assumption of linear mixing of source data simplifies equation 3.3 to

$$\mathbf{x}(t)=\mathbf{A}\mathrm{s}(t)+\mathbf{n}(t), \tag{3.4}$$

where **A** is the linear mixing matrix as the same as in equation 3.1. The linear ICA technique is simple and works efficiently in linear mixing environments. In brain EM signal processing applications, a reasonable assumption can be made that the underlying sources are mixed by using an instantaneous linear superposition of them at the measurement channels. Data signals based on such an assumption are, for example, fMRI and EM brain signals.

3.2.3 Linear noiseless ICA

A linear noiseless ICA is based on the assumption that the mixed observation data signals $\mathbf{x}(t)$ are noiseless or the noise $\mathbf{n}(t)$ is small enough to be ignored, then equation 3.4 reduced to equation 3.1.

$$\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t), \tag{3.1}$$

Obviously this assumption reduces the complexity of the mixing model, but it also makes the mixing probably less realistic. However this noise-free model may be thus considered a tractable approximation of the more realistic noisy model since the noise here during the data recording can be regarded a source of its own in nature. In this regard, this approximation of using the simpler model seems to work.

3.2.4 Nonlinear ICA

In a nonlinear ICA model, the assumption is made that the mixing of the sources does not need to be linear. The nonlinear mixing theory is proposed due to the basic linear model often being too simple for describing the observed data adequately. In this aspect, nonlinear ICA would be suitable for the applications of nonlinear mixtures. However in general this nonlinear ICA is often intractable and difficult to apply, since the indeterminacies in the separating solutions are much more severe than in the linear case. [132] [133] [134] [135]. Therefore the separation of the nonlinear ICA problem is usually highly non unique and generally the separating processing requires addition prior information.

This study mainly concentrates on the linear noiseless ICA model for the reasons that the recording data are EM brain signals (especially EEG signals) which are assumed to be instantaneously and linearly mixed by underlying brain sources; the noise added to the observation can be assumed to special underlying “sources” participating in the mixture; most outputs of existing ICA research are based on this simplified ICA model and work well for certain linear mixed real data.

3.2.5 ICA for convolutive mixing problems

Many ICA algorithms have been proposed to solve BSS supposing that the observations are instantaneous mixtures of the sources. However sometimes the mixing is known as convolutive mixing. This is normally due to the fact that the source signals have different time delays in each observed signal due to the finite propagation speed in the medium. Moreover, each observed signal may contain time-delayed versions of the same source due to multipath propagation caused typically by reverberation. That is to say observations are the convoluted mixtures of the sources and not instantaneous mixtures.

The convoluted mixing problem can be solved by extending ICA algorithms developed for instantaneous mixtures. Basically the convolutive BSS problem can be solved by two types of approaches. One approach is to work in the time domain [136], however working in the time domain has the disadvantage of being rather computationally expensive, because of the need to calculate many convolutions. More commonly the other approach works in the frequency domain and transform the convolution in the time-domain into multiplication in the frequency domain and then to apply ICA methods for instantaneous mixtures [137], [138], [139]. Some applications using the convoluted mixing model for particular areas (such as the enhancement of mixed speech signals) have been published elsewhere [140], [141].

3.3 ICA estimation principles

In order to estimate statistical independence, some basic definitions and terminology are needed to introduce a more formal understanding of ICA estimation.

3.3.1 Cumulative distribution function

In statistics, the cumulative distribution function (CDF), also known as the probability distribution function or just distribution function [142], represents the probability distribution of a real-valued random variable. Probability distributions are typically defined in terms of the probability density function (PDF). For a continuous random variable, a PDF represents a probability distribution. The definition of the probability that the variable x is less than or equal to $x0$ is:

$$P(x) = p(x \le x0) = \int_{x=-\infty}^{x0} p_x(x)dx\,, \tag{3.5}$$

where $P(x)$ represents the CDF of a given random variable x, and $p_x(x)$ is the PDF of x. For example, Figure 3.3 shows that the calculation of $P(x)$ is also equivalent to finding the shaded area underneath the PDF curve.

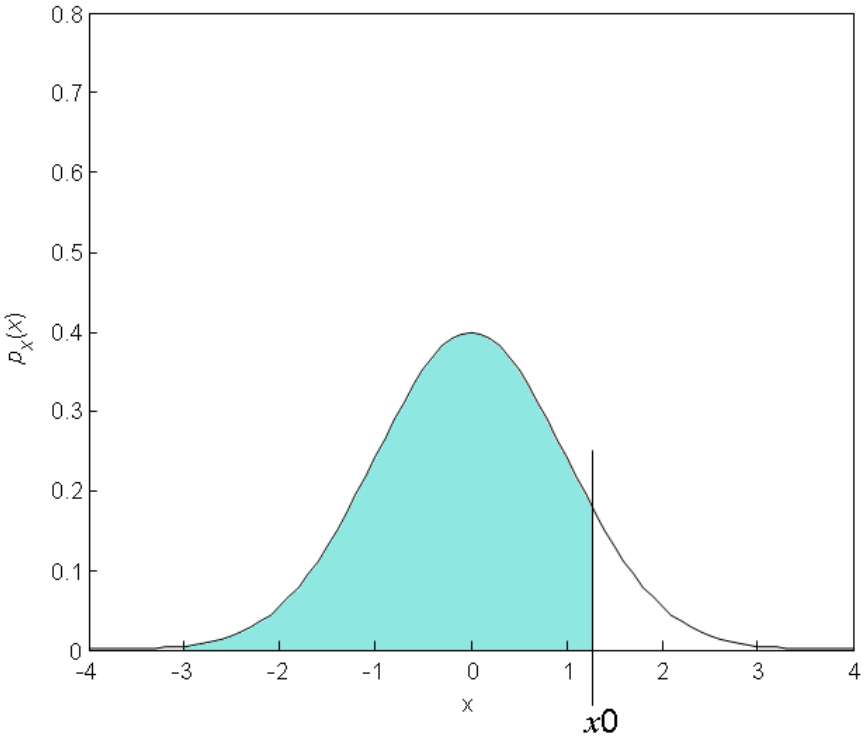

Figure 3:3 *A PDF function with a Gaussian distribution. The calculation of P(x) is equivalent to finding the shaded area underneath the PDF curve.*

3.3.2 Moments

The concept of moments in mathematics initially comes from physics [143]. A general definition in mathematics for n-th moment of a real variable is

$$\mathrm{E}[x^n] = \int_{-\infty}^{+\infty} x^n p_x(x) dx, \tag{3.6}$$

where E indicates the expectation operator, $p_x(x)$ is the PDF. In statistics some lower order moments represent the measurement of particular properties.

(1) First moment

The first moment corresponds to the mean value (normally represented by μ) of a random variable x which is also known as the expected value of variable x. It can be defined as:

$$\mathrm{E}[x] = \int_{-\infty}^{+\infty} x p_x(x) dx \,, \tag{3.7}$$

The mean determines the centre of probability distribution of variable x.

In some cases it is convenient to consider an alternative form of moment: the central moment. In general, the n-th central moment of a variable can be written as:

$$\mathrm{E}[(x-\mu)^n] = \int_{-\infty}^{+\infty} (x-\mu)^n p_x(x) dx \,, \tag{3.8}$$

(2) Second moment

The second moment $\mathrm{E}[x^2]$ of variable x can be formed as,

$$\mathrm{E}[x^2] = \int_{-\infty}^{+\infty} x^2 p_x(x) dx \,, \tag{3.9}$$

The above can also be shown as

$$\mathrm{E}[x^2] = \mathrm{E}[x]^2 + \mathrm{E}[(x-\mu)^2] \,, \tag{3.10}$$

where $\mathrm{E}[(x-\mu)^2]$ is the variance of the variable x (it is normally represented as σ^2). Variance is a measure of the averaged squared distance of its possible values from the mean. For example, assume there is real-valued random variable x which has a Gaussian/normal distribution with mean μ =0 and variance σ^2 =1. Figure 3.4 visually indicates the mean and variance of the given distribution.

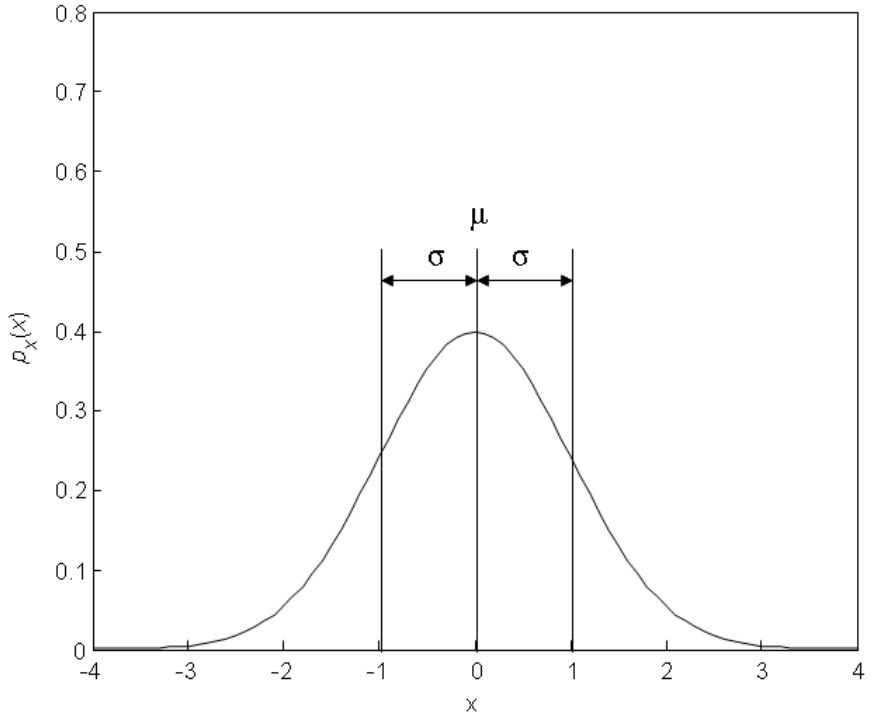

Figure 3.4: *An illustration of mean and variance of a given Gaussian/normal distribution.*

(3) Third moment

The third moment $\mathrm{E}[x^3]$ of variable x can be defined as

$$\mathrm{E}[x^3] = \int_{-\infty}^{+\infty} x^3 p_x(x) dx, \tag{3.11}$$

The third central moment of variable x is known as the skewness of x. Skewness is a measure of the "lopsidedness" of the distribution.

(4) Fourth moment

The fourth moment $E[x^4]$ of variable x can be defined as

$$E[x^4] = \int_{-\infty}^{+\infty} x^4 p_x(x) dx, \tag{3.12}$$

The fourth moment is commonly expressed through its Kurtosis [144], which for a zero mean process is defined as:

$$\text{Kurt(x)} = E[x^4] - 3E[x^2]^2, \tag{3.13}$$

Kurtosis is a measure of whether the distribution is "tall and skinny" or "short and squat" of the probability distribution of variable (more details given in Section 3.5.1). For example, Kurtosis is zero for a Gaussian random variable. Kurtosis can also be positive or negative. A random variable with a super-Gaussian distribution has a 'spiky' PDF and a positive Kurtosis. Whilst a random variable with a sub-Gaussian distribution has a 'flat' shaped PDF, then Kurtosis is negative. Figure 3.5 illustrates examples of Gaussian, super-Gaussian and sub-Gaussian distributions.

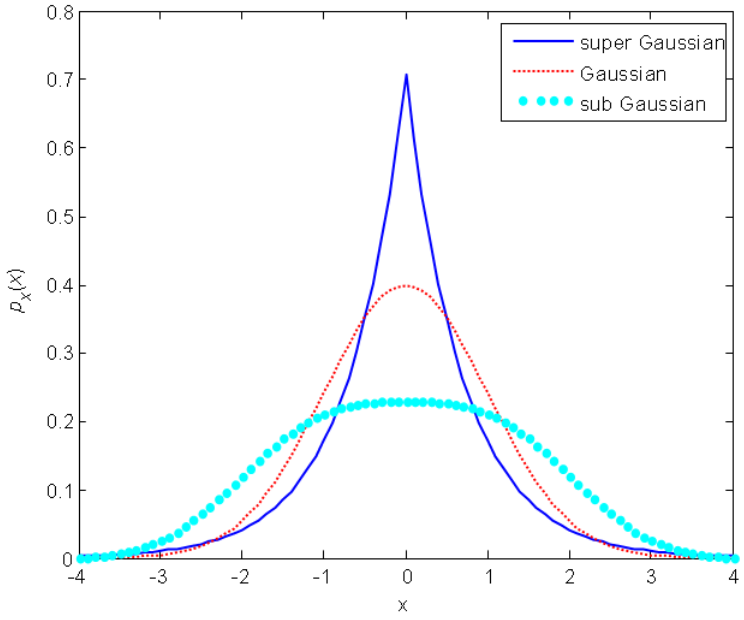

Figure 3.5: *An illustration of the shape of* Gaussian, super-Gaussian and sub-Gaussian distributions.

3.3.3 Cumulants

The terms, cumulants [145] of a distribution are closely related to the moments of the distribution. For $n = 1$, the n-th cumulant is just the expected value or mean; for $n =$ either 2 or 3, the n-th cumulant is just the n-th central moment; for $n \geq 4$, the nth cumulant is an n-th-degree monic polynomial in the first n moments (about zero), and is also a (simpler) n-th-degree polynomial in the first n central moments. For example Kurtosis is the fourth-order cumulant.

3.3.4 Independence and correlation

Correlation [146] is one of the most common statistics which describes whether and how pairs of variables are related. Correlation is often measured as a correlation coefficient

$$\rho_{x,y} = \frac{\mathrm{E}[x-\mu_x]\mathrm{E}[y-\mu_y]}{\sqrt{\mathrm{E}[(x-\mu_x)^2]}\sqrt{\mathrm{E}[(y-\mu_y)^2]}}, \tag{3.14}$$

where $\rho_{x,y}$ indicates the correlation coefficient between two random variables x and y. μ_x and μ_y represent the mean values for x and y.

From the above definition, when the correlation coefficient is close to 1, this indicates a strong correlation between this pair of variables, -1 means a strong negative correlation and values in between denote a certain degree of linear relationship between the variables. If the correlation coefficient is close to 0, then it indicates no correlation (uncorrelatedness) or a weak correlation between pair of variables.

If the variables are independent then the correlation is 0, but the converse is not true. For independence, two random variables x and y must meet the condition that if and only if the following exists

$$\mathrm{E}[x,y] = \mathrm{E}[x]\mathrm{E}[y], \tag{3.15}$$

then the two are independent. As equivalent, if x and y have PDF $p_x(x)$ and $p_y(y)$, then independence between x and y is equivalent to [147]

$$p_{x,y}(x,y) = p_x(x)\ p_y(y), \tag{3.16}$$

where $p_{x,y}(x,y)$ is the joint probability density function of x and y.

Intuitively independence is in general a much stronger requirement than uncorrelatedness: it is interpreted as a stricter condition than the condition of uncorrelatedness.

3.3.5 Obtaining a measure of independence

The key assumption for the ICA to be successful is that the source signals need to be non-Gaussian. ICA uses non-Gaussianity to measure the independence of the signals. Given a set of mixtures, ICA finds each source signal by finding that unmixing vector which extracts the most non-Gaussian source signal for the set of mixtures. Therefore the measurement of non-Gaussianity is defined as a separation strategy of the signals. For example, considering a simple example, assume there are two random sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ with uniform distributions (non-Gaussian) and that they are mixed linearly to form observed signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ by the following mixing matrix

$$\mathbf{A}_{\mathbf{mix}} = \begin{pmatrix} 0.5 & 0.8 \\ 0.4 & 0.2 \end{pmatrix},$$

By plotting the amplitude of one signal at each time point against the corresponding amplitude of the other signal, Figure 3.6 shows the joint distribution of the two random sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ with uniform distribution. Figure 3.7 illustrates the joint distribution of the mixed signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. To make ICA work, a first step called whitening (more details in section 3.4.2) is applied as a preprocessing to the data. The whitening process is simply a linear change of coordinates of the mixed data. This means that any correlations are removed in the data. For the same example, Figure 3.8 shows the effect of whitening on the mixed data. The square as seen in the graph indicates that the distribution of the mixed data after whitening clearly is a rotated version of the original square. After whitening, ICA then further rotates the whitened signals to try to identify the original measurements by finding the unmixing matrix that maximizes the non-Gaussianity. Obviously the assumption of non-Gaussian sources is an important factor in the ICA estimation. This begs the question: *what if the sources all have Gaussian distributions?* Assume that two random sources $\boldsymbol{s}_1$ and

$\boldsymbol{s}_2$ with uniform distribution are linearly mixed by the above same mixing matrix. Figure 3.9 shows the joint distribution of the two random sources $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ with Gaussian distributions and Figure 3.10 presents the joint distribution of the mixed signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. After the whitening step, the distribution shown in Figure 3.11 appears nearly symmetric. This indicates that there is insufficient directional information to help ICA rotate the signal back to the original measurement. Therefore if there is more than one Gaussian source underlying the measured/mixed data, then ICA can not separate the signals into their underlying sources. In the case of only one Gaussian source in the signals, after the non-Gaussian sources are extracted, the residual is the Gaussian source. That is to say ICA still works as long as there is only one source with a Gaussian distribution.

To use the non-Gaussianity as a metric, the ICA model needs to define a quantitative measure of non-Gaussianity of a random variable, for example, Kurtosis and Negentropy (more details in Section 3.5.2).

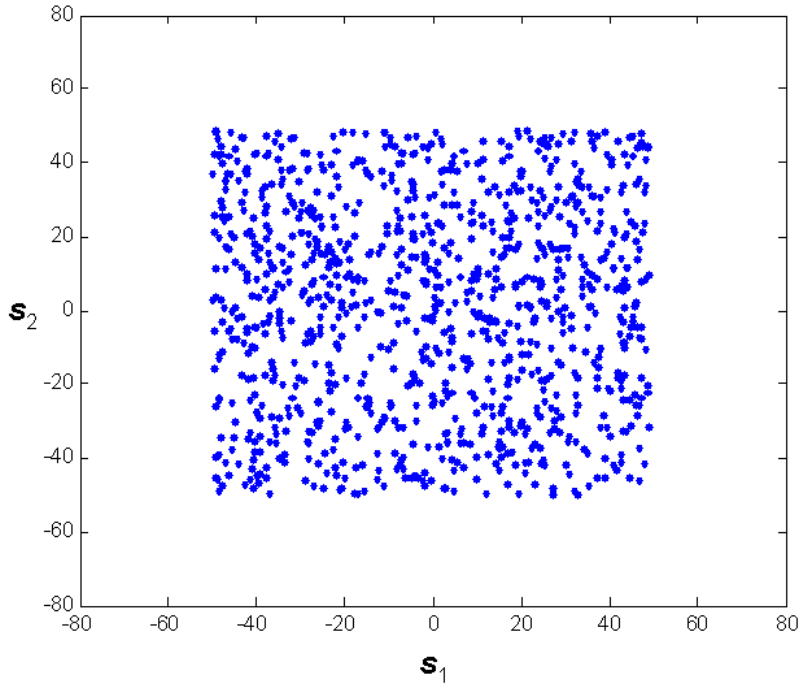

Figure 3.6: *The joint distribution (variables) of sources* $\mathbf{s}_1$ *and* $\mathbf{s}_2$ *with uniform distributions.* $\mathbf{s}_1$ *is in the horizontal axis and* $\mathbf{s}_2$ *is in the vertical axis.*

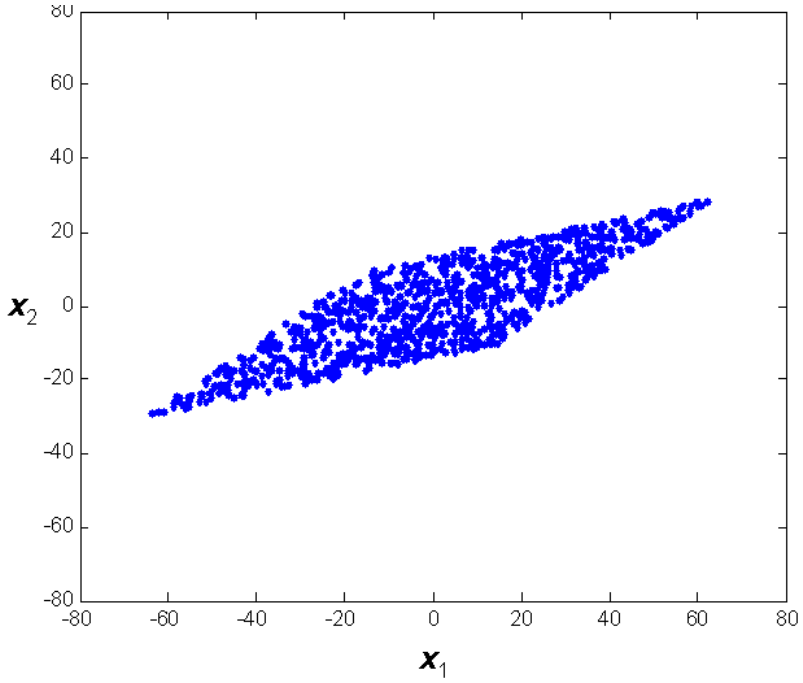

Figure 3.7: *The joint distribution of these two mixed signals,* $\mathbf{x}_1$ *is in the horizontal axis and* $\mathbf{x}_2$ *in the vertical axis.*

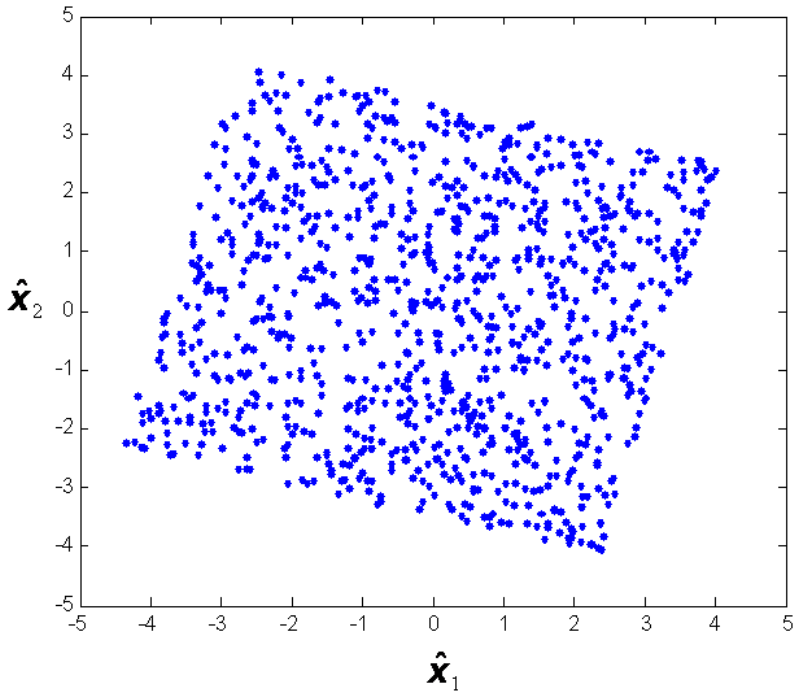

Figure 3.8: *The joint distribution of the two whitened data mixtures.* $\hat{x}_1$ *is in the horizontal axis and* $\hat{x}_2$ *in the vertical axis.*

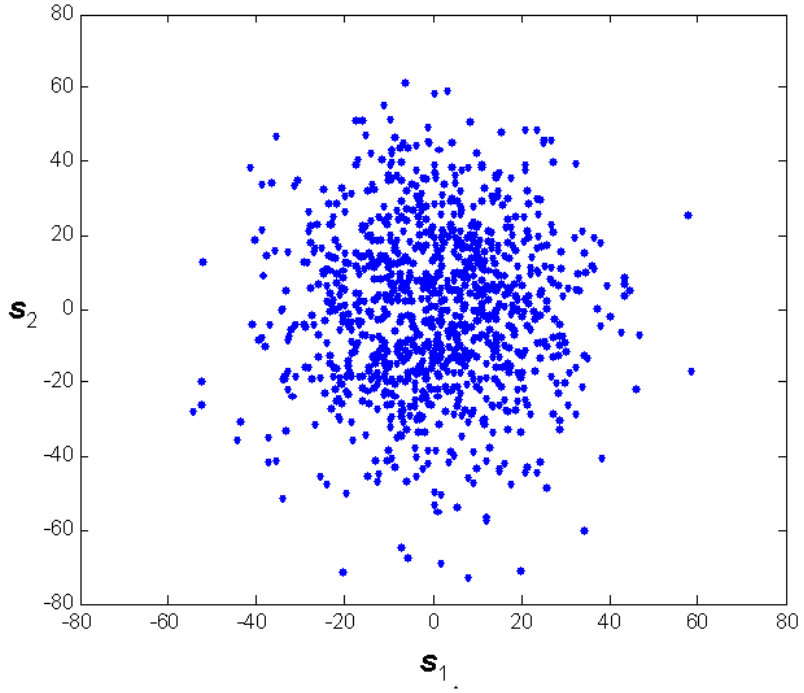

Figure 3.9: *The joint distribution (variables) of sources* s_1 *and* s_2 *with Gaussian distributions.* s_1 *is in the horizontal axis and* s_2 *is in the vertical axis.*

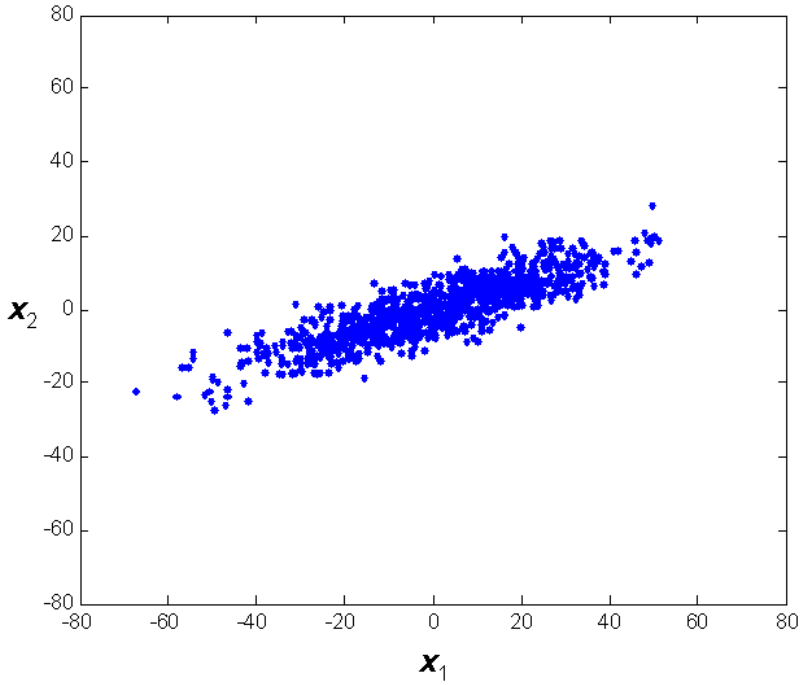

Figure 3.10: *The joint distribution of these two mixed signals, x_1 is in the horizontal axis and x_2 in the vertical axis.*

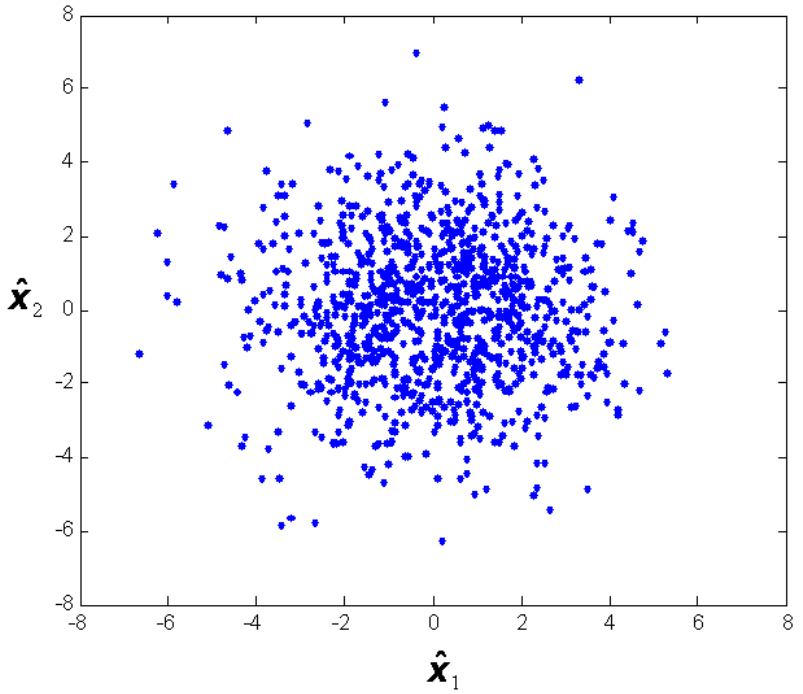

Figure 3.11: *The joint distribution of the two whitened data mixtures. $\hat{x}_1$ is in the horizontal axis and $\hat{x}_2$ in the vertical axis.*

3.4 ICA Preprocessing

In order to simplify and reduce the complexity of the problem for the actual iterative algorithm, generally two common steps of preprocessing are applied before the ICA algorithm: centering and whitening.

3.4.1 Centering

The most basic and necessary preprocessing step is to center the observed signal **x** [148], for example, to subtract the mean vector so that signal **x** becomes a zero-mean variable. Obviously source **s** is zero-mean too after this centering process. This preprocessing can simplify the ICA algorithms. However it does not indicate that the mean could not be counted in the processing. After estimating the mixing matrix **A** with centered data, the mean vector is added back to the centered estimates of **s**.

3.4.2 Whitening

Another important step is to whiten (or sphere) the data [148]. In a mathematical model, the observed data **x** can be linearly transformed to be a new dataset $\hat{\mathbf{x}}$ which is white by performing a linear transformation **V**, for example, $\hat{\mathbf{x}} = \mathbf{V}\mathbf{x}$ so that covariance matrix of $\hat{\mathbf{x}}$, $E\{\hat{\mathbf{x}}\hat{\mathbf{x}}^T\} = \mathbf{I}$. This can be easily accomplished by setting $\mathbf{V} = \mathbf{C}^{-1/2}$, where $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$ is the correlation matrix of the data **x.** It is easy to check $E\{\hat{\mathbf{x}}\hat{\mathbf{x}}^T\}$ is now transformed to $E\{\hat{\mathbf{x}}\hat{\mathbf{x}}^T\} = E\{\mathbf{V}\mathbf{x}\mathbf{x}^T\mathbf{V}^T\} = \mathbf{C}^{-1/2}\mathbf{C}\ \mathbf{C}^{-1/2} = \mathbf{I}$. After whitening, the components of $\hat{\mathbf{x}}$ are uncorrelated and their variances have equal unity.

Figure 3.3 – 3.5 show a simple example of whitening process on the data. In ICA literature, there are a number of different techniques to extract independent sources.

3.5 ICA algorithms

Here the most popular algorithms for ICA are separated into different groups: the first group relies on batch computations minimizing or maximizing some relevant criterion functions and the second uses adaptive algorithms based on stochastic gradient methods which is often implemented in neural networks. The algorithm based on the time structure of the sources [148] is considered as a separate case.

3.5.1 FastICA

FastICA [149] is one of the more popular and referenced ICA techniques in the literature. It attempts to separate underlying sources from the mixed data set based on their 'non-Gaussianity'. A classical measure of non-Gaussianity is kurtosis or the fourth-order cumulant. The kurtosis of **x** is classically defined by

$$\text{kurt}(\mathbf{x}) = \text{E}\{\mathbf{x}^4\} - 3(\text{E}\{\mathbf{x}^2\})^2, \tag{3.13}$$

where **x** is a zero-mean random variable. For a Gaussian **x**, the fourth moment $\boldsymbol{E}\{\mathbf{x}^4\}$ equals $4(\boldsymbol{E}\{\mathbf{x}^2\})^2$. Thus, kurtosis is zero for a Gaussian random variable. For most non-Gaussian random variables, kurtosis is nonzero. Typically non-Gaussianity is measured by the absolute value of kurtosis. The square of kurtosis can also be used. These are zero for a Gaussian variable, and greater than zero for most non-Gaussian random variables. There are non-Gaussian random variables that have zero kurtosis, but they can be considered as very rare.

Kurtosis, or rather its absolute value, has been widely used as a measure of non-Gaussianity in ICA and related fields. The main reason for this is its simplicity, both computational and theoretical. Computationally, kurtosis can be estimated simply by using the fourth moment of the sample data. Theoretical analysis is simplified because of the following linearity property: If $\mathbf{x}_1$ and $\mathbf{x}_2$ are two independent random variables, it holds that

$$\text{kurt}(x_1 + x_2) = \text{kurt}(x_1) + \text{kurt}(x_2) \tag{3.17}$$

and

$$\text{kurt}(ax_1) = a^4 \text{kurt}(x_1) \tag{3.18}$$

where a is a scalar. FastICA is to use a fast fixed-point iterative algorithm (to compute the iterated function, for example, the second step of the below FastICA algorithm in the sequence of given points) to find the local extrema of the kurtosis of the linear observed variables. In other words, as kurtosis is equal to zero for Gaussian distributed signals, the algorithm aims to maximize the magnitude of the kurtosis to make the estimated sources as non-Gaussian (independent) as possible.

The fixed-point algorithm of FastICA can be described as the following steps:

1. Set a random initial vector $\mathbf{w}(0)$ of norm 1 and k=1;
2. Let $\mathbf{w}(k)$=E$\{\mathbf{x}(\mathbf{w}(k\text{-}1)^T\mathbf{x})^3\}-3\mathbf{w}(k\text{-}1)$. The expectation can be estimated using a large sample of $\mathbf{x}$ vectors;
3. Divide $\mathbf{w}$(k) by its norm.
4. If $|\mathbf{w}(k)^T\mathbf{w}(k\text{-}1)|$ is not close enough to 1, let k=k+1 and go to step 2. Otherwise, output the vector $\mathbf{w}(k)$.

The final output of vector $\mathbf{w}(k)$ equals one of the columns of unmixing matrix. The corresponding non-Gaussian source signal can be obtained through $\mathbf{w}(k)^{\mathrm{T}}\mathbf{x}(t)$, t=1, 2, … To estimate n independent components, we need to run this algorithm n times.

In the downloadable FastICA toolbox [150], the fixed-point parameter is set as 1000 points. From the experiments, a remarkable property of FastICA is that only a small number of iterations usually 5-10, is usually enough to obtain the maximal accuracy allowed by the data set.

3.5.2 Infomax ICA

Another method of measuring non-Gaussianity is through negentropy. Negentropy is based on the information-theoretic quantity of differential entropy which is the basic concept of information theory [143] [151]. The entropy of a random variable can be interpreted as the degree of information that the observation of the variable gives. The entropy **H** of a random variable x with density $p_x(x)$ is defined as

$$\mathbf{H}(x) = -\int p_x(x) \log p_x(x) dx \tag{3.19}$$

For random variables of equal variance, Gaussian random variables have the largest entropy which means least information. Entropy is small for distributions that are clearly concentrated on certain values, for example, if the PDF is very spiky (non-Gaussian). This indicates that entropy could be used as a measure of non-Gaussianity. In practise, it more likely uses a measure of non-Gaussianity which is zero for a Gaussian variable and always nonnegative for a non-Gaussian variable. Furthermore, differential entropy – or Negentropy – is defined as the difference between the entropy of a Gaussian random variable with the same variance as the observed random variable, and the entropy of the observed variable [152]. Negentropy **J** is given as follows

$$\mathbf{J}(x) = \mathbf{H}(x_{gauss}) - \mathbf{H}(x_{observed}), \tag{3.20}$$

where x_{gauss} is a Gaussian random variable of the same covariance matrix as $x_{observed}$. Negentropy is always non-negative. It is zero when x has a Gaussian distribution and positive when x has a non-Gaussian distribution.

Infomax ICA [153] is such an algorithm that measures the non-gaussianity of sources by using negentropy. It works based on a neural network gradient-based algorithm whose learning rule is based on the principle of information maximization (so called Infomax), and it maximizes the output entropy of a neural network with nonlinear outputs. By doing that, ICA is able to recover the original sources which are statistically independent. However the main problem with Infomax ICA is that it involves a gradient training algorithm so that the speed of the convergence is varied and the convergence depends crucially on the correct choice of the learning parameters.

3.5.3 Temporal Decorrelation based ICA

Unlike the above two ICA techniques, a totally different technique to perform ICA is to consider the time structure of the sources. The idea behind this approach is to capture the dependency structure of the observed signals using a set of square matrices (or a stack of matrices) and then find the unmixing matrix [154]. One of the practical methods based on this time structure can be achieved through temporal decorrelation (TD) [155] [156]. For sources with stationary waveforms and unique power spectra, the time structure is adequately captured by temporal cross-covariances [157] [158]. The decorrelation operation in time structure ICA methods involves the joint diagonalization of a set of symmetric matrices which reflect the spatio-temporal covariance structure of the source mixture.

$$\mathbf{E}\left\{\mathbf{x}(t,\tau)\,\mathbf{x}^{\mathrm{T}}(t,\tau)\right\} = \mathbf{C}_{\tau}^{\mathbf{x}} = \mathbf{A}\mathbf{C}_{\tau}^{\mathbf{s}}\mathbf{A}^{\mathrm{T}}, \tag{3.21}$$

τ indicates the time delay, normally τ are a set of delays, for example, 1, 2, 3, …. T denotes matrix transpose $\mathbf{C}_{\tau}^{\mathbf{x}}$ is the signal cross-covariance matrix and $\mathbf{C}_{\tau}^{\mathbf{s}}$ is the source cross-covariance matrix. The TD approaches is defined based on the statistical independence of the sources. Furthermore, algorithms have recently been developed for non-orthogonal joint diagonalisation that process signal covariances directly with no need for pre-whitening, one such algorithm is used here and is called $\mathrm{LSDIAG_{TD}}$ [159].

$$\mathbf{C}_{\tau}^{\mathbf{s}} = \mathbf{W}\mathbf{C}_{\tau}^{\mathbf{x}}\mathbf{W}^{\mathrm{T}}, \tag{3.22}$$

The coefficients of the unmixing matrix **W** is optimized in such a way to transform the signal covariances matrix stack into the source covariances matrix stack as diagonal as possible since diagonal is assumed as independence. Figure 3.12 indicates the transformation between two covariance matrix stacks. Once **W** is estimated, each of the independent sources can then be separated by **s**=**Wx**.

Due to the specific assumptions, TD based ICA works fine on the stationary source waveforms with unique power spectra. However, the applications of the long-term biomedical signals recordings are unlikely to remain stationary. However if the assumption is made that the auto-correlation function of the source is slowly varying in time, so that the sources are basically stationary over short time windows, then the TD approach can be adapted to process non-stationary signals by using such short time windows and further applied to estimate the mixing matrix in the usual way [154]. Another issue worth noticing in this technique is the appropriate choice of the number of time lags which are used to describe the spatio-temporal covariance of the data. There could be an automatic method to select the average number of time lags of

the data by using statistical the selection model, but it may cost a huge computation time in practice [154]. Most of the present work prefers to set the number manually, based on experience and knowledge of the application domain (this may amount to trail and error) [160] [161].

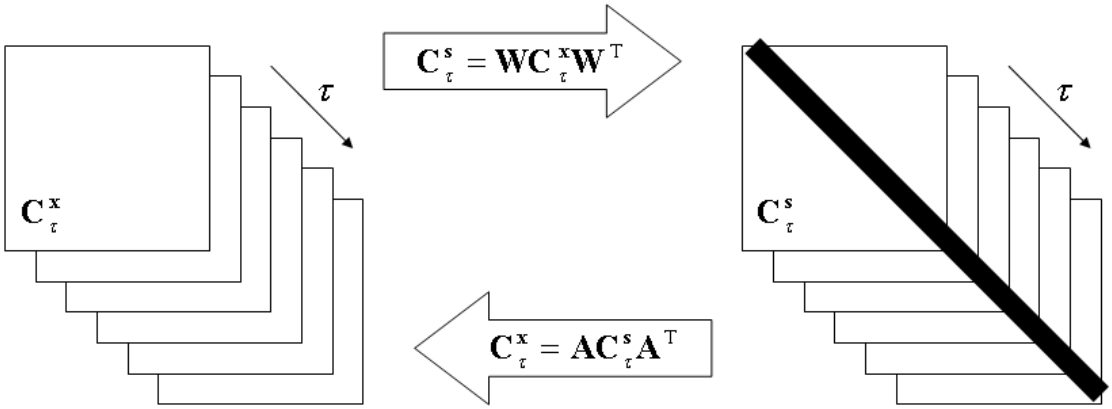

Figure 3.12: *The transformation between two covariance matrix stacks of* $\mathbf{C}_\tau^{\mathbf{s}}$ *and* $\mathbf{C}_\tau^{\mathbf{x}}$. *The mixing matrix* $\boldsymbol{A}$ *transforms the covariance stack of the sources to the covariance stack of the observations and vice versa with the unmixing matrix* $\boldsymbol{W}$.

3.5.4 Dimensionality reduction

Mostly it is true that the data are represented in a high dimensional form but the actual number of sources could be low dimensional. Especially when the number of channels exceeds the number of sources, the ICA approach applied on such a high dimensional space could waste much effort in dealing with irrelevant features. Therefore it is usually advantageous to reduce the high dimensional data into lower dimensional space as a first step. Singular value decomposition (SVD) [162] or principal component analysis (PCA) [163] are popular techniques for the dimensional reduction problem.

The SVD approach for dimensionality reduction condenses most of the information in a dataset to a few dimensions using eigenvectors of the transformed space. The equation for SVD of **X** (*m* x *n* matrix) is the following:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}}, \tag{3.23}$$

where **U** is an *m* x *n* orthogonal matrix, **S** is an *n* x *n* diagonal matrix with positive or zero elements, and $\mathbf{V}^{\mathbf{T}}$ is a transpose of the *n* x *n* orthogonal matrix. The diagonal elements of matrix **S** are necessarily equal to the singular values of **X**. The columns of **U** and **V** are, respectively, left- and right-singular vectors for the corresponding singular values. The dimension reduction process through SVD is to keep its first *k* ($k<m$) singular values. These *k* singular values are ordered in decreasing order along the diagonal of **S** and this ordering is preserved when constructing **U** and $\mathbf{V}^{\mathbf{T}}$, therefore keeping the first *k* singular values is equivalent to keeping the first *k* rows of **S** and $\mathbf{V}^{\mathbf{T}}$ and the first *k* columns of **U**. after this, Equation 3.12 is reduced to Equation 3.13

$$\hat{\mathbf{X}} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^{\mathbf{T}}, \tag{3.24}$$

$\hat{\mathbf{U}}$ is an *m* x *k* matrix, $\hat{\mathbf{S}}$ is a *k* x *k* matrix and $\hat{\mathbf{V}}$ is a *n* x *k* matrix. Then $\hat{\mathbf{X}}$ is the dimension reduced version of **X.**

There is a direct relation between PCA and SVD in the case where principal components are calculated from the covariance matrix. Performing PCA is similar to performing SVD on the covariance matrix of the data. Given a set of data, PCA finds the linear lower-dimensional representation of the data such that the variance of the reconstructed data is preserved. When we project our data onto such a low-dimensional hyperplane, the variance of our data is changed as little as possible.

For the PCA algorithm, let $\mathbf{X}_m^T$ represent a data matrix with zero mean. The covariance matrix (Σ) of $\mathbf{X}_m^T$ is then calculated and the eigenvectors and eigenvalues of Σ are found. The basis of the data and the application could help us to select a number of first k eigenvectors with largest eigenvaules which would decide the dimensions of projected hyperplane.

As the techniques don't have a fixed number to decide the basic vectors. Therefore the dimensionality reduction is somewhat of an arbitrary process. A carelessly selected value of k to truncate the data could lead to the so-called "dimensionality reduction curse" [164] which could affect the performance.

3.5.5 Source selection

With the nature of the BSS problem and the techniques used in ICA, one assumption is that the number of observed mixtures must be at least as large as the number of estimated components, but the exact number of underlying sources is unknown. Therefore correct determination of the number of sources becomes a major problem. Generally the most common methods to select the number of sources are based on cumulative and relative variance thresholds. However these methods don't offer the help to select the source(s) of interest after implementing ICA algorithms at all.

In the literature [165] [166], the source selection can be done as a subjective process by the authors. The selection usually follows some subjective criterion related to the expected outcomes of the analysis. Especially in biomedical applications we may possibly have certain knowledge about the nature of some of the source signals which could be extracted from the recorded data in mind, for example, many physiological signal patterns (heart beat waveform and rhythmic brain activity) have particular temporal, spectral or time–frequency characteristics. In some multi-channel

applications, the scalp topography which tells the components' physiological origins forms another common method on source selection, such as eye blinks, saccades and bursts of muscle activity. However all of the above methods to choose sources of interest remain highly subjective.

In a previous study a solution was proposed for source selection, which is to use prior information in the ICA model [167]. The idea behind it is to use only minor modifications of the estimation procedures, essentially by introducing constraints on the model, which can act on the spatial projections, or work on the temporal dynamics of the source waveforms. By adding prior knowledge into the model and letting ICA estimate the unknown parts, these modifications are presumed to guide the ICA solution to find an expected outcome. This may certainly help interpret the results meaningfully, although it is recognised that this may lead to sub-optimal results in some instances.

3. 6 Constrained ICA

Once a set of sources is determined through ICA, the ICs of interest must be identified. This is made difficult as the nature of the square mixing matrix means that a great many more sources will be identified over the expected (smaller) number of sources underlying the measurement set. In most specified signal processing applications, it is very often the case that the source signals which are aimed to extract are known to the applicants. Many expected signals or patterns have certain temporal, spectral or time–frequency characteristics. A practical way to extract only the sources of interest automatically is to use such prior knowledge or additional constraints on the source model – constrained ICA (cICA). Such prior information/knowledge is desirable and possible to add into the ICA model by only minor changes of the estimation procedures. These modifications can be achieved by imposing constraints on the model, for example to work on the spatial projections, or on the temporal dynamics of

the source waveforms. Such constraints are expected to guide the ICA solution to extract the ICs of the observations (if they are there). Some constraint selection techniques, for example, have been presented in a free ICA toolbox – ICALab [168]. Basically constrained ICA algorithms can be separated into two stages: the first stage is to apply an ICA algorithm to extract the suitable references signals; the second stage is to use some selected components as references signals in the constrained ICA applications. In order to extract more physiologically meaningful components, the selected references are better to give better signals for example the responses corresponding to pre-designed stimulus or events.

3.6.1 Spatially-constrained ICA

One example is to apply the spatial constraint on the mixing matrix which relies on specific prior knowledge regarding the spatial topography of some source sensor projections. This forms an advanced ICA algorithm called spatially-constrained ICA [154]. The idea of using spatial constraints in BSS was initially proposed in automated artifact removal in EEG. For example, the spatial topographies for eye-blink and eye movement can be included as an initial guess in the first two columns of the mixing matrix.

The spatial constraint operates on initializing columns for mixing matrix **A**. A set of predetermined constraint sensor projections forms as reference columns denoted by $\mathbf{A_c}$. Thus, the spatially constrained mixing matrix comprises two types of columns

$$\mathbf{A} = [\hat{\mathbf{A}}_\mathbf{c}, \mathbf{A_u}] \tag{3.25}$$

where $\hat{\mathbf{A}}_\mathbf{c} \approx \mathbf{A_c}$ are columns subject to the constraint, and $\mathbf{A_u}$ where are otherwise unconstrained columns (random values).

Depending on the application, the predetermined sensor projections forming the spatial constraint may be obtained from a source decomposition of a previous data segment using conventional BSS (ICA) methods [169]. Moreover in the algorithm, it is possible to choose one of three types of spatial constraint, namely hard, soft and weak spatial constraints which reflect the degree of certainty about the accuracy of predetermined source sensor projections. The columns of a hard constraint remains fixed, whereas the weak constraint only provides a starting guess for unconstrained subsequent estimation. Soft constraints limit the divergence between the constrained columns and their corresponding reference topographies [170]. Whilst column selection of the mixing matrix does not guarantee to offer good performance in the solution, it can help to separate the sources when they are present in the data. More information on the use of constraints can be found in [167] and [171].

3.6.2 Temporally-constrained ICA

Theoretically the ICA algorithm first converges to a single source with the largest negentropy of all the underlying ICs. However the algorithm is not guaranteed to converge to the global maximum due to random initialization of the ICA algorithm and other factors. The idea of using a temporal constraint is to guide the ICA algorithm to obtain a source output which is statistically independent of other sources *and* is closest to some reference signal [172]. This constraint (reference signal) does not have to be perfect, but it should be at least good enough to point the algorithm in the direction of a particular IC spanning the measurement space. The closeness constraint can be written as

$$g(\mathbf{w}) = \varepsilon(\mathbf{w}) - \xi \leq 0, \tag{3.26}$$

where $\mathbf{w}$ indicates a single unmixing weight vector, such that $\mathbf{y}=\mathbf{w}^{\mathrm{T}}\mathbf{x}$; $\varepsilon(\mathbf{w})$ denotes

the closeness between the estimated output **y** and the reference signal ***r***, and ξ some closeness threshold. The measure of closeness can be achieved by the methods such as mean-square-error or correlation, etc. For example using the correlation as a measure of closeness one can rewrite equation 3.15 as:

$$g(\mathbf{w}) = E\{\boldsymbol{r}(\mathbf{w}^{\mathbf{T}}\mathbf{x})\} - \xi \le 0, \tag{3.27}$$

where ξ becomes the threshold that defines the lower bound of the optimum correlation. The temporally-constrained ICA is now modelled as the following:

$$\text{Maximize: } f(\mathbf{w}) = \rho[E\{G(\mathbf{w}^{\mathbf{T}}\mathbf{x})\} - E\{G(\mathbf{v})\}]^2, \tag{3.28}$$

$$\text{Subject to: } g(\mathbf{w}) \le 0, h(\mathbf{w}) = E\{\mathbf{y}^2\} - 1 \text{ and } E\{\boldsymbol{r}^2\} - 1 = 0, \tag{3.29}$$

where $f(\mathbf{w})$ denotes the contrast function; ρ is a positive constant; **v** is a zero mean, unit variance Gaussian; $G(.)$ can be any nonquadratic function; $g(\mathbf{w})$ is the closeness constraint; $h(\mathbf{w})$ constrains the output **y** to have unit variance; and the reference signal ***r*** is also constrained to have unit variance [172].

3.6.3 Spectrally-constrained ICA

In some applications, rhythmic EEG signals within the EEG recordings (specifically μ rhythm activity) are of interest, a predefined spectral reference is proposed to use as the constraint. This spectral constraint then allows only those source activities with the same power spectrum to be extracted via the cICA algorithm. In [160] a reference channel was added as an extra row to the measurement matrix $\mathbf{x}(t)$, such that a new matrix $\hat{\mathbf{x}}(t)$ is created with

$$\hat{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{c}_1(t) \end{bmatrix}, \tag{3.30}$$

where $\mathbf{c}_1(t)$ is a suitable reference vector. In order to observe changes in rhythmic activity in specific frequency bands, a filtered white noise with a particular power spectrum is often used as this reference vector. The ICA problem is now such that the extra row in the measurement space due to the reference vector results in an extra row in the IC space after the ICA step (as well as a corresponding extra column in the mixing matrix). For an n-channel system, the first n elements of the extra mixing matrix column $[\mathbf{a}_1^{n+1}, \mathbf{a}_2^{n+1}, \ldots, \mathbf{a}_n^{n+1}]$ depict the spatial distribution (topography) of the new IC given by the row vector $\mathbf{s}_{n+1}(t)$. Furthermore, each of the elements of the $(n+1)^{\text{th}}$ row of the mixing matrix reflects a weighting of each corresponding IC. This row vector, $\mathbf{a}_{n+1}$, can in fact be used to depict the contribution of each topography described by the columns of the mixing matrix, due to the reference channel $\mathbf{c}_1(t)$. In this way ICA now provides the desired convenient spanning basis, and can also be used to obtain the topography of interest (extracted by summing the weighted contribution of each column of the mixing matrix). Furthermore, the weighting value of each IC provides us with a spectrum of values that can be interpreted to gain some insight into the complexity for a given reference. The above technique can be readily extended to more than one reference. The extraction of rhythmic EEG signal components (such as epileptic seizures) by this spectrally-constrained ICA method have been shown in [173].

3.7 Single channel ICA

In traditional multi-channel ICA analysis, the observation data must fulfil the following two primary conditions: 1) the underlying sources must be statistically independent, with static distributions; 2) the number of observations must be greater

than (or equal to) the number of sources expected to be extracted. From previous sections, the multi-channel ICA algorithms and applications are all based on these requirements. However since the data available for the analysis changes from multi-channel to single channel, the second condition no longer exists. One way to overcome this problem is to reform the single-channel into a multi dimensional representation via the method of delays [174] (also known as dynamical embedding [175]). Such an ICA algorithm on multi-dimensional data representation is called single-channel ICA (SC-ICA)

3.7.1 The matrix of delays

Dynamical embedding (DE) was firstly introduced by Takens [175]. It assumes that due to the non-linear interaction of just a few degrees of freedom, with additive noise, the measured signal exists as an unobservable deterministic generator of the observed data. This allows reforming the unknown dynamical system by constructing a new state space based on successive observations of the time series.

The basic idea of DE is to structure an *m*-dimensional embedding matrix for an *n*-valued scalar data set. Assuming a single data channel with *n* elements: $\{x_i\}_{i=1,\ldots,n}$, then delayed vectors in the constructed matrix are given as $\upsilon_k = \{x_k, x_{k+1\tau,\cdots,} x_{k+(m-1)\tau}\}$. The delay matrix υ is formed by obtaining υ_k for successive values of *k*, and combining these to form

$$\upsilon = \begin{bmatrix} x_t & x_{t+\tau} & \cdots & x_{t+N\tau} \\ x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(m+1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t+(m-1)\tau} & x_{t+(m)\tau} & \cdots & x_{t+(m+N-1)\tau} \end{bmatrix}, \tag{3.31}$$

where τ is the lag term, m is the number of lags (or the dimension of the matrix of delays) and N is the length of the matrix. After construction with these delay vectors, the matrix of delays is able to provide a topologically consistent representation of the underlying system so that the second primary condition is feasible. Therefore popular ICA algorithms can now be applied to the reformatted single channel data. Those techniques include FastICA, infomax ICA and other types of ICA algorithms described in the earlier sections of this chapter.

In order to construct the matrix of delays or DE system, the parameters such as m and τ need to be set to appropriate values. Obviously the choice of these parameters is an optimization process. In Takens' theorem, if the number of degrees of freedom of the underlying system is given by D, then the Euclidean embedding matrix dimension $\hat{m}$ must be at least as large as D, but in practice must be such that,

$$\hat{m} > 2D + 1, \tag{3.32}$$

Since in the real world, the value of D is unknown, another alternative method [176] is more practical in the studies. This method sets up the minimum size of m based on the sampling rate and the lowest frequency of interest as,

$$m \geq \frac{f_s}{f_L}, \tag{3.33}$$

where f_s denotes the sampling frequency, and f_L the lowest frequency of interest in the measured signal. Depending on different applications, there could be a variety of methods to try to estimate the value of τ, in the literature on similar studies, the popular value was set to 1 [176].

3.7.2 Projection of the ICs

After constructing a delay matrix, ICA is then applied which decomposes the delay matrix into a series of statistically independent components just as the multi-channel ICA application on multi-channel data does.

Normally the number of ICs is the same as the dimension of the delay matrix m. As the delay matrix is simply composed of time-shifted versions of the original time series, each individual IC is hard to interpret in the source space. In this study in order to assess the significance of each IC neurophysiologically, it must be projected back to the measurement space in isolation such that

$$\mathbf{Y}^i = \mathbf{a}_i \mathbf{u}_i^T, \tag{3.34}$$

where $\mathbf{u}_i$ is the ith IC (i = 1, 2, …, m), $\mathbf{a}_i$ the corresponding column of the mixing matrix $\mathbf{A}$ and $\mathbf{Y}^i$ the resulting 'matrix of delay vectors'. From $\mathbf{Y}^i$ it now becomes possible to extract the projected time series, $y_i(t)$, by performing an average of the adjusted rows of the matrix $\mathbf{Y}^i$, in order to recover the time series, i.e.,

$$y_i(t) = \frac{1}{m} \sum_{k=1}^{m} \mathbf{Y}^i_{k,(t+k-1),} \tag{3.35}$$

where t is equal to 1, 2, …, N, and $Y^i_{k,(t+k-1)}$ refers to the element of $\mathbf{Y}^i$ indexed by row k and column $t + k - 1$.

3.7.3 Selection of the ICs

Whereas in multi-channel ICA the columns of the mixing matrix $\mathbf{A}$ are interpreted as spatial distributions of their respective ICs, in SC-ICA these are interpreted as shifted

versions of the mixing filters (i.e. no spatial information but rather temporal information). As the independence of components is used as the important notion in multi-channel ICA applications, and it is not generally possible to linearly decompose a single time series into independent components. This study has to relax the notion of component-wise independence and turn to find and group the components of interest. Depending on different applications, specific signals such as particular epileptic activity or event-related evoked potential can be used for the purpose of grouping.

Recent studies [177] [178] have applied SC-ICA to extracting such activities. The basic idea is to apply ICA to the matrix of delays, then project each resulting IC back to the measurement space. The selection of the most relevant ICs was based on the subjective analysis of wave morphology and on derived spectrograms of each projected IC. Results demonstrated that single channel analysis can extract meaningful information and achieve good signal resolution [179].

However the subjective selection of relevant ICs has disadvantages. Obviously it is a slow process and cannot form an automatic robust system since it involves manual work. An approximate method [174] was introduced which intends to avoid the IC selection or the clustering step. The idea is to begin with a deflationary ICA algorithm on the delayed vector matrix. Followed by a step that forms a separation filter by using just a single basis vector from the unmixing matrix, one of the sources can be separated by applying this filter to the first extracted component. Depending on the data and the application, the algorithm can be repeated on the remaining signals to extract further related sources. From the results it is believed that this method performs efficiently, especially when there are only a small number of independent processes to be extracted. The algorithm mentioned in [174] is summarised as:

a) create the delayed vector matrix from the selected data set, temporally whiten the signal and reduce the dimension by PCA;

b) apply a deflationary ICA algorithm to learn the mixing matrix **A**;

c) create the separation filter $f_i(t) = \boldsymbol{a}_i(-t) * \boldsymbol{w}_i(t) \quad i \in \gamma_p \ \gamma_p$ is the selected subset of basis vectors; where $\boldsymbol{a}_i(t)$ is the column vector of **A,** $\boldsymbol{w}_i(t)$ is the row vector of **W** and * is defined as convolution;

d) calculate the scaling parameter a_p to adjust and rescale the filtered signal, a_p is given as: $a_p = \left\langle r_p(t), f_i(t) * r_p(t) \right\rangle \Big/ \left\langle r_p(t), r_p(t) \right\rangle$, where $\langle \cdot, \cdot \rangle$ is the usual vector inner product and initially $r_0(t) = x(t)$; $a_p f_i * r_p(t)$ is the extracted source component;

e) if further sources are needed to extract, calculate the residual $r_p(t) = r_{p-1}(t) - a_p f_i * r_{p-1}(t)$ and go back to Step (b).

3.8 Space-time ICA

When the input data is formed from a delay vector of samples, x(t) = [x(t); x(t-1), …, x(t-N +1)]^T, source separation is still possible and the resulting SC-ICA can be seen as a special instance of multi-channel ICA. However this model carries a rather restrictive separability requirement. [174] introduces the notion of space-time ICA (ST-ICA) to relax this requirement. The data structure is treated in the same manner as the SC-ICA. Moreover the dimension of the observation is augmented by including a number of delayed copies of observations. The definition of the $O \times T$ - dimensional space-time vector $\tilde{\mathrm{x}}(t)$ is as:

$$\tilde{\mathrm{x}}(t) = [x(t), \quad x(t-1), \ldots, x(t-T+1)]^T \tag{3.36}$$

where N is the number of copies in the delay vector. This can now be treated as an $O \times T$ dimensional multi-channel ICA application. In this way the problem is now translated and performed to be an example of a standard ICA algorithm applying on

this multi-dimensional matrix. This is called ST-ICA, and can also use the same component projection and selection techniques as in SC-ICA.

3.9 Applications of ICA in the literature

The data analyzed by ICA could originate from many different kinds of application fields, including digital images and document databases, as well as economic indicators and psychometric measurements. In many cases, the measurements are given as a set of parallel signals or time series; typical examples are mixtures of simultaneous speech signals that have been picked up by several microphones, brain waves recorded by multiple sensors, interfering radio signals arriving at a mobile phone, or parallel time series obtained from some industrial process.

One of ICA decomposition example is to use ICA to find filters for natural images and, removing noise from images corrupted with additive Gaussian noise [180]. Another emerging application area is telecommunications. A good example of a real-world communications application where ICA techniques are useful is the separation of the user's own signal from other interfering signals in CDMA mobile communications. In the field of communication networks a phenomenon called multipath [181] is the propagation that results in signals' reaching the receiving antenna by two or more paths. Affected by the multipath, the observed signal is a convolutive mixture of the source signals which have different time delays due to the finite propagation speed in the medium. In order to solve this convoluted mixing problem and separate independent sources, some publications [182] have shown that the ICA algorithm is able to effectively decompose the received signals into the independent paths and noise term and hence provides information about the delay estimate of these paths.

Not very long after the appearance of ICA, it has been broadly applied to biomedical signal and image processing, such as the analysis of electrocardiography (ECG) [183] [184], EEG [185], MEG [119], and fMRI [186]. When using an electromagnetic record as a research or clinical tool, the investigator may face a problem of extracting the essential features of the neuromagnetic signals in the presence of artifacts. The amplitude of the disturbances may be higher than that of the brain signals, and the artifacts may resemble pathological signals in shape. ICA techniques can separate the underlying activities from noisy signal. Moreover advanced constrained ICA has been used in the artifactual waveform identification [173] as well as the EEG rhythmic activity separation such as to extract alpha activities or epileptic seizures [160]. Single-channel ICA techniques have also showed the ability to extract particular epileptic activities [178] [187].

Moreover the application of EOG artefact removal has been embedded into several commercial devices, for example, EEG acquisition systems (g.tec® acquisition system [188] and the BIOPAC® system [189]) so that EOG artifact can be automatically removed from the recordings just by clicking the button. A commercial piece of software called Curry® [190] is used for brain source analysis and display, through EEG/MEG source localization packages. One of its features is to use ICA to visualize the spatio-temporal features of EEG and MEG data for the purpose of obtaining the maximum accuracy of electrical source analysis (this is manually performed and the ICA decomposition and subsequent analysis is left in the hands of the user of the software).

3.10 Summary

This chapter introduces the concept of ICA which is a method for performing BSS on time series. ICA techniques make two important assumptions – one is based on the

type of mixing of the independent sources and the other on how the statistical independence of those sources is measured. Of the many possible algorithms towards solving the BSS problem, ICA is popularly solved through the use of higher order statistic techniques which basically try to separate statistically independent sources based on their non-Gaussianity. Another ICA technique is based on using spatio-temporal and spatial-time frequency information. The main difference between these two techniques is that the latter technique uses the information inherent in the time-sequence of the measured data, but the former doesn't. This chapter shows that adding prior knowledge and letting the ICA estimate the unknown portions which makes the source selection process much easier. This can be interpreted that ICA would generally be more useful if some more assumptions are made within the BSS model. This chapter also introduces the concept of using ICA on a single channel recording or on a few channels. Such applications of ICA are called single-channel ICA and space-time ICA. To make it work, a single channel or a few channel recordings need to be reformed to a multi-channel data representation by the method of delays. Such techniques bring the opportunity to use ICA in applications with the channel restriction. In the literature, many ICA applications can be found in the biomedical signal processing field as this very popular application domain.

CHAPTER 4

A BCI pilot study: preliminary analyses

4.1 Background

The Southampton Brain-Computer Interfacing Research Programme (http://www.bci.soton.ac.uk/) plans to examine the effect of spoken and written language on brain activities. The research programme brings together biomedical engineering and the clinical sciences within the life sciences interfaces and neuroscience initiatives in the University of Southampton. However, before such hypothesis-driven clinical work can be undertaken, a pilot study was proposed on normal subjects. It aimed to a) conduct P300 ERP paradigms to establish feasibility and baseline data for the Southampton BCI research laboratory; b) repeat protocols from previous BCI research [76], involving spontaneous EEG recordings during cognitive tasks, using both motor and non-motor imagery tasks and c) explore preliminary signal processing techniques to analyse the obtained signals produced by experiments conducted for the first two aims above, improving accuracy and reliability of previously used methods. The pilot study included both ERP (P300) paradigm and cognitive tasks (motor imagery and non-motor imagery tests). The interests here are to examine basic and advanced signal processing techniques on the data from the pilot study so that the necessary knowledge may be accumulated and

contribute towards follow-on studies. The final overall goal is to build user-friendly BCI systems to widen accessibility and compliance for use in rehabilitation.

This chapter first introduces details of the pilot study regarding the choice of proposed tests, with a full description of the proposed test paradigms. The chapter then develops and demonstrates several proposed signal processing techniques related to different tests which include: averaging either in time or in the frequency spectrum; traditional ICA and spatially constrained ICA, and event related desynchronization/synchronization (ERD/ERS) mapping. Conclusions are then made after the proposed methods are compared and discussed.

4.2 Methods

4.2.1 Subjects

Seven healthy subjects aged 20–60 years were involved in this pilot study (three males). They were recruited from the staff and students at the University of Southampton. None of the subjects had attended any experiment similar to our proposed tasks before. Subjects were given an information sheet (Appendix A) and written, informed consent was obtained by each subject signing a consent form (Appendix B). The studies were approved by the School of Health Professions and Rehabilitation Sciences Ethics Committee.

4.2.2 Data acquisition

The non-invasive technique of surface EEG was used to record signals from the brain during different test paradigms. The electrodes were the silver plated adhesive disk type. The EEG recordings were acquired by NeuroScan EEG apparatus in the EEG Laboratory in the School of Psychology. The recordings were digitized at 250Hz from

24 scalp electrodes (VEOG_1, VEOG_S, HEOG_L, HEOG_R, A1, A2, FP1, FP2, F3, Fz, F4, FT7, FT8, T3, C3, Cz, C4, T4, TP7, TP8, P3, Pz, P4, Oz all based on the International 10/20 system). A1 and A2 were used as the reference electrodes. From 24 channels, 18 channels were selected (FP1, FP2, F3, Fz, F4, FT7, FT8, T3, C3, Cz, C4, T4, TP7, TP8, P3, Pz, P4, Oz) as those electrodes cover the active area of brain signals in this study. The discarded channels are EOG channels which contain mainly EOG noise. The visual stimulus programs are coded by the Presentation ™ software [191].

4.2.3 Experimental paradigms

Six types of experimental paradigms or protocols were explained and the standard instructions were given to subjects for each test. During the testing session, subjects sat still in a chair. Baseline EEG recordings were made during two 2-minute periods when the subject sat resting, first with eyes open and then with eyes closed.

Test 1: Auditory oddball task (approximately 2 minutes)

This experiment was a simple auditory ERP task which is expected to detect P300 responses [59]. Auditory stimuli of high frequency and low frequency tones were delivered randomly through external speakers and the interval between stimuli was approximately 1s. Subjects were instructed to click a button on a computer mouse when noting the infrequent high frequency tone. During the procedure approximately 50 high frequency tones out of 120 tones in total were presented. Subjects were asked to fixate on a cross in the middle of the screen during the test.

Test 2: N400 sentence semantic congruent test

The N400, another component with potential clinical use, is a negative going deflection that can be obtained in language tasks [192]. A typical situation for the reliable elicitation of the N400 entails the presentation of words that either match or

do not match a preceding context. This test can be done either in a sentence or a pair of words. For example, a sentence using context matching could be like the following: “I drink my coffee with cream and sugar/mud”; a pair of words context matching is like the following: “boat, ship/crater”. In the examples, “cream”, “boat” serve as previous context and the following word “sugar” or “mud”, “ship” or “crater” are the candidate words to match/un-match the previous context. Words that do not match the context give rise to a more negative waveform starting approximately 250ms after stimulus onset and peaking at about 400ms (Figure 4.1). The effect is widespread, normally having a maximal amplitude in the centroparietal area.

For the test in this pilot study, several auditory sentences were presented, some of which terminated with a semantically congruent one-syllable word and others with a semantically incongruent one-syllable word. The onset of the final word served as the stimulus (with an investigator pressing the button immediately when the last word was presented). The N400 ERP component evoked by the incongruent end word is expected to have a longer latency than one in a word matching experiment. For example, the latency may be spread in the range from 400ms to 800ms. The test presented ten pairs of congruent/incongruent sentences while several seconds were used as the inter stimulus interval. The total duration of the test lasted about four minutes.

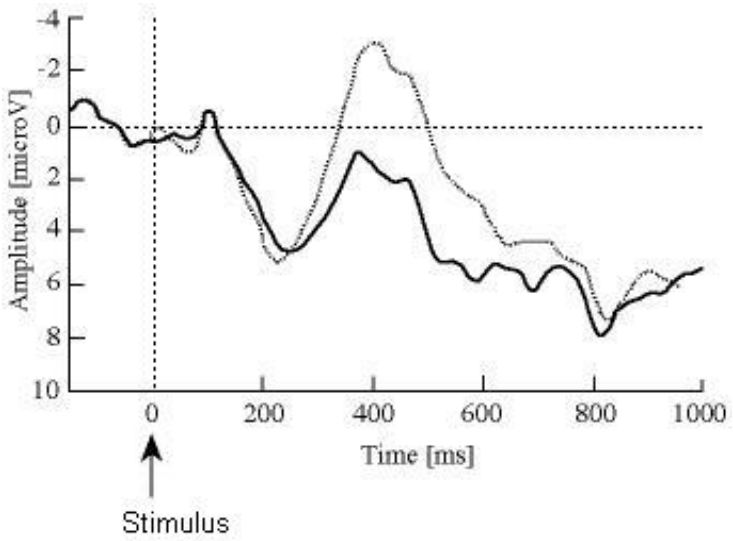

Figure 4.1: *Discourse-semantic N400 effect. Average ERP at Pz, elicited by a stimulus (critical word) that is matched (solid line) or mismatched (dotted line).*[192].

Test 3 & Test 4: Motor Imagery-left/right hand (approximately three minutes)

The subjects were asked to imagine the movement of grasping and releasing right or left hand without actually moving [193]. Before recordings began, subjects were asked to practice actually grasping and releasing right/left hand a few times and notice how this felt. They were then asked to imagine doing this, while making sure that their hand did not actually move. They were instructed by words on the screen to "start" and they imagined the movement for 15 seconds until they were instructed by the word "rest", when they stopped and rested for 5 seconds. There were 10 repetitions for each hand imagery test, following the protocol published in [76].

Test 5: Imagery-Spatial Navigation (approximately three minutes)

Subjects were asked to imagine being in the familiar surroundings of their own home, moving from room to room [194]. They were asked to imagine scanning the rooms, rather than think about actually walking around, to avoid motor activity. The protocol consisted of the subject following the instructions on the screen to start imagery (15 seconds) and then to stop and rest (5 seconds), involving 10 repetitions [76].

Test 6: Imagery-Music (approximately three minutes)

Subjects were asked to think of a favourite song or a familiar tune they enjoyed [194].

They were instructed to 'listen' to it in their head, without mouthing the words or moving any part of their body. This section had 10 repetitions.

4.2.4 The data quality

Despite standardised instructions being given to each subject and them remaining as still as possible while performing the tasks, there were some unexpected problems, including: presentation software crashed during the recording so that some trigger information was not captured; electrodes occasionally became detached from the skin and fell off; some electrodes produced high resistance and may have been faulty. These made some of the recordings impossible to assess and were discarded. The above means that the proposed analysis was only carried on the data which were practically available to us (Table 4.1).

Table 4.1: *The actual available dataset in this pilot data.* ✓ *indicates the data practically available to us and* × *means the unuseable data.*

Subjects ID	(1) P300	(2) N400	(3) Motor imagery right hand	(4) Motor imagery left hand	(5) Imagery music	(6) Imagery navigation
1	✓	✓	✓	✓	✓	✓
2	✓	×	✓	✓	✓	✓
3	×	×	✓	✓	✓	✓
4	×	×	✓	✓	✓	✓
5	×	×	✓	✓	✓	✓
6	✓	×	✓	✓	✓	✓
7	×	×	✓	✓	✓	✓

4.3 Signal analysis

To analyse the data from different tasks and show them into a meaningful presentation, appropriate processing techniques were proposed. For auditory ERP tests, the analysis used was the coherent averaging method and ICA techniques to enhance the SNR and enhance the P300 relative to the background noise. Coherent averaging and an improved averaging technique (see Section 4.3.2) were used to analyse the N400 data. For motor imagery data, the averaged power spectra and ERD/ERS maps were calculated to compare the power changes over the motor cortex. Averaged power spectra of imagination trials were compared for the two different non-motor imagery tests.

4.3.1 The analysis on ERP data

(1) Coherent averaging

As a traditional signal processing technique, coherent averaging is able to enhance ERPs' SNR since the technique assumes the actual EEG activities such as the P300s are invariant across data trails and background EEG noise is random noise and not time locked. The detailed coherent averaging method has been introduced in Chapter 2. This study applied the coherent averaging at Cz which is most active site for P300 activities. To examine the coherent averaging performance, different times of averaging are applied to the data.

(2) 'Standard' ICA

To extract P300 activity, the study also proposes to apply an ICA algorithm to the data. There are a number of algorithms available for the implementation of ICA. For this study, the FastICA algorithm which uses a practical approximation of negentropy

together with a fixed-point algorithm is applied on P300 data set. The details about the FastICA algorithm have been presented in the Chapter 3.

a. Component selection

After ICA decomposition, the underlying sources are separated into separate ICs. As we are interested in the extraction of P300's, those ICs with relatively larger amplitude in the latency range of the P300 should be considered. The columns of the $\mathbf{W}^{-1}$ matrix are denoted as the scalp topography of the components which provides evidence for the components' (spatial) physiological origins. According to the *a priori* physiological knowledge, the P300 appears at the vertex region (Fz, C3, Cz, C4, Pz), normally maximised at Cz. Therefore a manual selection of one P300 component based on viewing the topography map of the components and the IC itself is used in the study.

b. Component projection

The activity of the selected component needs to be projected back to the observed measurement space since the component in the source space is 'unit-less'. The detailed projection process is introduced in Chapter 3. After this projection, clearer P300 activities can be visually extracted from noisy EEG data.

(3) Spatially constrained ICA

The idea of spatially constrained ICA is already explained in the Chapter 3. If *a priori* knowledge about the spatial location of some of the sources is known, then it can be applied to the ICA algorithm by constraining the column(s) of the mixing matrix. As the typical P300 spatial distribution is known, such knowledge of P300 scalp topography can be used as a spatial constraint in the first column of the mixing matrix.

To construct this spatial constraint in the data all trials with potential P300 activities are manually selected and joined together in series to a new data set. The modification is expected to lead to the P300 relative projection strengths being enhanced. A normal ICA decomposes this data into ICs, and after viewing the time course and the corresponding spatial scalp topography of each IC, the scalp topographies associated with P300 components (corresponding column of the estimated mixing matrix) can be chosen as the selected spatial constraints. To simplify the problem a single constraint was considered in the study.

This single spatial constraint was used to initialize the first column of the mixing matrix **A** and set random values for the rest column for the spatially constrained ICA algorithm. After spatially constrained ICA on the original P300 data, the corresponding component was projected to the measurement space. The coherent averaging method was then applied on the extracted data at Cz.

(4) Correlation with a template

In order to assess the performance of ICA applications, we compared the correlation with a predefined template before and after applying ICA. The template was obtained from a 1.5s averaged P300 activity selected from averaging the above P300 trials.

4.3.2 Analysis on N400 data

Two simple methods were proposed: a normal coherent averaging method on the data and the averaging with a few samples shifted while the number of shifting steps was calculated from the maximum cross correlation values between one random selected epoch as a reference and the other epochs. The assumption made to apply the shifted averaging is based on the fact that the recorded stimulus marks for response activities in this data are not automatically set by the presentation program but manually by

clicking a button and the time delay between the actual stimulus start and the recorded one is going to be approximate (to within about one second). Therefore, shifting the response epochs a few samples to match the maximum cross correlation values between an individual epoch and a given reference in a range of one second recording will tune and find a best position to coherently average the data trials.

4.3.3 Analysis on imagery data

(1) Averaged power spectra

The exact nature of the manifestation in the brain of mental imagery is still uncertain. At this stage, there still lacks a clear solution to decide what exact EEG characteristics are most suitable to apply in the BCI field. One assumption is that different mental imagery tasks may lead to power changes over the active regions. A basic and efficient technique was applied to calculate the power spectra in frequency space.

The power is normally defined to be the average of the square of the signal's amplitude while in the frequency domain it is equivalent to the square of FFT´s magnitude [195]. Power spectra indicate the quantity of power at particular frequencies. Practically it can be computed through the power spectral density which presents power per Hz vs frequency. In this study, a 1024-point FFT was applied on ten 15-second response trials for each task. Depending on the different active locations, the averaged power spectra over ten repetition trials were compared for the position at T3, T4, P3 and P4 for non-motor imagery tasks and at C3 and C4 for motor imagery tasks.

(2) Event related desynchronization/synchronization (ERD/ERS)

In addition, different to non-motor imagery, motor imagery of (for example) hand movements is accompanied by the suppression of alpha-range activity on the contralateral hemisphere and with enhancement on the corresponding hemisphere. This phenomenon is called event related desynchronization/synchronization (ERD/ERS). Therefore this study applied the ERD/ERS mapping on hand grasping tasks as an extra analysis.

The calculation of ERD/ERS time courses can be performed in different ways. A standard ERD/ERS calculation [196] was performed by bandpass filtering of the data of each trial, squaring of samples and subsequent averaging over trials and over sample points. The ERD/ERS is then defined as the proportional power decrease (ERD) or power increase (ERS) in relation to a specific reference interval which is usually placed several seconds before trigger onset. The procedure is performed as follows:

$$y_{ij} = (x_{ij} - \bar{x}_j)^2, \tag{4.1}$$

$$A_j = \frac{1}{N-1}\sum_{i=1}^{N} y_{ij}, \tag{4.2}$$

$$R = \frac{1}{k}\sum_{r_0}^{r_0+k} A_j, \tag{4.3}$$

$$ERD_j = \frac{A_j - R}{R}, \tag{4.4}$$

where N is the total number of trials - (equal to ten in this study), x_{ij} is the j^{th} sample of the i^{th} trial of the bandpass filtered (cutoff frequency is between 8~15Hz) data, and $\bar{x}_j$ is the mean of the j^{th} sample averaged over all bandpass filtered trials. R is the

average power in the reference interval [r_0, r_0+k]. The length of the interval was five seconds.

4.4 Results

4.4.1 Results on P300 data

The proposed techniques were applied to three available sets of data from Subject 1, 2 and 6 as shown in Table 4.1.

Figure 4.2 shows the two selected topographic maps on the data from Subject 1, which correspond to two selected components after running a standard ICA. Figure 4.3 plots the averaged P300 activities against the number of coherent averaging on raw data and extracted data after normal ICA. A total of ten pairs of averaged results are shown. For each plot, the upper signal shows the averaged activity on the raw data and the lower one shows the averaged activity with the same averages as the upper signal but on the ICA extracted data. The first vertical dashed line represents the auditory stimulus onset and the second vertical dashed line marks 300ms where the P300 activities should occur. After ICA extraction and the coherent averaging, some noise - especially eye blinks - were reduced whilst P300 peaks were enhanced. Most extracted P300 signal trials are easy to identify visually. However the indentifiation of P300 ICs based on visual observation is limited. Here a simple measurement based on the calculation of the correlation between a trial and a P300 template is employed to assess the performance.

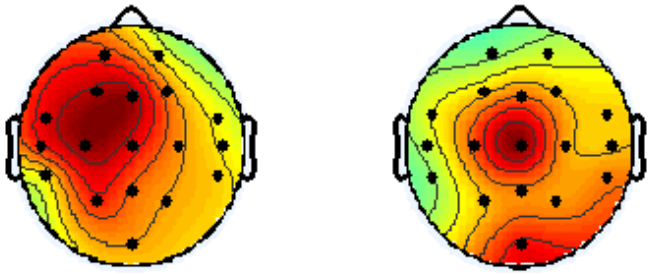

Figure 4.2: *The two selected topographic maps from the data of Subject 1 after the normal ICA. Each map is corresponding to an independent component. Therefore there are two selected components projected to the original measurement space.*

Figure 4.4 shows the template used to compute the correlation coefficient which was obtained from averaged selected P300 trials. These P300 segments were subjectively selected based on wave morphology from the raw data. Figure 4.5 shows the performance based on the correlation between the P300 template and the averaged P300 activities before and after standard ICA. The height of these two colour bars represents how well the raw and extracted signals are correlated with the predefined template. The results after the standard ICA showed a higher correlation than on raw data to the averaged P300 template. Moreover results also indicated that more averages did not necessarily achieve better correlation.

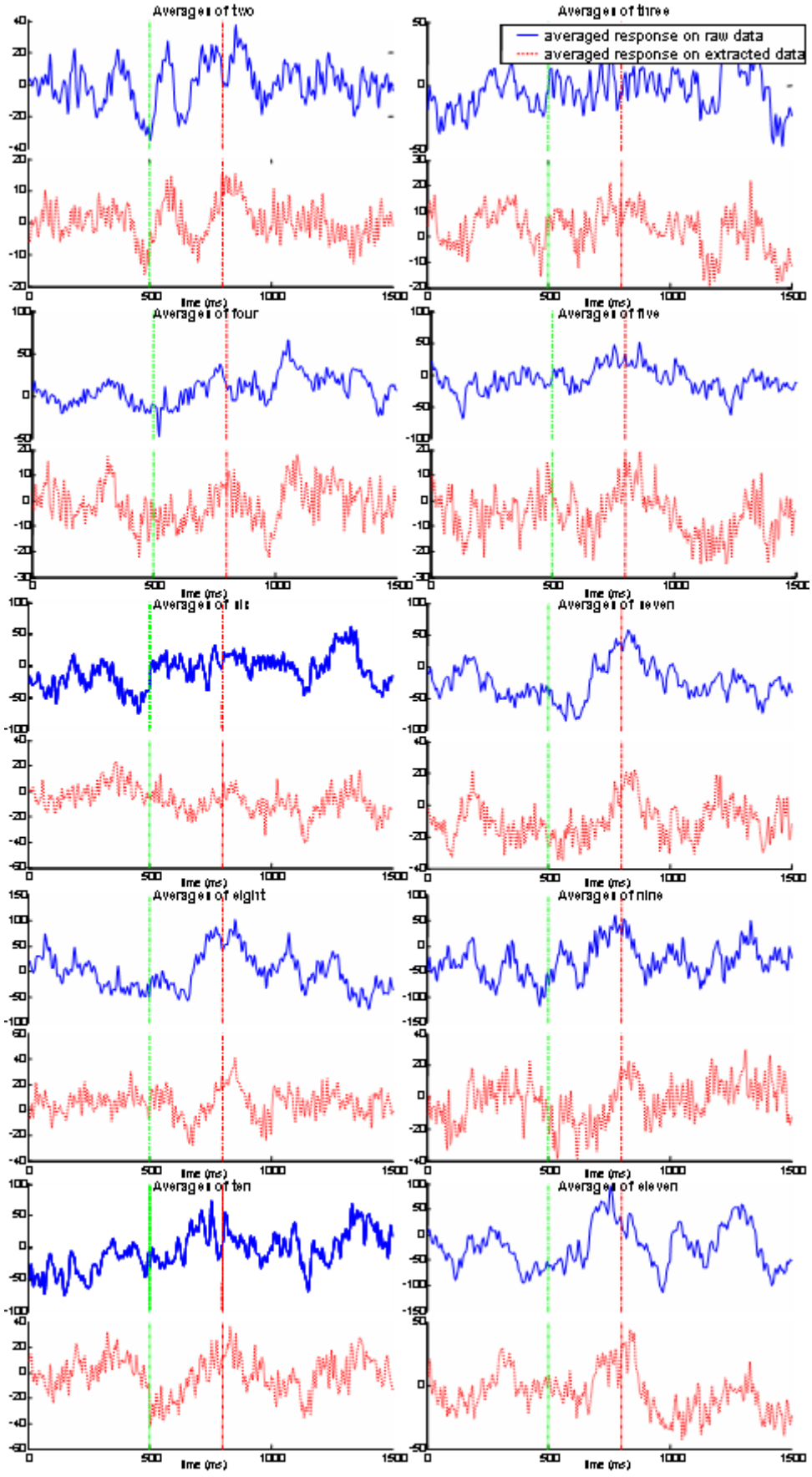

Figure 4.3: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 1 after normal independent component analysis (ICA). For each plot, the upper signal shows the averaged activity on raw data, and the bottom one plots the same average but on the extracted data.*

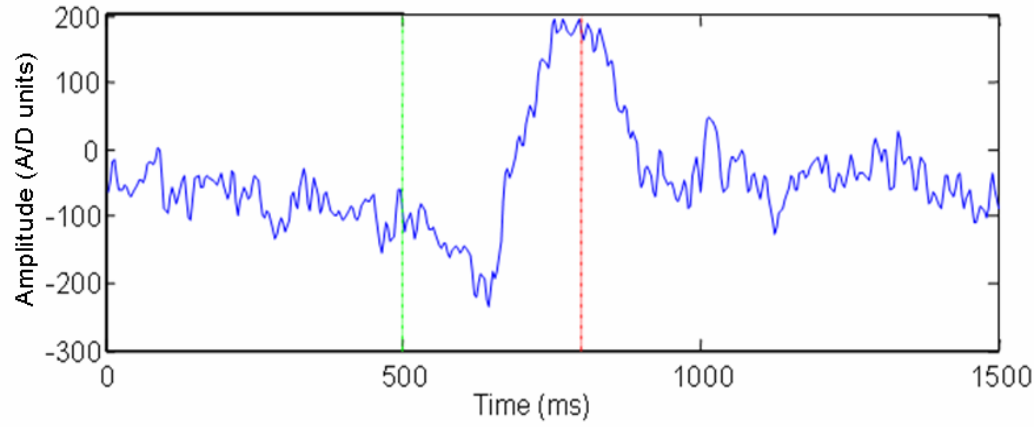

Figure 4.4: *The template used to compute the correlation coefficient with the averaged P300 activities. The template was an average of selected P300 epochs.*

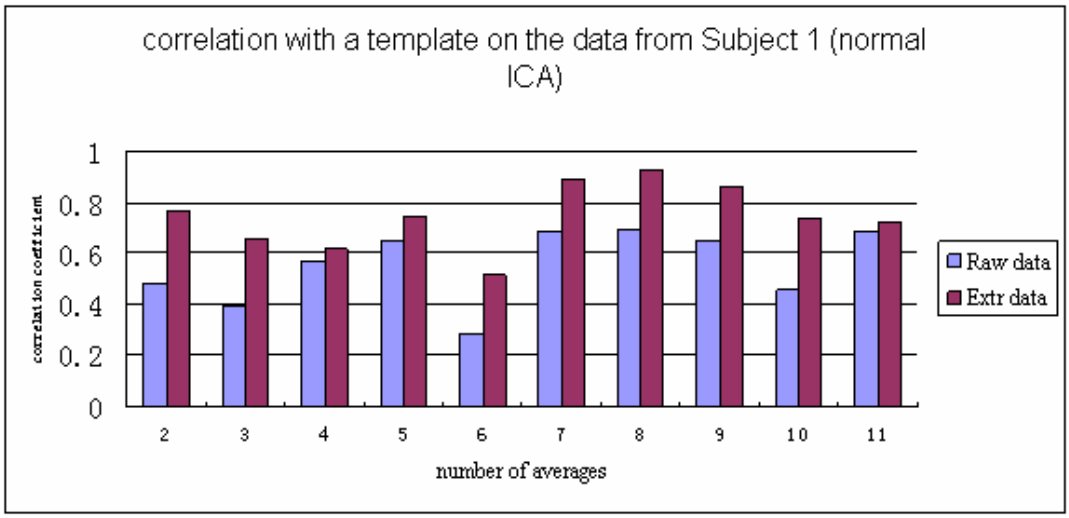

Figure 4.5: *The performance of the correlation between the P300 template and the averaged P300 activities before and after normal ICA. There were a total of 10 pairs of random averages of P300 activities.*

Figure 4.6 shows a single constraint for the proposed spatially constrained ICA on the data from Subject 1. This constraint was a selected column of the mixing matrix after a standard ICA on a selected data portion which includes all the selected P300 epochs. This recomposition of the data reinforced the existing P300s in the data so that ICA is able to maximally extract components of interests.

Figure 4.7 plots the randomly selected averaged P300 activities against the number of averages on raw data and extracted data after spatially constrained ICA.

Figure 4.8 shows the performance based on the correlation between the P300 template and the averaged P300 activities before and after the spatially constrained ICA. Similar to the results as the standard ICA show previously, some noise was reduced and the P300 response was clearly enhanced. The extracted signal showed a clear version of its raw signal. In terms of correlation performance, results after spatially constrained ICA showed higher correlation than their raw data counterpart.

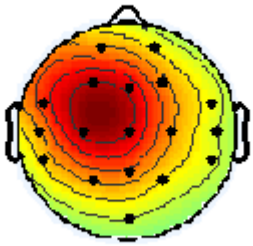

Figure 4.6: *The constraint topographic map for Subject 1. This constraint was a selected column of the mixing matrix after a normal ICA on a special data portion which includes all the selected P300 epochs*

Results of the P300 data analysis on the other two subjects in whom recordings were obtained using the same methods (Subjects 2 & 6) are shown in Appendix C. After the analysis on these three available sets of P300 data, the results indicated that both standard ICA and spatially constrained ICA techniques are able to improve the SNR. However the standard ICA involves a manual component selection which is dependent on user's knowledge and cannot form an automated signal processing system. Although there is the requirement of *a priori* knowledge to initialise the mixing matrix for the proposed spatially constrained ICA, the technique is rather straightforward, and the correct component is projected back to the measurement space each time. The performance from both techniques on these three datasets is almost identical. In comparison with the correlation measurement between the raw data and their enhanced version; a better performance was shown by the averaging on the extracted data than by the averaging on the raw data.

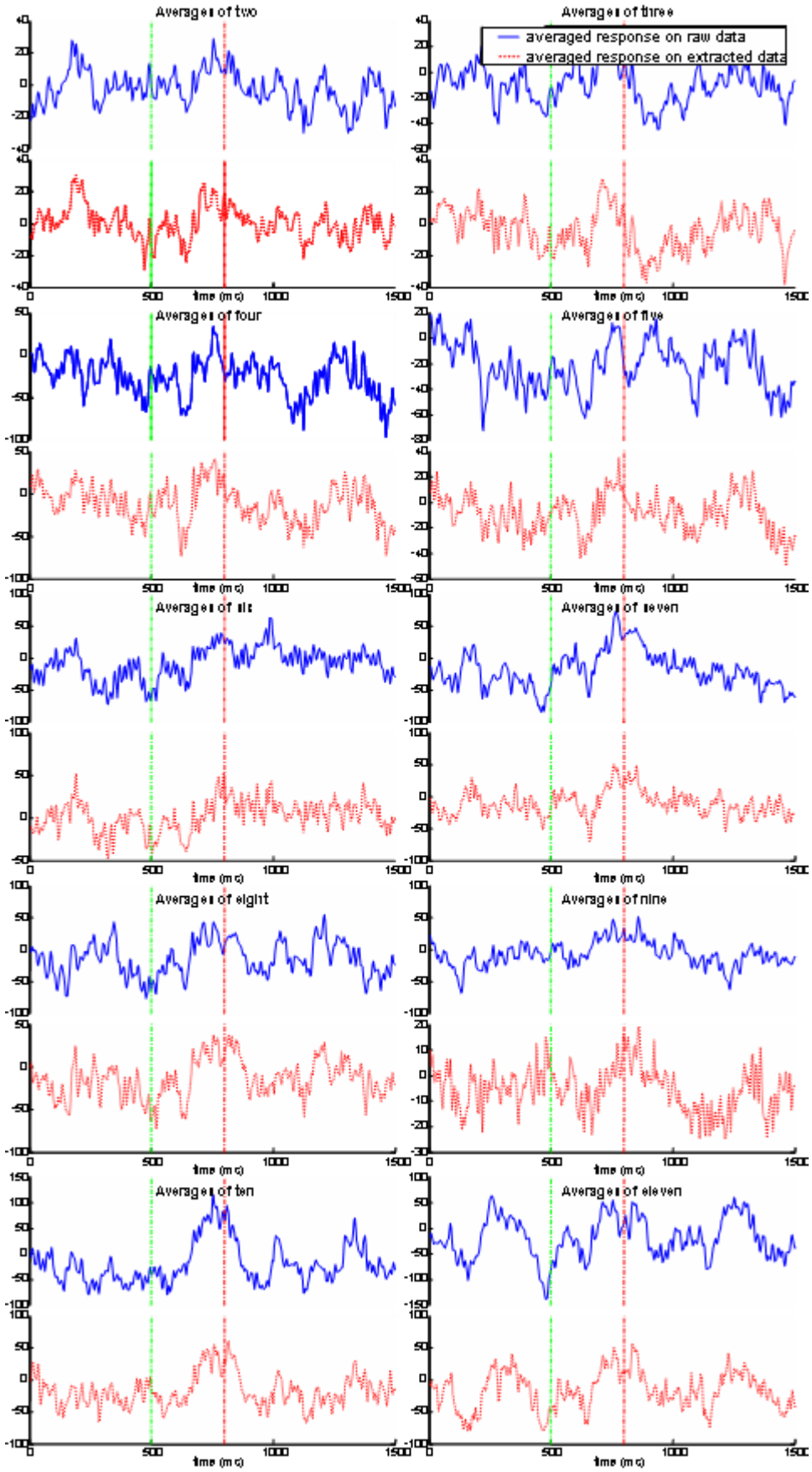

Figure 4.7: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 1 after spatially constrained ICA. For each plot, the upper signal shows the averaged activity on raw data, and the bottom one plots the same average but on the extracted data.*

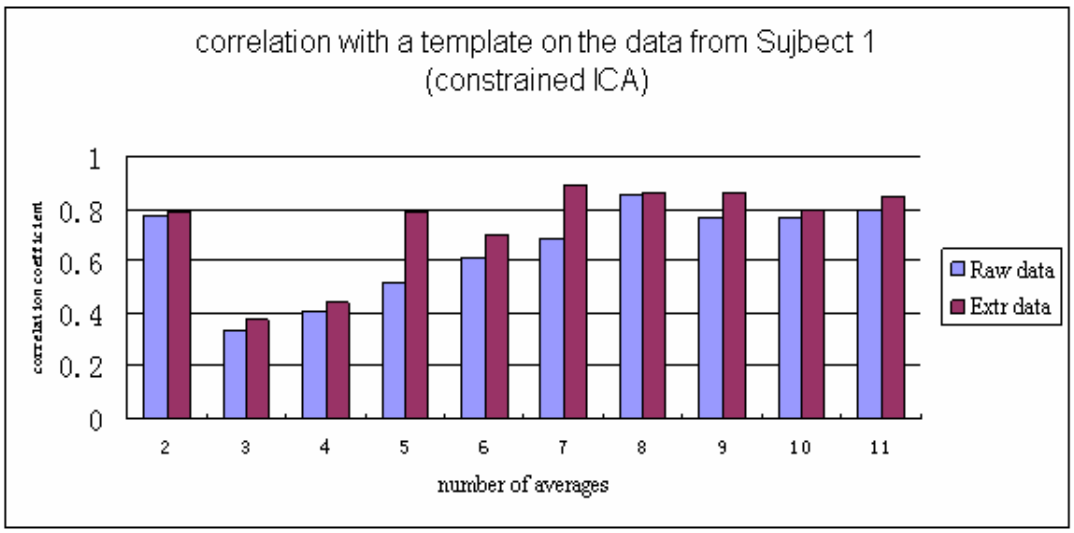

Figure 4.8: *The performance of the correlation between the P300 template and the averaged P300 activities before and after the spatially constrained ICA on the data from Subject 1.*

4.4.2 Results on N400 data

From the literature, a negative peak should occur around 400ms from the centroparietal area for averaged activities from (two-word pair) incongruent word matching; and a decay to the negative (not necessary a peak) should be delayed around 800ms in the case of sentence incongruent matching [197]. The present findings are only for the one subject in recordings from Pz.

Figure 4.9 shows the averaged activity for the sentence congruence/incongruence test after a coherent averaging on the data from Subject 1. The dashed wave was presented as the averaged activity for the sentence correct matching and the bold wave was for the sentence incorrect matching. The stimuli occurred at time 'zero'. However from the results the averaged response for incongruent words did not show changes to the negative from 400ms to 800ms in time. A possible reason for this may be because the stimulus marks manually set for the data were not exactly time-locked to the actual stimulus onset. Based on an assumption that the time between the estimated and the actual value may be delayed within one second, an averaging method with a few samples shifted was proposed on this N400 data. Therefore the process of

averaging is extended to individual components being shifted forward/backward within 250 samples (one second recording) to match the best cross-correlated with a reference activity. In Figure 4.10, the plots show the averaged activities after a few steps shifted. The number of shifting samples were found in the range of 250 samples (within one second of recording) to match the maximum cross correlation value between each response activity and a random selected reference. Since the reference can be any activity from the data, Figure 4.11 shows another possible plot of averaged activities obtained by comparing the cross correlation to another randomly selected activity as a reference. The results after the proposed averaging indicated that the amplitude has a trend to change to the negative from 400ms to 800 ms for the averaged activities of the incorrect matching test.

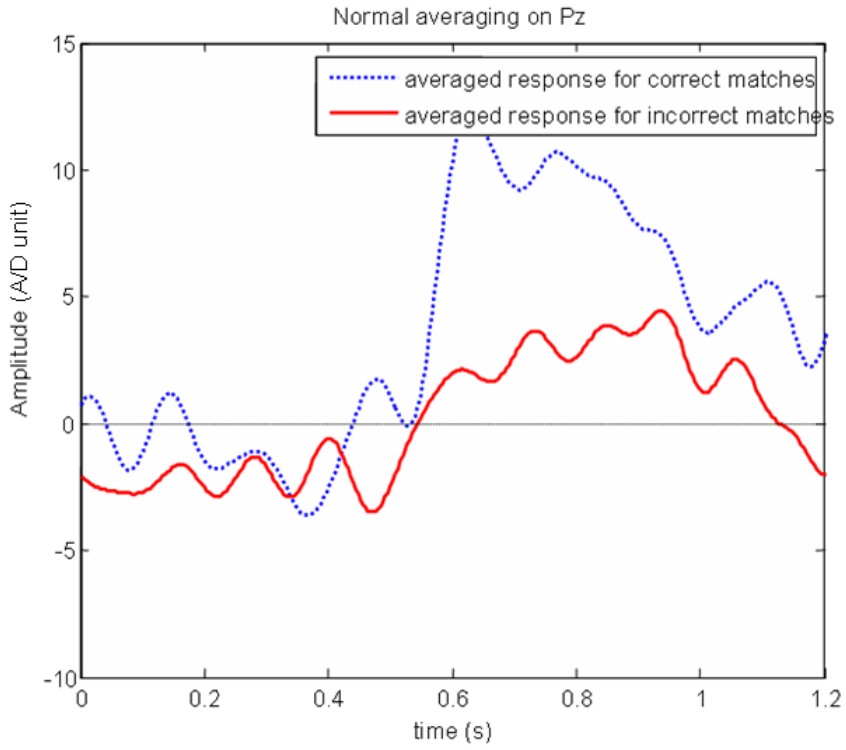

Figure 4.9: *The averaged activity for the sentence congruence/incongruence test after a normal averaging on the data from Subject 1.*

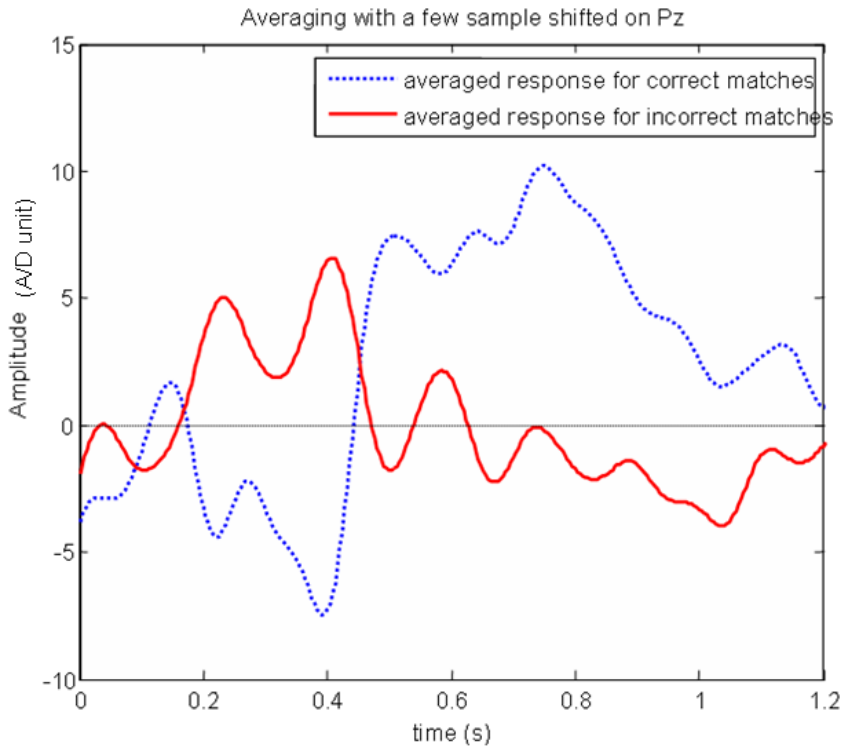

Figure 4.10: *The averaged activities after a few steps shifted. The number of shifting samples were found in the range of 250 samples (within one second recording) to match the maximum cross correlation value between each response activity and a random selected reference.*

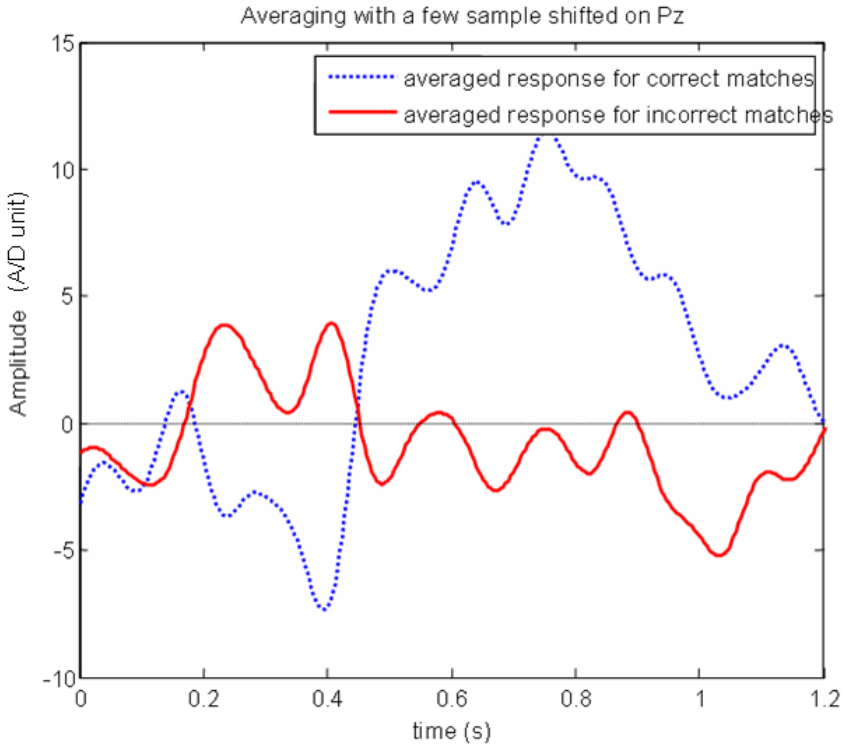

Figure 4.11: *Another possible plot of averaged activities was obtained by comparing the cross correlation to another randomly selected activity as a reference.*

4.4.3 Results on motor imagery data

Results of comparing the averaged power spectra between the response activity and the baseline signal on the contralateral hemisphere (C3 and C4) and ERD/ERS maps for right/left hand grasping imagination are presented below and discussed.

(1) Power spectra

Figure 4.12 shows the averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 7. The bold line presents the power spectra at C3, the bold dashed line is for the power spectra at C4, the thin line is for the power spectra of baseline signal at C3 and the thin dashed line is for baseline at C4. Baseline signals on C3 and C4 aims to represent brain activities when subjects think of nothing and the averaged power spectra of these two recordings was almost identical and stayed in the lower amplitude. Once the thoughts of motor imagery are involved, the output is such that a higher power than baseline signal is presented. Similarly Figure 4.13 shows the averaged power spectra for left hand grasping task from the same subject. However nothing different is shown for averaged power spectra of the baseline signal and left hand imagery across the frequency space. Figure 4.14 shows the above averaged power spectra for right/left hand imagination in one graph. Results using the same measurement for the data from the other six subjects are attached in Appendix C. From the results it is not possible to see a clear difference between these two hand imagery tasks which may suggest the measurement of power spectra on its own is not good enough to assess the performance of motor imagery. As an alternative approach, a mean power spectra within a certain range was proposed to measure the performance. The comparison of mean power spectra in the range from 8Hz to 35Hz over ten trials for all seven subjects together with the standard deviation of the averaged power spectra is shown in Figure 4.15. Ideally the power differences

should be observed on the same location for different hand grasping imagery tasks. However the results did not show much difference between the two different motor imageries. In addition, standard deviation values of power spectra variables for ten repetition trials might indicate that the mean power between 8 Hz to 35 Hz from the individual trial is far from the mean power. This may indicate that the power measurement also is not good enough in the motor imagery tasks.

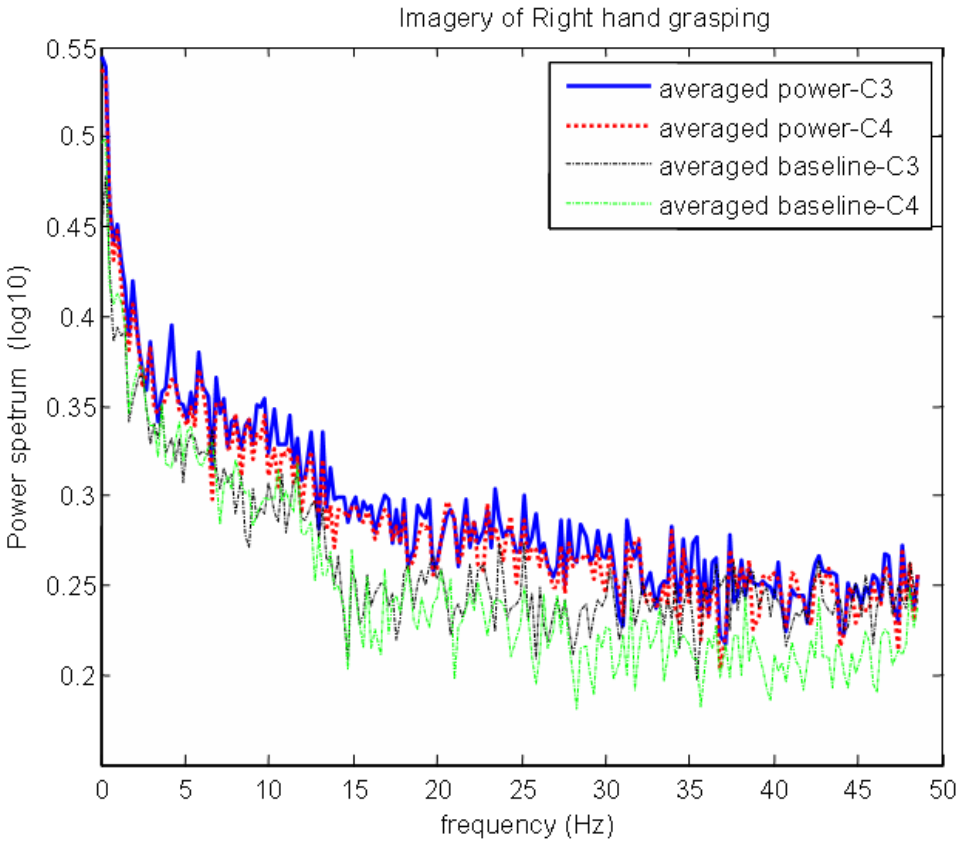

Figure 4.12: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 7.*

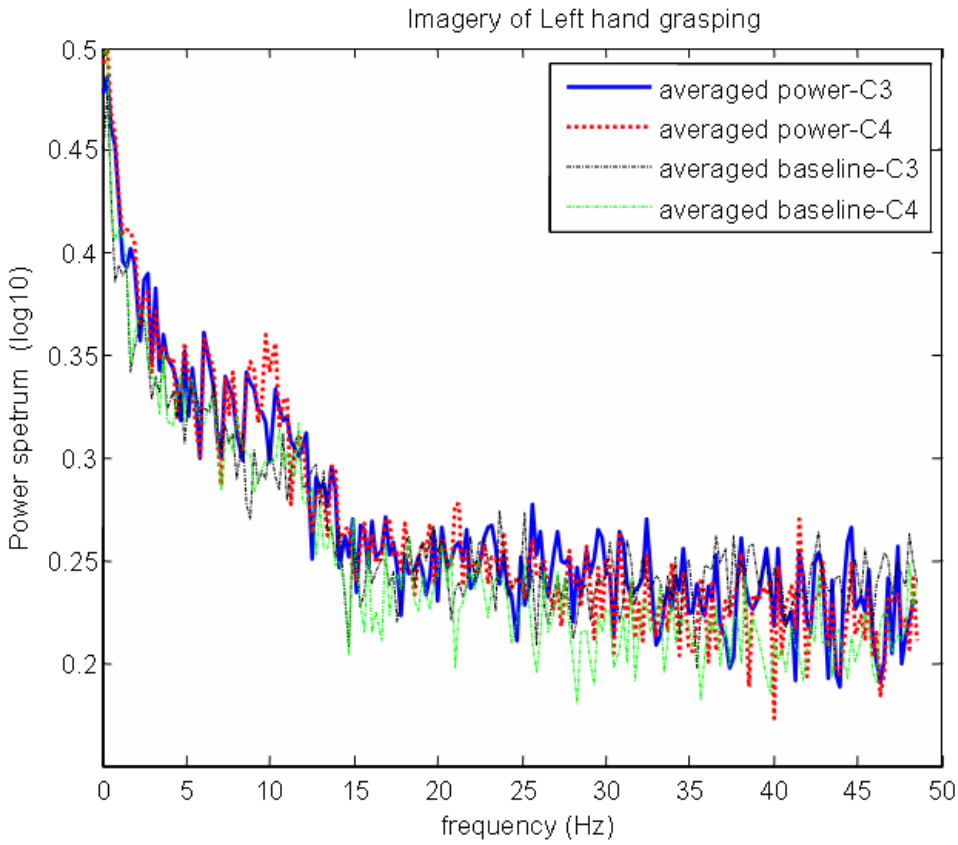

Figure 4.13: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 7.*

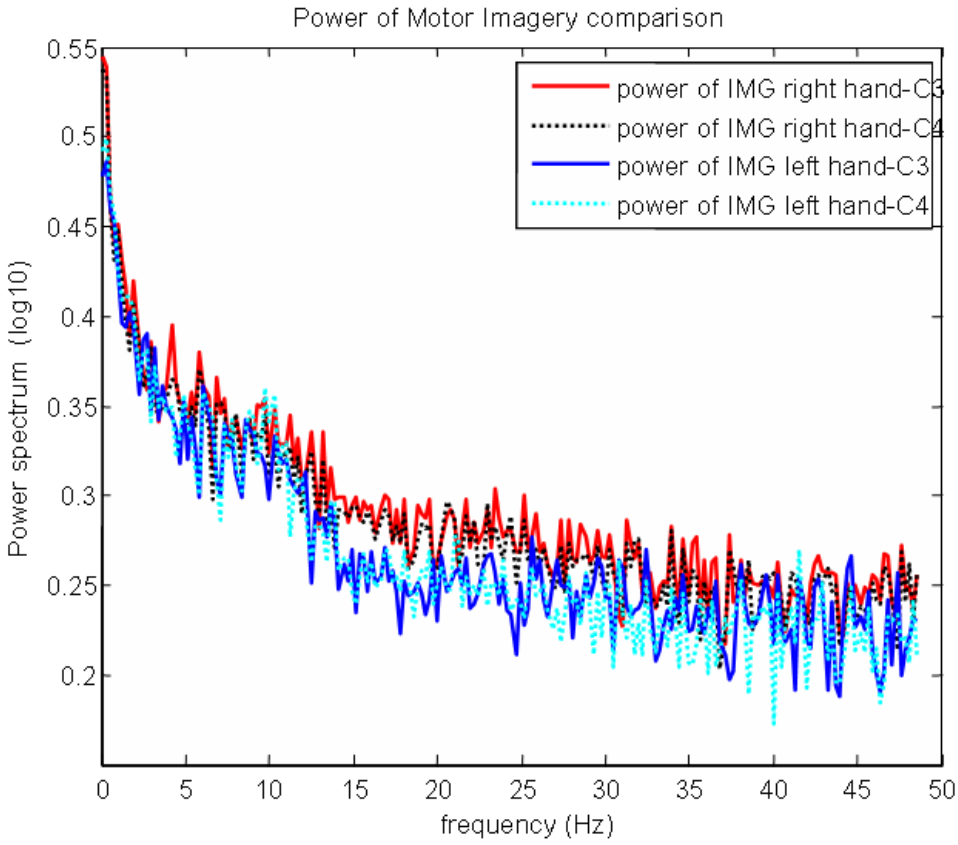

Figure 4.14: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 7.*

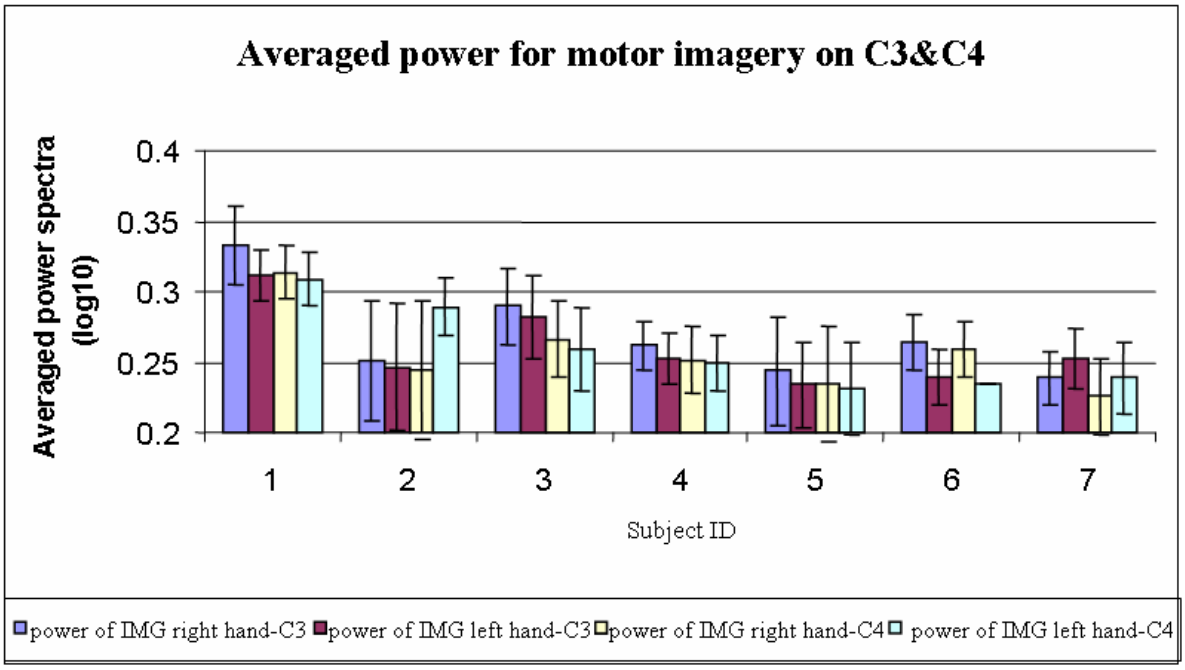

Figure 4.15: *The comparison of mean power spectra in the range from 8Hz to 35Hz over ten trials for all seven subjects together with the standard deviation of the averaged power spectra.*

(2) Event related desynchonisation/synchronisation (ERD/ERS)

As another measurement of motor imagery tasks, ERD/ERS maps may be a good choice to present the changes between two different tasks. Figure 4.16 shows the ERD/ERS maps for right hand grasping imagery on the data from the same Subject 7. The solid line presented the ERD/ERS for the right hand grasping imagination at C4. The dashed line indicated the ERD/ERS for the right hand grasping imagination at C3. The dashed vertical line indicated the stimulus started. For example, right hand imagery would restrain the alpha-range activity on the left hemisphere, as a result, a dashed declined curve after the stimulus was showed to confirm this phenomenon. Figure 4.17 shows the ERD/ERS maps for left hand grasping imagery for Subject 7 too. The results show the distinct ERD curve for about 2~3 seconds after stimulus when the subject imagined the right hand grasping on C3 and an ERS curve presented at the same time at C4; a similar ERD curve occurs in the left hand grasping imagination at C4 and an ERS curve achieved at C3. Results using the same

ERD/ERS measurement for the data from the other six subjects are attached in Appendix C. A summary of the actual ERD/ERS presentations for all the available data was shown in Table 4.2. However from the table, ERD/ERS occurred on both hand imagery tests only in the data for Subject 7. ERD/ERS either on right or on left hand imagery alone were present in four dataset from Subjects 1, 3, 4 and 6. The rest of the signals from Subject 2 and 5 did not show anything useful at all.

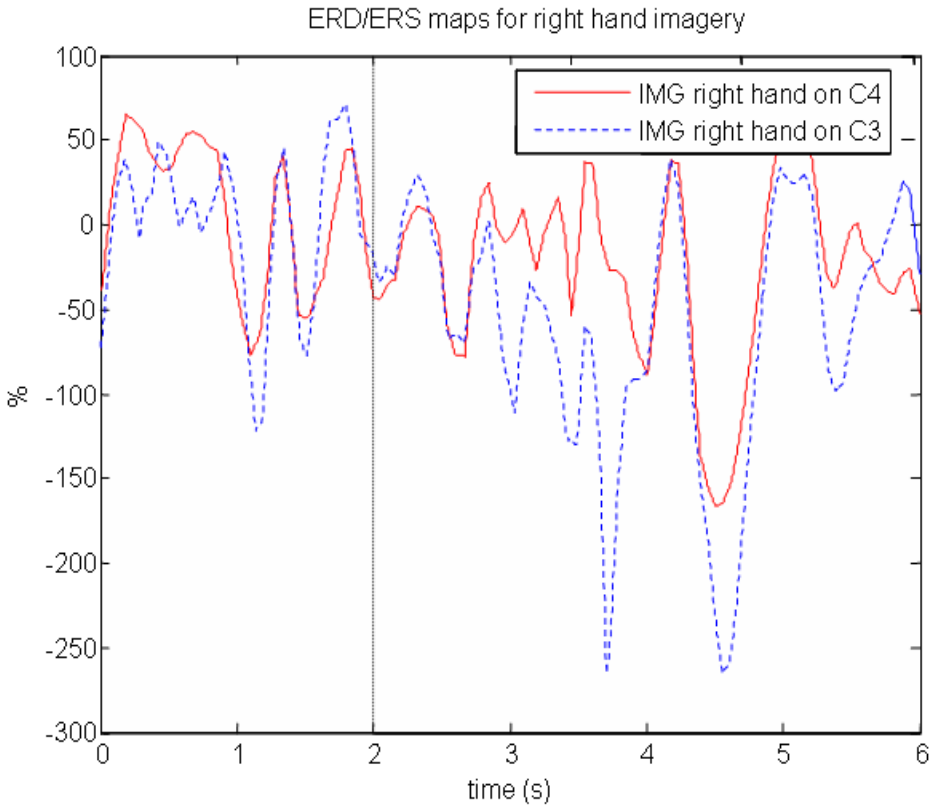

Figure 4.16: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 7. The dashed vertical line represented the stimulus onset.*

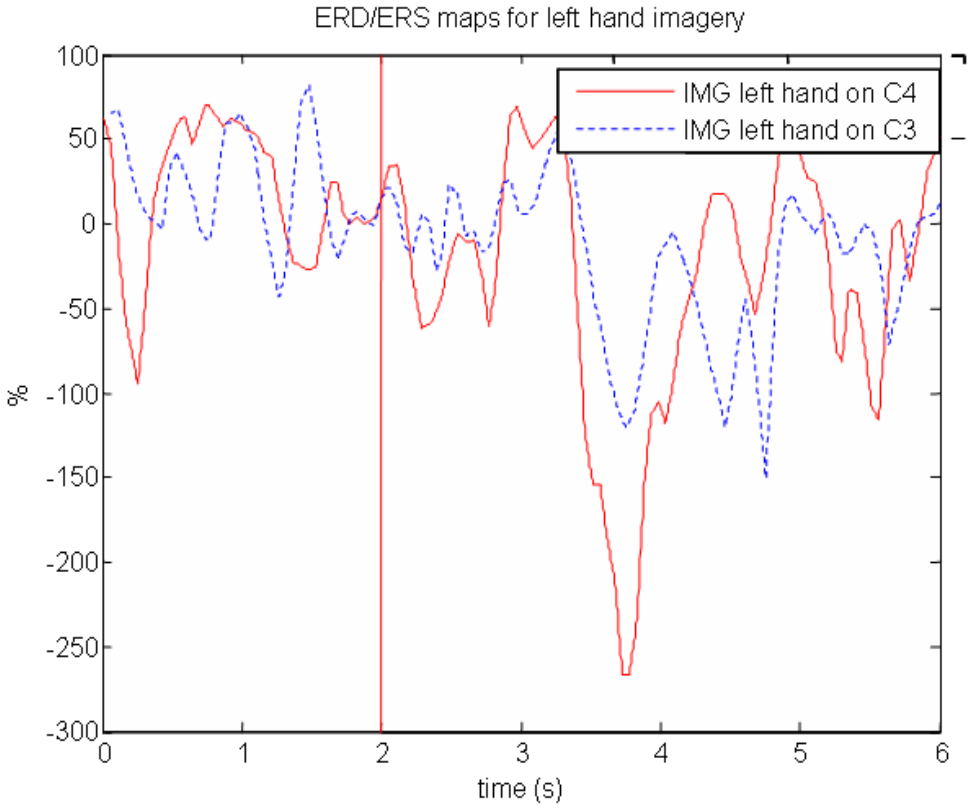

Figure 4.17: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 7. The dashed vertical line represented the stimulus onset.*

Table 4.2: *A summary of the actual ERD/ERS presentations for all the available data.*

Subjects	ERD presents in Motor imagery right hand task	ERD presents in Motor imagery left hand task
1	✓	×
2	×	×
3	✓	×
4	×	✓
5	×	×
6	×	✓
7	✓	✓

4.4.4 Results on non-motor imagery data

In the literature, spatial imagery and auditory imagery tasks caused people to generate the response at the tempero-parietal area [76]. For the spatial navigation imagery, the stronger responses are expected to occur at the temporal area (T3 and T4) while stronger responses can be viewed at the parietal area (P3 and P4) for the music imagery. Results of the averaged power spectra for the navigation imagery and music imagery tasks on the focus area are compared and discussed next.

Figure 4.18 shows the averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 1. Especially for frequencies greater than 8Hz, four power spectra for the navigation imagery and music imagery on sites of T3 and T4 were shown as separable. Figure 4.19 shows the averaged power spectra on the same data but over the parietal area. Similar to results on Figure 4.18, the power spectra can be seen to be different from each other. Results using the same measurement for the remaining six datasets are included in Appendix C. In order to estimate differences in the power spectra, Figure 4.20 and Figure 4.21 compared the mean power spectra in the range from 8Hz to 35Hz over ten trials for all seven subjects together with the standard deviation of these power spectra over the temporal area and over the parietal area respectively.

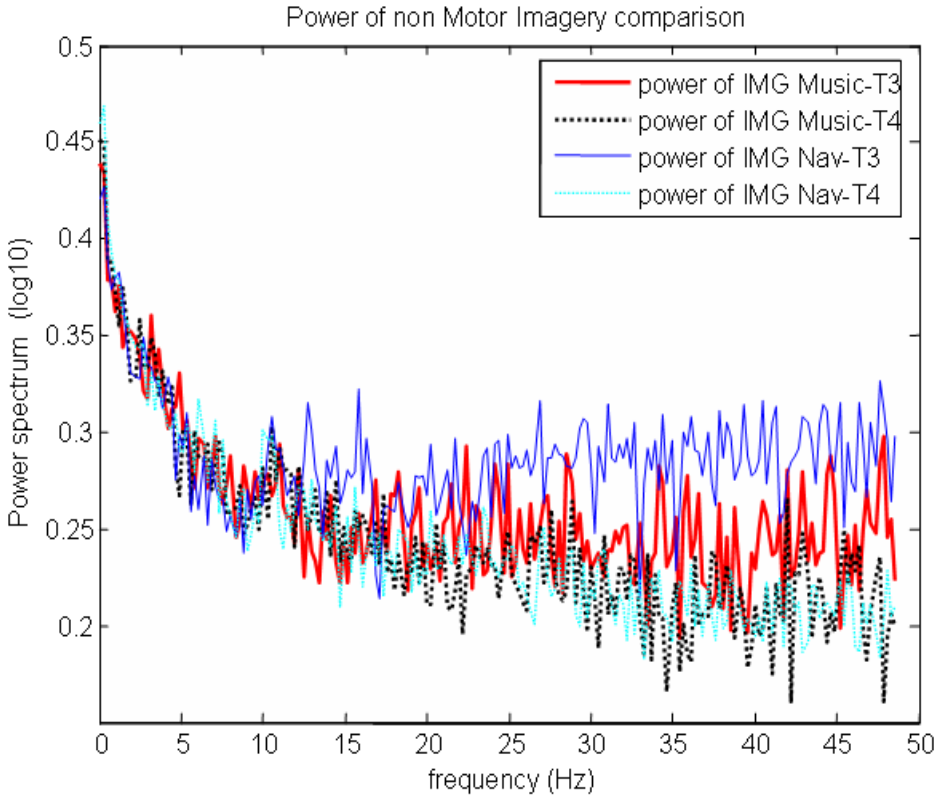

Figure 4.18: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 1.*

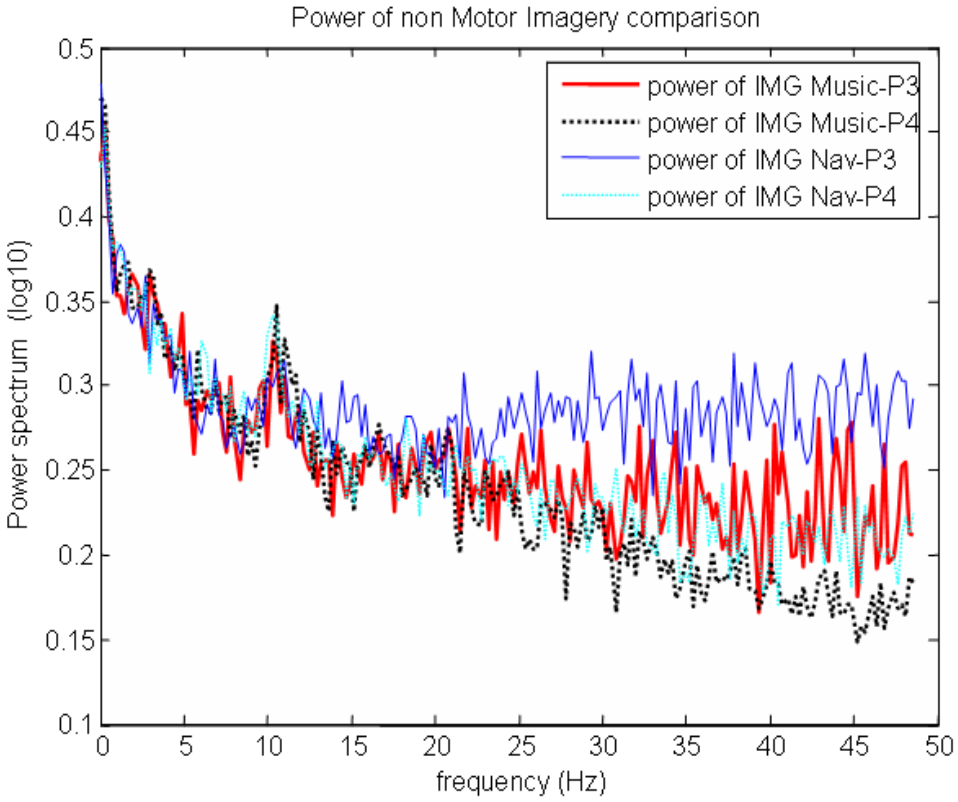

Figure 4.19: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 1.*

From results in all seven datasets, distinct mean power changes where a stronger mean power either occurred in the navigation or in the music task at T3 of the temporal area were observed. In addition, the standard deviation values in Figures 4.20 and Figure 4.21 indicate that how close the band of power from a single trial is to the mean power.

This finding is not as expected as the stronger mean powers should be observed at the parietal area for music imagery task, but the actual mean powers measured showed a distinct magnitude over the temporal area.

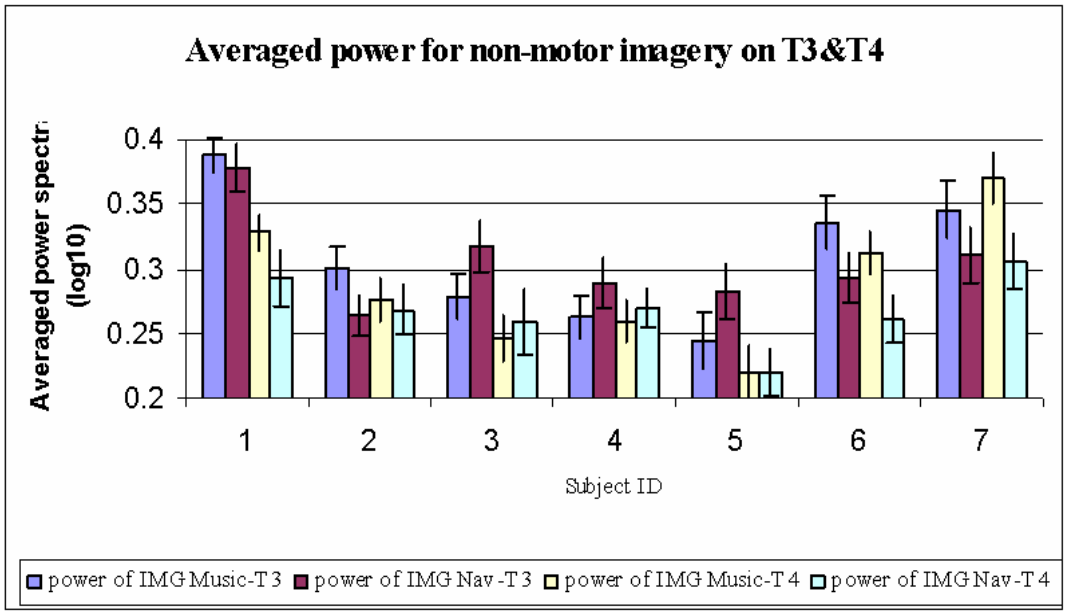

Figure 4.20: *The comparison of the mean power spectra for both music and spatial navigation (nav) tasks in the range from 8 Hz to 35 Hz over ten trials for all seven subjects together with the standard deviation of these power spectra on the temporal area.*

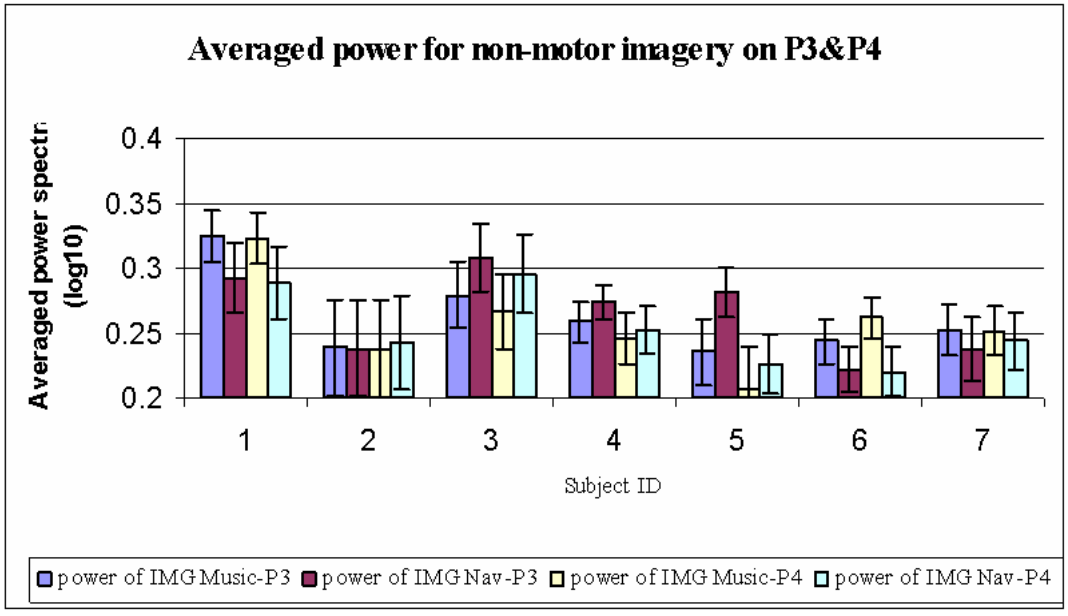

Figure 4.21: *The comparison of the mean power spectra for both music and spatial navigation (nav) tasks in the range from 8 Hz to 35 Hz over ten trials for all seven subjects together with the standard deviation of these power spectra on the parietal area.*

4.5 Discussion

The current findings from this pilot study have demonstrated an ability to run ERP experiments. The tasks in the pilot study include: auditory P300 test, motor imagery tests, non motor imagery tests and N400 test. Seven normal subjects volunteered and contributed several datasets for this study. Depending on the type of task, several signal processing techniques were proposed to analyse the available data.

In the literature, the P300 has become a common choice for psychological tests for clinical and scientific researchers since it is a natural and training free activity for most people [198]. The analysis started on the P300 oddball paradigm where four datasets were available. The signal to noise ratio was significantly enhanced and some artifacts, such as eye movement and blink, were removed after the proposed standard ICA and spatial constrained ICA methods were used through all the available data. However clear and distinct P300 responses could not be detected by proposed methods in the data except for the data from Subject 1. After comparing three datasets, unexpectedly strong noise and possibly imprecise stimulus marks were found in those two unsuccessful data which might be the major reason to cause the P300 detection failures. Furthermore other problems, for example, in the ratio between the number of infrequent and frequent stimulus used in this study was set imprecisely at about 42%, which was much higher than the normal settings (at most 30%) in the P300 literature [199]; the sound hearing level of the auditory stimuli might also not be calibrated within a suitable range and so could also influence the subject's attention to generate related activities.

A normal coherent averaging and an improved averaging technique were proposed and applied on the only available N400 data contributed by Subject 1 for the language (sentence) congruence/incongruence test. The averaged responses for the sentence incongruence test are supposed to show changes to the negative around 400ms to

800ms in the time scale. However due to the fact that the stimulus marks were set manually and therefore not set synchronously to the actual stimulus, the results after the normal coherent averaging did not show distinct changes as expected. After assuming the time delay of setting marks caused by human reaction is possibly around one second, the improved averaging applies averaging with a few samples shifted to match the best cross correlation value to a selected reference activity. Results after the proposed averaging with a few samples shifted (within 250 samples) show a better tendency to the negative from 400ms to 800ms in time. Compared to the results in the literature which show a good negative decay in a similar sentence incongruence test [197], the number of the averages applied in that study was about fifty times whilst in our case the number was ten times. However the disadvantage of this proposed averaging method should be noted that this "peak to peak" matching is to find a best averaging point of the data trial which is maximally cross correlated to a selected reference signal. For example a poor randomly selected reference could mislead to a completely wrong decision in the final averaged activity.

In the literature, cognitive tasks, such as motor imagery have been applied in BCI applications [65] [66] [68]. They can produce a natural response in EEG signals when a movement planned and the response can be detected relatively straightforwardly. The measurement of power spectra and ERD/ERS mapping method were proposed on seven available motor imagery (left/right hand grasping tasks) datasets. Results after calculating the mean power in a particular frequency band showed obscure changes at the same location for two different imagery tasks. However the standard deviation values of power in certain frequency bands for single trials between two different tasks at the same location indicated the pairs of power did not change consistently but variably. Therefore the power spectra alone were not good enough to classify the responses of single trials. After applying ERD/ERS mapping method on the data, results still did not show significant ERD/ERS change for two different tasks. Only

one subject presented the ERD/ERS on both tasks. Two subjects presented ERD/ERS on right hand imagery task while two subjects showed ERD/ERS on left hand task.

The other types of cognitive task, using non-motor imagery, such as spatial navigation imagery and auditory music imagery have been included in this pilot study. The same power spectra measurement was applied on the seven available datasets. In the physiological study, significant changes for the spatial navigation imagery should occur over the temporal area and strong changes for the auditory music imagery should present over the parietal area [200]. Results of mean power in a particular frequency band did show stronger changes in the navigation imagery task rather than in the music imagery task over the temporal area. However in the case of the music imagery there were no significant changes over the parietal area. Despite the fact that the results did not follow the expected physiological hypothesis, significant changes especially at T3, the left temporal area between two different tasks were still encouraging. The standard deviation values of power for a particular frequency band for single trials between two different tasks at T3 indicated each individual power component was tight and close to its mean power. That is to say that even from the power of single trials it is possible to decide their classification. Moreover, results may suggest that it is possible to use power from just one single recording channel (for example at T3) to decide the binary classification for a BCI application.

As a performance comparison between motor and non-motor imagery tasks, this study showed the spatial navigation and auditory music imagery tasks were significantly better discriminated than left/hand grasping tasks. Perhaps the reason is that different ways of performing the cognitive tasks employ different mental processes which make it easier for people to generate the related responses. Poor results of motor imagery tasks may be because left/hand grasping tasks were not functional. More functional motor tests (for example “steer wheel to the left or right”) could perhaps provide more reliable results.

4.6 Conclusion

This pilot study followed the paradigms in the literature to examine the possibility of using evoked potential and spontaneous activity within the Southampton BCI program. The study demonstrated that ICA techniques are able to reduce noise and reconstruct the enhanced P300 components to obtain a clean version of the original signal. Although P300 evoked potential experiments did not succeed in providing good quality data, several P300's advantages, such as, less training and easy processing, are of great benefit for use in BCI system. This study also suggests that the proposed cognitive tasks may be used to drive the BCI system. However the effectiveness of processing may depend on individual subjects and in different applications, for example, one would like to type words at a certain speed or want to control a cursor, say, to turn on/off a TV. Such variability has already determined that a universal BCI system for everyone does not exist. To design a practically working BCI system for the use outside of clinical laboratory, a broader range of reliable tasks should be considered and made available to different subjects to choose for different applications. To achieve this target, more work of enhancing the performance of cognitive tasks in current BCI systems and developing new signal processing techniques to improve current BCIs is required.

This was the first study performed within the BCI research program and served to allow us to get started applying various BCI paradigms as list in the literature.

CHAPTER 5

Multi-channel ICA algorithms for ERP based BCI

5.1 Introduction

The use of ERPs in BCI, especially P300 based word spellers, has become one of the most popular systems in BCI. However due to the difficulty of enhancing the SNR of these recordings, the performance for such systems is lacking. As shown in the previous chapter, ICA has the ability to extract the relevant information within noisy signals quite accurately. Applications of these ICA techniques on P300 based systems provide the opportunity of increased performance.

This chapter proposes three slightly different ICA approaches to solve this problem: (i) 'standard' ICA, (ii) ICA assisted by *a posteriori* template matching and (iii) spatially-constrained ICA. These methods are then applied to the datasets being used and the results are then analysed and discussed. These ICA approaches are able to enhance the SNR such as to improve the final character identification accuracy which is an indicator of how well the system can translate the brain activities. Furthermore, the results indicate that it is possible to reduce the number of epochs required to perform stimulus locked averages, whilst still maintaining good performance measures. This has the potential of speeding up the word speller and has

further implications for use on similar ERP based systems both within BCI and elsewhere.

5.2 The BCI Dataset

To validate and improve signal processing and classification methods for BCIs, some world leading BCI groups organized an online BCI data bank — known as the BCI competition data sets [201]. These data sets are obtained from several popular BCI systems such as SCP based BCI, spontaneous activity based BCI and P300 based BCI. Each data set consists of continuous single-trials of EEG activity, one part labeled (training data) and another part unlabeled (test data). Initially for the purposes of the competition, the labels for test data were not available. After the competition, the labels for testing sets were released and the data sets became available for developing new methods towards improving BCI studies. This chapter uses the BCI competitions 2003 IIb dataset which are recorded from a P300/ERP based BCI word speller [202].

As shown previously in Chapter 2, the ERP has already played an important role and is widely used in cognitive tasks in psychology as well as in BCI research. Within BCI, P300 potentials can provide a means of detecting a user's intentions concerning the choice of objects. Basically a subject is shown a character matrix, the rows and columns of which flash randomly at high speed. Large P300 waves are observed in response only to the flashing of the chosen character.

In this P300 based BCI, the user was presented with a 6×6 character matrix with 36 symbols on a screen. The user then focuses attention on letters of an expected word, one by one. The rows and columns of the matrix are successively and randomly flashed. For each letter of the word, there are 12 illuminations (6 rows and 6 columns) which provide the visual stimulus. Two (one row and one column) out of the twelve intensifications decide a desired character. Figure 5.1 depicts the steps taken in

recognising a character. Each small vertical notch represents each of 12 stimuli. Suppose the user is focusing on the character 'C', then the P300 responses are expected to be evoked by the stimuli 3 and 7 respectively. It is expected that the evoked waveforms are different from those recorded where the stimuli that did not contain the desired character, i.e. a P300 is expected only for stimuli 3 and 7.

Signals are sampled at 240Hz and collected from one subject in three sessions (two labeled sessions and one unlabeled session). Each session consisted of a number of runs. Each run is stored in one Matlab file. In each run, the subject focused attention on a series of characters (Table 5.1). For each character, the word matrix was displayed for a 2.5 s period, and during this time each character was shown blank. Subsequently, each row and column in the matrix was randomly illuminated for 100ms and 6 rows and 6 columns' illuminations resulted in 12 different stimuli. After illumination of a row/column, the matrix was blank for 75ms. Sets of 12 illuminations were repeated 15 times for each character. Thus there were 180 total illuminations for each character. Each sequence of 15 sets of intensifications was followed by a 2.5 s period, and during this time the matrix was blank. This period informed the user that this character was completed and to focus on the next character in the word that was displayed on the top of the screen.

The goal of this study is to use the labeled data in the training data sessions to learn a certain subject oriented information to be used to predict the words in the testing session. This testing session was unlabeled for the purpose of the competition and after this BCI competition unlabeled words were released so that all data are fully labeled for the use by other BCI research group. A series of target characters in the testing data sessions are shown in Table 5.2.

The data was originally recorded from 64 scalp electrodes. The number of channels was manually reduced to simplify the analysis and also decrease the

computational complexity. 19 channels were selected based on the placement of standard International 10/20 System which include Channel C3, Cz, C4, Fp1, Fp2, F7, F3, Fz, F4, F8, T3, T4, T5, P3, Pz, P4, P6, O1 and O2. The reason to select such an electrode placement is that 10/20 system is one of the most common electrode setup methods and the selection adequately covers the central and parietal sites in particular which are the P300's active fields [59]. The dataset is not subjected to any additional preprocessing and after analysis the results are lowpass-filtered for visualisation purposes only.

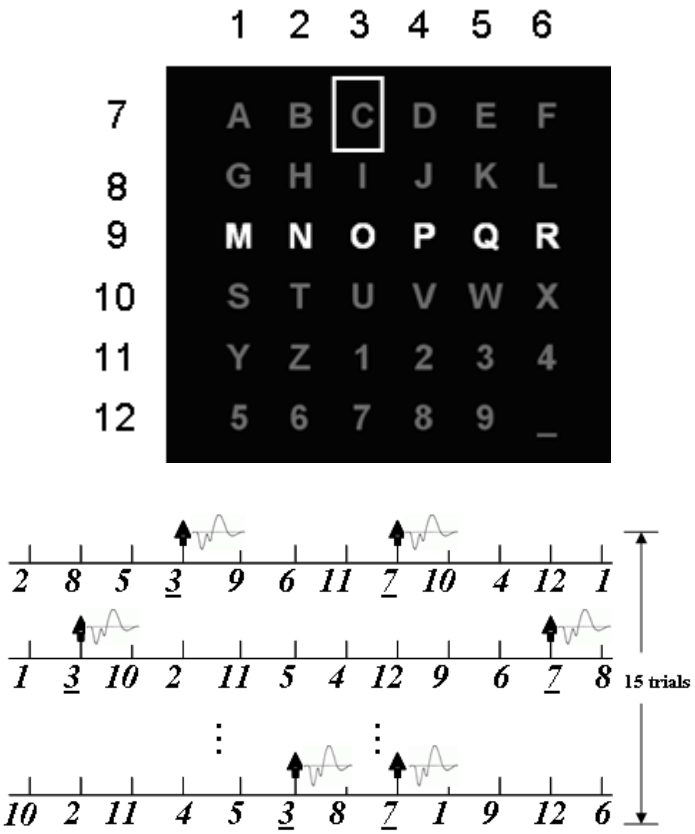

Figure 5.1: *A sketch of character 'C' recognition in the P300 based word speller. (a) Suppose the user is focusing on the character 'C' which is located in Column 3 and Row 7. Flashing Columns and rows form Stimuli. (b) Two responses evoked by stimuli 3 & 7 point to this character and 15 trials would reinforce this decision.*

Table 5.1: *The target word for each run in labeled (training) Session 10 and 11.*

Session	Run	Target character
10	1	CAT
10	2	DOG
10	3	FISH
10	4	WATER
10	5	BOWL
11	1	HAT
11	2	HAT
11	3	GLOVE
11	4	SHOES
11	5	FISH
11	6	RAT

Table 5.2: *The target word for each run in unlabeled (testing) Session 12.*

Session	Run	Target character
12	1	FOOD
12	2	MOOT
12	3	HAM
12	4	PIE
12	5	CAKE
12	6	TUNA
12	7	ZYGOT
12	8	4567

5.3 Signal enhancement methodologies

This section contrasts a basic signal enhancement technique, coherent averaging, with three new ICA based algorithms.

5.3.1 Coherent averaging

As already introduced in Chapter 2, coherent averaging is a classic signal enhancement method especially for time-locked signals [86]. The random flashing of each row and column for the word matrix in this P300 based BCI is considered as a time-locked stimulus, the information about when exactly the stimulus appeared is stored in the recorder data. The number of different stimuli is twelve, therefore there will be the potential of twelve evoked responses to these twelve stimuli. This set of twelve stimuli is repeated fifteen times in random order. Based on the assumption that the background noise is random and incoherent across data trials, through coherent averaging the brain responses for the same stimulus, it is possible to reduce the noise and enhance the SNR of the data. Here the coherent averaging is applied at Cz which is the most active site for P300 activities [59].

5.3.2 Standard implementation of ICA

A standard implementation of ICA is first applied to the data. The details about this ICA algorithm have been presented earlier in Chapter 3. The FastICA algorithm and an online downloadable toolbox are used in this experiment [150]. In the literature FastICA is one of the most popular ICA techniques since it has fewer parameters to choose and has a fast rate of convergence. The process of this standard ICA can be separated into three steps: component selection, component projection and character detection.

(1) Component selection

After ICA decomposition, the underlying sources are separated into separate ICs. Based on the knowledge from the scalp topography (the columns of the $\mathbf{W}^{-1}$ matrix) of the components and the knowledge of the P300's physiological origins in which

the P300 appears at the vertex region (Fz, C3, Cz, C4, Pz), normally maximised at Cz, one component containing the P300 pattern is manually chosen based on the topographic maps and the IC morphology itself.

(2) Component projection

The selected source activity needs to be projected back to the observed measurement space and the projection process brings the selected source back to the original data space and ensures that it is back to the correct scale. The detailed projection process is introduced in Chapter 3. After this projection, clearer P300 activities can be seen over the noisy EEG data.

(3) Character detection

There are many ways to detect the P300 potential in the 'de-noised' EEG and predict the characters in this P300 speller. In order to evaluate the ICA performance, a very simple linear detector within a window from 300 ms to 360 ms on Channel Cz is considered (Figure 5.2), in which two peaks with the maximum amplitude values out of twelve waveforms are selected as candidate P300's.

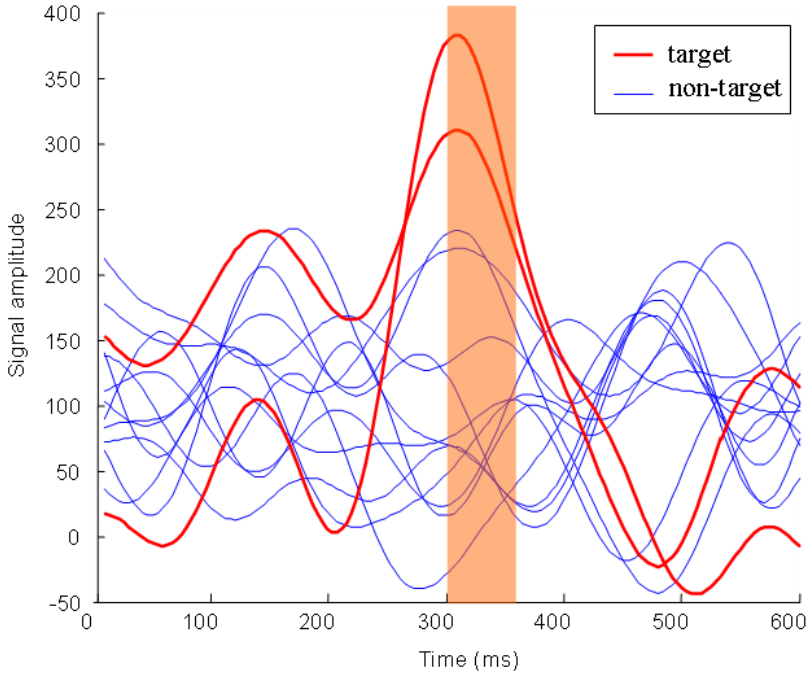

Figure 5.2 *An illustration of detection of target P300 activities over twelve evoked brain activities, the orange area represents the window used to detect the target activities. Two peaks with the maximum amplitude in the window are selected as candidate P300's.*

5.3.3 ICA with template assisted component selection

The method of Section 5.3.2 is proposed to select and project one related component only. However it is an ideal assumption that ICA can completely separate the different relevant information into the different independent sources. In practice relevant information can be distributed into one or several ICs, i.e. the ICA processing is not perfect, resulting in least dependent components. In this section a new algorithm is introduced to automatically select one or more related least dependent components. There are two main steps to achieve this algorithm: creating a template and computing the correlation.

(1) Creating a template

Since the columns of $\mathbf{W}^{-1}$ describe the topographical information of the each independent component, where the spatial information can help to identify related underlying sources, the template used here is a column selected from $\mathbf{W}^{-1}$ which is obtained from running ICA on a balanced data set. As two training datasets are available fully labelled, the target/non-target ratio is balanced by replacing non-target epochs with target epochs and so increasing the original ratio of target to non-target from 1:5 to 1:1 (20 epochs of P300 response signals and 20 epochs of non-P300 signals). This modification leads to the P300 relative projection strengths being enhanced. After running ICA on this data, a topography which presents maximal signal intensity at the vertex region is selected manually as the template among the scalp topographies.

(2) Computing the correlation

After applying ICA to the original testing data and the inverse unmixing matrix $\mathbf{W}^{-1}$ is obtained, the correlation values between the template obtained from the first step and each column of $\mathbf{W}^{-1}$ is used to assist in choosing the related P300 components. After computing the correlations, the set of correlation values which describe the degree of spatial relationship between the template and each IC are found and rank ordered. The first three highly ranked components are automatically selected and projected into the measurement space. The fixed number three was tuned and found empirically in the training data sessions. The enhanced data is here classified by using the same simple linear character detection method described above.

5.3.4 Spatially-constrained ICA

As discussed in Chapter 3, if *a priori* knowledge about the spatial location of some of the sources is known, then it can be applied to the ICA algorithm by constraining the column(s) of the mixing matrix. The concept of spatially-constrained ICA which uses this idea has already been explained in Chapter 3. Here, as the P300 is used and its typical spatial distribution is known, and then P300 scalp topography information could be applied as a spatial constraint to initialise a column of the mixing matrix for spatially-constrained ICA algorithm.

As is known from the spatially-constrained ICA algorithm, the spatial constraint may be obtained from a source decomposition of previous data by using standard ICA methods or any other suitable method. In this case, the template obtained in section 5.3.3 is used as a single spatial constraint to initialize the first column of the mixing matrix **A** and set random values for the rest of columns. After spatially-constrained ICA, the corresponding component is projected to the measurement space and the same simple linear character detection is applied to the enhanced data.

5.4 Results

Denoising and decomposition by ICA and the final classification were all processed on the raw (unfiltered) data. The results presented in this section are averaged for 15 times. Moreover purely for visualization purposes (i.e. after all processing is complete) the results are also presented after the application of a low-pass FIR filtered (cutoff at 10Hz) version. In order to evaluate the different performances fairly, all the examples in the following figures are shown on the same character. Figure 5.3 shows 12 response activities corresponding to the 12 stimuli (coherently averaged). The two thick red lines represent the targets and the other blue lines represent non-target responses. It is clear that it is difficult to distinguish P300 targets directly from this

data. The distinct P300 waveforms after extraction using the manual component selection method are shown in Figure 5.4, the two target waveforms clearly show P300 responses such that the two targets can be identified clearly from the other ten responses. Figure 5.5 illustrates the performance of ICA with the template assisted selection method; the three topographic maps are most likely correlated to the (automatically selected) P300 template components. Figure 5.6 depicts the performance of ICA using the fixed spatial constraint. As can be seen from the figures, the three types of ICA implementation extract the related responses very effectively.

Due to the fact that ICA is able to reduce the noise and extract the components of interest from the signal at the same time, this means ICA has also optimized the feature extraction stage. A very simple classification method on the extracted data set was used. Using these techniques a maximum accuracy of 96.8% (30 out of 31 characters) was achieved on the test set. The classification accuracy is defined as the percentage of the number of correct classified characters over the total number of characters. It is also equivalent to the sensitivity statistical measures [203]. One of the winners for this BCI competition dataset achieved 100% accuracy by using an advanced classifier based on SVM [204] technique. The method uses a new formed training dataset to train the proposed SVM classifier. This dataset is manually selected two P300 target response signals and two non-P300 signals of every character in provided training datasets (in total 168 epochs). Moreover in order to obtain a good performance, it has to carefully select parameters for this SVM classifier based on the knowledge about the application and the SVM classifier itself. Whilst the standard ICA does not need the training data, it is straightforwardly applied on the testing datasets. As for the template assisted ICA and spatially constrained ICA, the selected training dataset only has just 40 epochs and the step of P300 topography (template and constraint) selection is very simple (visually choose one with a strong activation over the P300 physiological origin). Therefore ICA techniques are comparable in speed and efficiency with the SVM methods. Moreover on looking into the

misclassified character after the experiment, a misclassification appears to identify the target character of 'E' to 'C'. In the speller matrix, these two characters are placed at the same row (the seventh) but at different columns ('E' for the fifth and 'C' for the third). Therefore the averaged evoked response for Stimulus 3 appeared to be the one of the target activity with high amplitude. Since the proposed simple classifier straightforwardly determines two activities with highest amplitude within a window, so the classifier detected the correct row activity but missed the correct column one.

The performance was also examined when reducing the number of averaging trials before detection takes place. Figure 5.7 and Table 5.3 summarize and compare the performance of the three different ICA-based methods with those obtained by using simple coherent-averaging whilst varying the number of epochs used in the averaging process.

In the experiments, signal enhancement based on ICA plays a very important role. After ICA, a simple linear classifier is used to classify target and nontarget responses. A time window around the 300ms is applied and simply detected two maximum responses to estimate a character. Three proposed ICA methods achieved the source separation goal well. In BCI applications, especially for an online system, manual selection of the P300 component trial by trial obviously is not possible. With the assistance of a template obtained from the training data, the P300 related components can be automatically selected by estimating the correlation as shown in Section 5.3.3. Spatially-constrained ICA uses the template previously obtained as the constraint to initialize the mixing matrix which will increase the rate of convergence of the algorithm to the related sources. After Spatially-constrained ICA, the component(s) containing P300 patterns are separated and placed in the first position directly by the algorithm. It offers a more practical method to implement an online BCI system and shows the same high accuracy as all of the ICA based methods.

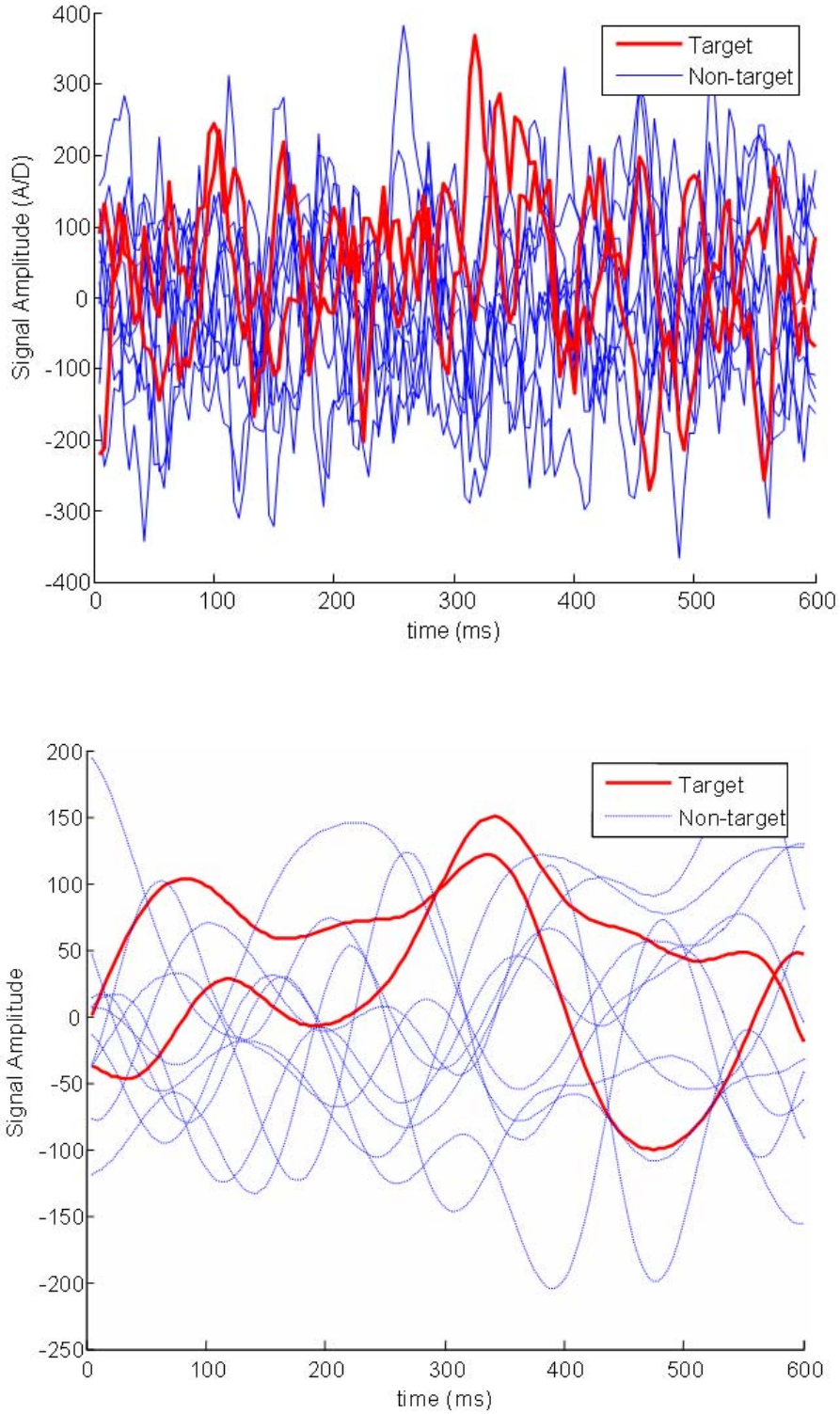

Figure 5.3: *The above plot is the 15-epoch averaged response curves on Cz on raw data. The two thick lines (red) are the targets. The bottom plot is the lowpass filtered version. They are difficult to distinguish directly from the data.*

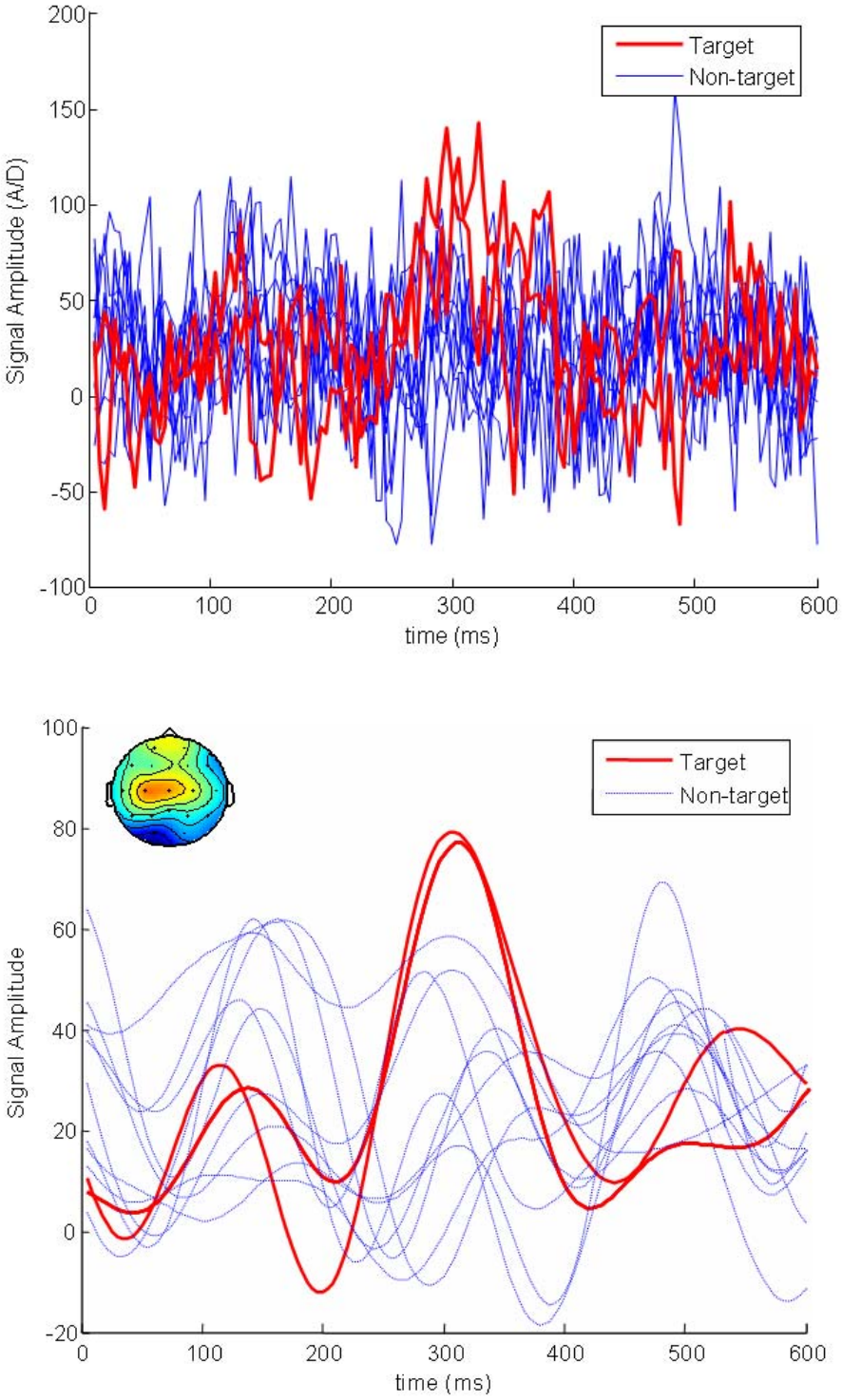

Figure 5.4: *The above plot is the 15-epoch averaged response curves on the extracted data after ICA. The bottom plot is the lowpass filtered version and the topographic map (above left) represents the single project component. Two targets can be identified from the other ten responses.*

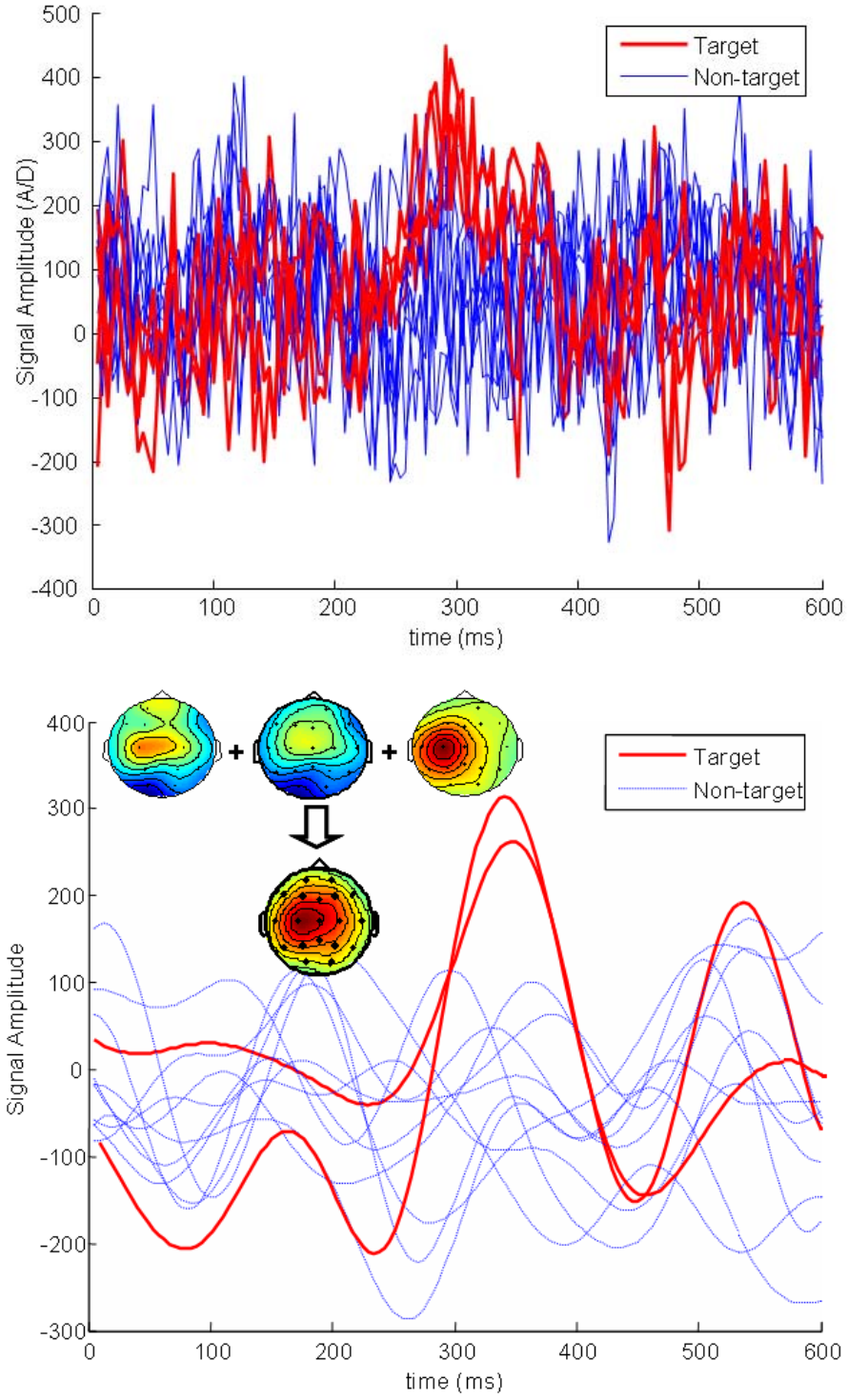

Figure 5.5: *The above plot is the 15-epoch averaged response curves on the extracted data after ICA with template method. The bottom plot is the lowpass filtered version and the inset three topographic maps represent the three projected components. These three form a superimposed map which shows the P300 patterns in this combination making them clearer to see.*

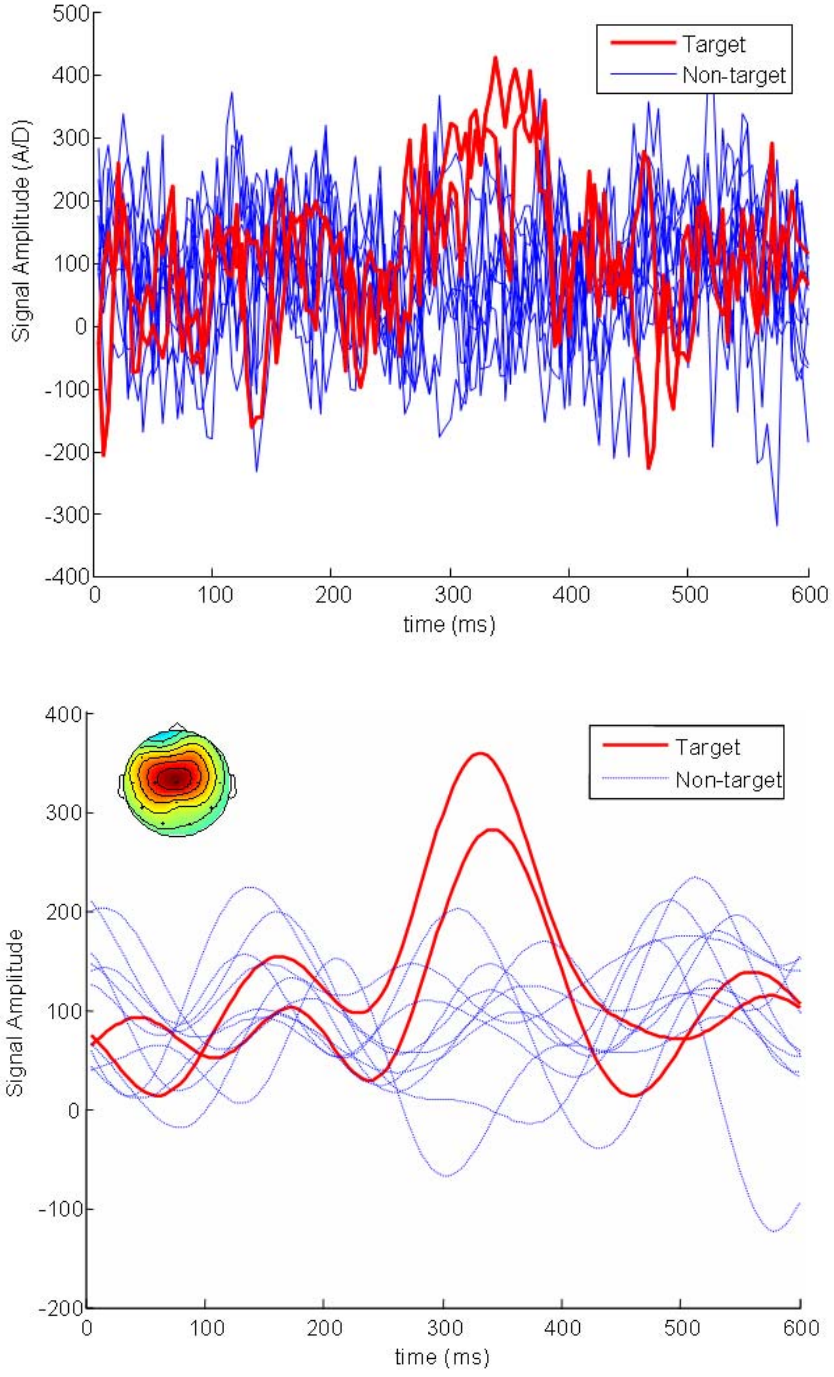

Figure 5.6: *The above plot is the 15-epoch averaged response curves on the extracted data after spatially-constrained ICA. The bottom plot is the lowpass filtered version. The spatial template (represents the projected component) is shown inset at the top left.*

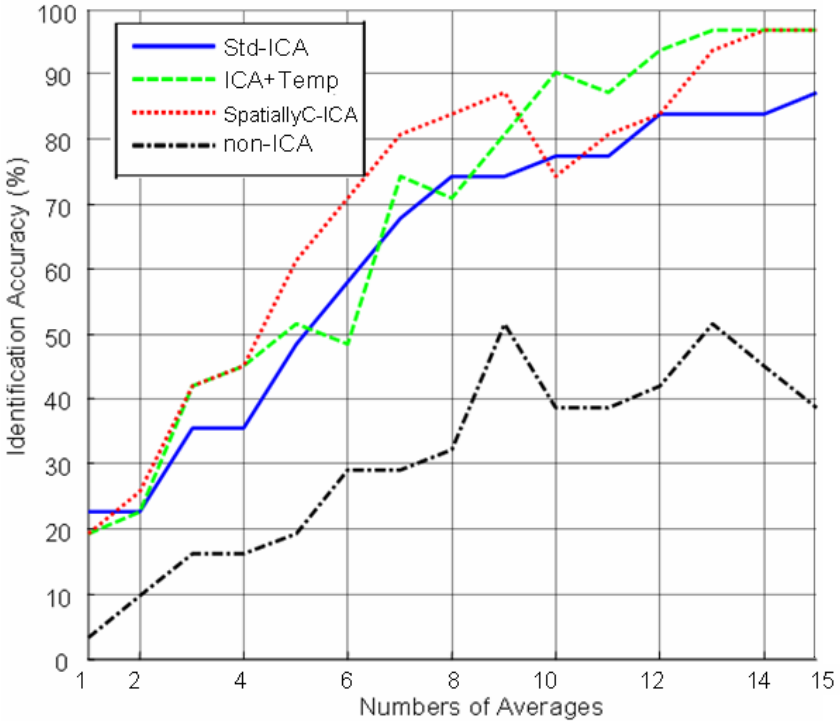

Figure 5.7: *Comparison of classification accuracy with different averages by using averages (Non-ICA) and three proposed ICA methods. The blue line shows the classification performance by standard ICA, the green line represents the performance by ICA plus a template assisted, the red line indicates the classification accuracy by spatially constrained ICA and the black line is for the non-ICA.*

Table 5.3: *Comparison of classification accuracy (%) with different averages by using averages (Non-ICA) and three proposed ICA methods.*

Numbers of Averages / Methods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
non-ICA	3.23	9.68	16.13	16.13	19.35	29.03	29.03	32.26	51.61	38.71	38.71	41.94	51.61	45.16	38.71
Std-ICA	22.58	22.58	35.48	35.48	48.39	58.06	67.74	74.19	74.19	77.42	77.42	83.87	83.87	83.87	87.10
Spatially C-ICA	19.35	25.81	41.94	45.16	61.29	70.97	80.65	83.87	87.10	74.19	80.65	83.87	93.55	96.77	96.77
ICA+Temp	19.35	22.58	41.94	45.16	51.61	48.39	74.19	70.97	80.65	90.32	87.10	93.55	96.77	96.77	96.77

5.5 Summary

This chapter introduces the use of ICA in P300 (ERP) based multi-channel BCI systems. The results clearly show a significant improvement after ICA when compared with the classification results obtained from the raw coherent averaged data, even a very simple classifier can achieve very good classification accuracy. The results also show that even for fewer averages the ICA based techniques still exhibit quite good performance – this, coupled with the ability to use a simple linear classifier, means that this has the great potential for speeding up the word speller paradigm and has further implications for use on after ERP based systems – both within BCI and for other clinical use.

The next chapter introduces an ICA application on spontaneous activity based BCI. This BCI system uses spontaneous rhythmic activity rather than time locked evoked responses as the control features. However due to artefacts and other problems within the EEG recordings, the performance of such a BCI system is usually quite mediocre. A novel ICA method is proposed to work towards solving these problems.

CHAPTER 6

A multi-channel ICA algorithm for spontaneous EEG based BCI

6.1 The BCI paradigm

As discussed in Chapter 2, spontaneous activity based BCI has shown great potential. It uses more natural rhythmic brain activities evoked by imagining (say) limb movement which is believed to be easy to learn and understand during the training sessions. The power changes within a specific rhythmic band can be used as the control features, for example, where the user generates these patterns to control the movement of a cursor on a computer screen. In this way the system offers the user at least two or even four degrees of freedom to move a cursor on a screen. The features can also be extended and applied to other assistive systems such as to drive a powered wheelchair or a robot arm etc. However, as has already been said, the EEG is recorded from surface electrodes, and artifacts contaminate the raw EEG recordings and render the unprocessed recording virtually unusable. Moreover, as the prediction of imagery is dependent on the individual's ability to generate a good ERD, such a BCI will mutually exhibit variable performance.

The studies shown so far have applied ICA to BCI applications [205] [206] [207], for example, ERP based BCI and spontaneous activity based BCI. The results indicate that ICA performed well in extracting *time-locked* ERPs. However since spontaneous activity based BCI does not use time-locked activity but rather relies on rhythmic activities as features, applications of ICA are unable to track the changes in power spectra among the different sources. However using time structure based source decomposition methods, the sources with stationary waveforms and unique power spectra can be isolated. Furthermore, when the power spectrum of the particular source activity is known, the spatial extent of the sources can be extracted by constrained ICA.

This chapter proposes to apply a time structure based spectrally-constrained ICA algorithm to a μ rhythm based BCI system (this type of BCI has been introduced in Chapter 2) and to describe the selection of power features from the provided data sets and the overall classification system used. The study then presents the results obtained and discusses the performance enhancements to be had from the use of this algorithm in this way.

6.2 The Data

This chapter uses two data sets: BCI Competition Data Data Set IIa and Data set IVa. Although these two experiments were designed in different ways, they both use the property of the ERD power spectrum adjusted by different specific activations as shown in Chapter 2.

6.2.1 BCI Competition Data Set IIa

The BCI Competition Data Set IIa (self-regulation of μ- and/or central β-rhythm) was provided by the Wadsworth Center [201]. The idea is that the subjects either increase

or decrease their μ or β rhythm amplitude power to control a cursor's vertical position aiming to the correct target through visual feedback. This data set contains a whole record of an actual BCI system from 3 trained subjects in 10 sessions (about 30 minutes per session) each. EEG was recorded from 64 scalp electrodes (10/20 system) sampled at 160 Hz and each electrode was referred to an electrode on the right ear. For this BCI to work, after a one second resting period during which the screen stays blank, a target appears at one of four possible positions on the right-hand side of the screen. One second later, a cursor appears at the middle of the left of the screen and starts moving at constant speed across the screen from left to right. When the cursor reaches the right-hand side, the screen is cleared and the next trial begins. The experiment includes visual feedback whereby the vertical position of the cursor on the screen is determined through brain activity. Three data subsets marked as AA, BB and CC are supplied. Each session consisted of 192 trials (48 trials for one of four targets: 'top', 'upper middle ', 'lower middle' or 'bottom'). The first six sessions are labelled as *training* sets. The remaining four sessions are *test* sets. This study only selects trials with the target position code: 'top' (Target 1) and 'bottom' (Target 2) to examine the proposed method.

6.2.2 BCI competition data set IVa

The BCI competition III data set IVa (motor imagery, small training sets) is from the Berlin BCI group [208]. For this BCI to work, the subjects imagine either right hand or right foot movements indicated by a visual cue on screen without feedback. The power in a particular frequency band is used to control the cursor movement. This data set contains 118 multi-channel (extended 10/20 system) EEG signals recorded from five healthy subjects (labelled 'aa', 'al', 'av', 'aw' and 'ay' respectively) at a sampling rate of 100 Hz. During the experiments, subjects were prompted by a displayed letter (R/right hand, or F/right foot) to imagine for 3.5 s either right hand (Target A) or right foot movements (Target B) without feedback. Each type of MI was

recorded 140 times, thus in total there are 280 trials for each subject. Between the trials, there was a random period of time (1.75 to 2.25 seconds) in which the subject could relax. This data set also brings with it a challenge in that only a small amount of training data are available, this allows one to examine the influence of using small training sets in order to reduce the overall training time. The task is to classify the type of the imagined movement for each trial in an offline fashion.

The data were originally recorded from 64 scalp electrodes for Data Set IIa and 118 electrodes for Data Set IVa. Since the activity in the motor cortex is the interest in this study, so the electrodes around the sensorimotor cortex were chosen manually, these included: C5, C3, C1, C2, C4, C6, CP5, CP3, Cp1, Cp2, Cp4, Cp6, P5, P3, P1, P2, P4 and P6. A small segment of EEG data is used for training in the proposed algorithm: for Data Set IIa, the data between 0.5-2 s of each trial is used after the cursor is displayed on the screen; for Data Set IVa, the data between 0.5-2.5 s is considered after the instruction is displayed on the screen. There are two main reasons to select such a short time segment for the data. One reason is to satisfy the assumption of the proposed time structure based ICA technique (the details is given in the following methodology section). The other reason is that the subjects generate strong responses only in a short time period after stimulus. The parameters of the short window are selected after examining the available training sets manually.

6.3 Methodology

6.3.1 Temporal Decorrelation based spectrally-constrained ICA

In this chapter it is proposed to use the spectrally-constrained ICA technique based on temporal decorrelation and a specific algorithm $\text{LSDIAG}_{\text{TD}}$ is selected to capture the dependency structure of the observed signals. The detailed TD based ICA algorithm has been introduced in Chapter 3. One basic assumption with this method is that the source waveforms should be stationary and have unique power spectra. Obviously this

method is not suitable to use for long-term EEG recording. In order to use TD based ICA, the assumption needs to be made that the signals are stationary over short time windows. In this chapter, since the imagery epochs last just a few seconds (about 4 seconds long), the auto-correlation function of the source activity is assumed slowly varying in time. The temporal correlations over short time windows can be computed and then the mixing matrix is estimated in the usual way [154].

The TD based ICA algorithm used in the study is called the $\text{LSDIAG}_{\text{TD}}$ technique [159], where

$$\mathbf{C}_{\tau}^{\mathbf{s}} = \mathbf{W}\mathbf{C}_{\tau}^{\mathbf{x}}\mathbf{W}^{\mathbf{T}}, \tag{3.11}$$

$\mathbf{C}_{\tau}^{\mathbf{x}}$ is the signal cross-covariance matrix and $\mathbf{C}_{\tau}^{\mathbf{s}}$ is the source cross-covariance matrix. $\mathbf{W}$ is the unmixing matrix. The algorithm aims to transform the signal covariances matrix stack into the source covariances matrix stack as diagonal as possible since diagonal is assumed to represent independence. After TD based ICA $\mathbf{W}$ is then estimated so that each of the independent sources can be separated by $\mathbf{s}=\mathbf{W}\mathbf{x}$.

Another issue of TD based ICA is to decide the number of time lags as shown in Chapter 3. Here, since the data are separated into two parts: a training session and a testing session, the number of time lags of use is tuned and found from the training session. It turns out that the use of 5 lags is able to achieve the best performance.

6.3.2 The Reference Channel

In this spectrally constrained model, prior knowledge of the spectral content of the sources can be introduced into the model by means of reference channels. If the power spectrum of a particular source activity is known, such reference(s) would consist of filtered noise with the desired power spectrum. This chapter applies the method using

just a single (μ rhythm) reference. To observe changes in rhythmic activity in specific frequency bands, band-pass filtered white noise are used to derive a reference signal. Particularly an 8^{th} order Butterworth BP filter with lower and upper corner frequencies set appropriately is used to set the desired constraint. Since the phase information of this added reference channel is meaningless (i.e. the study is not expecting the phase of the reference signal to be connected in any way to that of the desired brain response), the problem of matching the phase of the reference channel with that of the desired activity in the recordings can be overcomed through calculating the lagged covariance matrices that $\text{LSDIAG}_{\text{TD}}$ requires via the FFT and then removing the phase information of the signal in the frequency domain. Recall that the convolution of two functions $f(t)$ and $g(t)$ can be obtained by

$$h(t) * g(t) = F^{-1}\{H\{f\} \cdot G\{f\}\} \tag{6.1}$$

where * denotes convolution, F^{-1} denotes the inverse Fourier Transform, $H\{f\}$ and $G\{f\}$ the Fourier Transform of $h(t)$ and $g(t)$ respectively.

6.4 The Proposed Algorithm

The flowchart of this work includes four parts: 1) spatial filter generation, 2) spatial filtering, 3) power feature extraction, and 4) classification. This is depicted in diagrammatic form in Figure 6.1.

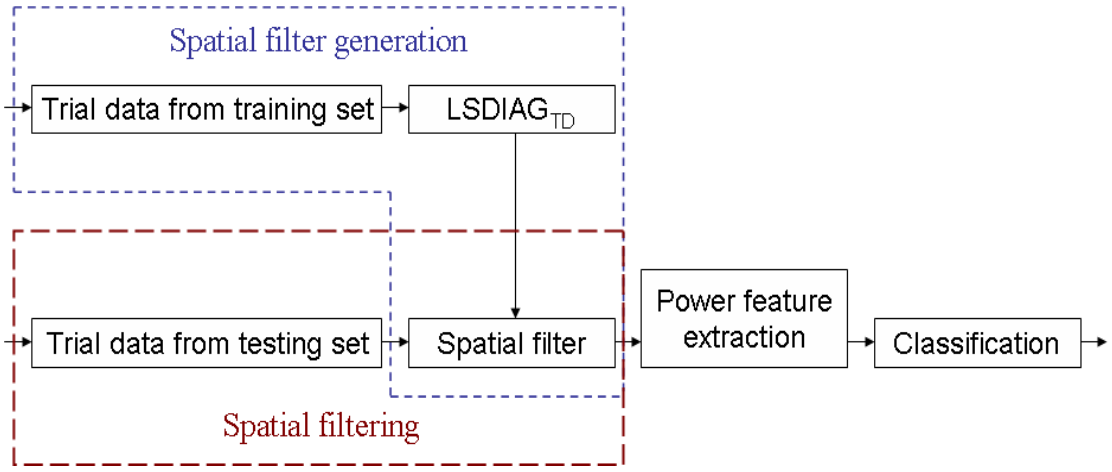

Figure 6.1: *A diagram depicting the proposed algorithm. It includes four parts: spatial filter generation, spatial filtering, power feature extraction and classification.*

6.4.1 Spatial filter generation

The power spectrum of each data trial is calculated and averaged across the frequency domain. For display purposes, the averaged trial power spectrum on C3 towards to two targets is shown in Figure 6.2. This confirms that two kinds of activities have different power amplitude at the μ band around 8Hz~13Hz which can be used as the control features. For the analysis, a number of data trials of two different targets in the training data set were used to estimate the lagged covariance matrix stack $\mathbf{C}_{\tau}^{x}$ the stack of matrices are treated as arising from two-part averaged lagged covariance matrix stacks $\mathbf{C}_{\tau}^{XT_1}$, $\mathbf{C}_{\tau}^{XT_2}$ in which each part is obtained from trial data corresponding to one of two targets. The reason to construct such stacks of matrices is that the stack of covariance matrices needs to capture as much information as possible so that for this two-target system the covariance matrix should be constructed from the data of both two targets.

$$\mathbf{C}_{\tau}^{\mathbf{x}} = \left[\mathbf{C}_{\tau}^{XT_1}, \ \mathbf{C}_{\tau}^{XT_2}\right] \tag{6.2}$$

$$\mathbf{C}_{\tau}^{\mathrm{XT}_i} = \left[\frac{1}{m} \sum_{\mathbf{x_k} \in \mathrm{XT}_1} \mathbf{C}_0^{x_k}, ..., \frac{1}{m} \sum_{\mathbf{x_k} \in \mathrm{XT}_1} \mathbf{C}_l^{x_k} \right] \tag{6.3}$$

$$\mathbf{C}_{\tau}^{\mathrm{XT}_2} = \left[\frac{1}{n} \sum_{\mathbf{x_k} \in \mathrm{XT}_2} \mathbf{C}_0^{x_k}, ..., \frac{1}{n} \sum_{\mathbf{x_k} \in \mathrm{XT}_2} \mathbf{C}_l^{x_k} \right] \tag{6.4}$$

where $\tau = [0,...,l]$ depicts the range of lags (here l=5 as determined in previous work [209]). $x_k \in [\mathrm{XT}_1, \mathrm{XT}_2]$ denotes that trial data are from training set corresponding to the labels: Target 1/A and Target 2/B. The number of trials in each data set for XT_1 and XT_2 are m and n respectively. Here the value of m equal to n is set to balance the proportion of trials for both targets. Once the unmixing matrix $\mathbf{W}$ is estimated from the training datasets, it then works as a simple spatial filter straightforward on the testing datasets.

The spectrum, $\mathbf{P}(i)$ is defined as a trial spectrum in the i^{th} channel by the sum of the ordinates of the frequency bins ($\mathbf{h}_d$) within the proposed frequency band, i.e.

$$\mathbf{P}(i) = \sum_{d=1}^{\mathrm{D}} \mathbf{h}_d \,, \tag{6.5}$$

where D denotes the number of frequency bins. After spectrally-constrained ICA decomposition, the EEG data are extracted into the ICs. Thus the power spectrum after spectrally-constrained ICA is defined as the sum of the weighted spectra of sources (ICs) within the μ band. So, for given source epochs, the power feature reflected in an individual channel is defined as

$$\mathbf{f}_p(i) = \sum_{j=1}^{k} \mathbf{a}_{k+1,j} \mathbf{P}_{ic}(j) \mathbf{a}_{i,j} \tag{6.6}$$

where k denotes the number of sources. As this implementation of

spectrally-constrained ICA assumes a square mixing matrix, then the number of sources is the same as the number of measurement channels, and $\mathbf{a}_{i,j}$ is an element in the mixing matrix **A**. $\mathbf{a}_{k+1,j}$ is a particular element in the last row of **A**. $\mathbf{P}_{ic}(j)$ denotes a trial spectrum in the j^{th} IC source.

6.4.2 Feature selection

In order to find discriminative power bands for each subject, here the study uses a technique called ─ the r^2 measure, also called the coefficient of determination [210], which is able to interpret the proportion of variability in a data set. A definition of the r^2 measure is as following:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - Y_i')^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}, \tag{6.7}$$

where Y_i represents an individual data point of Target 1 in the frequency domain, Y_i' represents an individual data point of Target 2 in the frequency domain and $\overline{Y}$ represents the average of the Y_i values. r^2 values range from 0 to 1, with 1 representing that the covariates can predict the outcome perfectly in the model, and 0 representing a complete lack of predictability of the outcome. The power spectra of two targets are calculated in these two data sets, and then combined the variables on each individual channel into r^2 values. By comparing to the averaged power spectra corresponding to two targets, this describes the relationship between power intensity and target labels. These parameters were slightly different due to differences in each individual recording. Figure 6.3 shows two discriminative power bands roughly around 10-15 Hz (μ band) and 23-28 Hz (β band) in Data set IIa. The reason that there exist two power bands is because both movement and imagery are associated with μ

and β rhythm together. Depending on which rhythm is used as the control pattern, in this work we choose 10-15 Hz as the working band. Increased power is taken to correspond to Target 1 which raises the cursor in Subjects AA and CC while it makes the cursor go down in Subject BB (Figure 6.4). In Data set IVa, the sub-band approximately around 8-15 Hz is selected to calculate power (Figure 6.5). Increased power is related to Target 2 which is the right foot imagination (Figure 6.6) in all subjects.

6.4.3 Classification

In order to evaluate the performance of the proposed algorithm, a simple one-dimensional linear classifier based on thresholding the power feature(s) is considered in the chosen frequency band for the final classification. The threshold value is selected by minimizing the number of trials misclassified in both classes from the training set for individual subjects, for example the threshold value for Subject CC is between $7.842\times10^9 \sim 8.031\times10^9$ in arbitrary units. In addition, as a comparison for the classification performance, a more complex classifier, a SVM based classifier is also applied. The study uses a downloadable SVM toolbox [211]. The main parameters are set as the following: the power of 40 data trials on three channels C3, CP1 and CP5 for both targets is used as the SVM classifier's training input; the kernel function is set to be linear.

The next procedure is to decide which power feature will be suitable to use for the classification. Based on the distribution of r^2 values across the topography spectrum in the previous section, a number of channels (between 1 and 3) around the left sensorimotor cortex were selected. The power on C3 was used in the threshold classifier and the power on C3, CP1 and CP5 for the SVM classifier as the use of these power features was found to give better classification accuracy in this study.

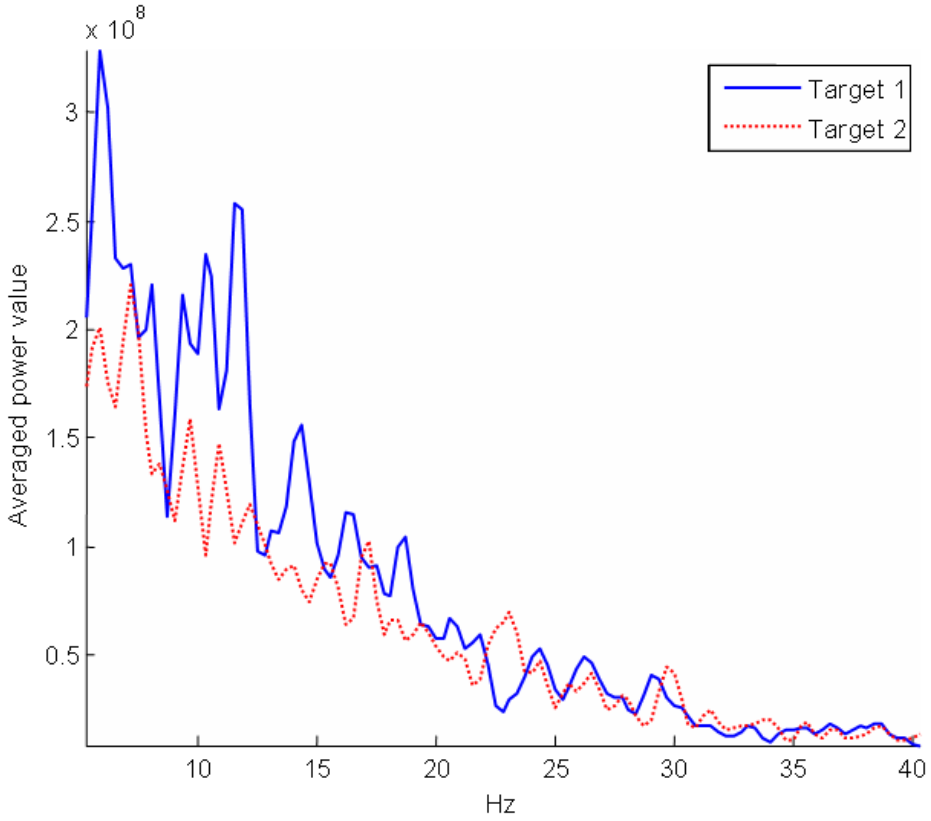

Figure 6.2: *The averaged trial power spectrum on Channel C3 towards to two targets in a training set for Subject CC.*

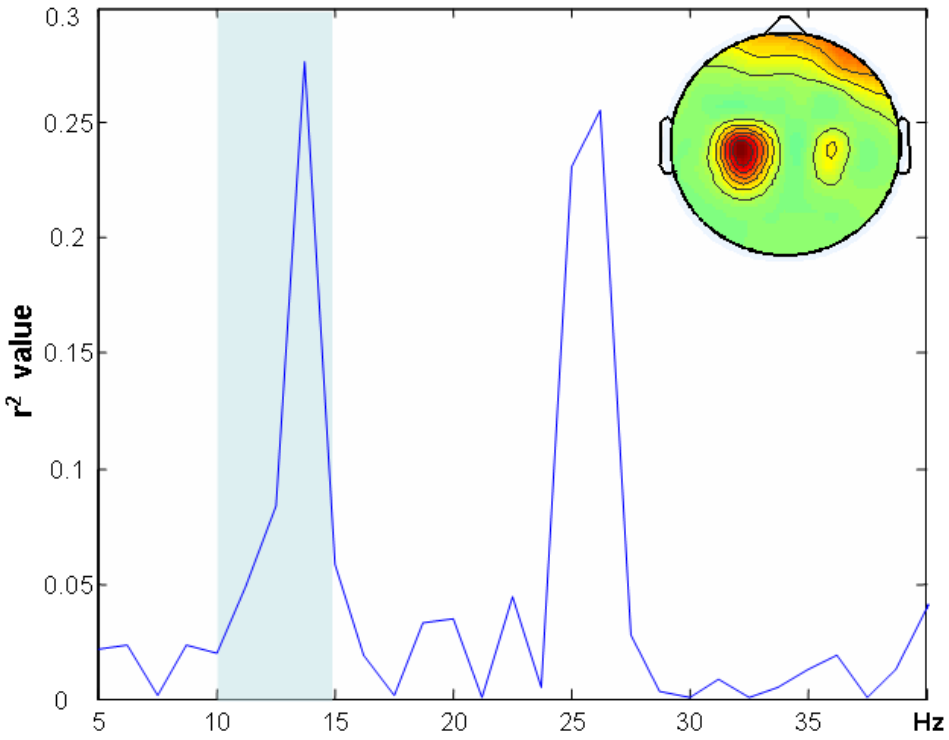

Figure 6.3: *r^2 values across the spectrum on Channel C3 for Target 1 & 2 (Subject CC). The shadowed frequency band was chosen here. Inset is the topography of the r^2 values at 13.75 Hz across all channels.*

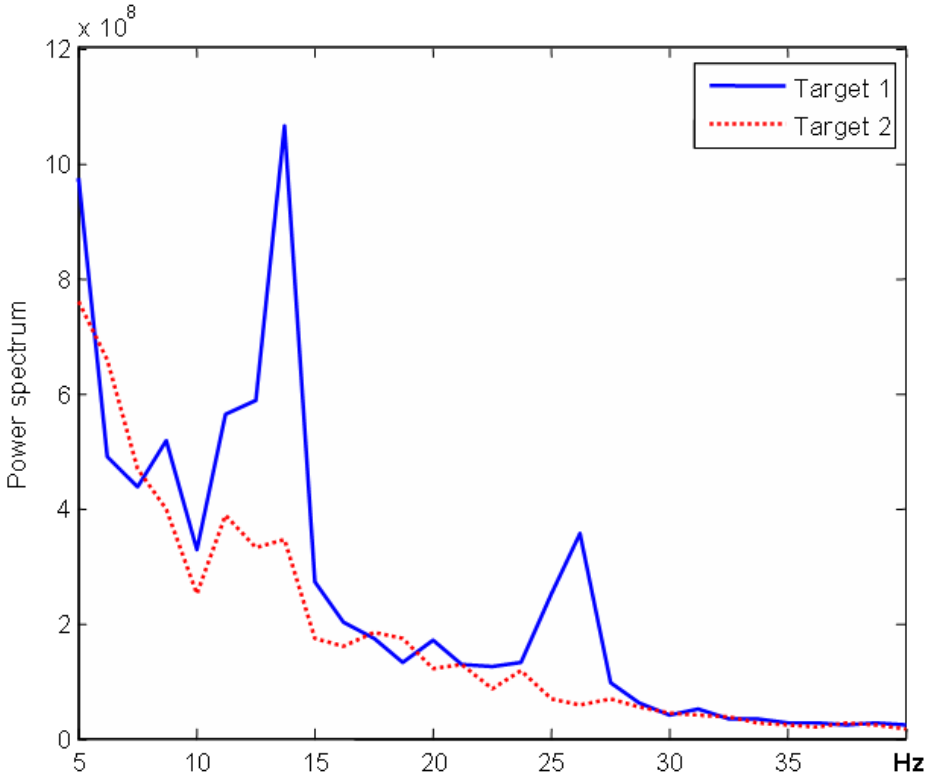

Figure 6.4: *Averaged power spectra of data trials corresponding to Target 1 & 2 (Subject CC). In this experiment, greater power in 10~15Hz (Target 1) implies the cursor going up and vice versa.*

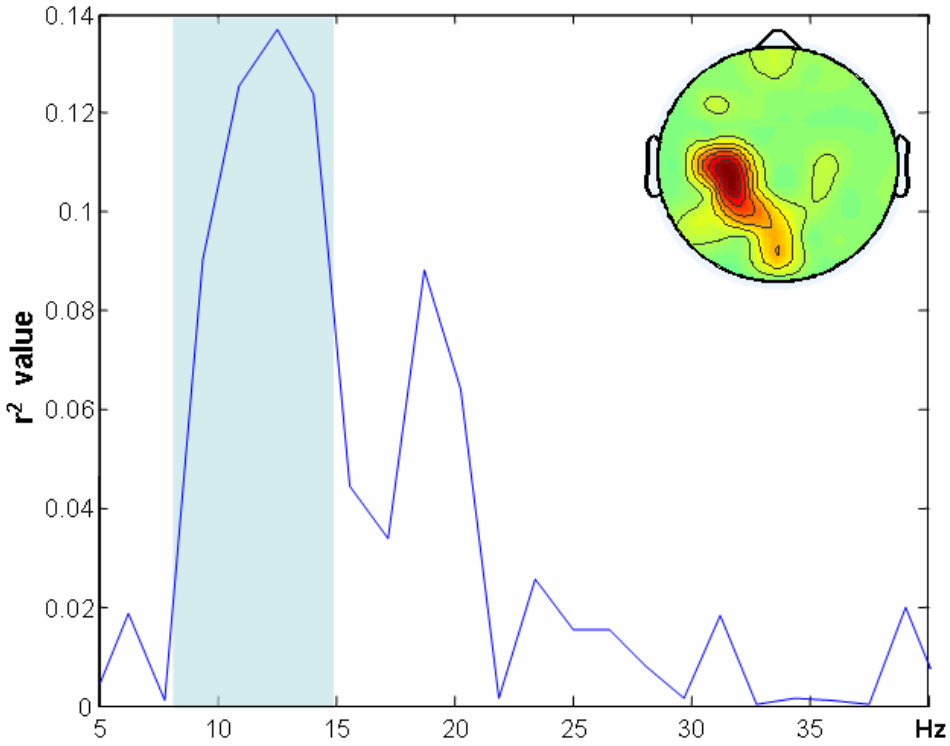

Figure 6.5: *r^2 values across the spectrum on Channel C3 for Target A & B (Subject 'ay'). The shadowed frequency band was chosen in this work. Inset is the topography of the r^2 values at 12.25 Hz across all channels.*

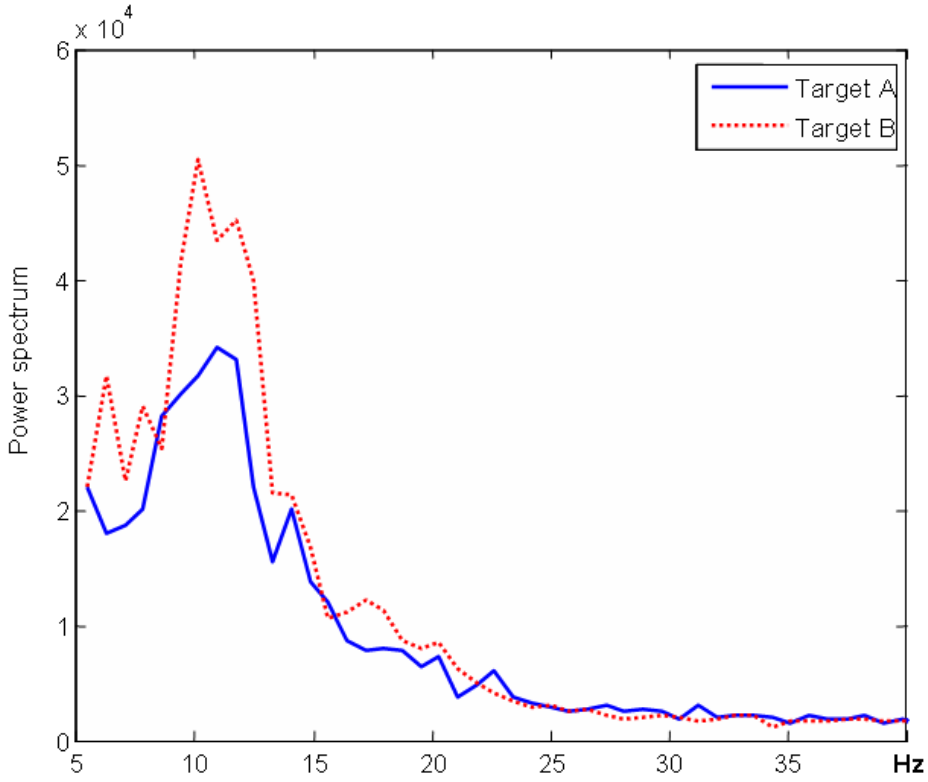

Figure 6.6: *Averaged power spectra of trials corresponding to Target A & B (Subject 'ay'). The averaged power for imagined foot movement (Target A) is greater than the power for hand movement imagination (Target B).*

6.5 Results

The proposed TD based ICA method generates the spatial filter from the training data for each individual subject. The spatial filter is then applied to the unlabeled test dataset. It can capture the relevant dynamics of the subject's brain state more robustly. Furthermore, the resulting time series will have optimised the spectra resulting in better discrimination between two different brain states. The results show that following this pre-processing, even a simple linear classifier can achieve very good classification accuracy.

To make the proposed ICA system work, a segment of bandpass filtered white noise is used as the reference signal (constraint). Moreover this reference needs to be modulated to match each individual's particular EEG frequency band. The sub-plot in Figure 6.7 illustrates the original white noise with arbitrary units used for Subject CC; the middle sub-plot shows the bandpass (10 ~ 15 Hz) filtered white noise modulated with the mean power of the signal recordings; the bottom sub-plot represents the same signal (in the middle sub-plot) in the frequency domain.

Figures 6.8 and 6.9 depict the power features related to different targets before and after the processing for channel C3. Figure 6.8 plots the power features of Testing session 10 for Subject CC in Data set IIa. Ideally, the higher power feature represents Target 1 and the lower power for Target 2 (Figure 6.3). However, without spatial filtering, the power features between the two targets from the original data appear overlapped, and a classifier based on either a simple linear method or a potentially complicated advanced method is hardly able to separate these patterns efficiently. After the spectrally constrained ICA processing, the weighted power values for two different targets are better separated than the power features from the unprocessed data. Figure 6.9 shows the power features of Subject 'ay' from Data set IVa. The

power related to the right hand movement imagination is marked as Target A and the power for right foot movement imagination marked as Target B. As shown in Figure 6.5, the averaged power for imagined foot movement is larger than the power for hand movement imagination, but powers correlated to two different targets do not show much different in the raw data. After processing, the power features are maximally separated into the different levels, which further demonstrate the improved separation achieved by using this spatial filter. The above examples suggest that the use of this spatial filter can help to extract different brain activities within a particular μ rhythmic band.

Table 6.1 lists the classification results on the test sets (most sessions have 52 trials for each target, several have 51 trials) in Data set IIa. For each subject, 80 randomly chosen trials in total (40 for each target) are used to calculate the spatial filter. The results are shown as three columns for each individual subject. The first column shows the results using the unprocessed data. The results of using a threshold based linear classifier with one power feature on C3 are shown in the second column. The third column is for the results from an advanced SVM based classifier using three power features on C3, CP1 and CP5. Table 6.2 shows the classification performance on the testing data in Data set IVa. There are five subjects contributing to individual sub-sets with different sizes of training and testing sessions. The numbers of available trial data for training/testing sessions are shown in the first column. To construct the spatial filter, the total number of training trials is selected from 28 and 80 (average of 65 trials was used) due to the different size of training sets. As before, one power feature on C3 is used for the final classification based on a threshold and a linear classifier. Moreover, three features on Channel C3, CP1 and CP5 were also applied to examine the performance of a SVM classifier. In addition, as a comparison, the last column lists the classification results from previous published work [212] which proposes a method based on dynamical system (DS) features together with a SVM classifier. The overall classification accuracy is about 85% by this DS+SVM method. From the two tables, we can see that spectrally constrained ICA implementation

extracts the related rhythmic information very effectively. After processing, the classification accuracy was of 82% for Subject AA, 69% for Subject BB and 90% for Subject CC in comparison with the average 62% accuracy before processing in Dataset IIa. In Dataset IVa, the classification accuracy was of an average of 82% through five testing sets which is 30% higher than the accuracy using the unprocessed data. It is worth noting that the more advanced SVM based classifier did not show a significant improvement in performance on the same data, although an increase of about 2% compared to the simple linear classifier was observed.

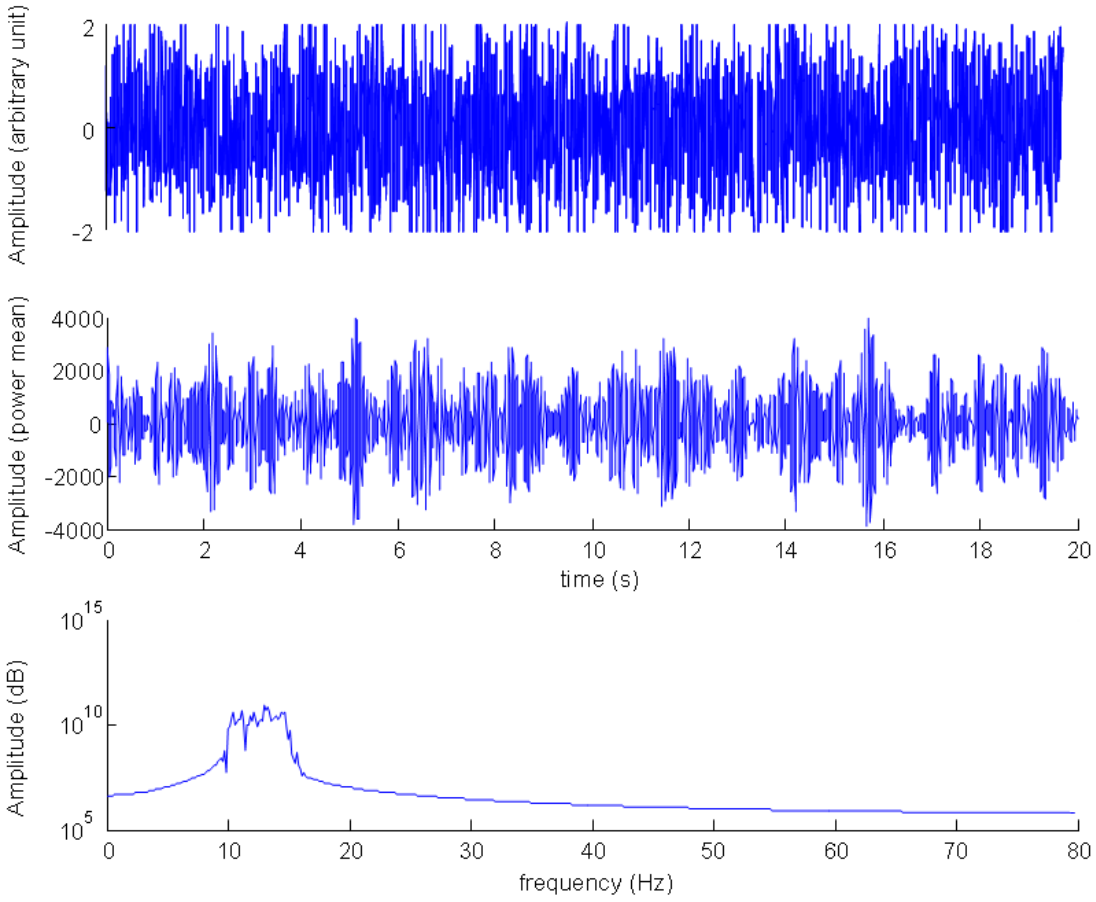

Figure 6.7: *(above) The original white noise with arbitrary unit for Subject CC; (middle) The bandpass (10 ~ 15 Hz) filtered white noise modulated with the mean power of the signal recordings; (below) The same signal in the middle shown in frequency domain.*

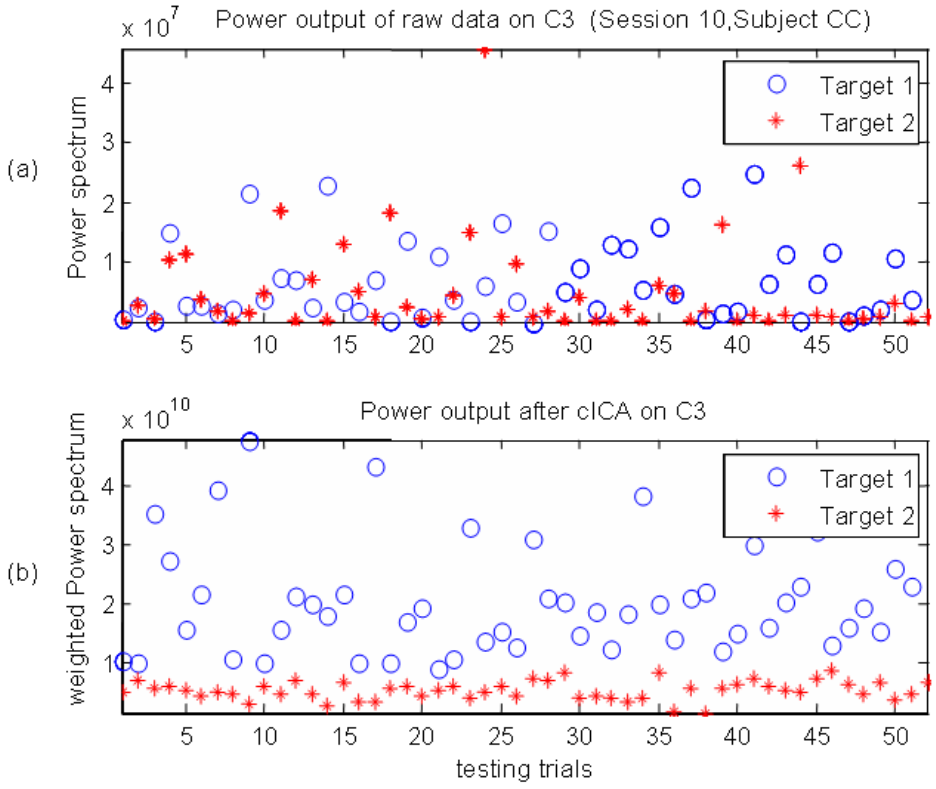

Figure 6.8: *The power feature outputs of Subject CC for testing Session 10. (a) shows the power features on C3 using the unprocessed data; (b) shows the power features on C3 after spectrally constrained ICA processing. A circle denotes Target 1 (cursor up); a star indicates Target 2 (cursor down).*

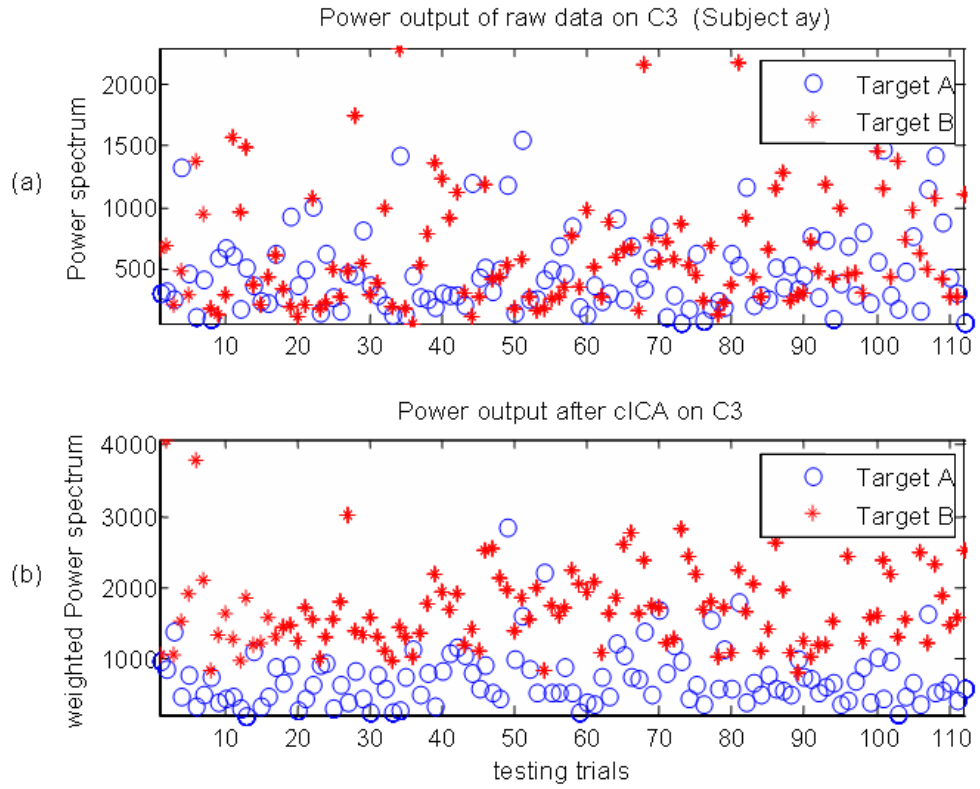

Figure 6.9: *The power feature outputs for Subject 'ay' on testing set. (a) shows the power features on C3 using the unprocessed data; (b) shows the power features on C3 after spectrally constrained ICA processing. A circle denotes the power feature for Target A (right hand imagination); a star indicates the power feature for Target B (right foot imagination).*

Table 6.1: Classification accuracy of the test set based on power feature(s) in Data set IIa. The three columns for each individual subject show the performance of linear classification on unprocessed data, linear classification and SVM classification on the processed data.

Testing data set	Data AA			Data BB			Data CC		
	Linear classifier on raw data	linear classifier on extracted data	SVM on extracted data	Linear classifier on raw data	linear classifier on extracted data	SVM on extracted data	Linear classifier on raw data	linear classifier on extracted data	SVM on extracted data
Set 7	64.6%	80.2%	85.4%	65.6%	72.0%	73.0%	58.3%	85.4%	87.4%
Set 8	59.4%	88.5%	89.6%	71.9%	72.9%	72.9%	62.1%	92.2%	90.3%
Set 9	61.5%	80.2%	79.2%	66.8%	63.5%	67.7%	60.1%	86.1%	88.1%
Set 10	65.6%	80.2%	80.2%	59.4%	68.8%	72.9%	61.2%	96.1%	98.1%

Table 6.2: *Classification accuracy of the testing set based on power feature(s) in Data set IVb. The columns depict the results using the three proposed classification schemes, and the last column lists published [212] for comparison.*

Data set	training/test trials	linear classifier on raw data	Linear classifier on extracted data	SVM on extracted data	SVM on DS features
al	224/56	48.2%	85.7%	89.3%	96.3%
aa	168/112	46.0%	83.0%	85.7%	83.3%
av	84/196	49.5%	75.0%	75.0%	72.7%
aw	56/224	55.4%	80.3%	85.3%	86.9%
ay	28/252	54.3%	85.0%	85.0%	89.0%

6.6 Summary

Two datasets described in this chapter being part of the BCI competition dataset have been used to examine the performance of the proposed algorithm. Both of these data sets use the characteristic that changes in the amplitude of sensorimotor rhythms over the right/left hemisphere act as the major control pattern. The difficulty here is to maximally and reliably identify at least two classes from single-trial data. The proposed ICA technique using spectral constraints has been developed and applied to isolate and extract the power spectrum in the rhythmic band of interest. In order to demonstrate the performance of the proposed spectrally constrained ICA, the power feature in the μ rhythm frequency band is used as the major classification pattern. The results using a simple linear classifier and a SVM to classify the ICA processed data show that the classification accuracy has considerably increased over processing the raw data. After the basic analysis the overall classification accuracy is improved by about 20% in Dataset IIa and 30% in Dataset IVa. As an additional comparison of classification performance to spectrally constrained ICA in Dataset IVa, the results of a method using DS features as well as a SVM classifier is reviewed and compared [212]. This method also includes two steps for data pre-processing (a temporal filter and a spatial filter). The accuracy of the techniques in [212] was about 3% more than the results of spectrally constrained ICA with a linear classifier and 1% more than spectrally constrained ICA with a SVM. However the use of a linear classifier and a simple spatial filter in this study aims to simplify the problem and is desirable from a computational complexity perspective.

As this work is an application to single trial classification, the sensitivity to artifacts in the EEG becomes a major problem. The $\text{LSDIAG}_{\text{TD}}$ ICA algorithm uses the covariance of the trial data to estimate the covariance stack matrices which are the essentials to calculate the unmixing matrix and hence the spatial filter. The random selection of training trials which may include artifacts can cause serious changes to the final filter. Therefore, most methods require that the data should be artifact free,

which can be achieved by several preprocessing steps such as filtering or manual artifact rejection. Here instead of applying any preprocessing methods before hand, the study estimates the stack matrices by using the averaged lagged covariance matrices from the data. The idea behind the process is that the influence of artifacts is reduced since the procedure of averaging the covariance matrices acts as a filter which could balance and minimise the random noise level. Moreover, the system includes a training phase used to tune the proper unmixing matrix (spatial filter) using the proposed ICA. Once the unmixing matrix has been computed, it works as a spatial filter to remove the additional artifacts by weighted spatial averaging the testing data trials and returns the processed time series patterns. After filtering, the different brain activities seen as power changes can be clearly extracted. It indicates that through the use of spectrally constrained ICA, it is possible to track the rhythmic changes of different brain states in the EEG. These results show a clear improvement for use in this kind of BCI system.

The next chapter proposes the application of the ICA algorithm to recordings from single (or few) channel(s) in a P300 BCI application. The aim is to drastically reduce the number of channels required; not a big number of channels as is required in multi-channel application, the actual recording channels here are reduced to just one or two channels. This would mean that simpler BCI systems could be designed using less setup time and with less reliance on accurate placement of the scalp electrodes.

CHAPTER 7

Single channel ICA algorithm for ERP based BCI

7.1 Introduction

Conventional BSS (such as multi-channel ICA) generally requires multi-channel data for the inputs. However there are many circumstances where only one recording channel is available or desired. Such BSS techniques are not directly applicable to the data which consists of only one single channel record. Fortunately by applying as explained in Chapter 3 the so-called method of delays, the single channel data can be rearranged and represented as a matrix of time-delayed vectors. In this way, the BSS analysis techniques can be used with single channel recordings and is able to solve the more sources than sensors problem. This ICA algorithm is referred to here as SC-ICA. Moreover, this single channel model can be extended to a more general algorithm called ST-ICA based on the combination of space-time vectors.

This chapter demonstrates the proposed SC-ICA, as well as ST-ICA methods, on the dataset from a P300 based BCI system. It also proposes an automatic method to speed up the component selection process. Results show that the proposed ICA methods can separate the single channel recordings into their underlying components. Moreover the results are also comparable in the final classification performances between Multi-channel ICA, SC-ICA and a lowpass FIR filter.

7.2 The dataset

The dataset used here is from the BCI competition Data set IIb as it explained in Chapter 5. There are a total of 180 brain responses for each presented character (12 responses multiply 15 reinforcements), 30 of which should contain P300 responses. The study firstly demonstrates the proposed methods on the data where each character is present in only three 1.5s-epochs (the recording of an epoch starts 0.5s before the stimulus presents) where possible P300 patterns are randomly selected and then concatenated to form a 4.5s trial. To examine the performance when using different channel locations, three channels including C3, C4 and Oz are selected to form several data combinations: C3, C4 and Oz alone, and C3 & C4, C3 & Oz and C4 & Oz combined. In the literature the P300 activity dominates over the parietal electrode sites (around C3 and C4), whereas Oz is located further way, towards the occipital region. However as the EEG is believed to be a linear mixture of underlying brain sources, even the recordings from a far electrode, for example Oz, is supposed to contain P300 patterns at a lower SNR. The above combination sets aim to explore this fact. The second purpose is to examine and compare the performance of the character identification on the single channel Cz.

7.3 P300 EP extraction

7.3.1 The proposed ICA techniques

(1) Single channel ICA (SC-ICA)

To examine the possibility of extracting P300 EP activities on just one channel by ICA, SC-ICA is applied. The details have already been introduced in Chapter Three. However for SC-ICA single channel the data need to be reformatted as a

multi-dimensional data representation which is done through the method of delays. Such a multi-channel representation is also called the matrix of delays. Some parameters, for example dimension of the matrix of delay *m* and delay lag τ need to be set. To set parameters for the embedding dimension *m*, the analysis introduces and uses Equation 3.22 as shown in Chapter 3. The sampling rate is set at 240Hz, and the lowest frequency of interest is assumed to be at about 2 Hz. Therefore *m* is set to 120; and τ is set to 1 as the default value. In short this SC-ICA algorithm includes following steps:

a) create the delayed vector matrix and temporally whiten the signal;

b) apply an ICA algorithm to learn the mixing matrix **A**;

c) project all the components back to the measurement space;

d) select the relative components based on observing the positive peaks presented around 300ms after the stimulus onset (subjectively);

e) project the selected components together to the measurement space and form the extracted signal.

For the ICA algorithm, both FastICA or Infomax ICA can be chosen in this analysis. However due to the high dimension of this delayed input data and the length of the data, the FastICA algorithm sometimes fails to converge for the exceptionally low achieved SIR problem [213]. Generally dimensional reduction is applied to discard some weak components so that the possibility of convergence could be improved. However Infomax ICA is able to achieve a better global convergence [214]. For this reason, Infomax ICA (the software in the EEGLab package) is selected for this analysis.

The proposed SC-ICA method tests the performance on different chosen channels. The analysis starts on Channel C3, C4 and Oz separately (C3 and C4 are located over the P300 focus).

(2) Space-time ICA (ST-ICA)

A ST-ICA technique is developed to work on the two-channel data. Adding just one more signal channel would provide extra information on both temporal and spatial domains. Although such information is quite limited, it allows ICA to achieve a better extraction performance. The study is applied to the two-channel combinations: C3&C4, C3&Oz and C4&Oz. The new delay matrix is similar to the application of the single channel case. However, this matrix is constructed from the concatenation of two delay vectors for each channel. So the value of *m* for each set of delay vectors is 120 and then the dimension of this delayed matrix would be 240. For τ, the default value 1 is used here again. In the IC projection step, each IC has to be back-projected on to the measurement space for each channel separately. This also means that the number of ICs will be doubled. The IC selection is the same as the manual method shown above.

7.3.2 Dimensional reduction

In the above context, the dimension of the matrix of delay *m* has been set 'big enough' to capture the necessary information of the signal. In effect, it takes a single channel data and generates a multivariate data set. Because this data set is likely to be of extremely high dimensionality, the problem of interpretation of this data remains. To overcome the problem, the dimensionality of this data should be reduced before applying ICA. Here SVD is used to decompose each delay matrix into a set of orthogonal components. Dimensional reduction of the data is performed by truncating the SVD transformation after a certain number of eigenvectors. This step also provides an opportunity to reduce some of the low level noise [215] [216].

7.3.3 Manual IC selection

Since each individual IC is hard to interpret physiologically in the source space, to assess each IC's neurophysiological significance, it needs to be projected back to the measurement space. However it is not generally possible to linearly separate independent components from a *single* time series, for example, the resulting sources are not necessarily truly independent. Therefore the IC selection is about finding and grouping the components of interest. Since this application is to extract P300 EPs, the P300 wave morphology is used as the IC selection criteria.

7.3.4 Automatic IC selection

The current component selection is based on a subjective (manual) method. Obviously it is a functionally limited method and can be only really used in the lab for study purposes. To build a practical BCI system, the component selection has to be automated. Here the study proposes and demonstrates a simple P300 extraction application which involves an automatic method – an improved approximate method to take the place of manual selection. The basic idea is to apply SC-ICA on the selected data from Cz (these data are randomly selected from the P300 response epochs) and learn the parameters needed to construct a separation filter. The obtained filter then can be applied to extract the rest of the data. The original automatic algorithm can be repeated if there are more sources in the data. Here the study found that running this filter once on the data is good enough to extract the P300 components. The algorithm is summarised as:

a) create the delayed vector matrix from the selected data set, temporally whiten the signal and reduce the dimension by SVD (here reduce the dimension down to 40 by observing the convergence of the singular spectra of the matrix of delay);

b) apply a deflationary ICA algorithm to learn the mixing matrix **A**;

c) create the separation filter $f_i(t) = a_i(-t) * w_i(t)$, where $i \in \gamma_p \ \gamma_p$ is the selected subset of basis vectors; where $a_i(t)$ is the column vector of **A** and $w_i(t)$ is the row vector of **W**;

d) apply this filter to the available data set and then calculate the scaling parameter a_p to adjust and rescale the filtered signal, a_p is given as: $a_p = \langle x(t), f_i(t) * x(t) \rangle / \langle x(t), x(t) \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual vector inner product; the $a_p f_i * x(t)$ is the extracted source component.

Infomax ICA has been applied in the previous session, for the automatic IC selection, here the aim is to examine another popular ICA algorithm – FastICA. In order to prevent the failure of convergence, the data dimension is reduced by SVD (the number of dimensions is set to 40). The FastICA program is obtained from FastICA software package [150].

7.3.5 Comparison of extraction methods

In the digital signal processing field, many filtering techniques have already been developed with the aim of enhancing the signal's SNR. Of course, one of the primary applications is noise removal: removing unwanted parts of the signal, such as random noise, or extracting the useful parts of the signal, such as components lying within a certain frequency range. In this aspect, SC-ICA works like a filter. As a performance comparison of noise reduction and useful component extraction, the study uses a lowpass FIR filter on the same data. This FIR filter is setup as part of the EEGLab toolbox. The filter has a cutoff frequency at 10Hz (since the frequency band of interest for P300 EP is below 10Hz) and the filter order is set to 72 by an approximate method (3×(sampling frequency/cutoff)). In addition, the performance after this FIR filter is applied is compared with the performance after the proposed SC-ICA techniques with manual component selection and with automatic component selection.

Figure 7.1 shows the frequency response of the applied lowpass FIR filter. As the proposed automatic component selection after SCICA is equivalent to the filtering and the filter coefficient (order is 79 in this case) are found from $f_i(t) = a_i(-t) * w_i(t)$. Figure 7.2 shows the frequency response of the applied filter learned from SCICA. Comparing to the frequency responses of this lowpass FIR filter and the filter directly learned from SCICA, both two filters perform lowpass filtering the data. FIR filter reduces higher frequency components and smoothes the signal. However the filter from SCICA passes the low frequency components and also leaves partly higher frequency components.

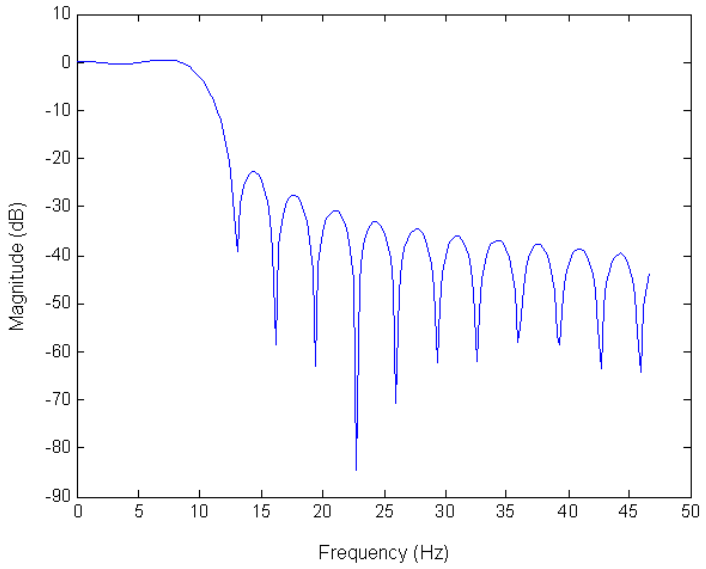

Figure 7.1: *The frequency response of the applied lowpass FIR filter. The filter order is set to 72 by an approximate method (3×(sampling frequency/cutoff), the sampling frequency was 240 Hz and cutoff was 10 Hz).*

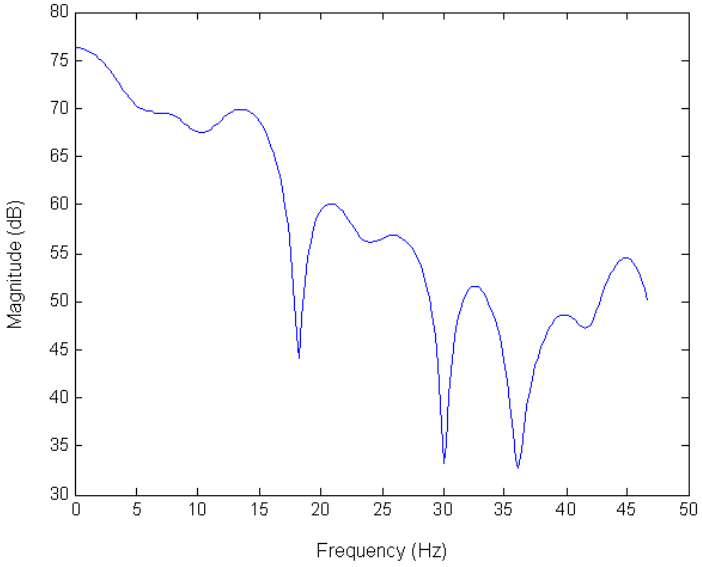

Figure 7.2: *The frequency response of the applied filter learned by SCICA. The filter coefficients (order is 79) are found by* $f_i(t) = a_i(-t) * w_i(t)$.

7.4 Results

Figure 7.3 depicts the results for the single character 'a' using SC-ICA on C3, Figure 7.3a shows the raw data and consists of three concatenated 1.5s P300 epochs. The solid vertical line represents the stimulus presentation time and the dashed line marks the point 300ms following the stimulus onset. It is clear in the raw data that the P300 representation is poor due to the low SNR. Figure 7.3b shows the manually selected ICs with P300 patterns in the measurement space after SC-ICA is performed. The signal recovered from the back-projected and summed ICs of Figure 7.3b is depicted in Figure 7.3c. For each epoch, a peak around 300ms after the stimulus is apparent.

Figure 7.4 depicts the same analysis steps on the same character ‘a’ but as measured from the opposite electrode C4. The raw data shown on Figure 7.4a still shows a poor SNR level. Figure 7.4b shows the manually selected ICs in the measurement space after ICA. Figure 7.4c depicts the recovered signal by the back projection of selected ICs of Figure 7.4b. The extracted data have the clear P300 peak for each epoch. Comparing both of the extracted signals on C3 and C4, the extracted P300 patterns are nearly identical and also exhibit similar amplitude.

Figure 7.5 demonstrates the same analysis on the same character ‘a’ but far from P300 focused area – on channel Oz this time. It is even more difficult to identify any P300 patterns from the raw data shown in Figure 7.5a. The extracted signal in Figure 7.5c is much cleaner and depicts similar peaks as in those extracted on C3 and C4 – albeit at a lower SNR. These results demonstrate that the SC-ICA method is able to extract the information from the single recording on the related central cortex region and even in the locations far from it.

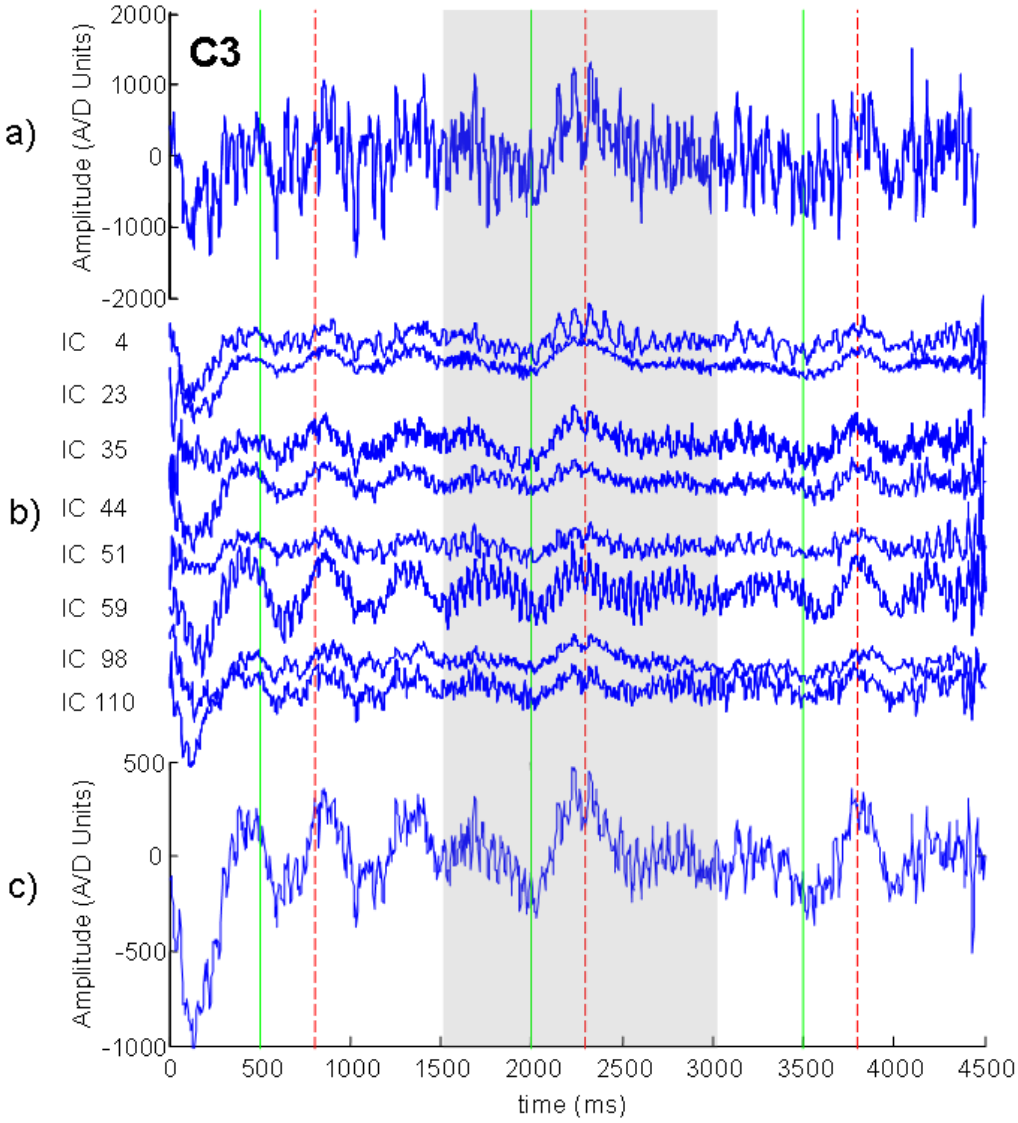

Figure 7.3: *(a) Raw data for character 'a' using single channel C3 consisting of 3 P300 epochs. (b) The selected ICs in the measurement space. (c) Extracted signal from the back-projected selected ICs.*

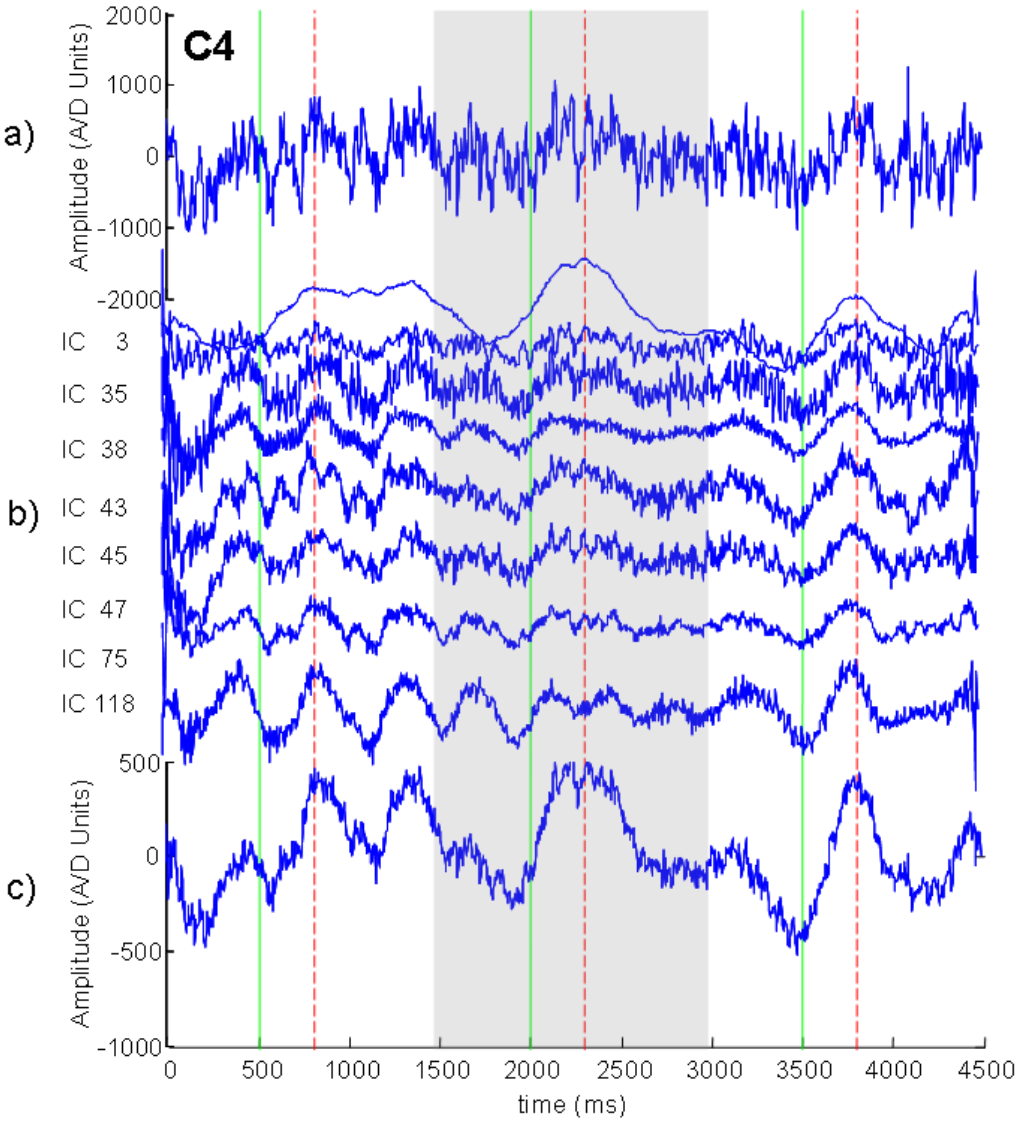

Figure 7.4: *(a) Raw data for character 'a' using single channel C4 consisting of 3 P300 epochs. (b) The selected ICs in the measurement space. (c) Extracted signal from the back-projected selected ICs.*

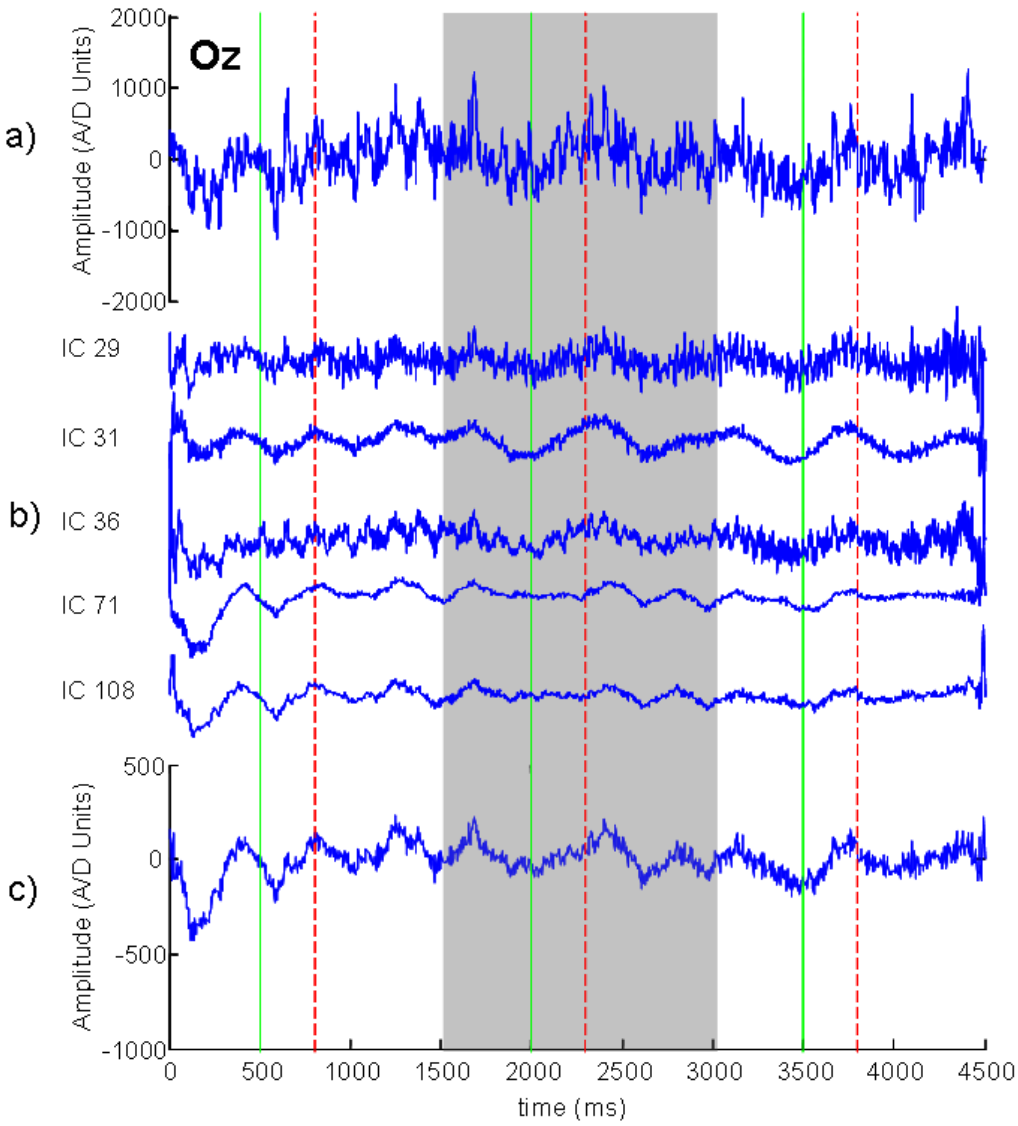

Figure 7.5: *(a) Raw data for character 'a' using single channel Oz consisting of 3 P300 epochs. (b) The selected ICs in measurement space. (c) Extracted signal from the back projected selected ICs.P300 peaks (at a lower SNR signal before) are apparent in the trace in (c).*

Figure 7.6 shows the analysis as before but using ST-ICA on both channels C3 and C4. Both raw signals are depicted in Figure 7.6a. After ST-ICA, the selected back-projected components for each channel are shown in Figure 7.6b. These selected components are finally projected back to the measurement space together to form an extracted signal. In Figure 7.6c the results show that using only one additional channel leads to a similar to results in Figure 7.3c.

Figure 7.7 depicts the same ST-ICA analysis but on the channel combination C3 and Oz and Figure 7.8 presents the results on the combination C4 and Oz. These tests are designed to determine whether it is possible to extract components for a pair of recordings: one from near the P300 site and the other further away. From the extracted signals are shown in Figure 7.7c and 7.8c, with the assistance of recordings from C3 and C4 in these two examples, the final extraction gives a better visual improvement compared to the previous extractions on a signal recording channel.

Figure 7.9 uses the same two channels on C3 and C4 as Figure 7.6 but this time the middle epoch is replaced by a non-P300 epoch. This aims to test the reliability of the ICA method in conditions where the response is alternately present. After applying ST-ICA, as expected the recovered signal (Figure 7.9c) depicts P300 peaks in the 1st and 3rd epoch but none in the 2nd. For the purpose of visualization, Figure 7.10 plots two of the extracted signals, of Figure 7.6c and Figure 7.9c, superimposed on top of each other. In order to measure the similarity of extracted signals, one way of doing this is to calculate the correlation values between these two signals for each epoch (there are three 1.5 second epochs in the signal and the correlation is assessed through each pair of epochs across the signals). The values for each pair of epochs are 0.93, 0.15 and 0.89, correctly depicting a strong correlation when P300's are present and very little correlation at epoch 2.

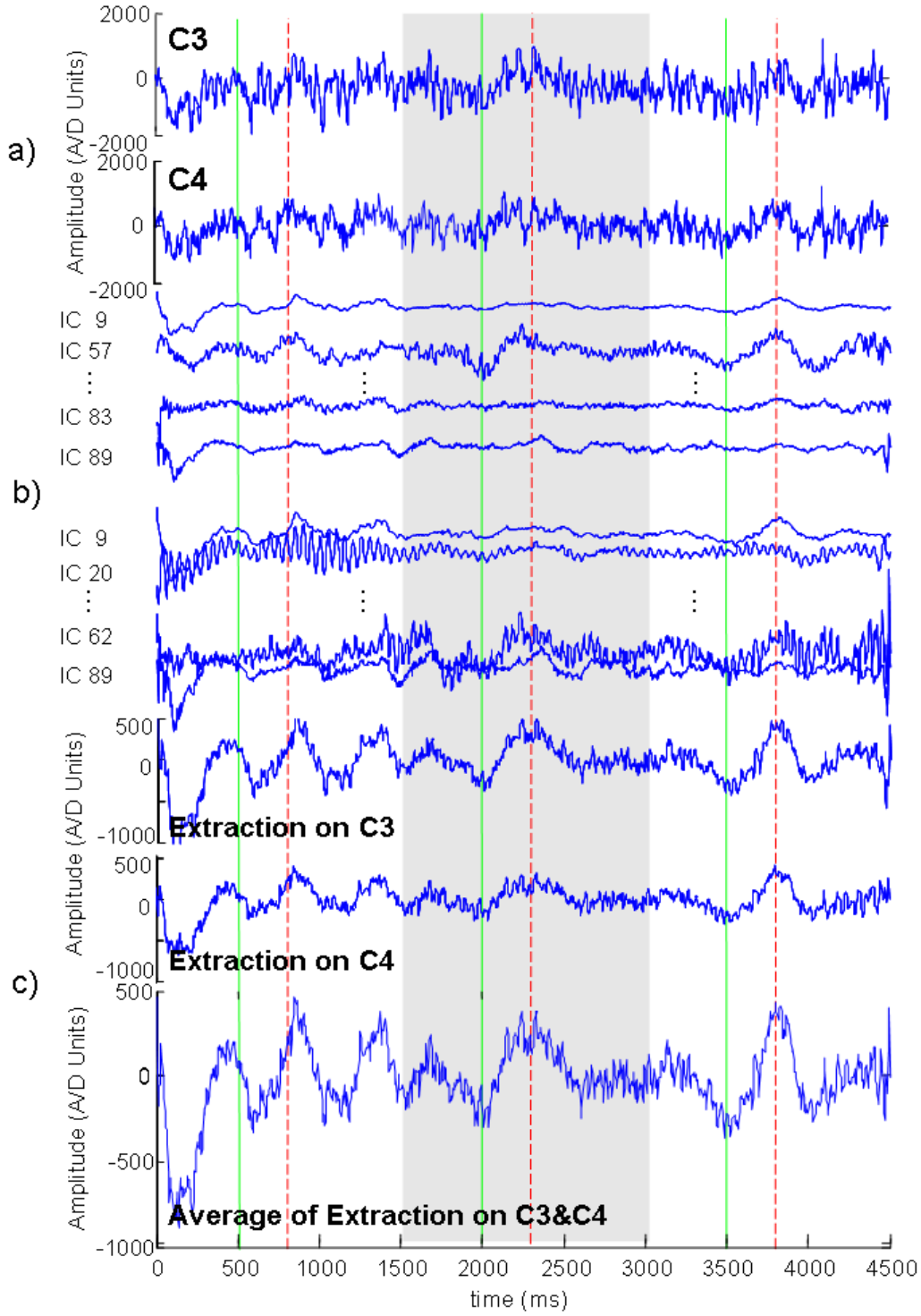

Figure 7.6: *(a) Raw data for character 'a' using channels C3 and C4 consisting of 3 P300 epochs. (b) The selected ICs for each channel in the measurement space. (c) Extracted signal from the back-projected selected ICs for C3 and C4 and the average of these two extracted signal.*

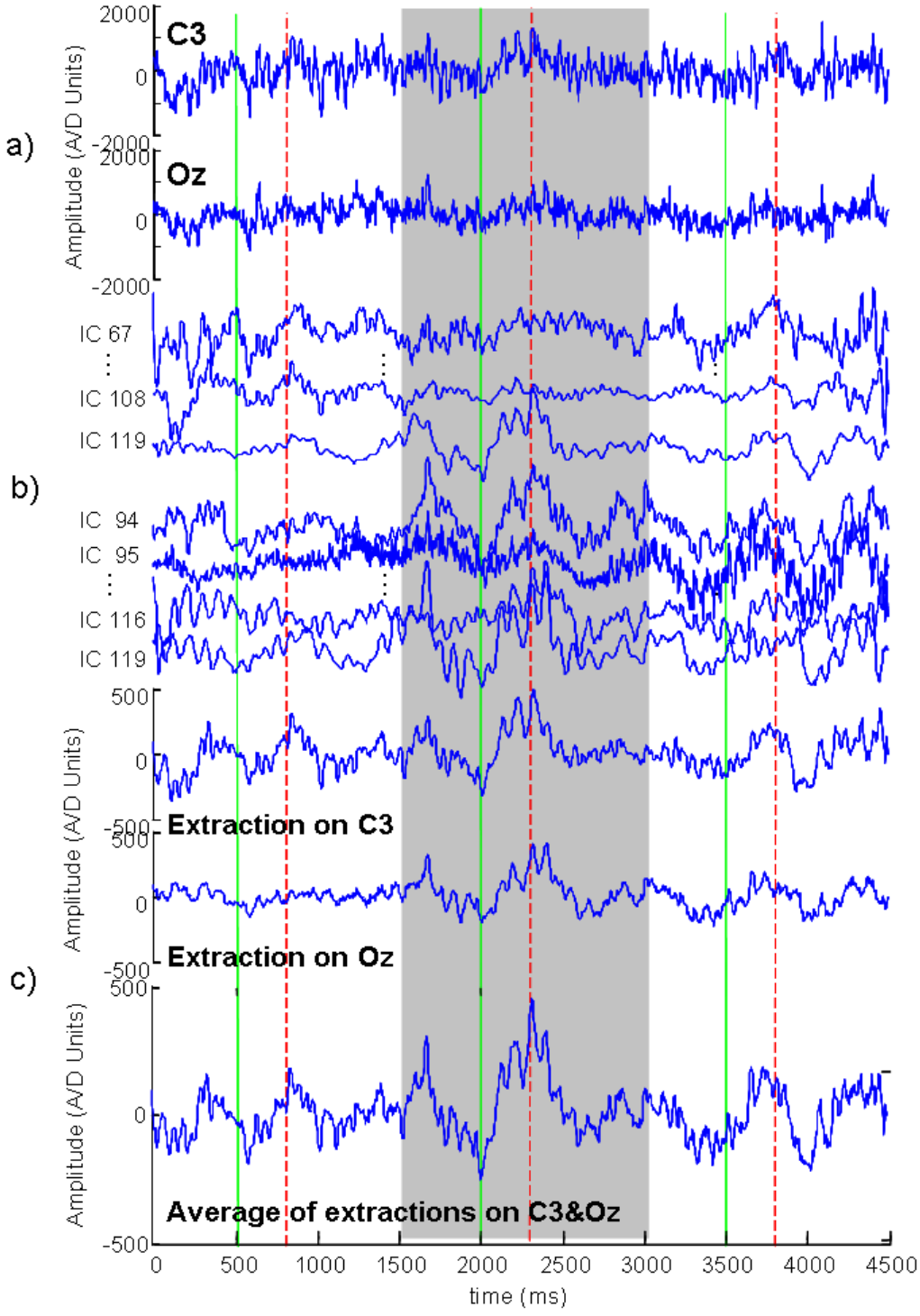

Figure 7.7: *(a) Raw data for character 'a' using channels C3 and Oz consisting of 3 P300 epochs. (b) The selected ICs for each channel in the measurement space. (c) Extracted signal from the back-projected selected ICs for C3 and Oz and the average of these two extracted signal.*

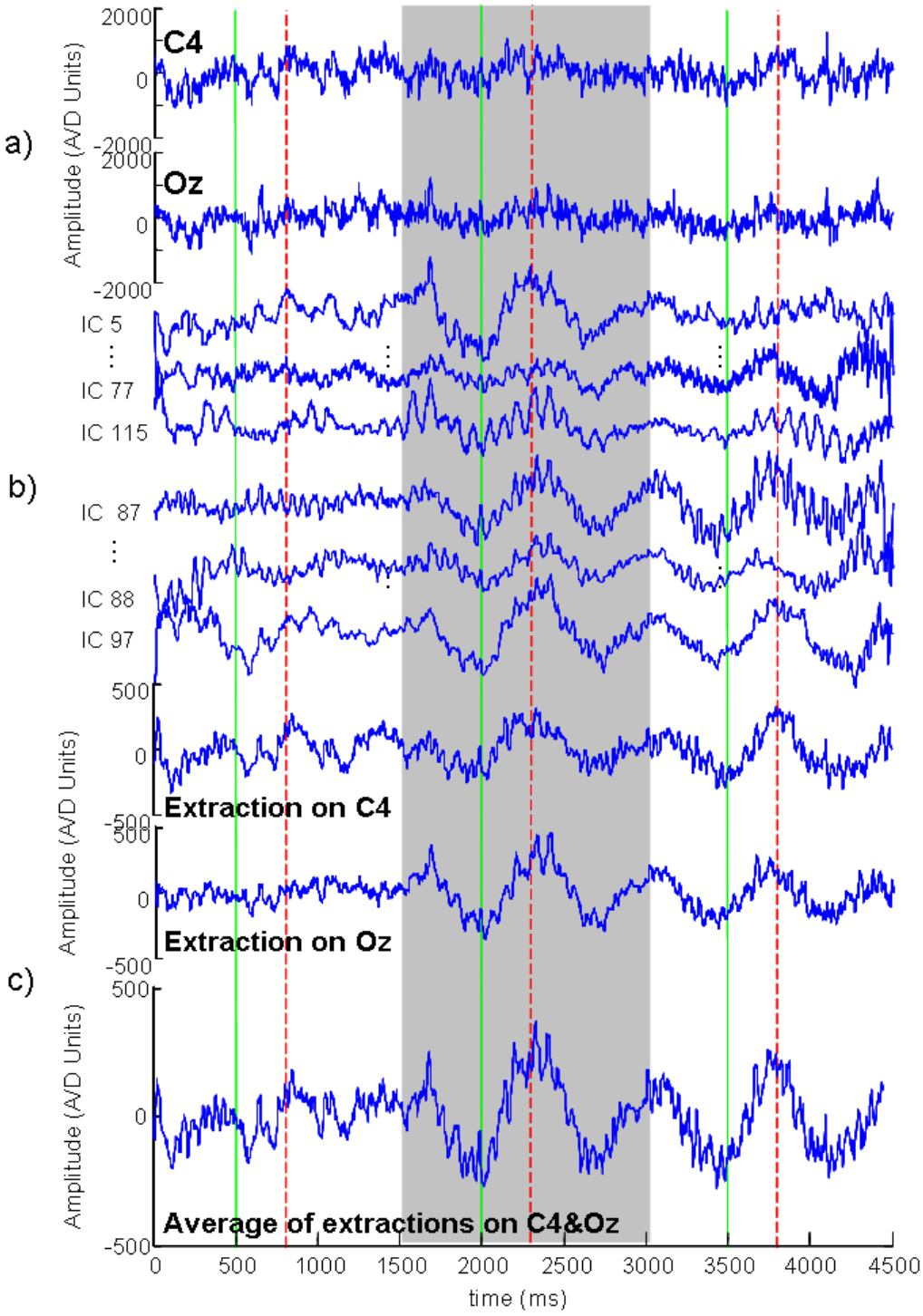

Figure 7.8: *(a) Raw data for character 'a' using channels C4 and Oz consisting of 3 P300 epochs. (b) The selected ICs for each channel in the measurement space. (c) Extracted signal from the back-projected selected ICs for C4 and Oz and the average of these two extracted signal.*

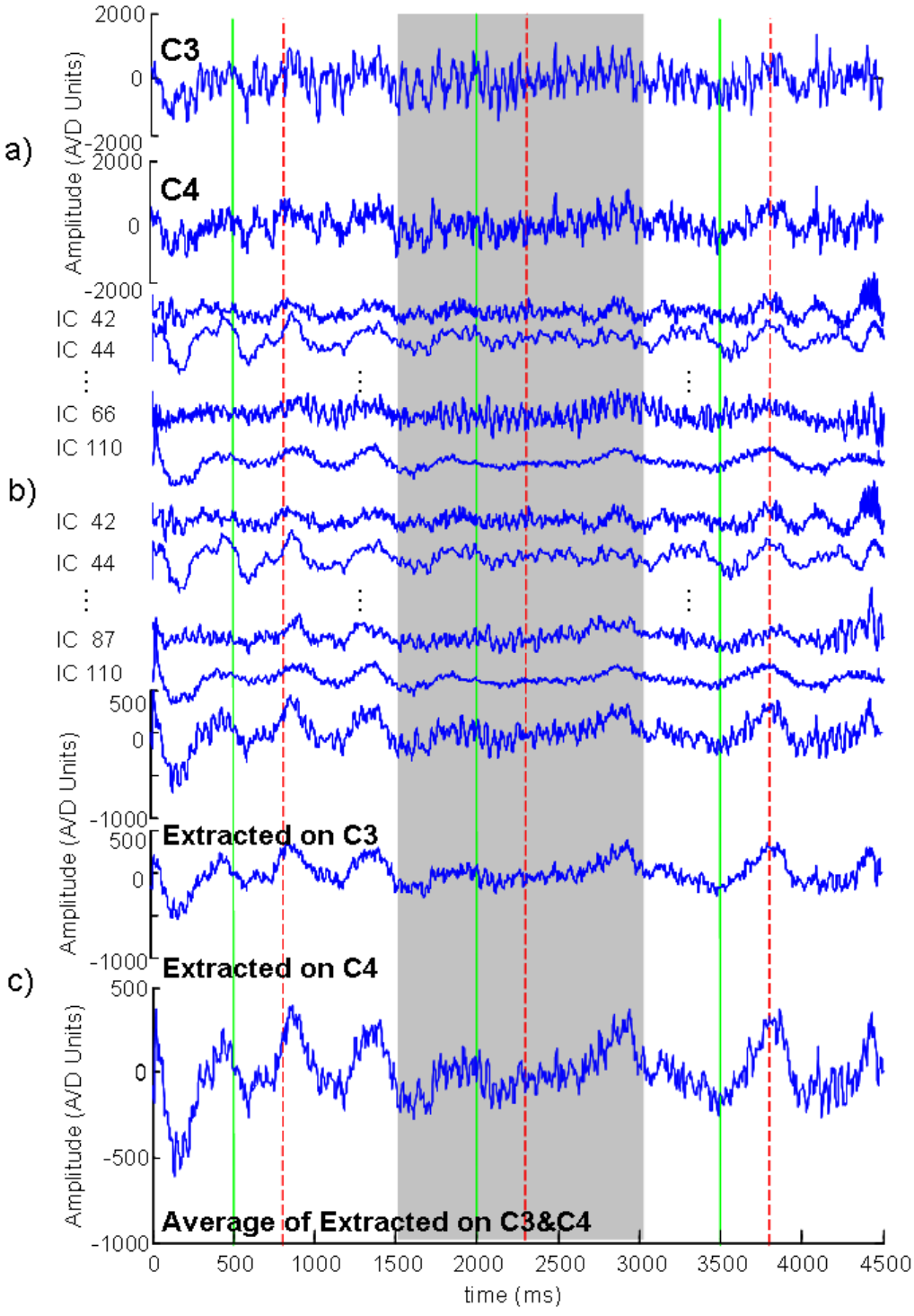

Figure 7.9: *(a) Raw data for character 'a' using channels C3 and C4 – epoch 2 is a non-P300 epoch. (b) The selected ICs for each channel in the measurement space. (c) Extracted signal from the back-projected selected ICs for C3 and C4 and the average of these two extracted signal.*

One way to assess the performance of extracted P300 signal is to compare the correlation with a predefined template. Here two P300 templates are prepared: Template A (Figure 7.11a) is a 1.5s averaged P300 which is obtained from averaging a number (200) of raw P300 responses on C3; Template B (Figure 7.11b) is 4.5s of activity which is repeated Template A three times in series to the same length of data trial. After that, it is easy to calculate values of correlation between the outputs of the SC-ICA analysis and these two P300 templates. It is worth noting that P300 is a subject dependent signal and can be affected by habituation in which the peak amplitude could be decreased. Moreover the P300 peak appears somewhere near to 300ms, and not exactly at 300ms. After a number of averages, the averaged P300 peak could become wider and the amplitude level could be lower compared to an individual P300 wave. Therefore the templates from coherent averaged raw EEG are not necessarily the most desirable way to construct a 'gold standard'.

The graph of Figure 7.12 shows the correlation values from raw data and extracted data after SC-ICA together with Template B, it indicates that SC-ICA over C3 consistently outperforms its raw signal counterpart. Furthermore, the same analysis on Oz yields good results, again always exceeding the raw signal counterpart and sometimes even exceeding the performance of the raw channel located over the P300 focus.

In order to build an online real time BCI system, the component selection has to be automatic. Here this study examines the possibility of applying an automatic method. Figure 7.13 depicts the results for the single character 'e' using SC-ICA on C3. Figure 7.13a shows the raw data and consists of three concatenated 1.5s P300 epochs as before. The solid vertical line represents the stimulus presentation time and the dashed line marks the point 300ms following the stimulus onset. The study applies the approximate component method which has been introduced in Section 3.6.3, Chapter 3. The proposed algorithm (depending on the application) can repeatedly be

used to extract more than one component of interest. But in this case, the first extraction has almost extracted the entire relevant P300 component. Figure 7.13b presents this extracted component in the measurement space. As a comparison, Figure 7.13c shows the extracted component for the same data but by using manual component selection instead. Both extracted P300 peaks between the manual and automatic methods are nearly identical. This confirms that the approximate method can perform at least at the same performance level as the manual selection method.

To be able to apply this automatic component separation method to this BCI system, the above approximation has been changed by a few steps to an improved version. For this new method, the P300 components are extracted by filtering the data with the filter that is obtained from the SC-ICA on the training data.

Figure 7.14 presents a set of 15 P300 repetition responses to one type of flashing stimulus for the character 'z'. The thick and dashed activity depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms. The raw P300 responses are quite noisy and it is not possible to view P300 patterns at all.

Figure 7.15 shows the extracted P300 components for the data as shown in Figure 7.14. The thick and dashed red line depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms. The extracted data are visually cleaner than the raw data. Some individual extracted components exhibit a similar shape as Template A.

The correlation values between Template A with raw P300 epochs and the extracted components are shown in Figure 7.16. For those original signals with low SNR or even without P300 pattern, the extraction cannot enhance the data quality. However once such extraction is applied on the signal containing P300 patterns, things change. By comparing the correlation values with and without extraction, the results (for example the correlation values between Template A and the sixth epoch

are 0.45 (on the extract data) and 0.32 (on the raw data) show clearly that such processing is able to enhance the quality of P300 in the recordings.

However applying an appropriate lowpass FIR filter could properly achieve the same outcome, this is discussed in Section 7.5.

Figure 7.17 highlights an unaveraged P300 response plotted from the raw and extracted data. The thick and dashed active depicts P300 Template A. Figure 7.18 shows the P300 response which benefits from two averages on the raw and extracted data. Figure 7.19 presents the P300 activity after three averages and Figure 7.20 shows the P300 activity after four averages. The results show that after a few averages the SNR of both raw signal and extracted signal is improved. Visually the extracted signal has shown a clearer P300 pattern than the raw signal. Figure 7.21 shows the correlation values between the averaged signals and Template A.

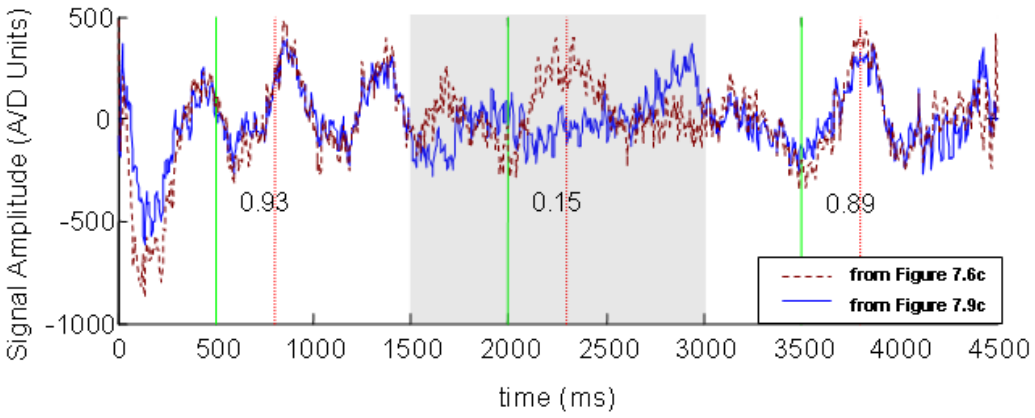

Figure 7.10: *Two extracted signals (from Figure 7.6c and Figure 7.9c), the values in bold indicate the correlation between each epoch across both channels.*

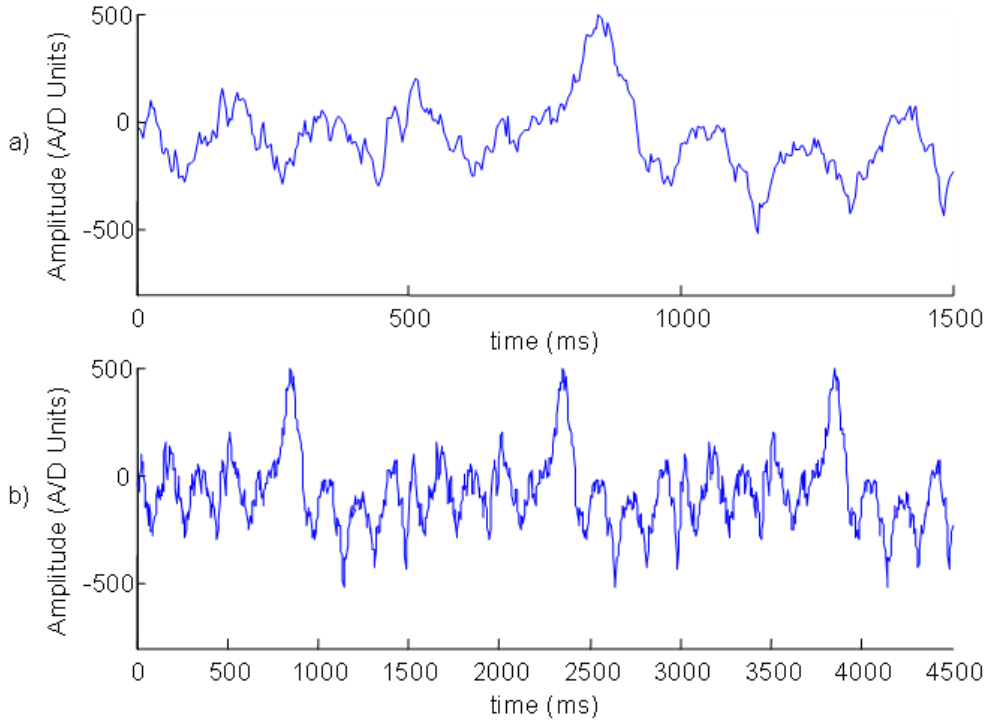

Figure 7.11: *a) P300 Template A is a clear 1.5s P300 response which is averaged from 200 P300 epochs; b) P300 Template B is a 4.5s activity which is repeated P300 Template A three times in series.*

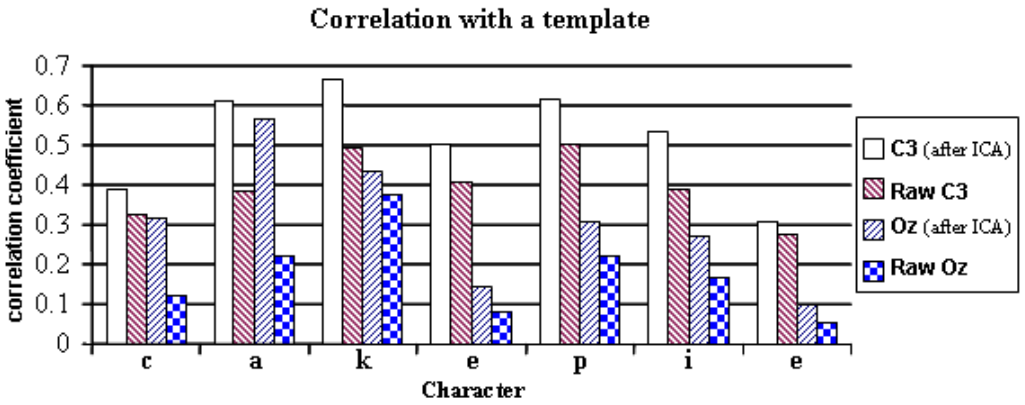

Figure 7.12: *Correlation values between signals (before and after ICA using single channel C3) and P300 Template A.*

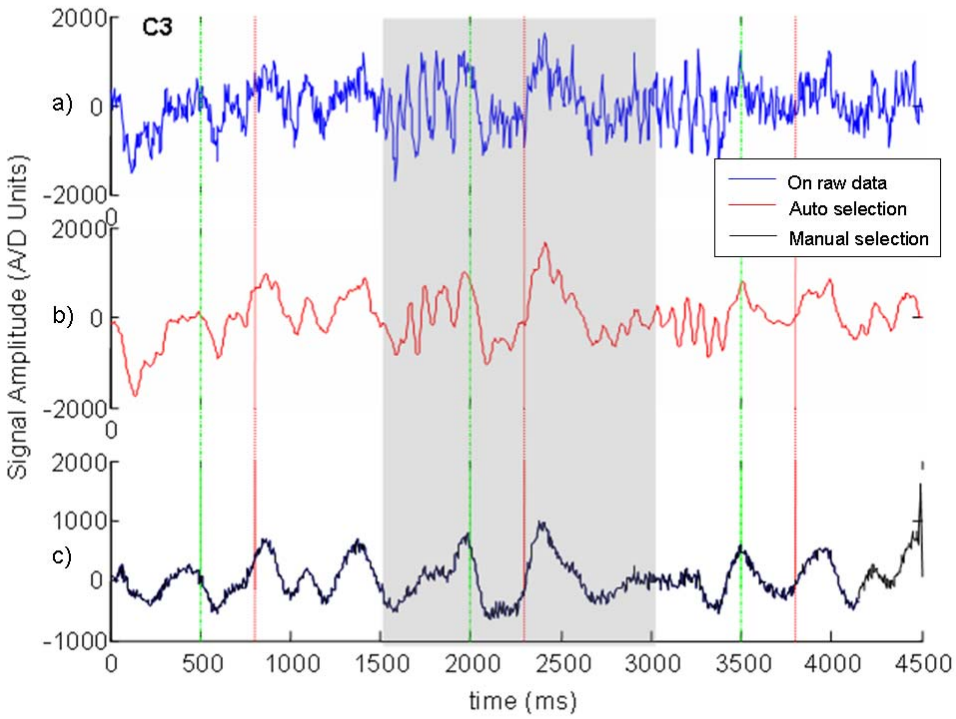

Figure 7.13: *(a) Raw data for character 'e' using single channel C3 consisting of 3 P300 epochs. (b) Extracted signal by applying the filter obtained from the approximate method. (c) The extracted signal by using manual component selection as a comparison.*

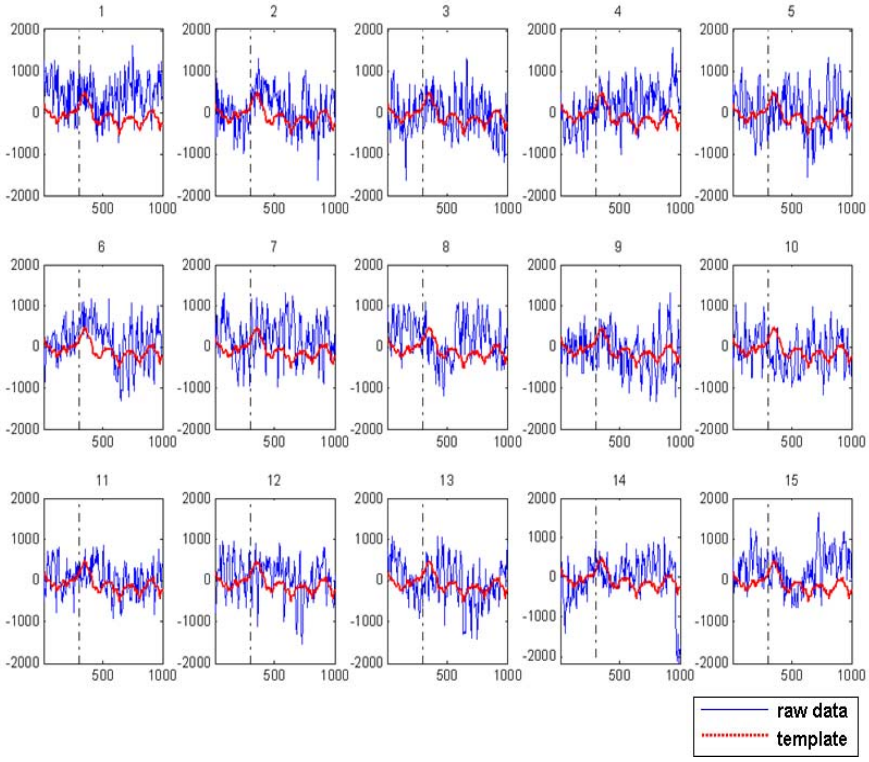

Figure 7.14: *A set of 15 P300 repetition responses to flashing stimulus for the character 'z'. The thick and dashed active depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

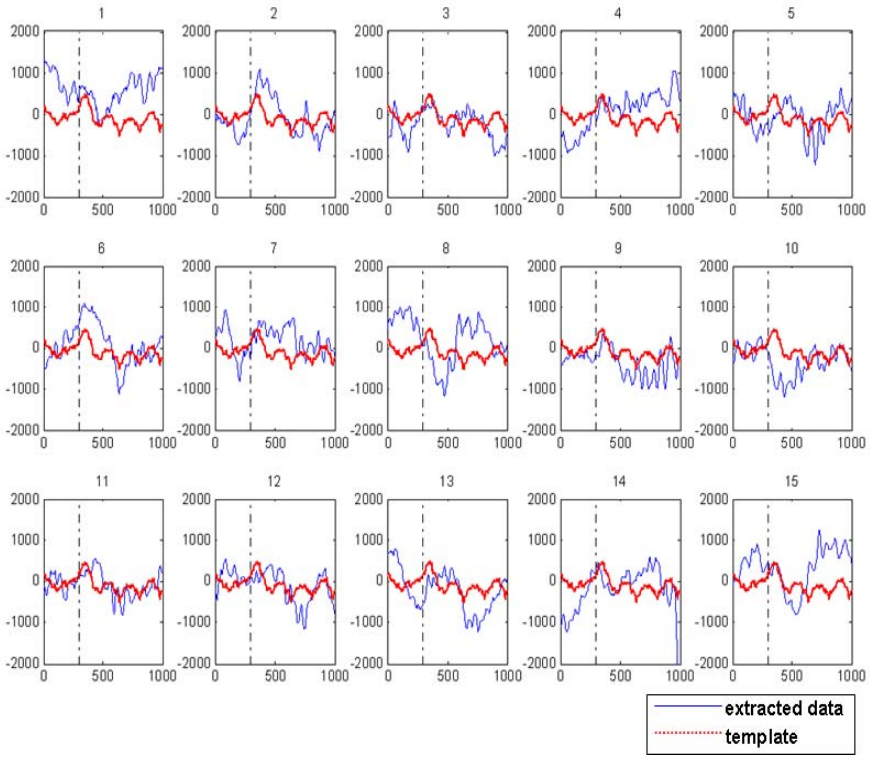

Figure 7.15: *15 Extracted P300 components from the same data shown in Figure 7.12. The thick and dashed line (red) depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

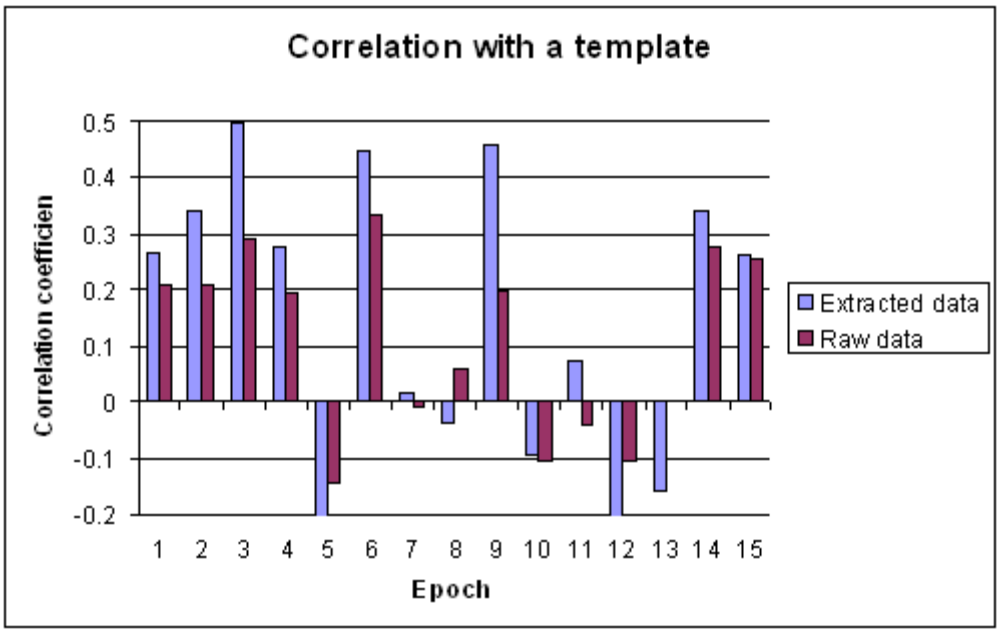

Figure 7.16: *The correlation values between Template A with 15 raw P300 epochs and the extracted version.*

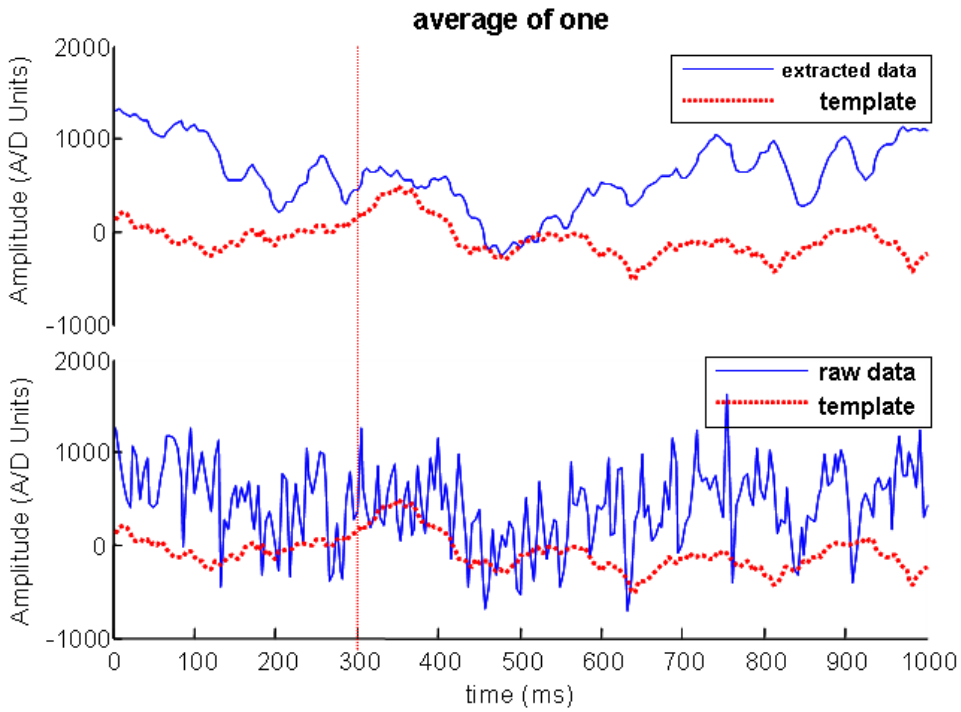

Figure 7.17: *A non-averaged P300 response plotted from the raw and extracted data. The thick and dashed active depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

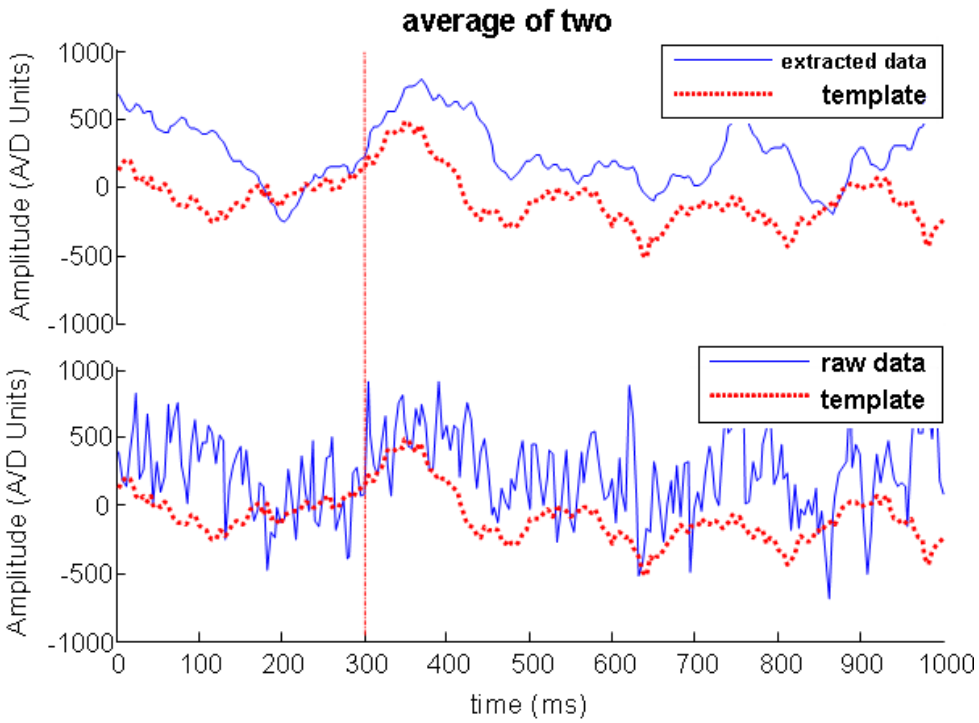

Figure 7.18: *A twice averaged P300 response plotted from the raw and extracted data. The thick and dashed active depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

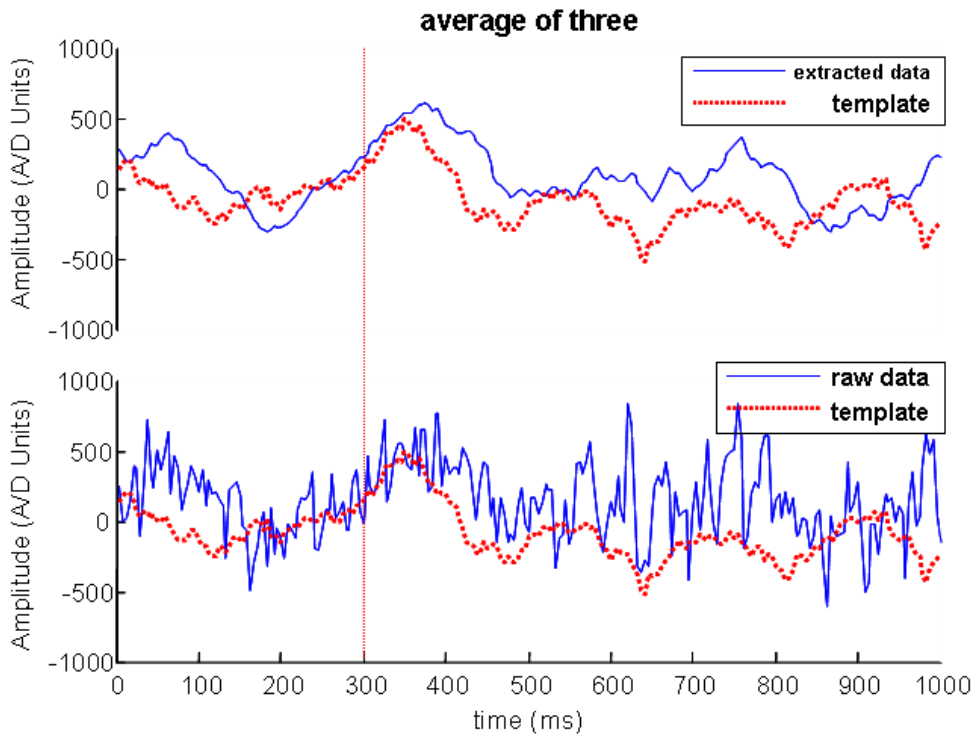

Figure 7.19: *A three-time averaged P300 response plotted from the raw and extracted data. The thick and dashed active depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

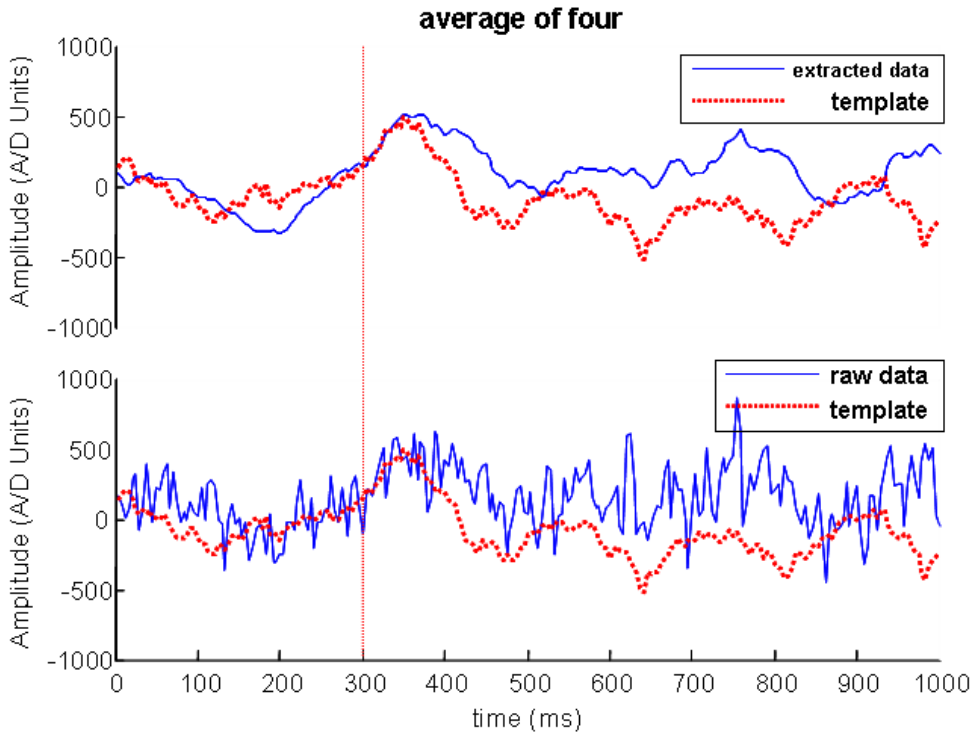

Figure 7.20: *A four-time averaged P300 response plotted from the raw and extracted data. The thick and dashed active depicts P300 Template A. The dashed vertical line represents the stimulus presentation at the time 300 ms.*

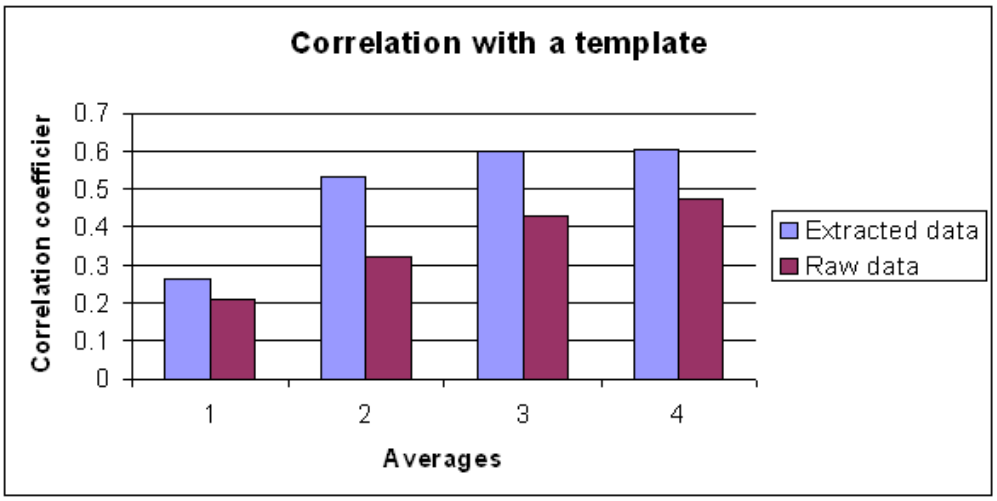

Figure 7.21: *The correlation values between Template A with raw averaged data and extracted averaged data.*

To compare the extraction performance between the traditional lowpass FIR filter and the proposed SC-ICA techniques, Figure 7.22 shows an example of filtering a signal segment consisting three 1.5s P300 signal trials. Results after SC-ICA with manual component selection, a lowpass FIR filter and the raw signal on its own are superimposed to give a better view for a close comparison. The results indicate both lowpass FIR filter and SC-ICA are able to extract the P300 peaks out of the noisy signal. The SC-ICA manages to preserve maximum wave appearance so that the signal amplitude nearly remains the same. As a drawback perhaps, the extraction still has some higher frequency noise left.

In order to examine and enhance the component extraction so to improve the final P300 word identification accuracy. Figure 7.23 shows twelve 8-time averaged activities for one character detection, these two bold waves are the targets containing averaged P300 EP. The first vertical line at 500ms presents the stimulus and the second vertical line at 800ms indicates the line 300ms after the stimulus onset. Figure 7.23 (a) shows results after the lowpass FIR filter. Components higher than 10Hz are filtered off, extracted signals are very smooth and P300 peaks in the target waves are

clear. As a comparison, Figure 7.23 (b) shows results after the proposed SC-ICA and the manual component selection method. Two clear bold target waves indicate SC-ICA is able to extract the component of interest. In both plots similar P300 peaks can be observed. Figure 7.24 demonstrates another good example of the performance difference caused by the same problem. Data shown here are twelve 9-time averaged activities for the detection of a character. After that coherent averaging is then applied to the extracted signals. Obviously the averaging does not bring much benefit to the lowpass filtered signals since they have already had high frequency noise reduced and smoothed through the filter. As the assumption of coherent averaging is that the noise is random and not fixed to the stimulus, so the noise can be reduced through averaging.

As a comparison of final classification performance for this P300 speller system, the classification accuracy of using the previous multi-channel ICA algorithm, and SC-ICA with manual/automatic component selection, FIR lowpass filter and raw data on its own. Table 7.1 depict the final classification accuracy whilst looking at various times of averaging on the data. Figure 7.25 gives a better visualization of such performance. Obviously the classification using multi-channel ICA is the best performed since multi-channel signals offer more information and allows ICA to construct a suitable spatial filter to extract component of interest from the mixing multi-channel data. Performance after FIR lowpass filtering gives a better result than one on the raw data, but not as favourable as in the multi-channel case. The SC-ICA together with manual component selection acts more likely a precise filter for each data portion (signals for one character form a data portion). This method reduces most of the noise and leaves the slow wave untouched so as to benefit the final classification accuracy during the processing of the averaging. SC-ICA methods end up extracting parameters that are equivalent to an FIR filter. SC-ICA with the manual component selection performs marginally better than the FIR filter, however the FIR

filter works, and since it is a much simpler method to implement, it should be the method of choice in this single channel application.

The most important thing is that this method can implement and provide an automatic and fast way to extract components of interest from the noisy signal which is more suitable to embed into an online BCI system.

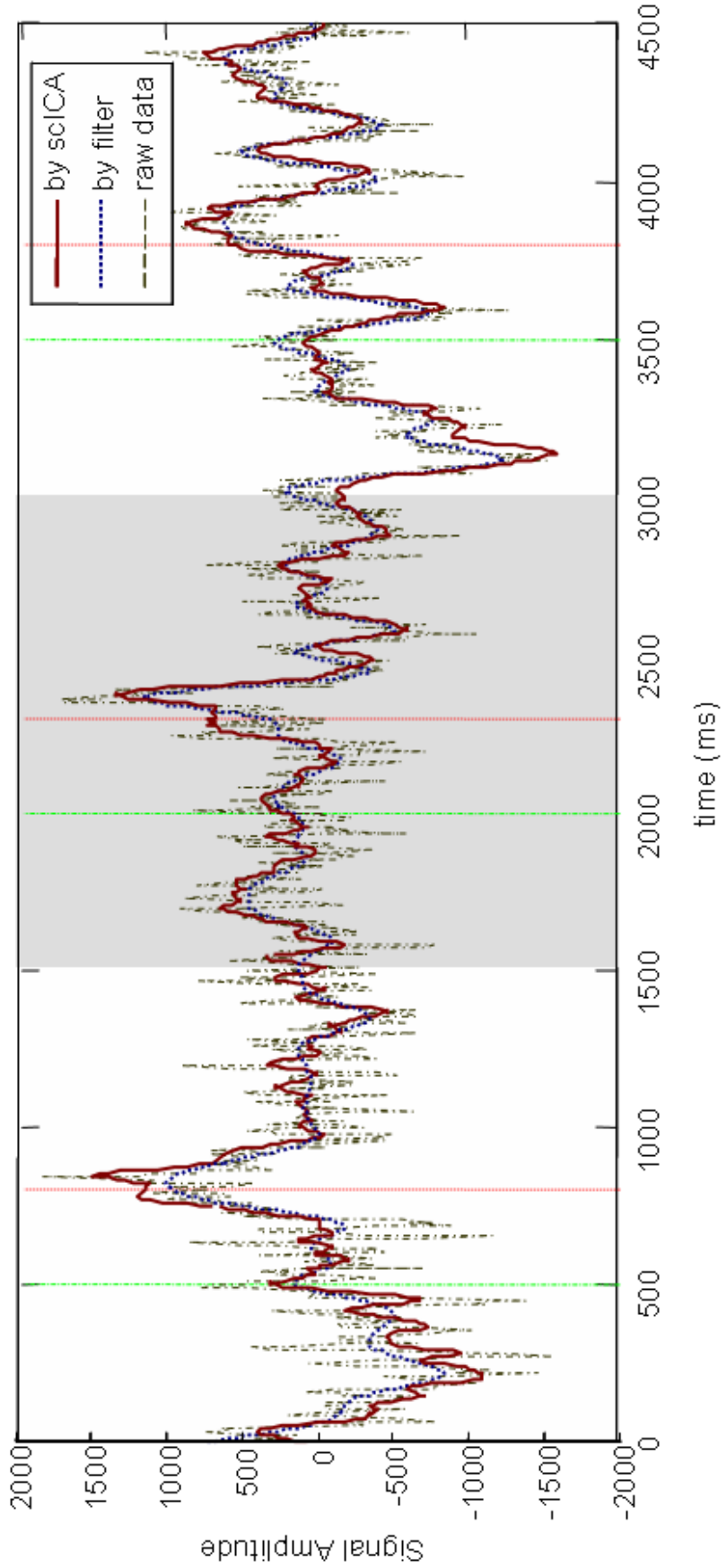

Figure 7.22. *An example of filtering a signal segment consisting three 1.5s P300 signal trials. Results after SC-ICA with a manual component selection, a lowpass FIR filter and the raw signal on it own are superposed to give a better view for a close comparison.*

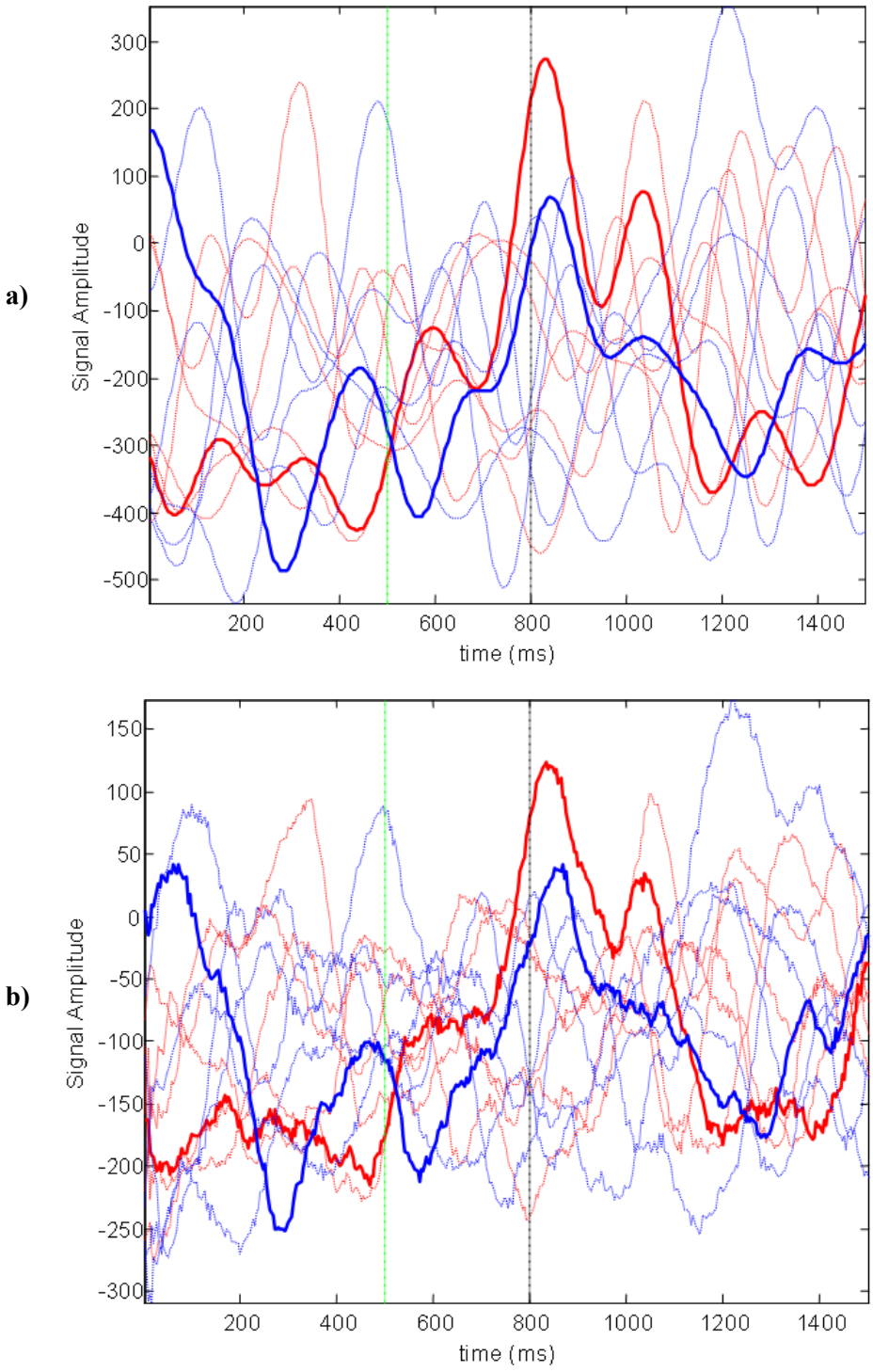

Figure 7.23: *Twelve 8-time averaged activities for one character detection, two bold waves are the targets containing averaged P300 EP. a) shows results after the lowpass FIR filter; b) shows results after proposed SC-ICA and the manual component selection method.*

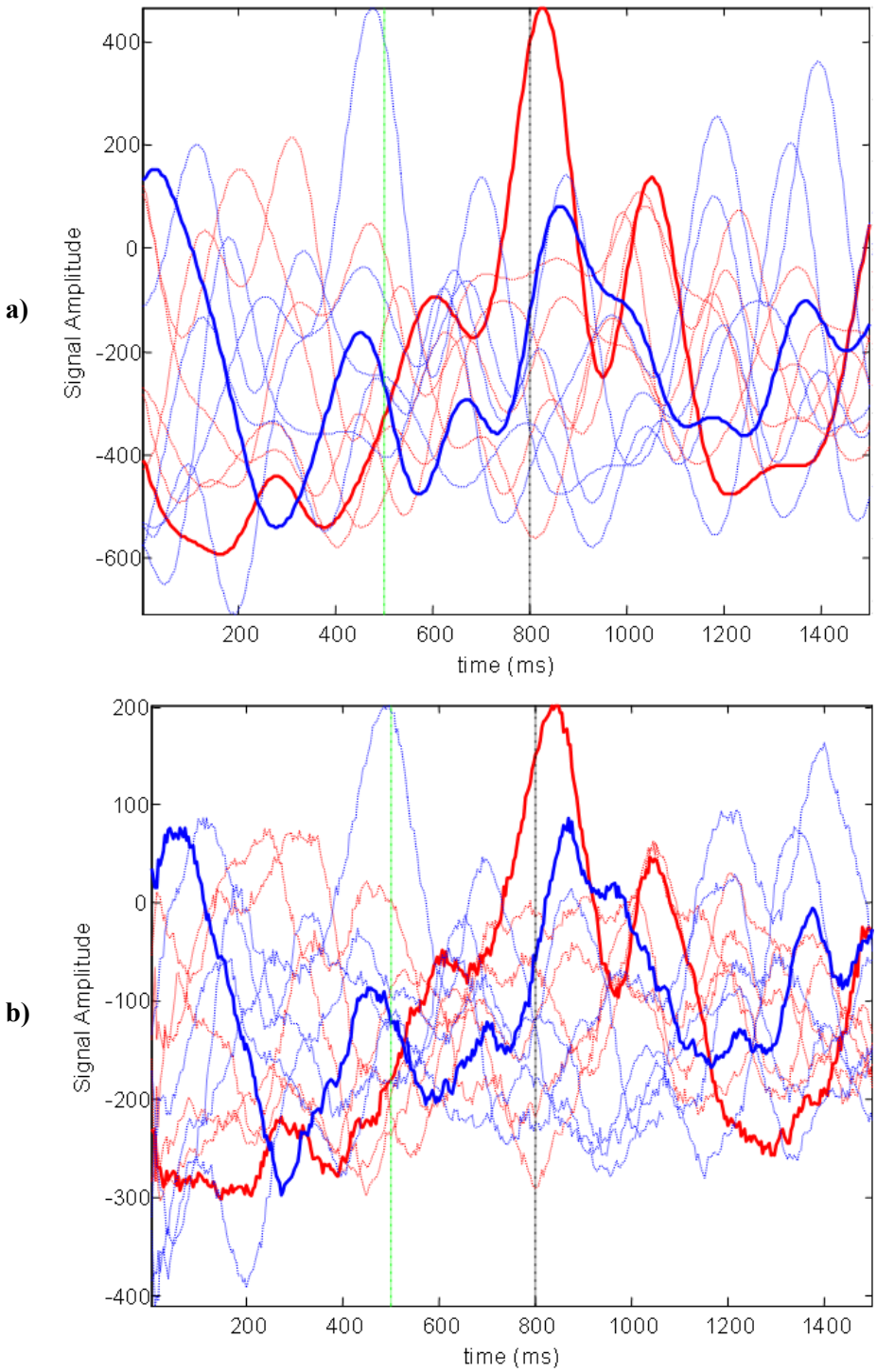

Figure 7.24: *Twelve 9-time averaged activities for another character detection, two bold waves are the targets containing averaged P300 EP. a) shows results after the lowpass FIR filter; b) shows results after proposed SC-ICA and the manual component selection method.*

Table 7.1: *The final comparison of the classification accuracy (%) using a previous multi-channel ICA method, the proposed SC-ICA algorithm, a lowpass FIR filter and a normal average method.*

Methods	Numbers of Averages														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
en-ICA	29.03	41.94	51.61	54.84	67.74	64.52	67.74	80.65	83.87	93.55	93.55	96.77	96.77	96.77	93.55
FIR filter	6.45	3.23	16.13	22.58	29.03	38.71	58.06	58.06	64.52	67.74	58.06	67.74	67.74	74.19	74.19
SC-ICA manu	12.90	16.13	25.81	29.03	41.96	45.16	61.29	58.06	74.19	70.97	70.97	70.97	70.97	80.65	87.10
SC-ICA auto	6.45	3.23	9.68	16.13	16.13	35.48	35.48	41.94	48.39	51.61	51.61	70.97	67.74	70.97	83.87
raw	3.23	9.68	16.13	16.13	19.35	29.03	29.03	32.26	51.61	38.71	38.71	41.94	51.61	45.16	38.71

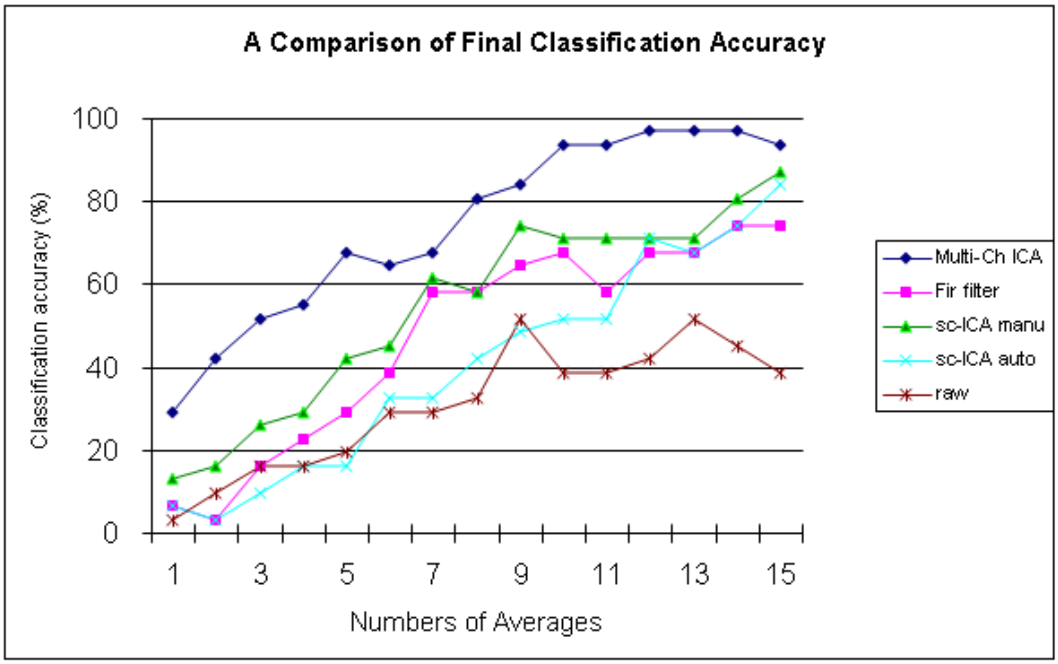

Figure 7.25: *The final comparison of the classification accuracy using a previous en-ICA method, the proposed SC-ICA algorithm and a lowpass FIR filter.*

7.5 Summary

This chapter proposes a new ICA technique on single or few channel(s) of recorded data. This chapter first demonstrates the proposed SC-ICA together with a manual component selection on just three P300 epochs from the total of thirty P300

repetitions for each character. The data is structured as three individual channels C3, C4 and Oz and also in another three combinations: C3&Oz, C4&Oz and C3&C4. The extracted data from one-channel studies show very clear P300 peaks with higher SNR compared to the raw data. Even from a relatively far channel on Oz, the technique is still able to trace P300 responses. Moreover with only a minimal requirement of an extra recording channel, this work shows that the proposed ST-ICA algorithm can increase the BCI performance. Since channels C3 and C4 are much closer to the P300 area, the original data SNR is already good so that the extraction performance from this pair of channels is much clearer and better than from others. By using this ST-ICA method, even the data from a far channel such as Oz has benefited from the assistance of extra recording from C3 or C4. The results have shown a better performance than those using Oz individually. Therefore with the proposed ICA algorithm on few or even a single channel, accurate spatial location of the recording electrode is not critical and the P300 responses can still be recovered to a usable degree of accuracy.

Furthermore since the IC selection algorithms in the literature currently are based on a manual selection which will negate a real time BCI system, here the study presents and demonstrates an application using an improved version of approximate automatic IC selection method to take the place of manual selection. The new method creates a filter by using the vectors of the mixing and unmixing matrixes from ICA. After that the system performs rather like on the signal filtering in one go rather than to select the relevant components. In this way the approximate method avoids the selection step and as a consequence performs substantially faster than the application with the manual method. The results show that this approximate component selection method has a similar performance as those from the manual method. As a comparison of classification performance after different extraction techniques are applied, this chapter presents results by using proposed multi-channel ICA method introduced in previous chapter, the SC-ICA method with manual/automatic component selection and lowpass FIR filtering. Results also show the classification accuracy along with

different amounts of averaging applied. The results indicate that naturally more information available to multi-channel ICA contributes to a better classification performance. As for the SC-ICA method, since in this situation only one recording is available, the current results are much better than results after lowpass FIR filter but are not good as those in the multi-channel applications. Whereas the appropriate number of averages improves the SNR of P300 patterns which lead to achieve a better classification.

In summary, these results on the P300 speller BCI dataset are extremely encouraging. The proposed ICA method can extract information from single or few channels with a reasonable accuracy even at low SNR channels. This method also has the potential of reducing the required number of repeated trials and speeding up the system – whilst maintaining accuracy. Furthermore this system can be easily extended to other, similar task-relevant EP based systems.

CHAPTER 8

Discussion and Future work

8.1 Background

Throughout the history of the development of BCI, although not long, people have realized that this may be the only way for patients who have suffered some severe debilitating diseases to increase their independence, leading to a dramatically improved quality of life and also reduce social cost. These diseases include ALS, brainstem stroke, brain or spinal cord injury and numerous other debilitating diseases. The diseases may result in paralysis of the entire motor system restricting both verbal and nonverbal communication, these also called locked-in syndrome. From a healthy person's point of view, the quality of life in such patients is low. However, the quality of life can be maintained regardless of the physical decline. By creating another communication output pathway, BCIs allow a person with restricted motor abilities to maintain communication and operate electrical devices and computers directly through their brain activity.

In theory, BCIs can be divided into dependent and independent types. In dependent BCIs, the brain's normal output pathways are not used to carry the message, but the pathways need to be active to carry certain generated brain activities pertaining to a given task. For example, a VEP based BCI depends on extraocular

muscles and EEG signals generated by the cranial nerves for the operation. Whereas for the independent case, relevant EEG signals will arise regardless of the activity pattern carried in brain output pathways. The generation of the EEG control signal depends mainly on the user's intent. Such control patterns include: SCPs, P300, μ and cortical neuron activity. Depending on the type of electrodes being used, in the literature BCIs are also separated into two classes, namely implanted BCIs and non-implanted BCIs. In the former, the micro electrode array is implanted directly into the grey matter of the brain through neurosurgery. As the electrodes are very close to the neuronal activity, this is able to produce the highest quality signals. However, there are safety and potential ethical issues with implanted methods in human subjects. Whether the implanted electrodes' performance remains optimal in the long term is also questionable at this stage. For non-implanted technologies, mainly surface attached EEG electrodes are used, which obviously are inexpensive and easy to wear, setup and maintain. But the drawbacks of non-implanted technologies could be such as incorrect electrodes location, electrodes falling off, time consuming to attach multi-electrodes and their characteristics varying with time.

EEG recording from non-implanted methods also could lead to the problem of poor signal resolution because the skull attenuates signals, dispersing and blurring the electromagnetic waves created by the groups of neurons. Eye movement, teeth clenching, frowning, varying facial expression, muscle contractions and other forms of artifacts can also spot the EEG signal. Moreover the noise contaminating the recordings is generally non-deterministic. To reduce these artifacts, filtering methods such as FIR, IIR filterings are usually applied to the recordings. However due to the diversity of artifacts, the performance of the filtering sometime is not effective.

In order to detect the relevant EEG patterns for BCI, one way of dealing with those artifact-containing segments is to simply discard them. However it is impossible to do this in real time BCI system. Based on the assumption that the noise is random and not reproducible, one possible way is to perform a coherent averaging, i.e. to

measure the EEG control signal more than once and average out all the measurements point-by-point. This can substantially improve the SNR. However too many times of averaging could slow down the speed of processing or even cause habituation. AR modelling is often used to remove EOG artifacts from EEG signals. Although one problem in practise is that when AR modelling removes EOG from EEG signals it could possibly remove part of the EEG signal too. Another method used for BCI application is spectrum analysis. It can be used to separate the relative contribution of the different frequencies and reduce the noise's troublesome frequency component from the signal. However if the frequencies of noise and components of interest are overlapped and fall in the same frequency band(s) it could be more difficult for the spectrum analysis to separate the component of interest from the signal. Since the EEG signals are so small in amplitude and sensitive to noise, to achieve an efficient detection the input signal (or features derived from it) should be selected with great care.

These techniques bring with them, however, two problems in developing an efficient BCI system: (1) The repetition of signal measurements for the use of signal reinforcement means longer recording times, and (2) multi-channel signal input for the use of spatial position optimisation produces huge amounts data. This really slows down the BCI signal processing capability and the detection accuracy is not guaranteed to be improved much anyway. For these reasons, the real-world application of BCI is somewhat limited. Therefore, the development of faster, more useful and stable BCI systems continues to remain a difficult challenge.

EEG signals measured on the scalp actually are the result of linear mixture of underlying cortical activities. The discovery of how the sources mix and extracting them from the observations would offer the great opportunity to enhance the signal SNR fundamentally. In the field of biomedical signal processing, the technique of ICA provides a tool to extracting a set of underlying sources or components from a set of random variables, measurements or signals. The ultimate aim is to extract

information underlying a set of signal measurements made over time. Fundamentally, the sources are assumed to be unknown and to be mixed linearly or even nonlinearly. ICA tends to demix these sources by exploiting this independence of the sources underlying the measured data. In the literature there are quite a number of publications about applications of ICA to biomedicine and even to BCIs. However most studies only tend to apply the existing ICA techniques to preprocess their data (e.g. filtering the data to reduce some simple noise). This work here expands the ICA concept and develops improved ICA techniques especially for use in the field of BCI research.

8.2 Objectives

The thesis presents two overreaching objectives in this BCI study: (1) to reduce the amount of reinforcement needed in order to reduce the processing time and hence increase the information transfer rate, whilst maintaining a high performance, and (2) to reduce the number of recording channels to make the application of BCI recording equipment less cumbersome and less reliant on accurate electrode placement in the context of increasing classification accuracy or at least with it remaining the same as the current classification performance.

8.3 Discussion

This thesis proposes and demonstrates several ICA techniques on two popular BCI systems: ERP based BCI and spontaneous EEG based BCI to deal with the EEG pattern detection problems and improve the overall performance. It starts with the preliminary analyses on the Southampton BCI pilot study. The purpose here is to examine the basic and advanced signal processing techniques on the pilot datasets so that the necessary knowledge can be accumulated and contributed towards follow-on studies. The pilot study includes ERP experiments (P300 and N400 tasks) and mental

imagery experiments (motor/non-motor imagery tasks). As this is a pilot study, the pilot data are not designed and structured as meaningful information carriers. Therefore it is difficult to measure and compare the performance with the results from the literature. However, irrespective, results show that the proposed ICA techniques are able to achieve an improvement between the extracted and raw ERP datasets so that the relevant P300 peaks are extracted more clearly and are much more easily identifiable. Results also indicate the mental imagery can be used in the Southampton BCI study if the suitable paradigm is available and the subjects are well trained.

In a multi-channel P300 based word speller BCI application, three ICA methods are proposed to improve the existing systems in the literature. After extracting and enhancing the desired responses through these ICA based techniques, the results show that much better performance can be achieved by using these techniques followed by a simple classifier, when compared to straight coherent averaging followed by simple classification, on the BCI competition data set. Furthermore, two of the proposed algorithms which benefit from the use of a spatial constraint and a predefined template mainly optimize and automate the components selection, meaning that this allows ICA to be run in an automated fashion. This is non-trivial as most ICA algorithms yield components which require interactive and subjective post processing. The final classification accuracy by an ICA technique has been increased to about 40% higher than one by coherent averaging. The study also shows that for fewer averages the ICA based post-processed techniques still exhibit quite good performance. For example, after ICA the data just applied to eight averages/epochs can achieve 83.9% classification accuracy whilst the data by coherent averaging can give only 32.3% accuracy.

Multi-channel spontaneous rhythmic activity based BCI applications are different to the above systems. In the above numerous repetitive trials could be performed, in this case the analysis must be based on “single trial” classification. Traditional methods need to select some trial recordings as the training datasets. However if the

training sets include artifacts, it can cause serious problems and lead to a poor performance. Therefore most methods require that the data should be artifact free, which can be achieved by several preprocessing steps such as filtering or manual artifact rejection. Here instead of applying any preprocessing methods beforehand, the study proposes an ICA technique based on time structure and spectral constraints to deal with this classification issue. It estimates the stack matrices by using the averaged lagged covariance matrices from the data. The idea behind the process is that the influence of artifacts is reduced since the procedure of averaging the covariance matrices acts as a filter which could balance and minimise the random noise level. Through the technique of spectrally constrained ICA, a spatial filter suited to each individual EEG recording is learned. This can effectively extract discriminatory information from two types of single-trial EEG data. Through the use of the ICA algorithm, the classification accuracy is improved by about 25%, on average, compared to the performance of the same classifier on the unpreprocessed data.

The thesis has already demonstrated how efficient ICA based techniques are able to enhance single-trial recording so that this means it is not necessary to apply many repetitions to obtain a reliable output. To examine the possibility of achieving the second objective— i.e. reducing the number of recording channels, an ICA technique working on single or few channel(s) recordings is proposed and demonstrated on the above same P300 word speller application. The results show that it is possible to extract single trial evoked potentials and to do so even on recording channels not sited over the event focus. Moreover, an automatic IC selection method is also proposed to take the place of manual source selection. The idea of manual selection is to simply check and choose the related signal patterns visually after ICA. But it is impossible to use it in an online BCI system which requires signal processing in real time. From the results the automatic IC selection performs substantially faster than the application of a manual method. The final classification performance indicates that more

information available to multi-channel ICA can benefit a better classification performance. However, as only one recording channel is available for the SC-ICA method, the current results are much better than results after lowpass FIR filter but are not as good as those in the multi-channel applications (unsurprisingly). Whereas the appropriate number of averages improves the SNR of P300 patterns which helps SC-ICA to achieve a better classification. Therefore the final classification by SC-ICA can be comparable to that obtained using multi-channel analysis techniques.

In this thesis, ICA has played an important role in signal enhancement (or signal de-noising). The results are extremely encouraging and show that the proposed algorithms are able to accurately and repeatedly extract the relevant information buried within noisy signals. Therefore the proposed ICA techniques are able to reduce the number of repetitions and reduce the number of recording channels in the BCI applications. Moreover the quality of the extracted signal is enhanced such that even a very simple linear classifier can achieve good classification accuracy. The use of a linear classifier in BCI applications simplifies the classification process and is desirable from a reduced computational complexity perspective. In order to maximally increase the final classification performance, more advanced classifiers can be applied in the BCI application. However results from this study did not necessarily show a significant improvement in performance on the same data used by the simple classifier. For example, a SVM on the data from a spontaneous EEG based BCI increased the performance by about 2% more than the simple linear classifier. Therefore a good classification rate coupled with low computational cost is designed in order to be able to achieve a reliable, on-line, system for BCI – especially for use outside of the clinical laboratory.

8.4 Future Work

Extensive future work should examine the possibility of using the single channel ICA concept to reduce the number of channels or even use a signal channel recording in the spontaneous EEG based BCI application. The work will discover the ICA application on identifying and isolating rhythmic component underlying single channel brain recordings.

Up to this point, the study has examined and analyzed improvements in several existing BCI systems and paradigms. When the development of BCI systems is reviewed, some BCI applications appear difficult to improve performance due to the limitation of the existing BCI paradigm themselves. Control signals like SCP requires long training time but its processing speed is still slow. μ rhythm based BCI again requires long time training and the performance is more dependent on the individual's skill.

An upgraded BCI paradigm which combines the P300 ERP paradigm and the VEP approach can possibly be a good solution to overcome the existing limitation. The advantage of grouping these two features in one system is that the P300 ERP is known to be a training free component for most subjects, such that in the literature the P300 word speller has been proved to be able to reach a relatively good information transfer rate and classification accuracy. However, one drawback is that the performance is much more dependent on subjects' intention and concentration. A VEP is a naturally occurring response and more dependent on the stimulus presentation than subject attentiveness. For example in a VEP application some objects are flashing with different frequencies and the frequency correlated signal pattern can be magnified and detected when a flickering stimulus is visually fixated. In the literature, VEP based BCI belongs to dependent BCI paradigms and requires little or even no training for the subjects, and it can reach the maximum information

transfer rate among all types of BCI systems. Furthermore a technique called a pseudorandom binary sequence (PRBS) [217] [218] can be introduced as a new feature in the presentation stimulation, VEPs can benefit from several advantages, such as, the recording time reduction and more efficient pattern detection. As a summary, visual stimuli with different frequency are involved in this new formed BCI paradigm, the user intends to response to low frequency flashings to generate P300 EP patterns; the natural response — the VEP is evoked by PRBS flashings regardless of the user's intention. All the characteristics are able to support each other and make the detection more efficient.

To make this paradigm work, the analysis could be divided into three steps:

Step 1: to construct P300 based word speller (it has been done already);

Step 2: to construct a VEP based system using a standard flashing stimulus at fixed (yet alterable) flashing rates. This will be followed by a VEP system with PRBS stimulus.

Step 3: to combine the P300 and VEP to build up the proposed BCI system. At the same time, the developed ICA techniques will be plugged into the system to maximally support the detection of the signal control patterns.

This new formed BCI system should offer these immediate benefits: fewer recording channels, less need for training, less (or no) averaging, less false detections and faster on-line classification.

Appendix A

Information sheet for Southampton BCI pilot study

University of Southampton

School of Health Professions and Rehabilitation Sciences

Brain Activity Associated with Language

Information for Research Participant (Version 2)

Purpose of the study
We are undertaking a study which involves collaboration between three schools within the University of Southampton (School of Health Professions and Rehabilitation Sciences {SHPRS}, the Institute of Sound and Vibration Research {ISVR} and School of Psychology). We are studying the effect of spoken and written language on brain activity. This research requires the involvement of healthy adults who do not have dyslexia, any skin diseases or allergies, or a history of a neurological disorder or head injury. We would like you to consider taking part in this study.

What the study involves
If you agree, you will be asked to come to the School of Health Professions and Rehabilitation Sciences for three appointments of approximately 1 hour each. During this time you will undertake some language assessments, during which the natural electrical activity of your brain ('brain waves') will be recorded from your scalp. This involves placing 14 sensors attached to leads on your scalp, and keeping them in place with a 'vaseline-like' gel and sticky tape. The leads are sufficiently sensitive to pick-up electrical activity from the brain, but they cannot send any information *to* your brain. Brain activity is recorded in the form of an EEG (electroencephalogram).

Once the electrodes are in position, you will be asked to complete some simple tasks involving: looking at pictures on paper or a computer screen; listening to certain sounds; thinking about specific things we ask you to (e.g. imagine you are sitting on a beach…). The sensors will then be removed, and we will try to remove as much of the gel as possible. You may, however, wish to wash your hair on the same day.

Personal information will not be released to, or viewed by, anyone other than the researchers named below. All results will be coded so that they are anonymous, i.e. they will not include your name or any other identifying information.

Your participation is voluntary and you can withdraw consent at any time, even once the data have been collected. If you chose not to participate there will be no consequences to your academic progression if you are a student or your employment if you are a member of staff. If you have any further questions, please email Morwenna Collins on moc@soton.ac.uk.

What we hope to achieve
By conducting this research, we hope to learn more about how people understand spoken language and form words in order to speak and hold a conversation. We then plan to use the information to better understand the problems faced by people who have difficulties speaking and understanding after strokes and head injuries, and to use the technology of recording brain activity to help in their treatment.

If you wish to participate in this study please contact Morwenna Collins directly. Participants will be asked to sign a consent form at their appointment.

Prof Maria Stokes – Director of Research & Professor of Neuromuscular Rehabilitation (SHPRS)
Morwenna Collins Research Fellow (SHPRS)
Dr Christopher James - Lecturer in Biomedical Engineering (ISVR)
Dr Alexandra Hogan - Lecturer in Developmental Neuroscience (School of Psychology)
Dr Avijit Datta - Consultant Respiratory Physician (Honorary Lecturer School of Medicine)

July 2005

University of Southampton

School of Health Professions and Rehabilitation Sciences

Brain Activity Associated with Language

Post-participation description of study (Version 2)

The aim of this research is to study healthy peoples' performance when they undertake commonly used language assessments. Specifically we are interested in brain activity associated with the understanding of written and spoken words. We need to collect this information in preparation for a study of people who have had a stroke or head injury and have difficulties speaking or understanding what is being said. We have experience of studying certain types of brain activity in response to processing simple and complex sounds and we are now hoping to apply this information to developing new technology for people with disabilities.

We were also interested to see if imagining or remembering certain things produces the same consistent pattern of brain activity across individuals and testing sessions. In order to study this we asked you to remember your favourite piece of music, imagine moving a part of your body and imagining scanning around your house. We were investigating the possibility of being able to record consistent brain signals while people think of certain things. We aim to make these recorded signals reliable enough to control a cursor on a computer screen, which is known as brain-computer interfacing or BCI. This would have implications for people who have damaged their language systems and are unable to understand or speak due to accident or illness. We did not tell you too much about the things you were going to be asked to think about, as this may have affected the way you performed during the tests and made the results invalid.

Brain activity recorded from one person does not give any meaningful information, so no individual feedback is given in this study. It is only when we average together brain activity from healthy subjects and compare them in a later study to those from people with language difficulties, that we see anything new or of interest. Although our brain activity study is similar to a clinical EEG study (e.g. for epilepsy), it does not provide the level of information necessary for a clinical diagnosis. If you are concerned about any aspect of your physical or mental health you should consult your GP in the first instance.

We thank you for your participation in this study, and remind you that you may withdraw consent for us to use your results even though you have already participated. If you have any further questions, please contact Morwenna Collins (moc@soton.ac.uk).

If you have any concerns that the researchers could not address any issue about your rights as a participant in this study, or if you feel that you have been placed at risk, you may contact the Chair of the School of Health Professions and Rehabilitation Sciences Ethics Committee, School of Health Professions and Rehabilitation Sciences, University of Southampton, Southampton, SO17 1BJ. Phone: 023 8059 2142.

July 2005

Appendix B

Consent form for Southampton BCI pilot study

University of Southampton

School of Health Professions and Rehabilitation Sciences

Brain Activity Associated with Language

Consent Form (Version 2)

I _____ (full name in block capitals) have read the information sheet (dated Version 2) and consent to participate in this study. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefits to myself. I understand that the data collected as part of this study will be treated confidentially, and that published results of this project will maintain my confidentiality. In signing this consent letter, I am not waving my legal claims, rights, or remedies. A copy of this letter has been offered to me.

I give consent to participate in the above study (circle yes or no): **YES** **NO**

I consent to the session being videotaped (circle yes or no): **YES** **NO**

If yes, I consent to the use of my videotaped image (face) being presented to students or other health professionals for the purpose of:

Education in this University (circle yes or no): **YES** **NO**

Education in any University other than Southampton (circle yes or no): **YES** **NO**

National and international conferences (circle yes or no): **YES** **NO**

Signature _____ Date _____

Name of Researcher taking consent: _____

July 2005

Appendix C

Results of preliminary analyses of a BCI pilot study

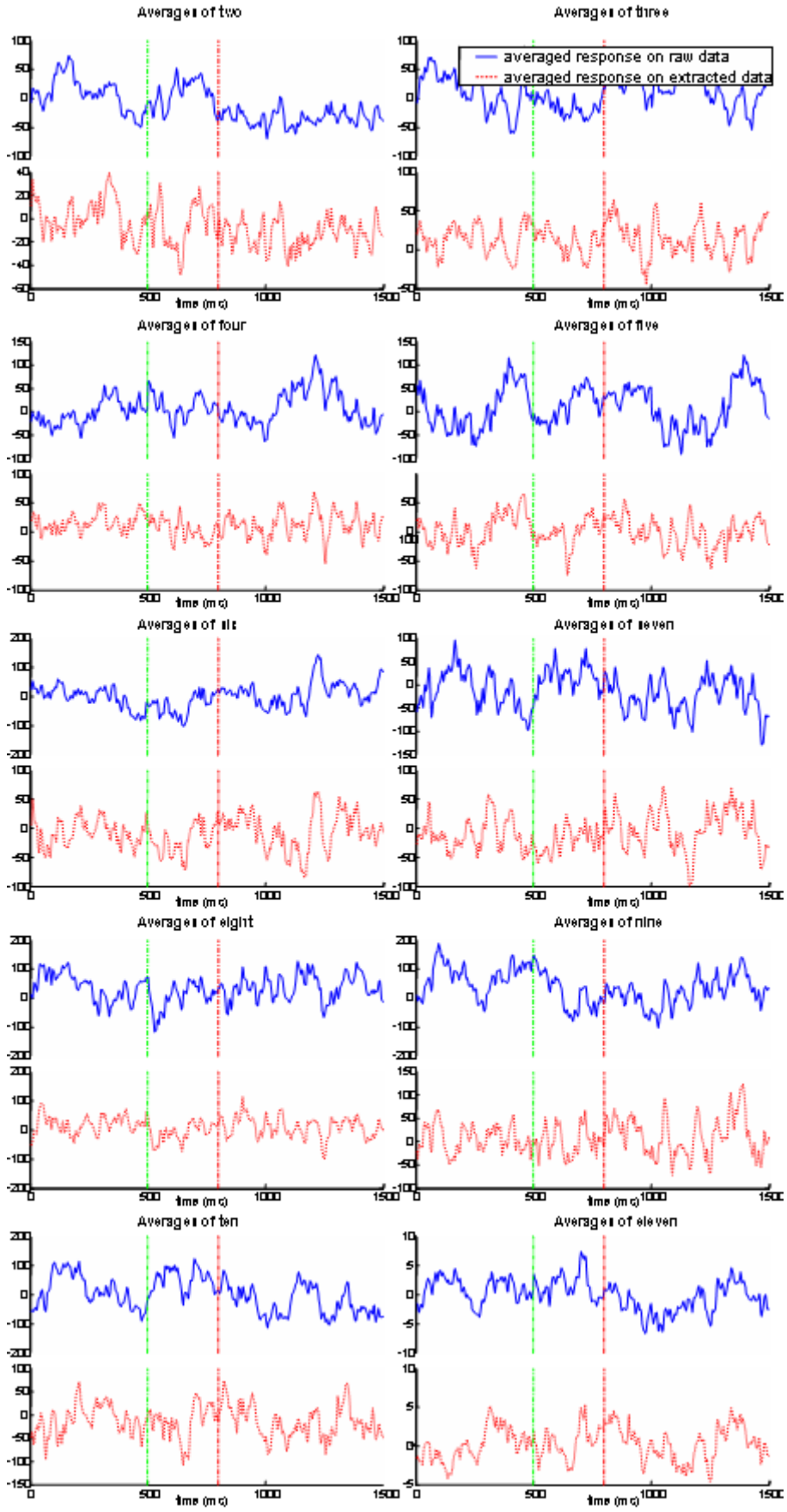

Figure C.1: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 2 after normal ICA. For each plot, the upper signal shows the averaged activity on raw data, and the bottom one plots the same average but on extracted data.*

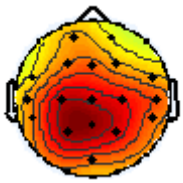

Figure C.2: *The one selected topographic map from the data of Subject 2 after the normal ICA. Each map is corresponding to an independent component. Therefore there are one selected components projected to the original measurement space.*

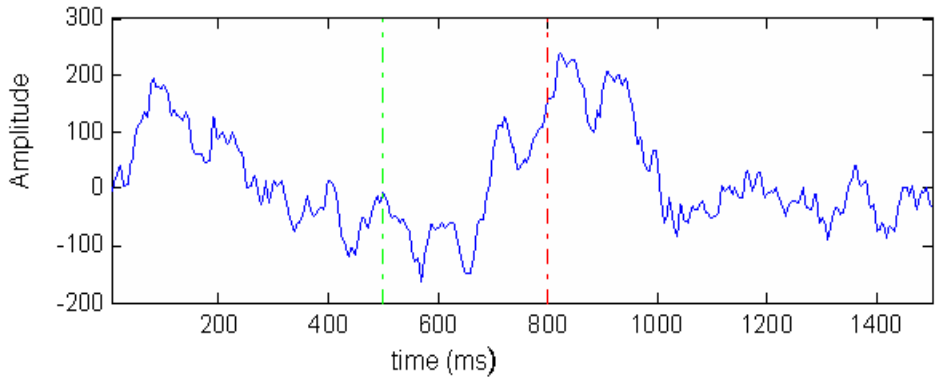

Figure C.3: *the template used to compute the correlation coefficient with the averaged P300 activities on the data from Subject 2. The template was an average of selected P300 epochs.*

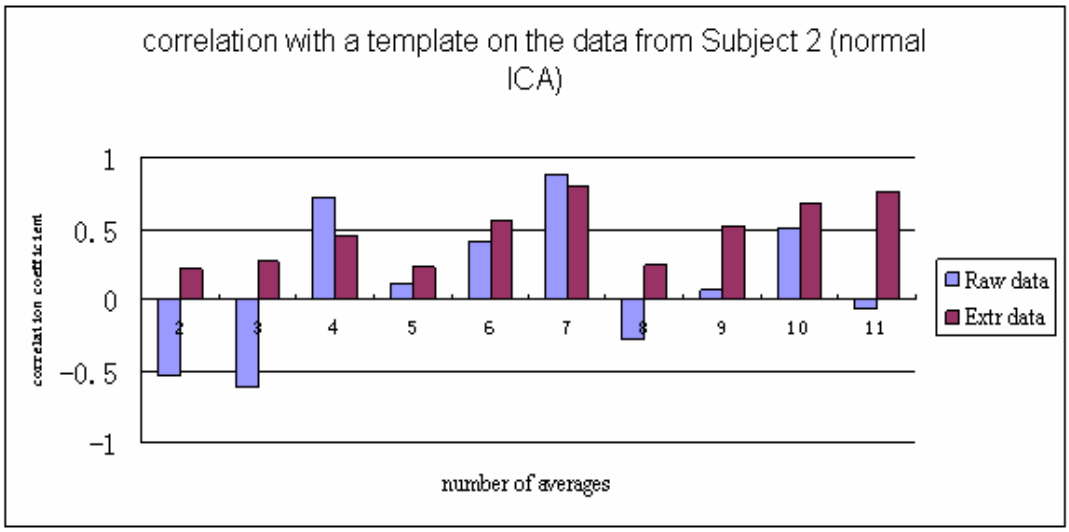

Figure C.4: *The performance of the correlation between the P300 template and the averaged P300 activities before and after normal ICA on the data from Subject 2. There were a total 10 pairs of random averages of P300 activities.*

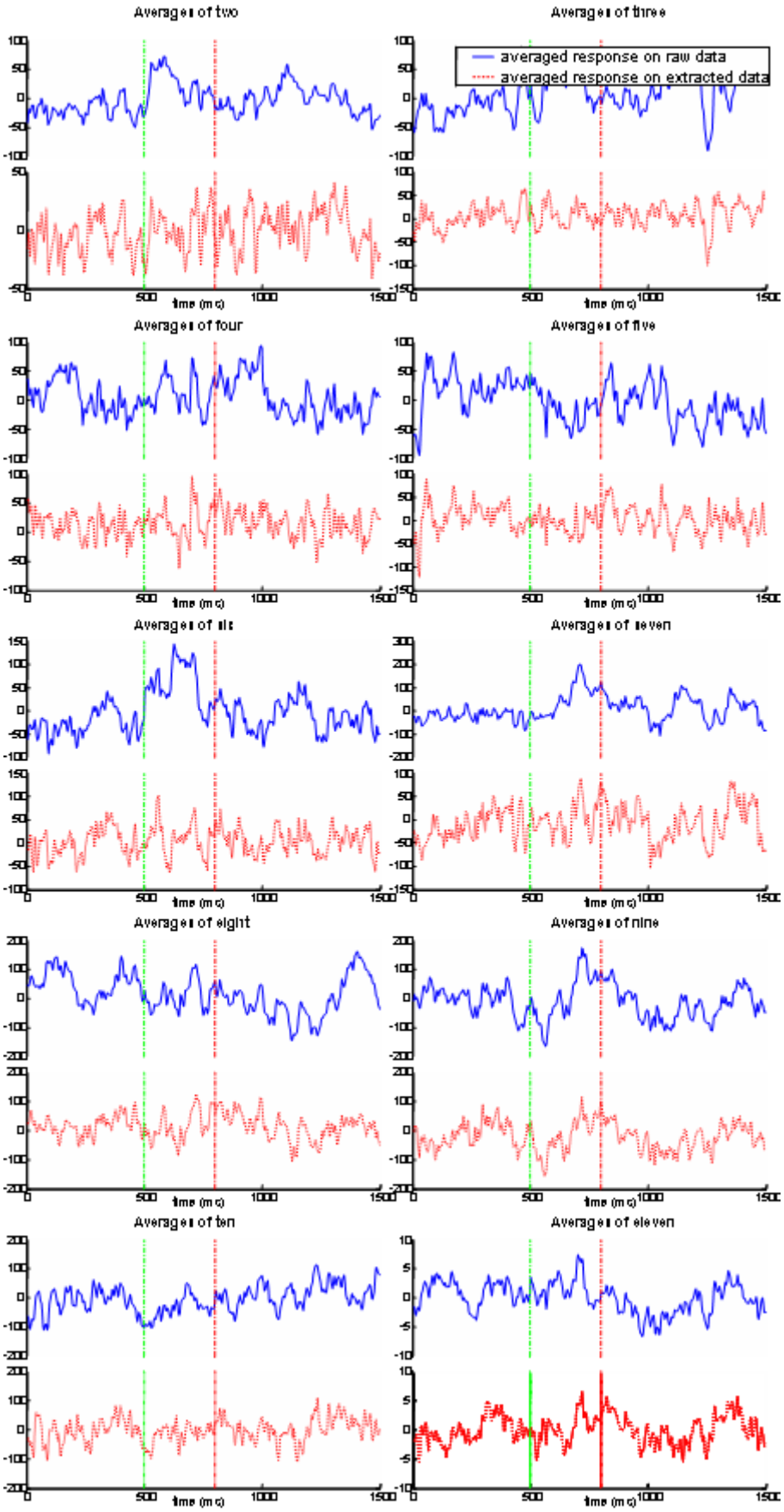

Figure C.5: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 2 after spatially constrained ICA. For each plot, the upper one shows the averaged activity on raw data, and the bottom one plots the same average but on the extracted data.*

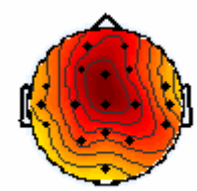

Figure C.6: *The constraint topographic map for Subject 2. This constraint was a selected column of the mixing matrix after a normal ICA on a special data portion which includes all the selected P300 epochs*

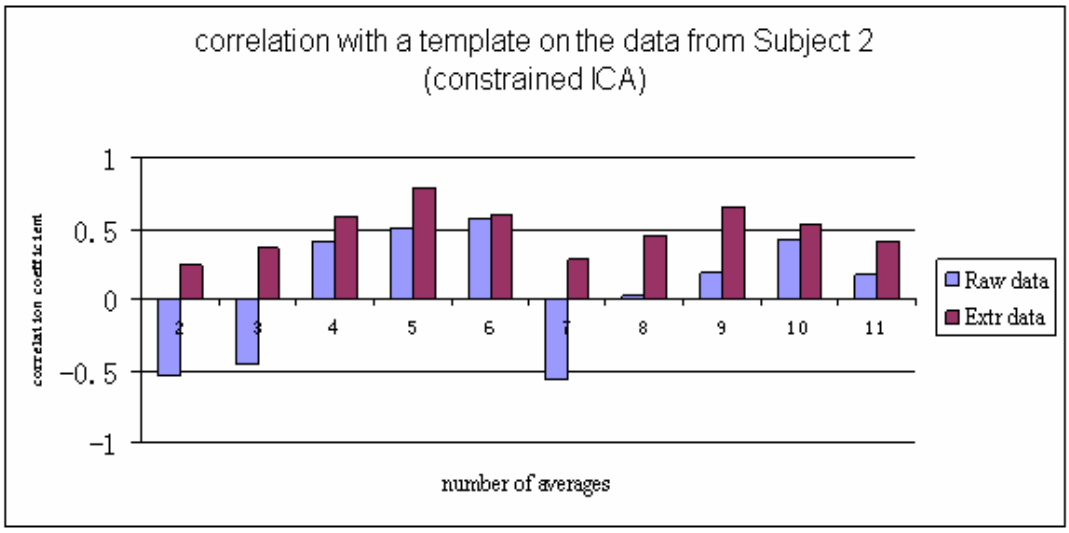

Figure C.7: *The performance of the correlation between the P300 template and the averaged P300 activities before and after the spatially constrained ICA on the data from Subject 2.*

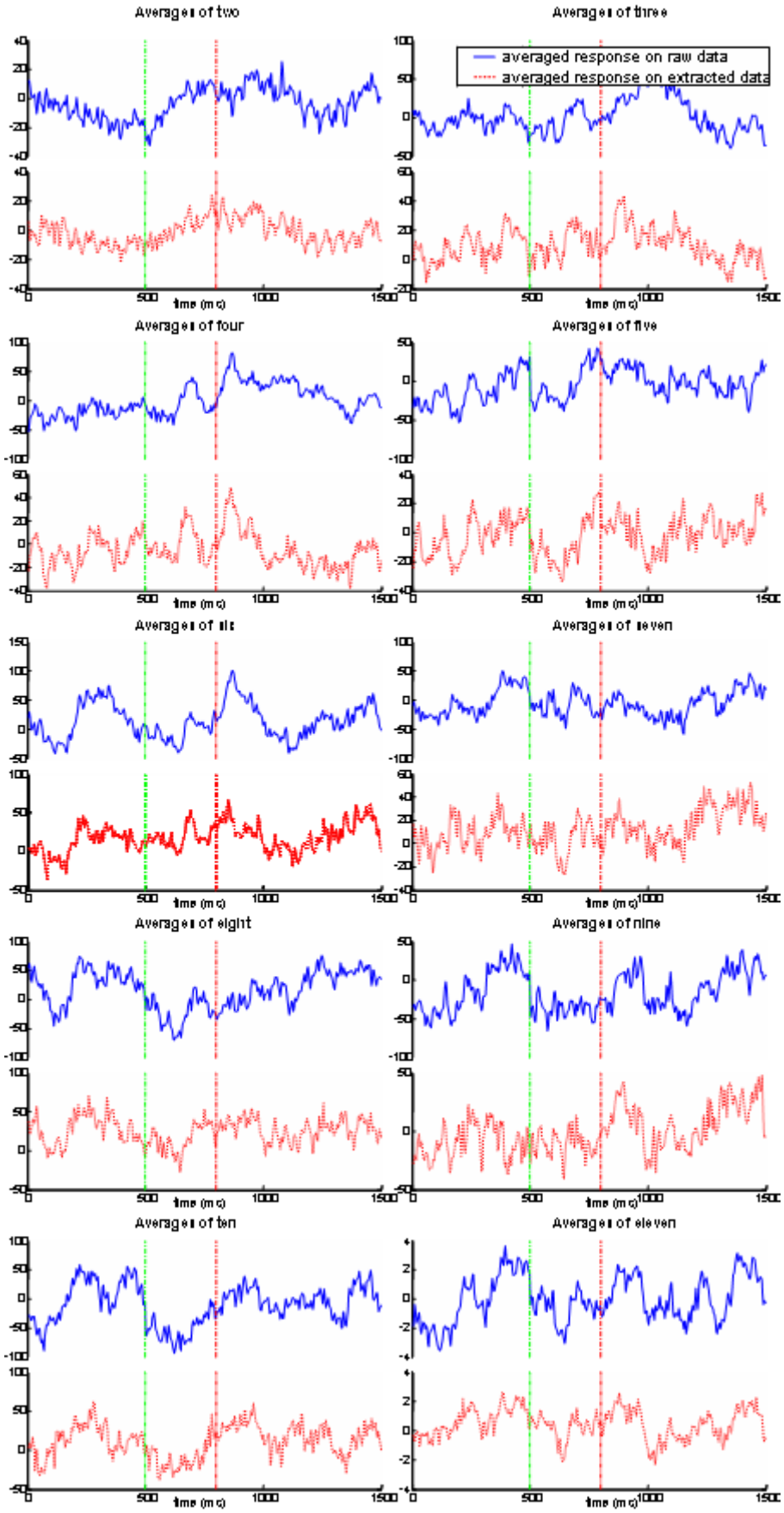

Figure C.8: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 6 after normal ICA. For each plot, the upper signal shows the averaged activity on raw data, and the bottom one plots the same average but on the extracted data.*

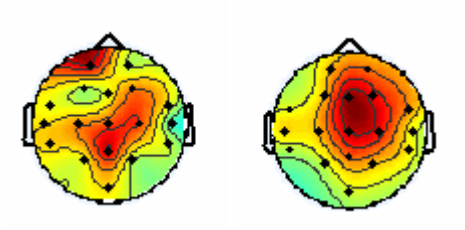

Figure C.9: *The two selected topographic maps from the data of Subject 6 after the normal ICA. Each map is corresponding to an independent component. Therefore there are two selected components projected to the original measurement space.*

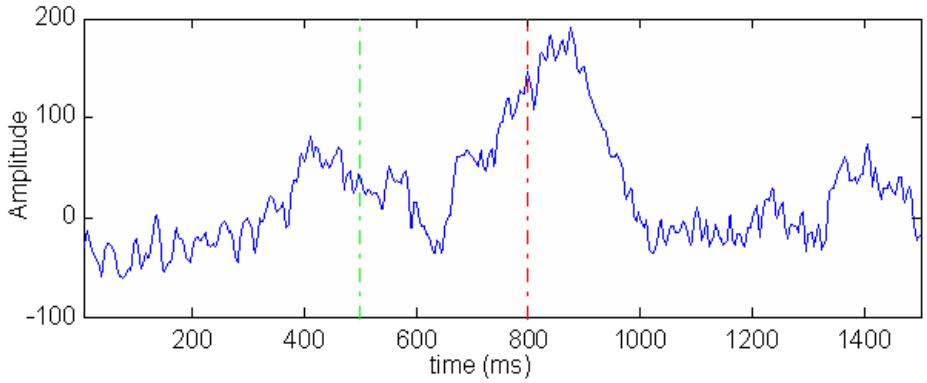

Figure C.10: *The template used to compute the correlation coefficient with the averaged P300 activities on the data from Subject 6. The template was an average of selected P300 epochs.*

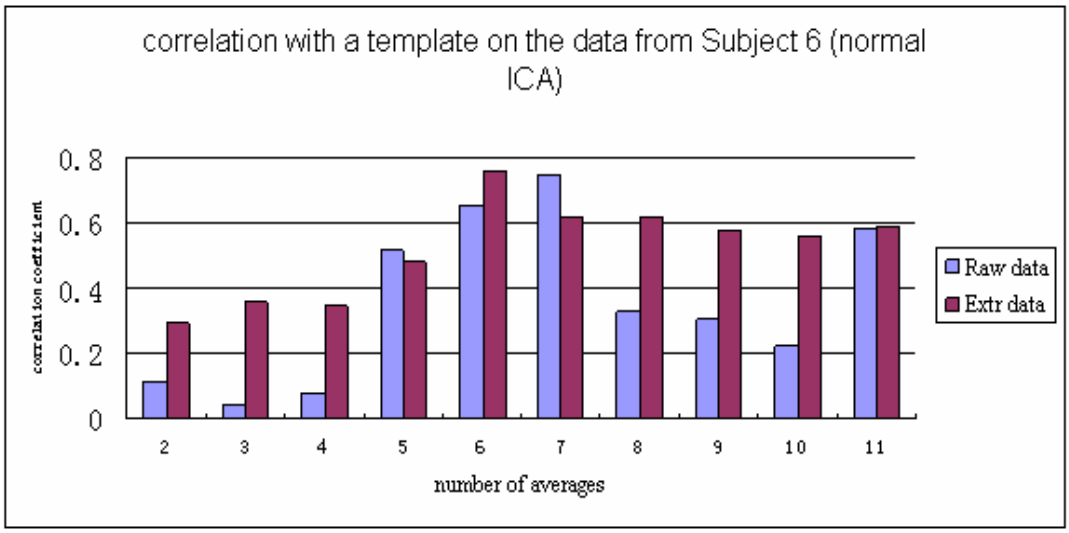

Figure C.11: *The performance of the correlation between the P300 template and the averaged P300 activities before and after normal ICA on the data from Subject 6. There were a total 10 pairs of random averages of P300 activities.*

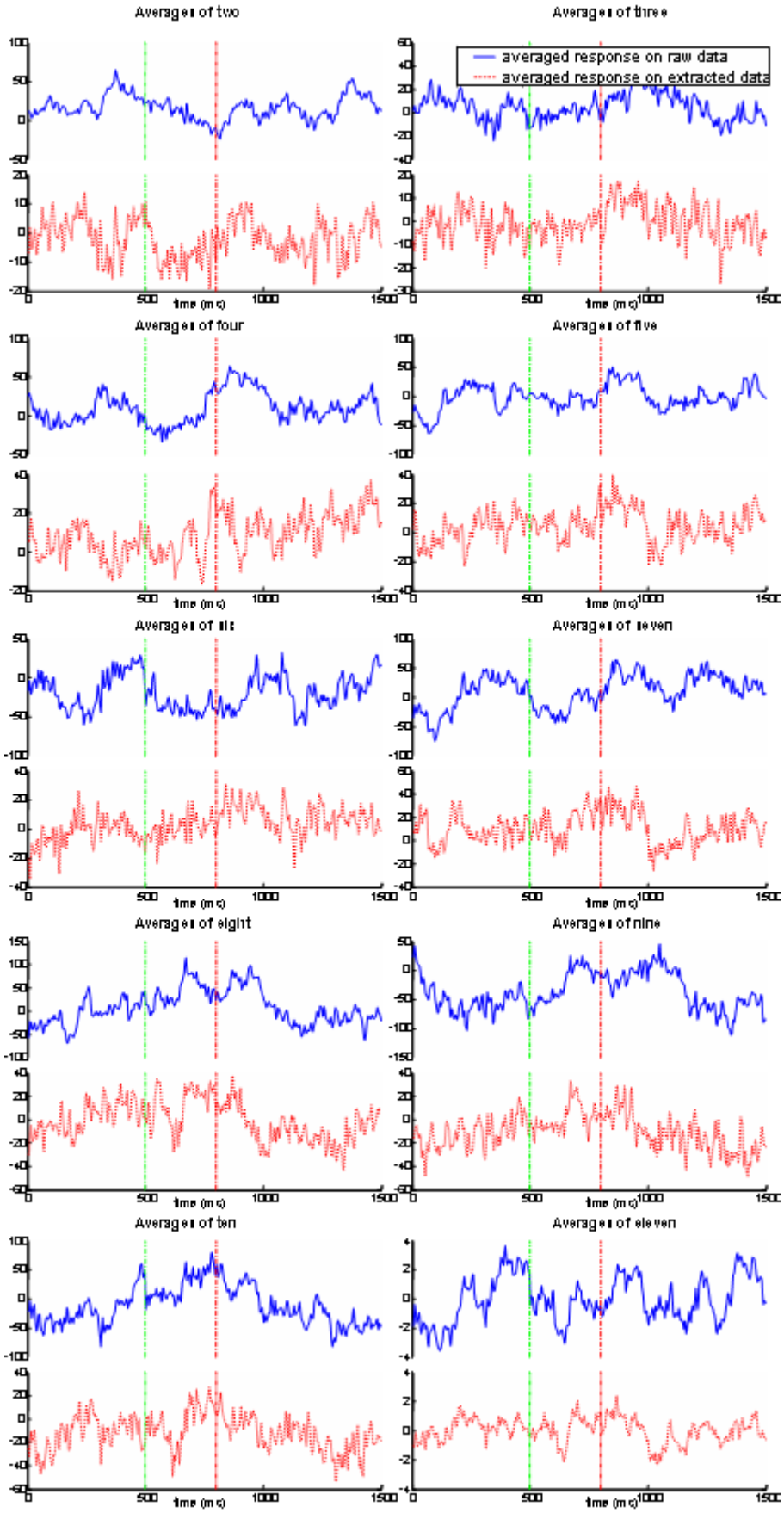

Figure C.12: *The averaged P300 activities vs the number of random averaging on raw data and extracted data for Subject 6 after spatially constrained ICA. For each plot, the upper signal shows the averaged activity on raw data, and the bottom one plots the same average but on the extracted data.*

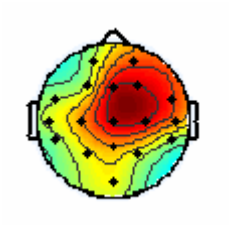

Figure C.13: *The constraint topographic map for Subject 6. This constraint was a selected column of the mixing matrix after a normal ICA on a special designed data with all the selected P300 epochs*

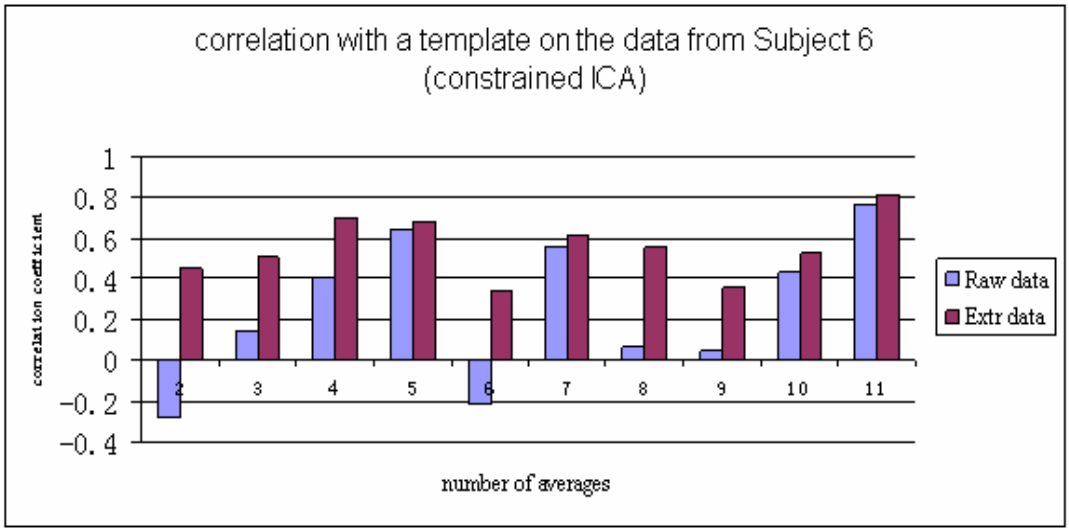

Figure C.14: *The performance of the correlation between the P300 template and the averaged P300 activities before and after the spatially constrained ICA on the data from Subject 6.*

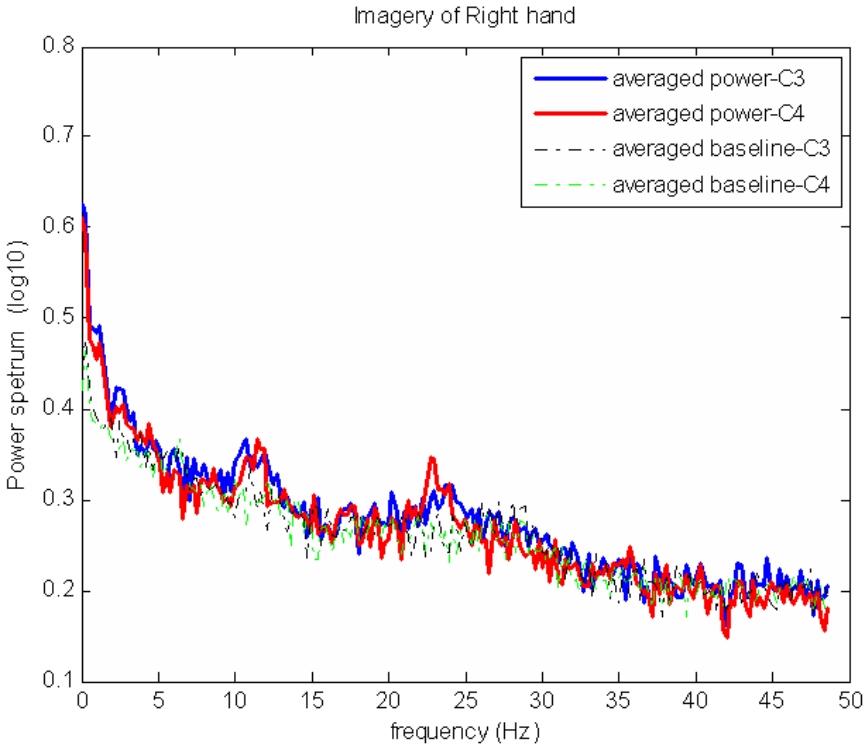

Figure C.15: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 1.*

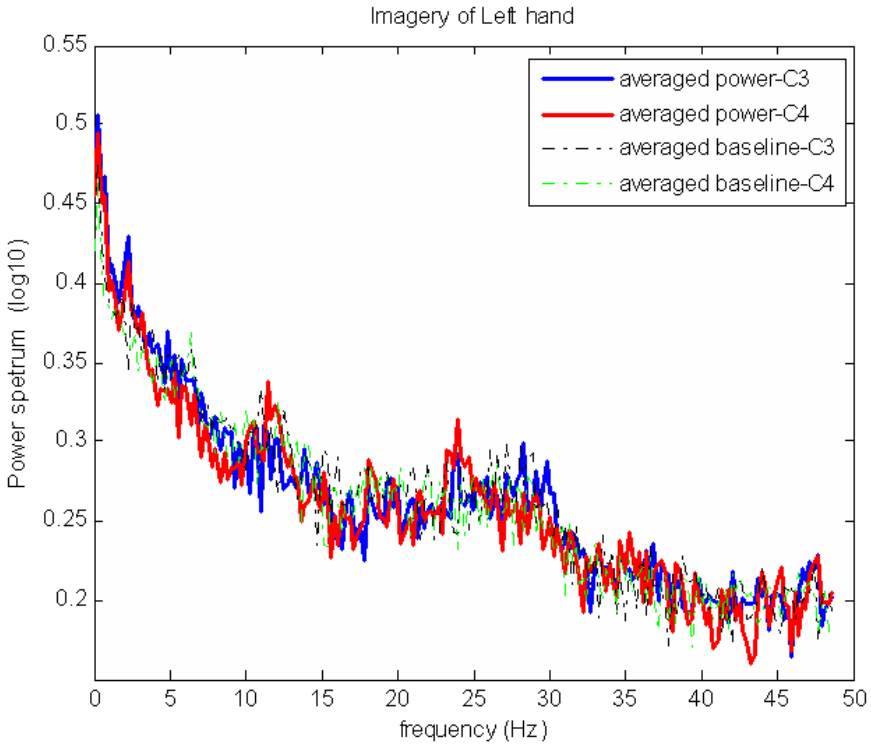

Figure C.16: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 1.*

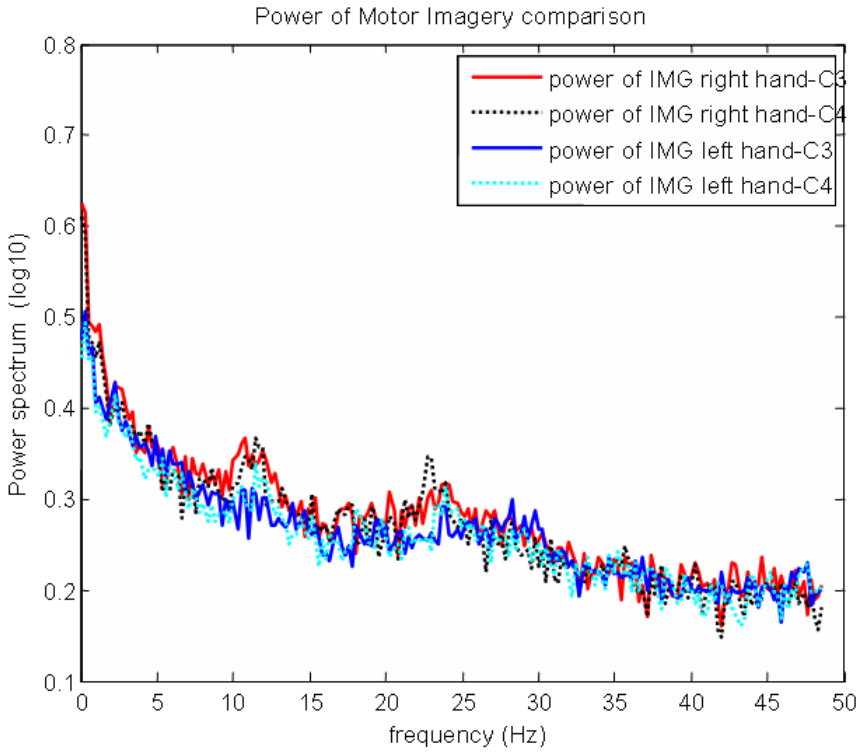

Figure C.17: *The averaged power spectra for right/left hand imagination in one graph on the data from Subject 1.*

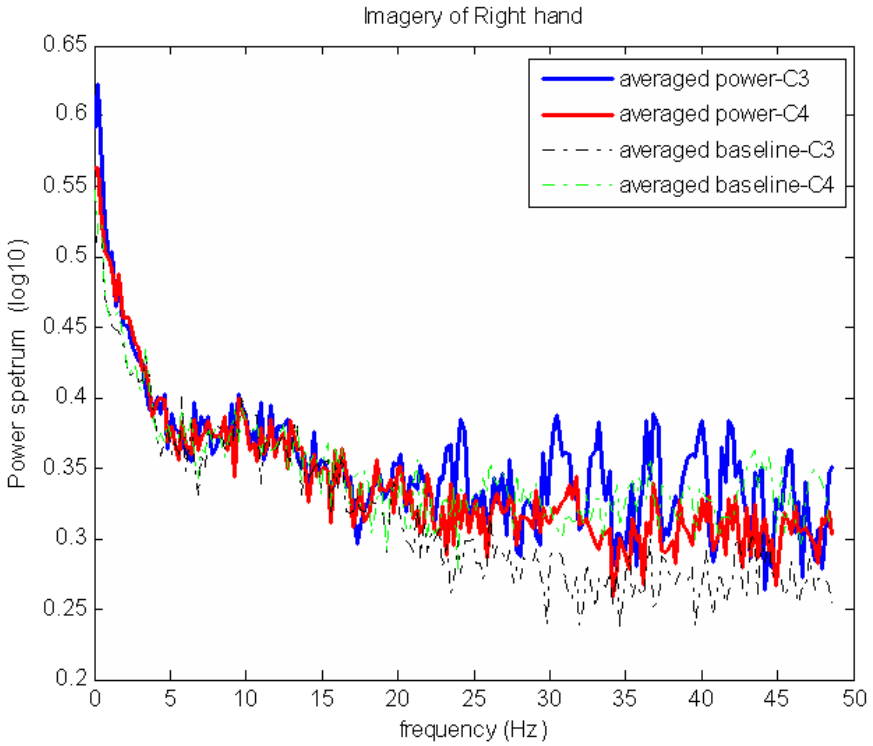

Figure C.18: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 2.*

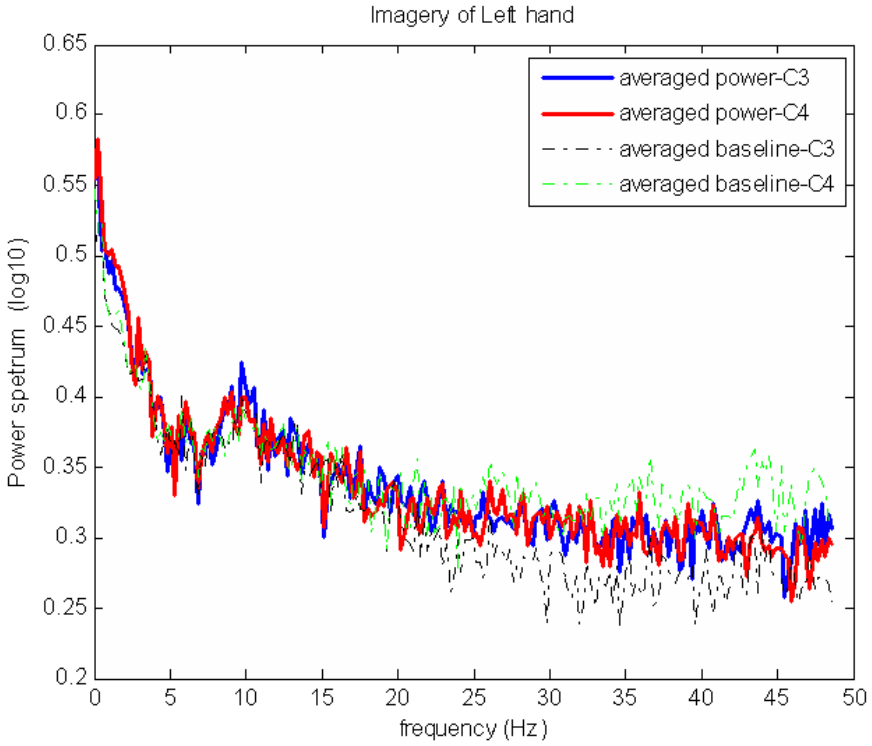

Figure C.19: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 2.*

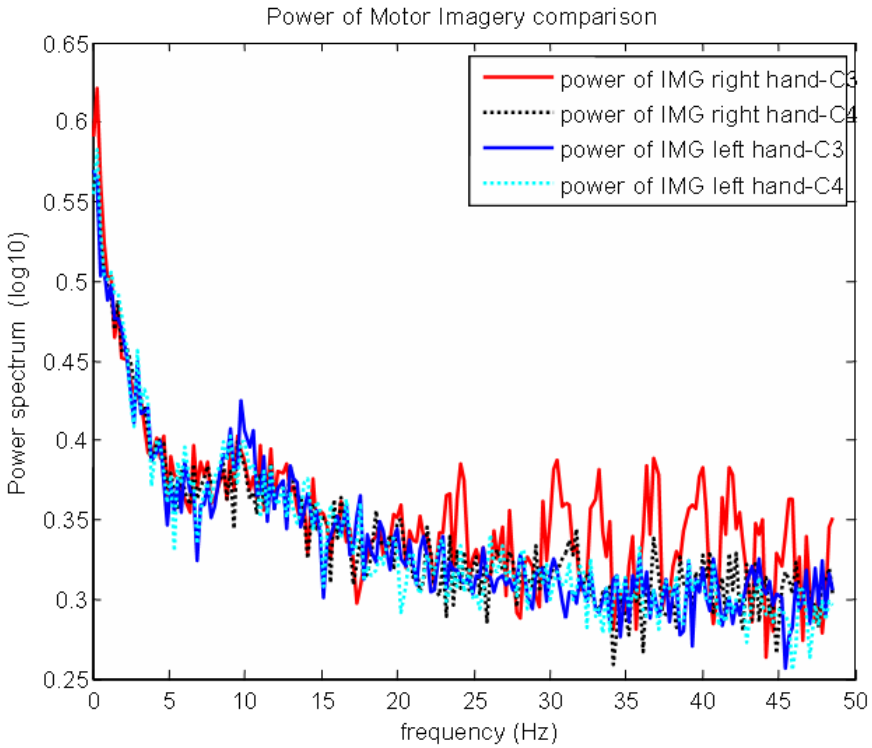

Figure C.20: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 2.*

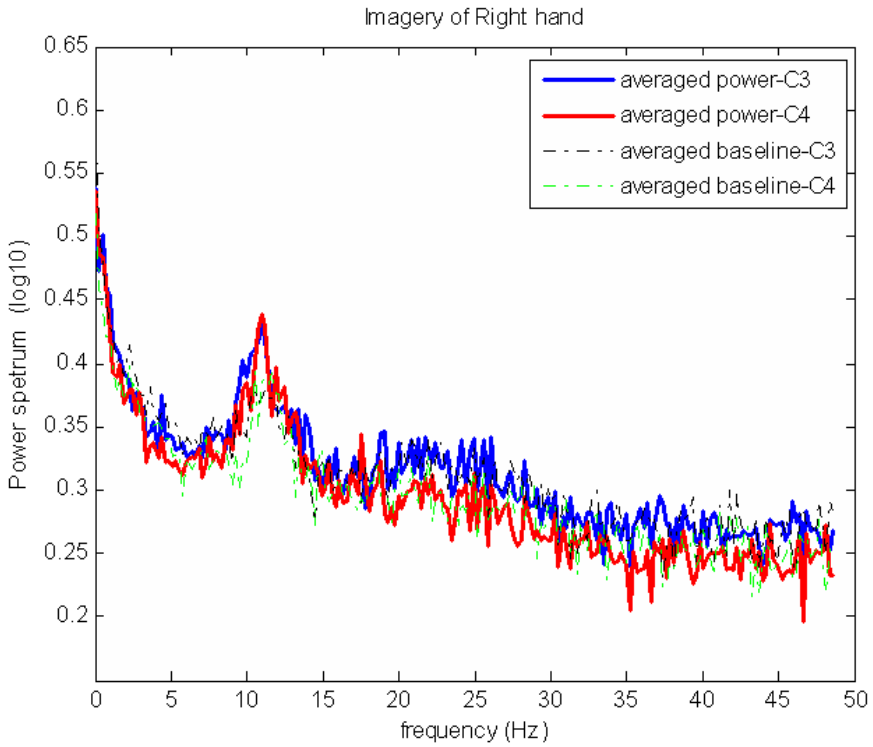

Figure C.21: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subiect 3.*

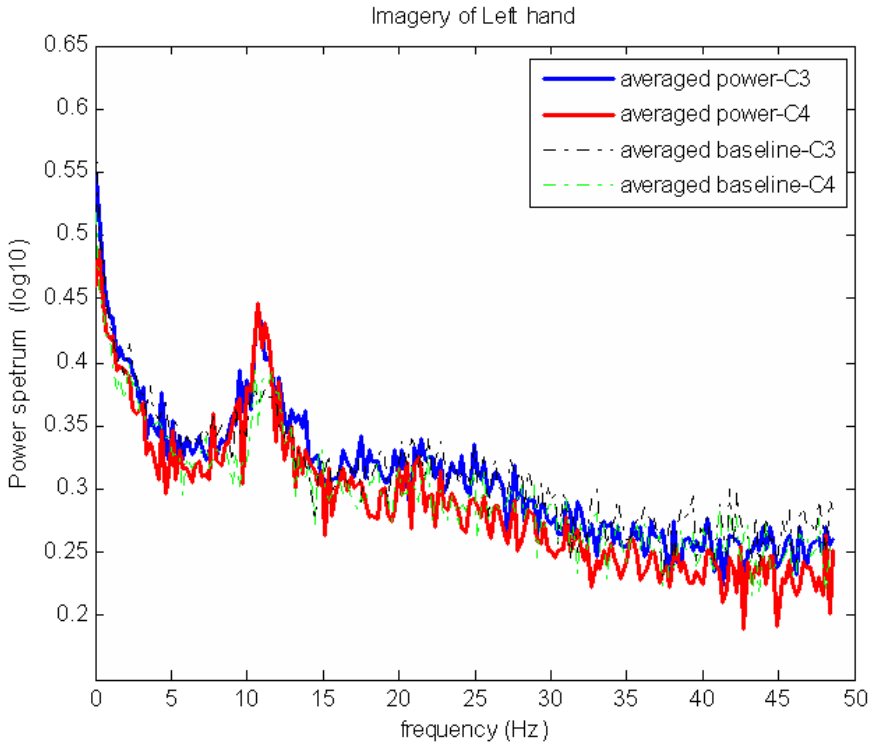

Figure C.22: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 3.*

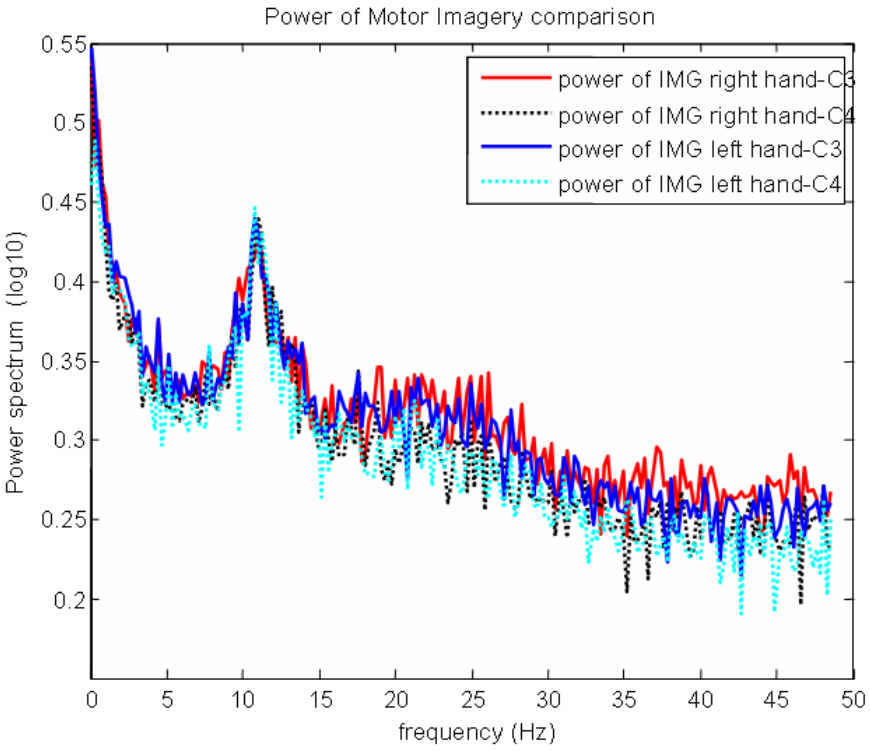

Figure C.23: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 3.*

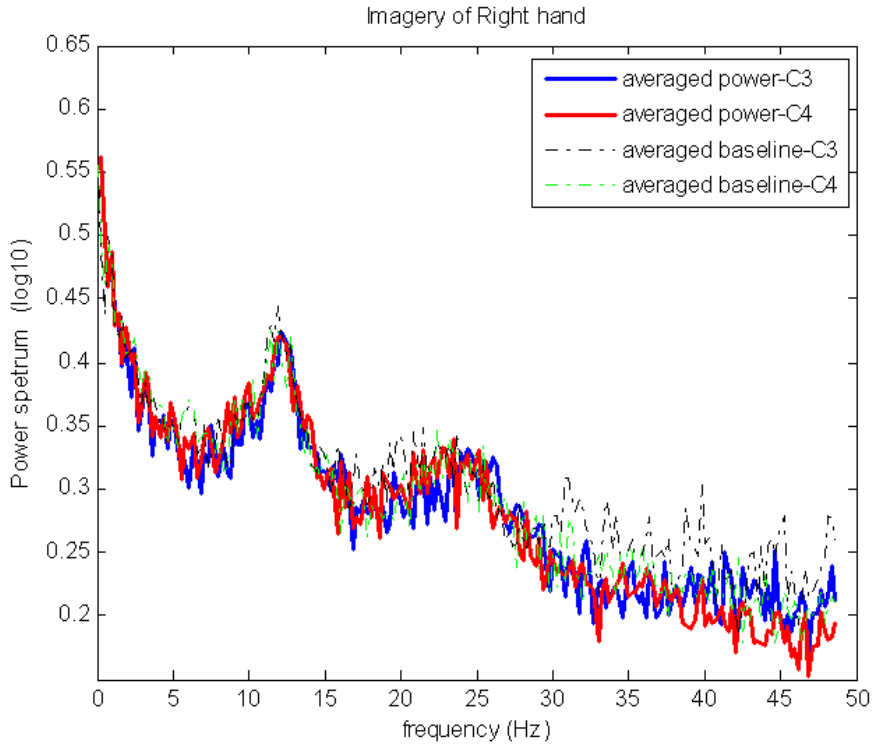

Figure C.24: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 4.*

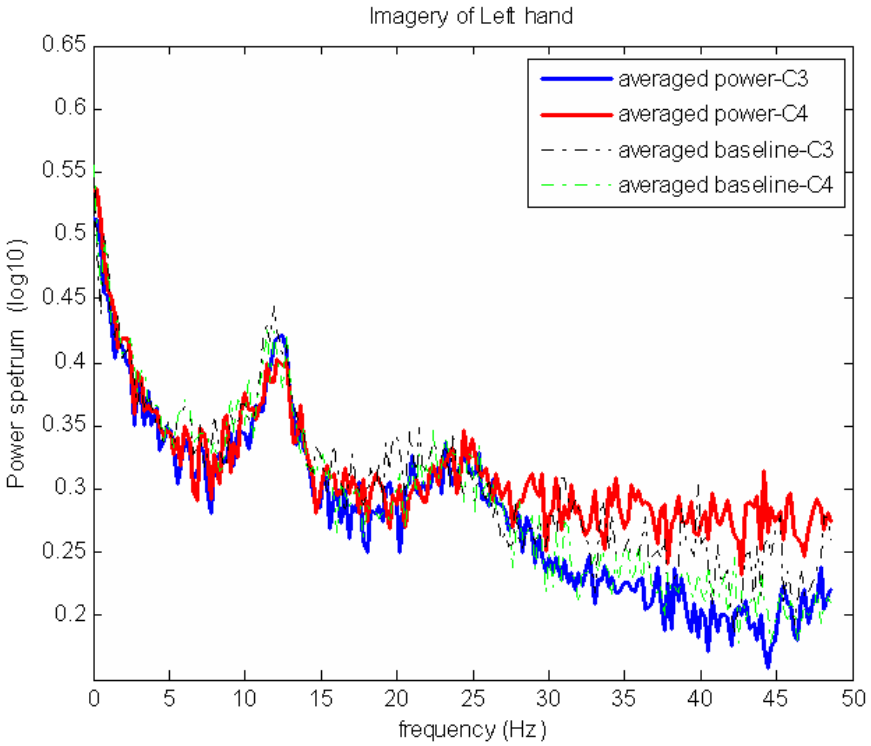

Figure C.25: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 4.*

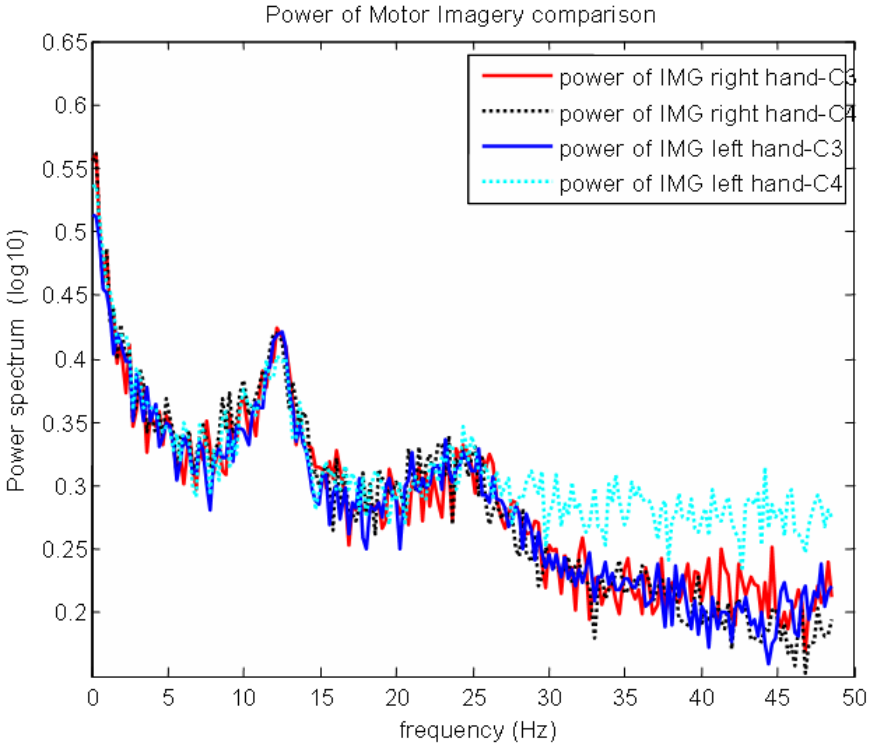

Figure C.26: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 4.*

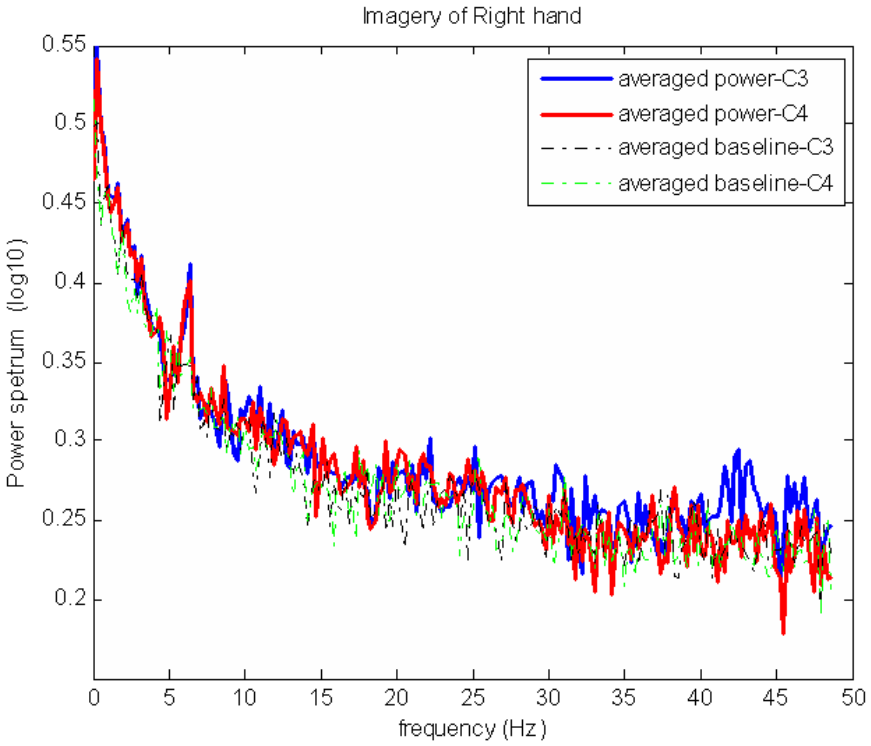

Figure C.27: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 5.*

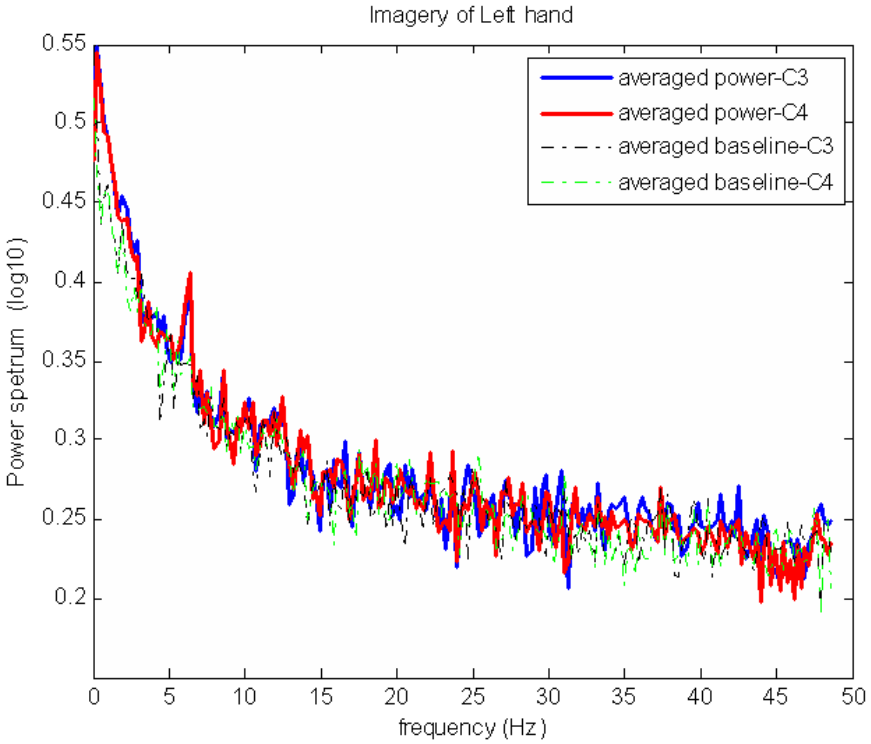

Figure C.28: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 5.*

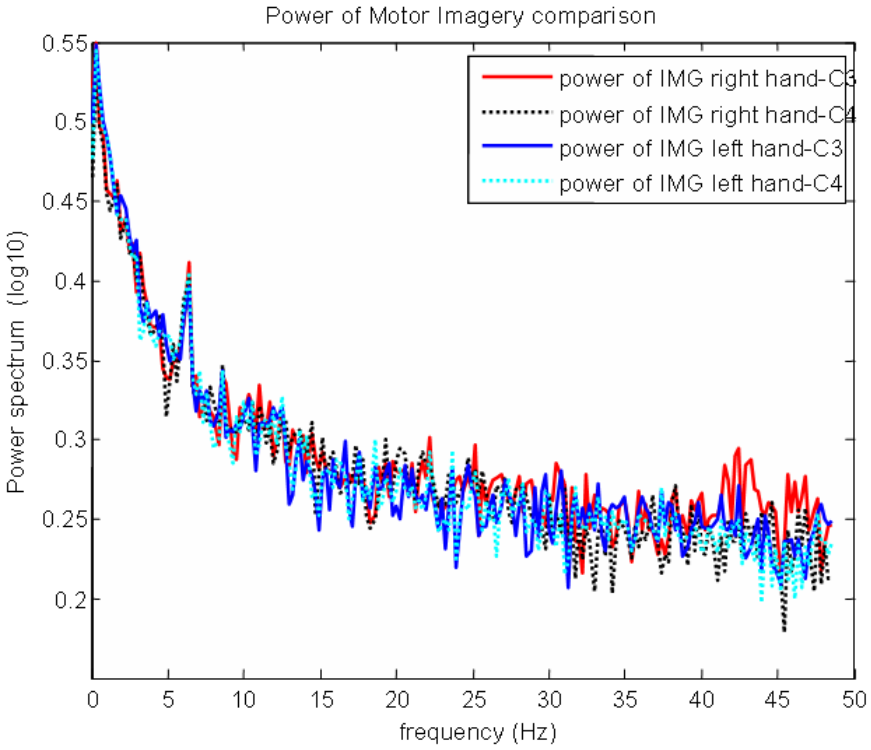

Figure C.29: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 5.*

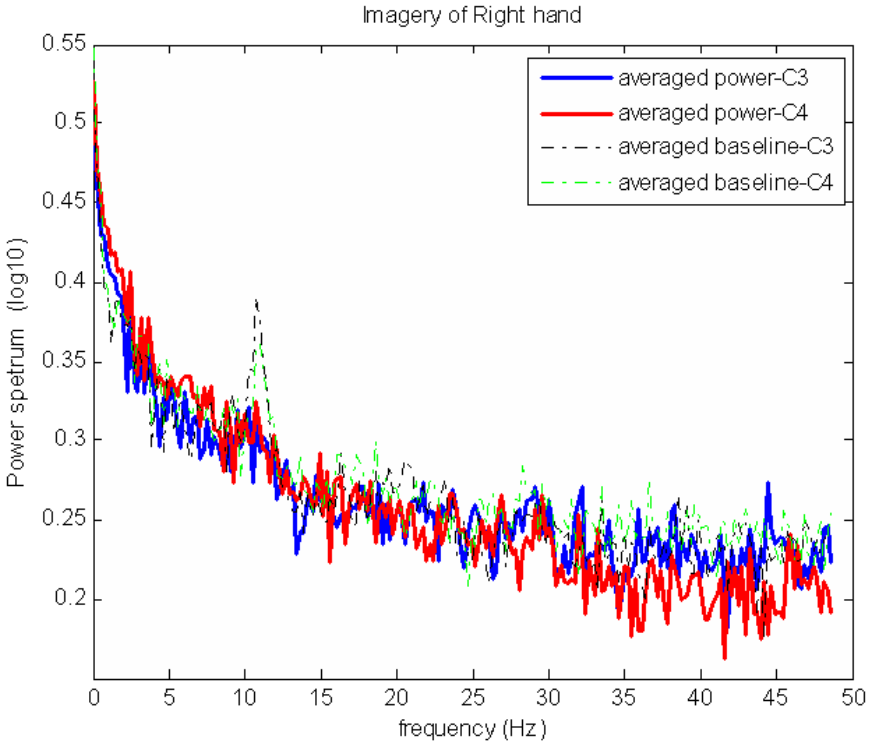

Figure C.30: *The averaged power spectra over ten trials for right hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 6.*

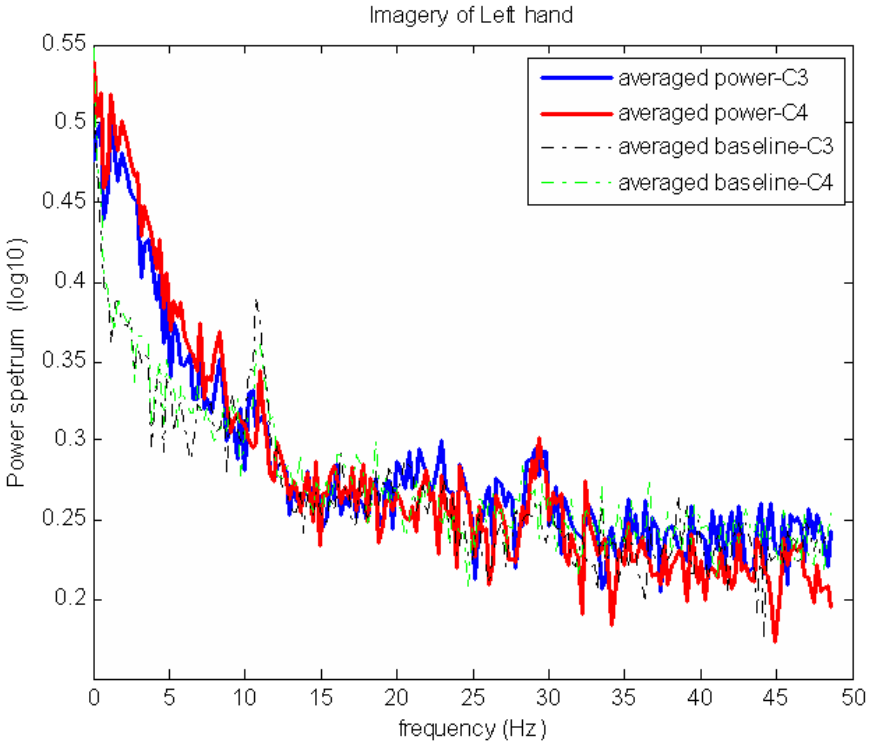

Figure C.31: *The averaged power spectra over ten trials for left hand grasping task together with the averaged power spectra of baseline signal on C3 and C4 in frequency domain on the data from Subject 6.*

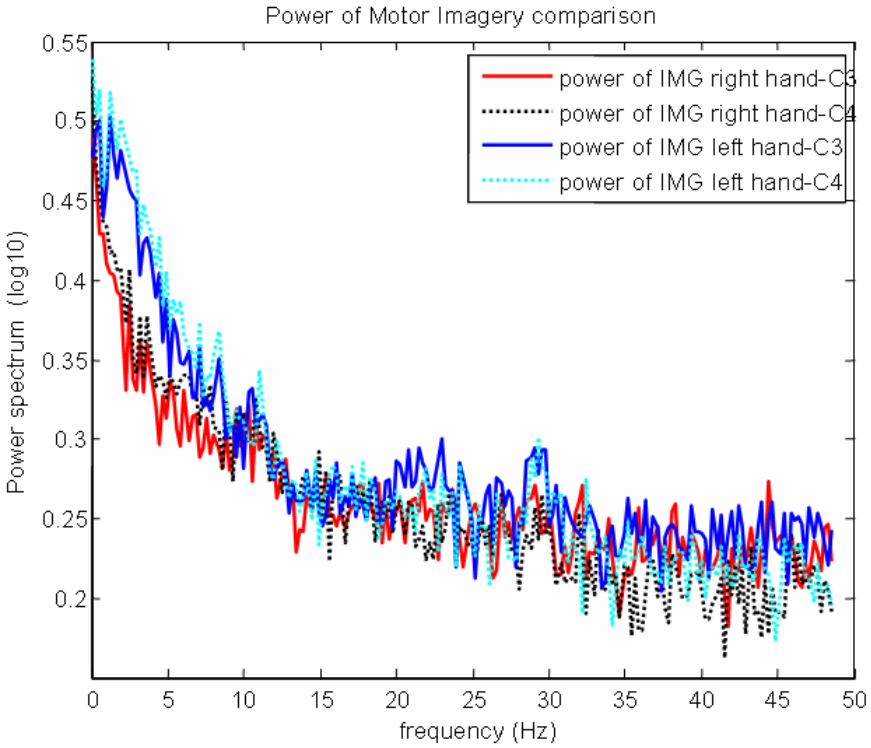

Figure C.32: *the averaged power spectra for right/left hand imagination in one graph on the data from Subject 6.*

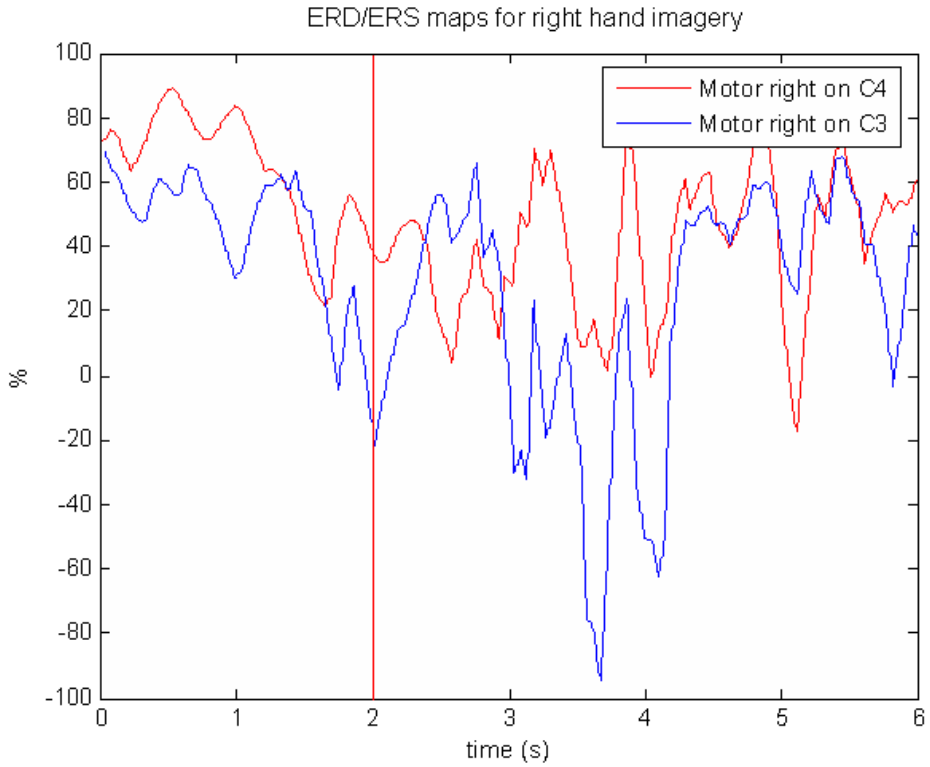

Figure C.33: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 1. The dashed vertical line represented the stimulus onset.*

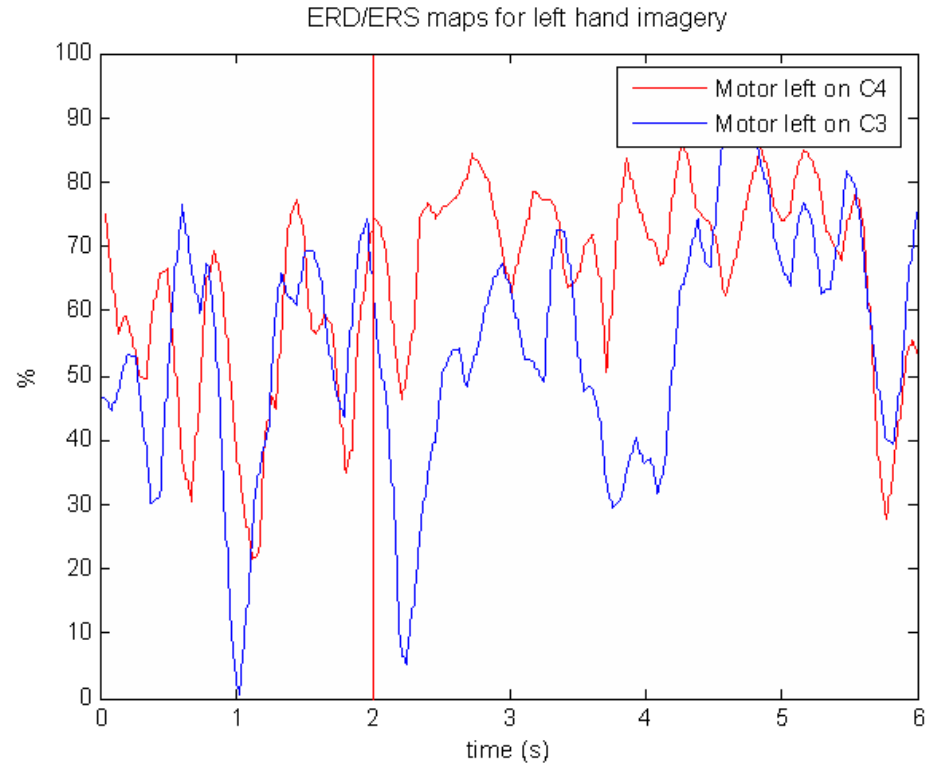

Figure C.34: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 1. The dashed vertical line represented the stimulus onset.*

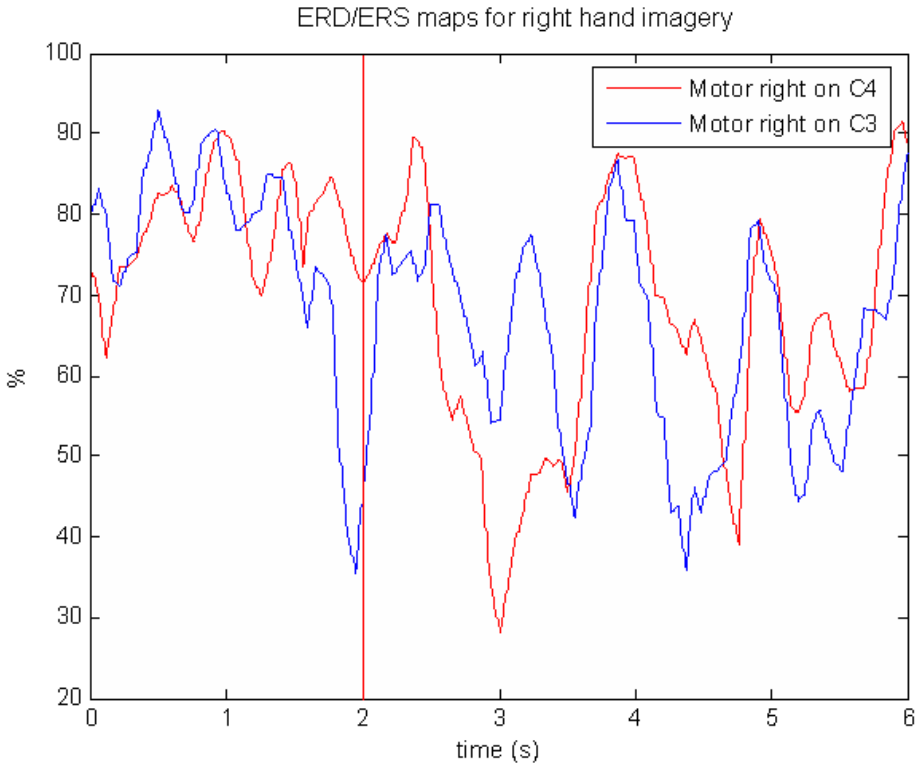

Figure C.35: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 2. The dashed vertical line represented the stimulus onset.*

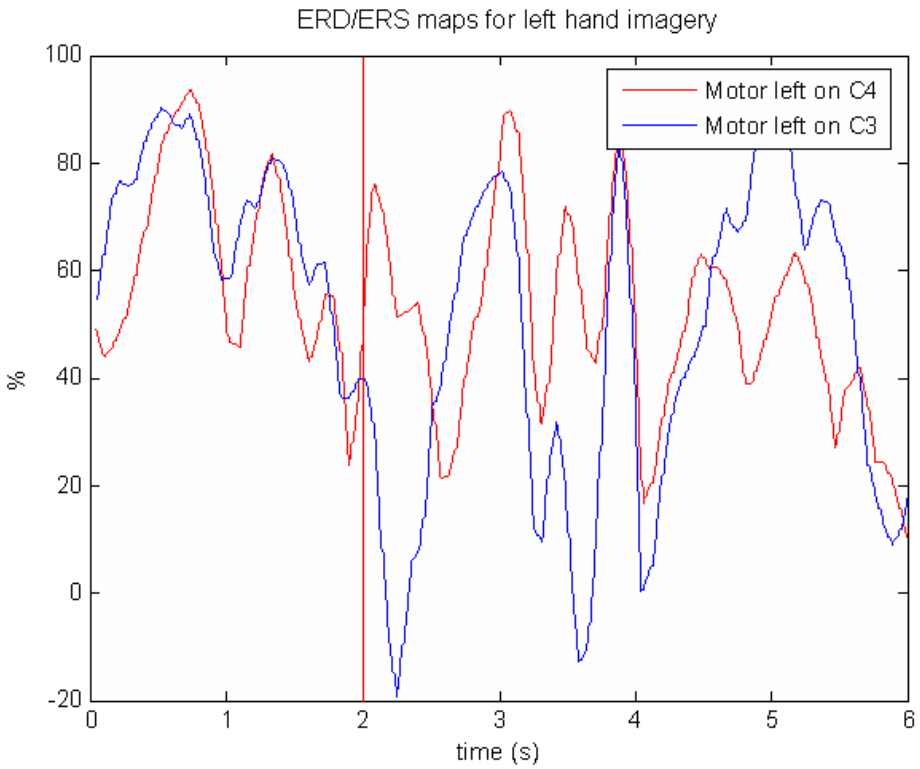

Figure C.36: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 2. The dashed vertical line represented the stimulus onset.*

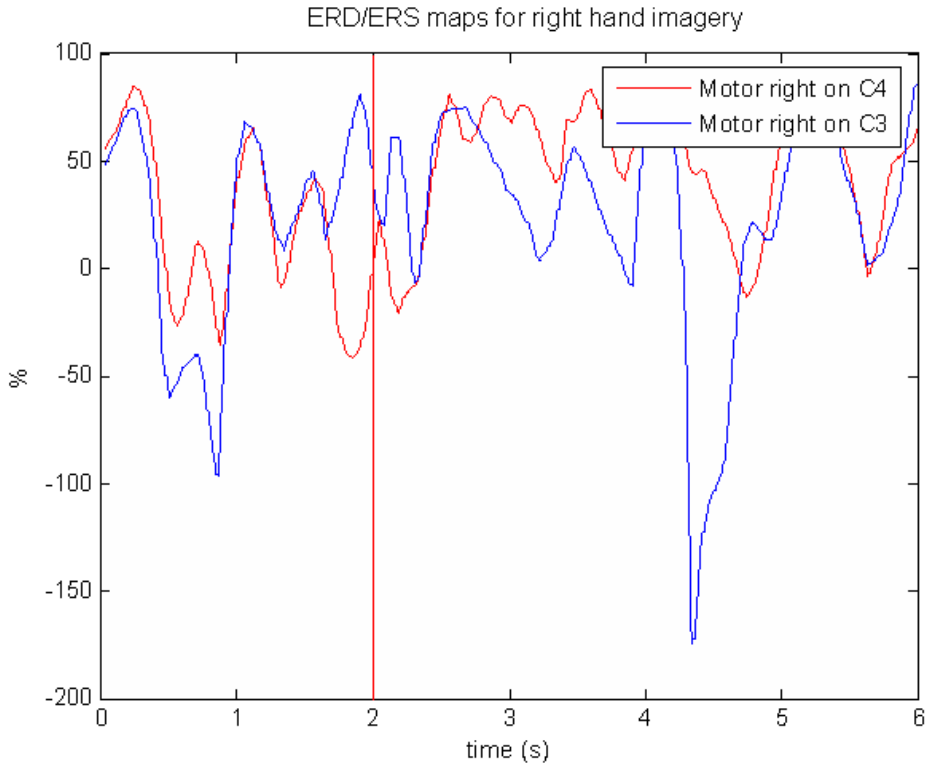

Figure C.37: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 3. The dashed vertical line represented the stimulus onset.*

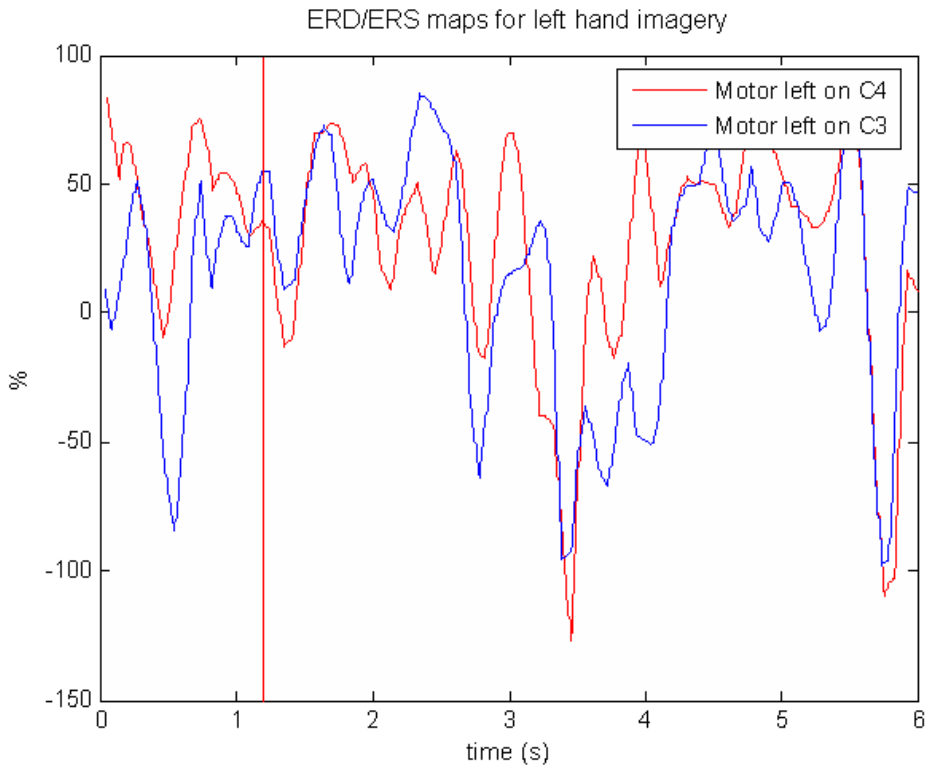

Figure C.38: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 3. The dashed vertical line represented the stimulus onset.*

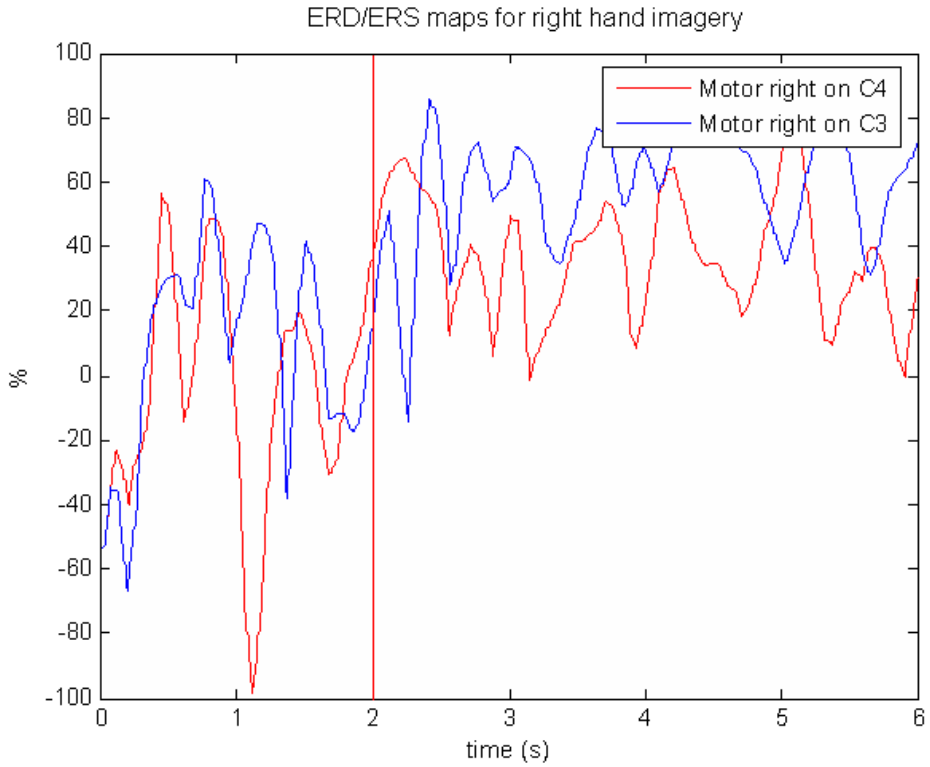

Figure C.39: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 4. The dashed vertical line represented the stimulus onset.*

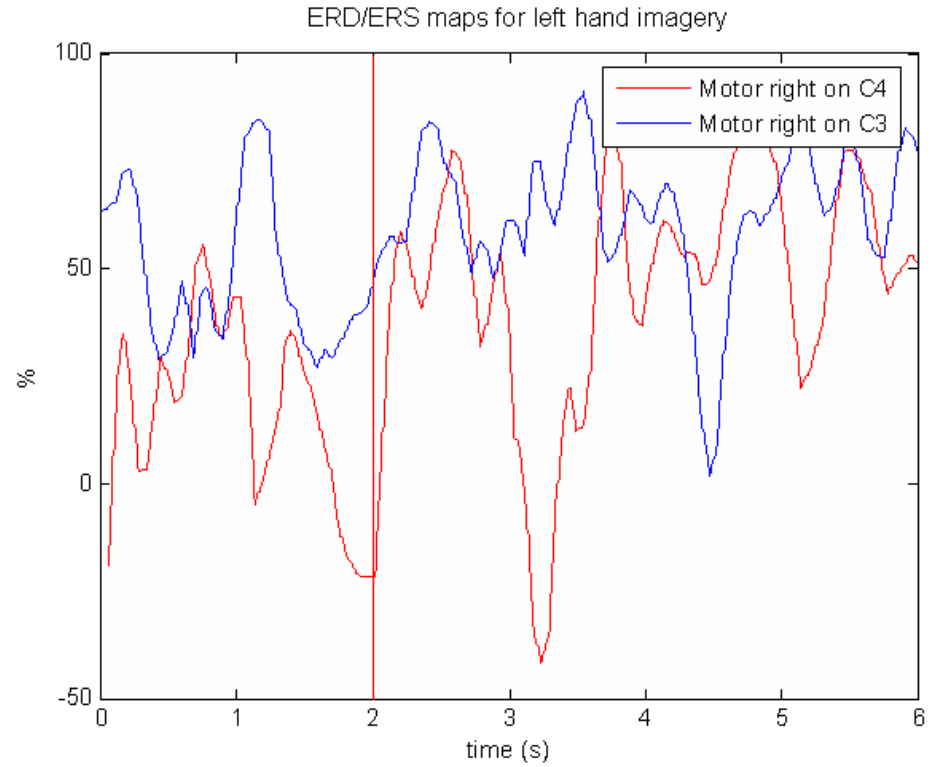

Figure C.40: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 4. The dashed vertical line represented the stimulus onset.*

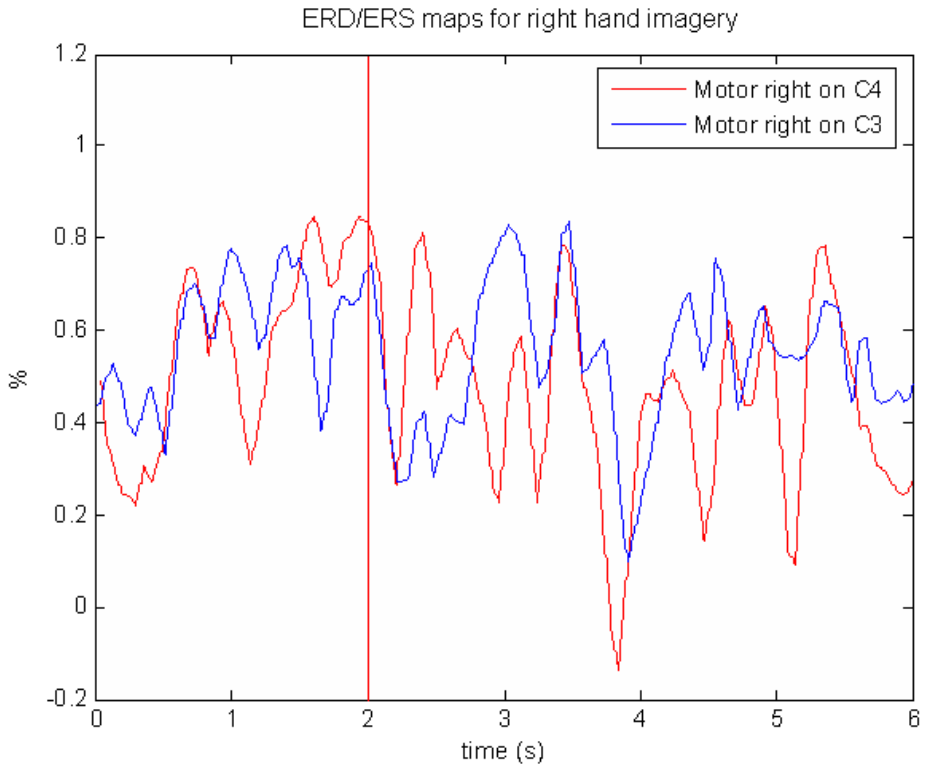

Figure C.41: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 5. The dashed vertical line represented the stimulus onset.*

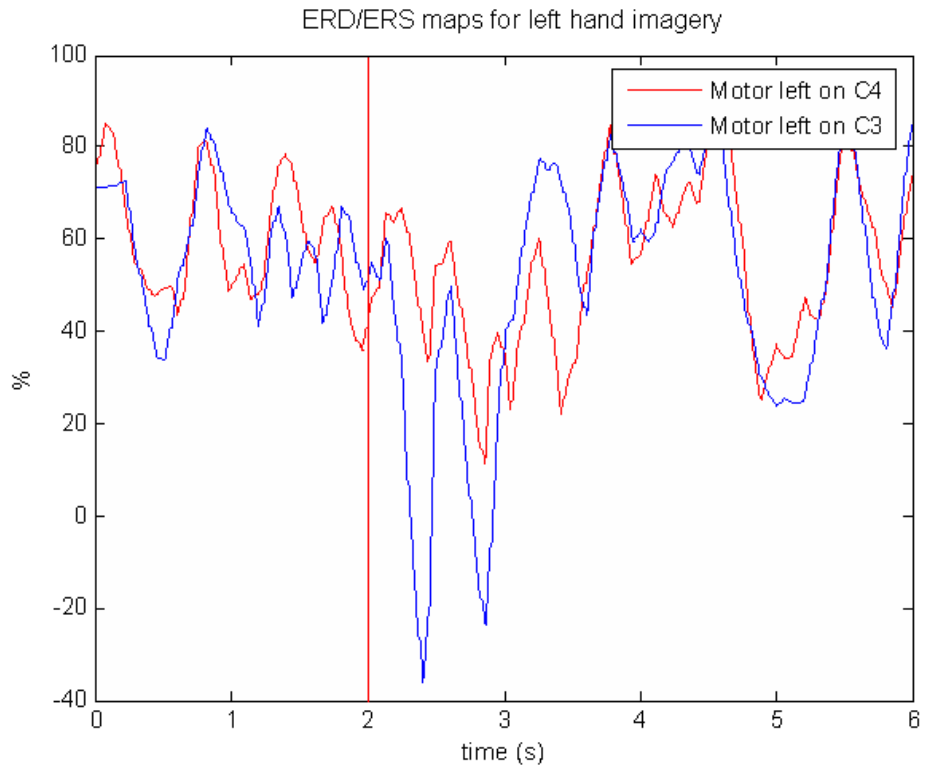

Figure C.42: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 5. The dashed vertical line represented the stimulus onset.*

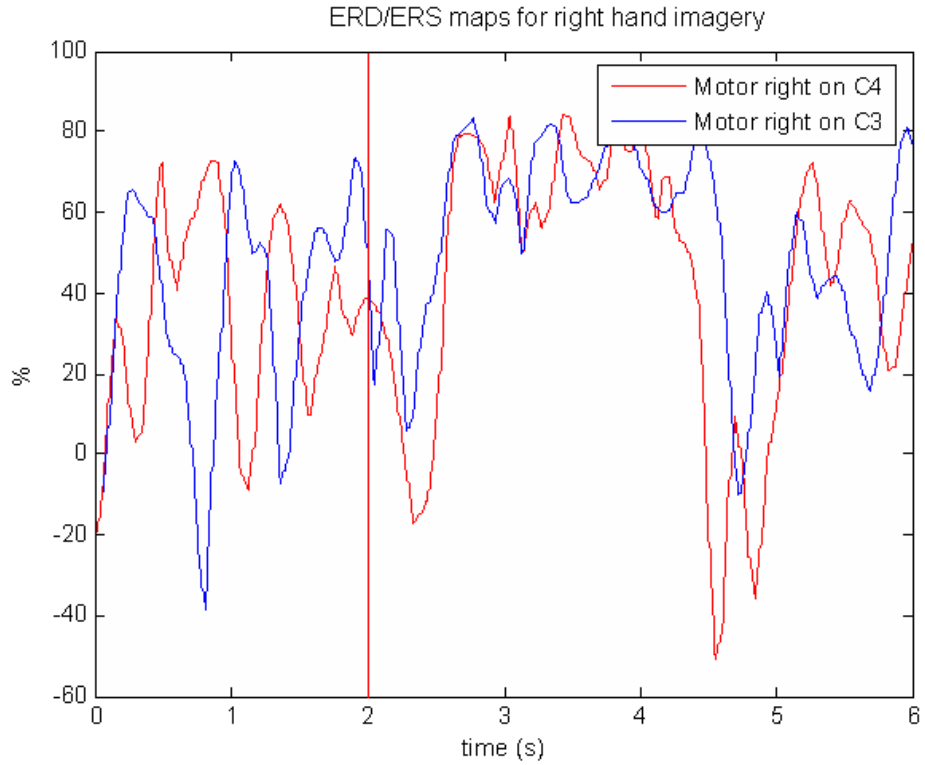

Figure C.43: *The ERD/ERS maps for right hand grasping imagery on the data from the same Subject 6. The dashed vertical line represented the stimulus onset.*

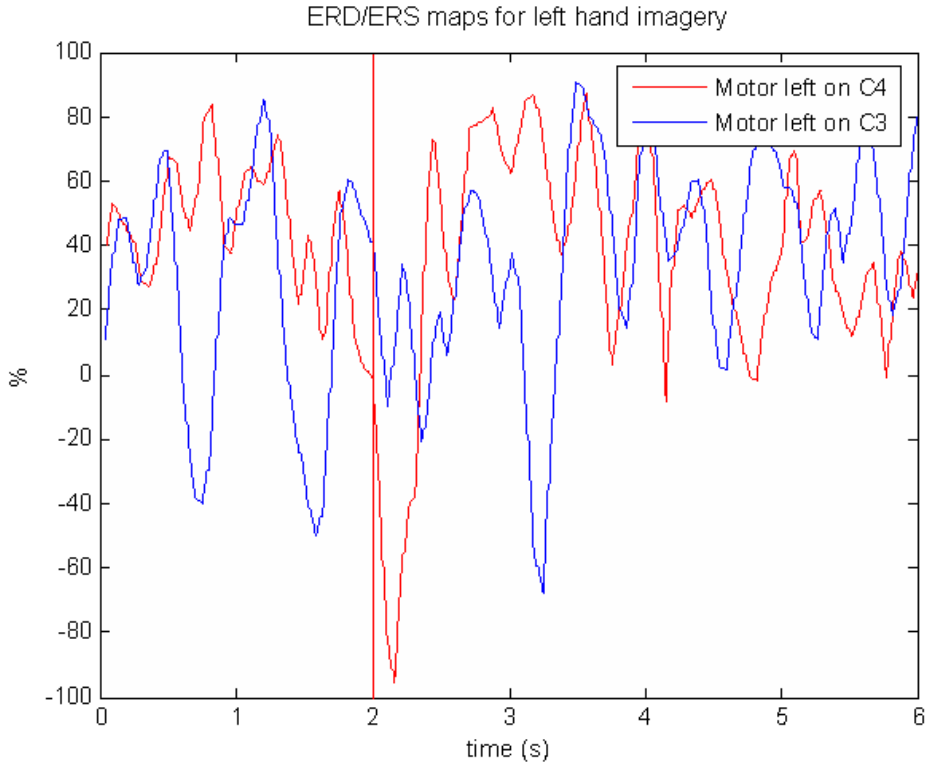

Figure C.44: *The ERD/ERS maps for left hand grasping imagery on the data from the same Subject 6. The dashed vertical line represented the stimulus onset.*

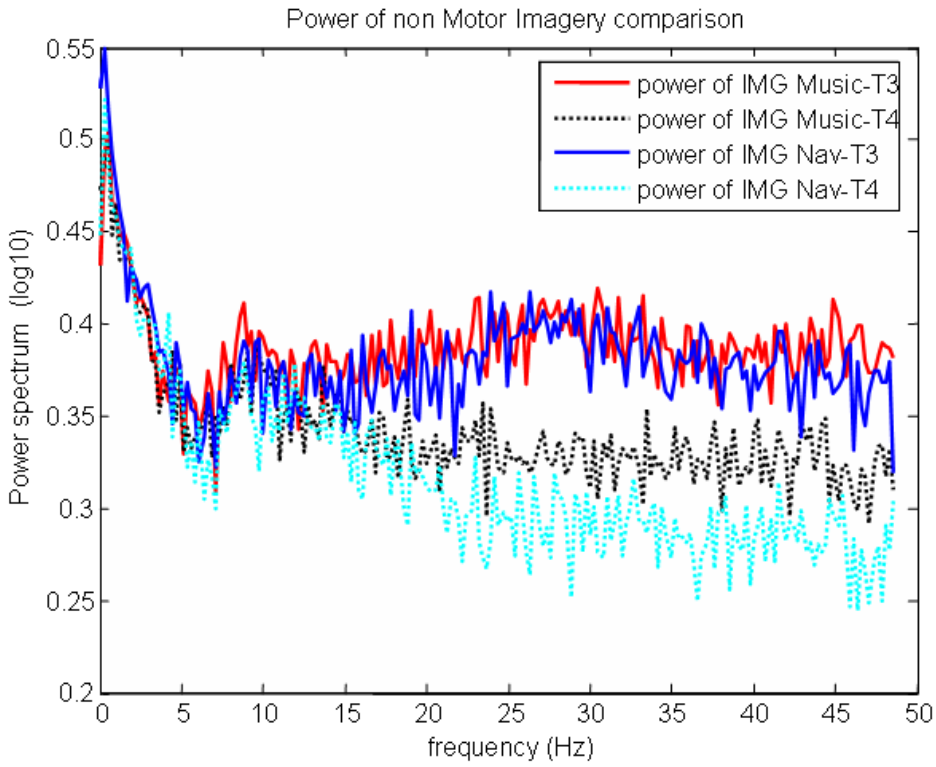

Figure C.45: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 2.*

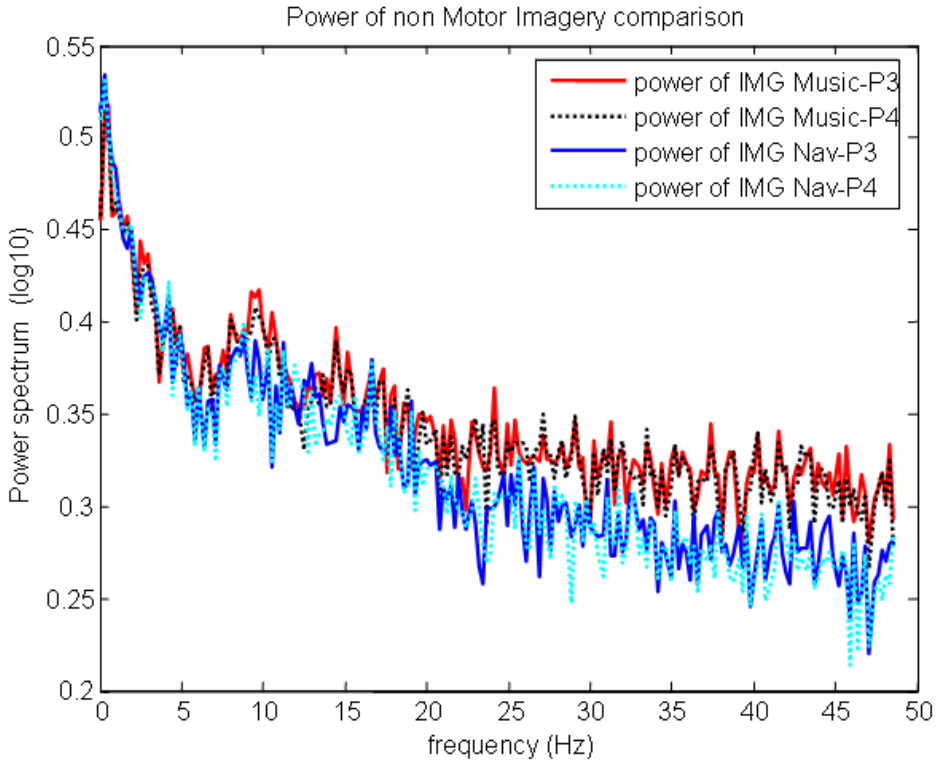

Figure C.46: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 2.*

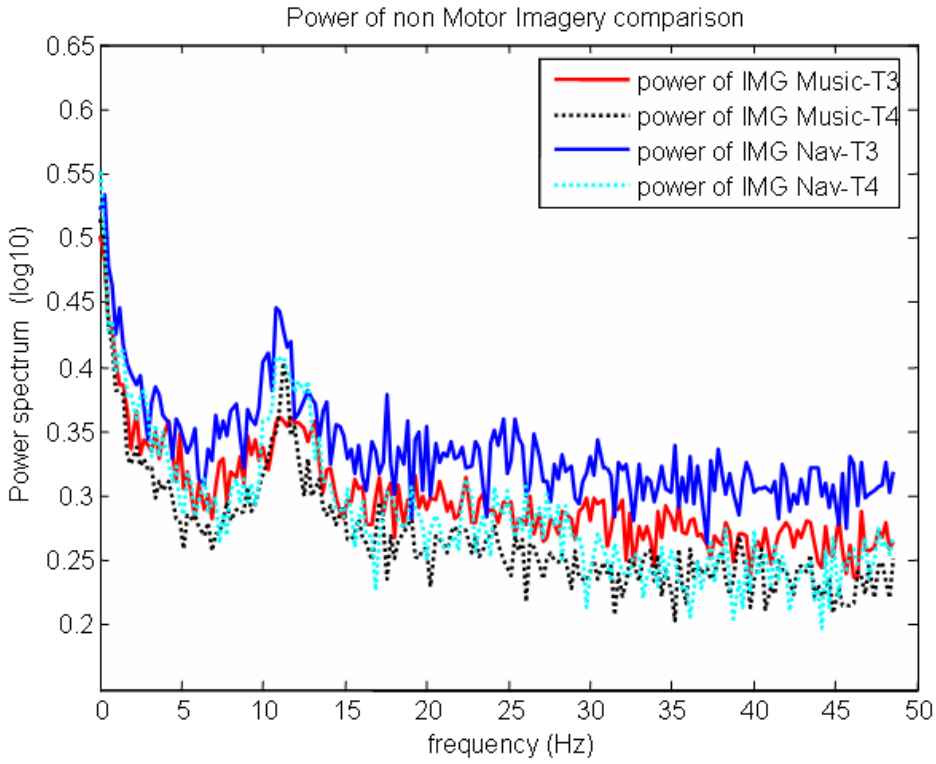

Figure C.47: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 3.*

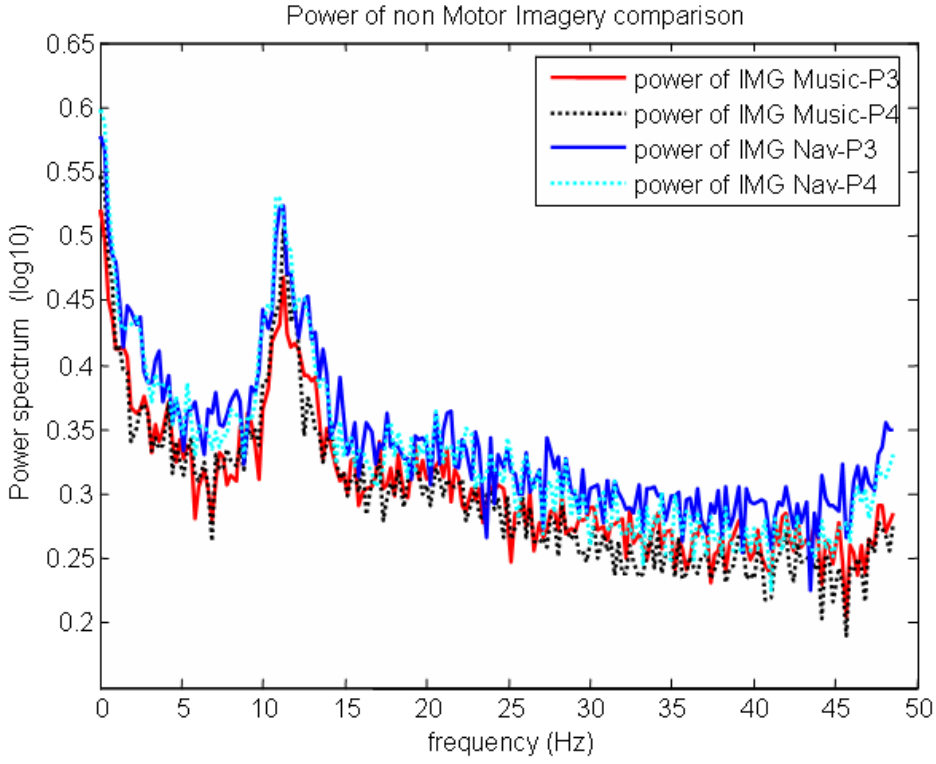

Figure C.48: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 3.*

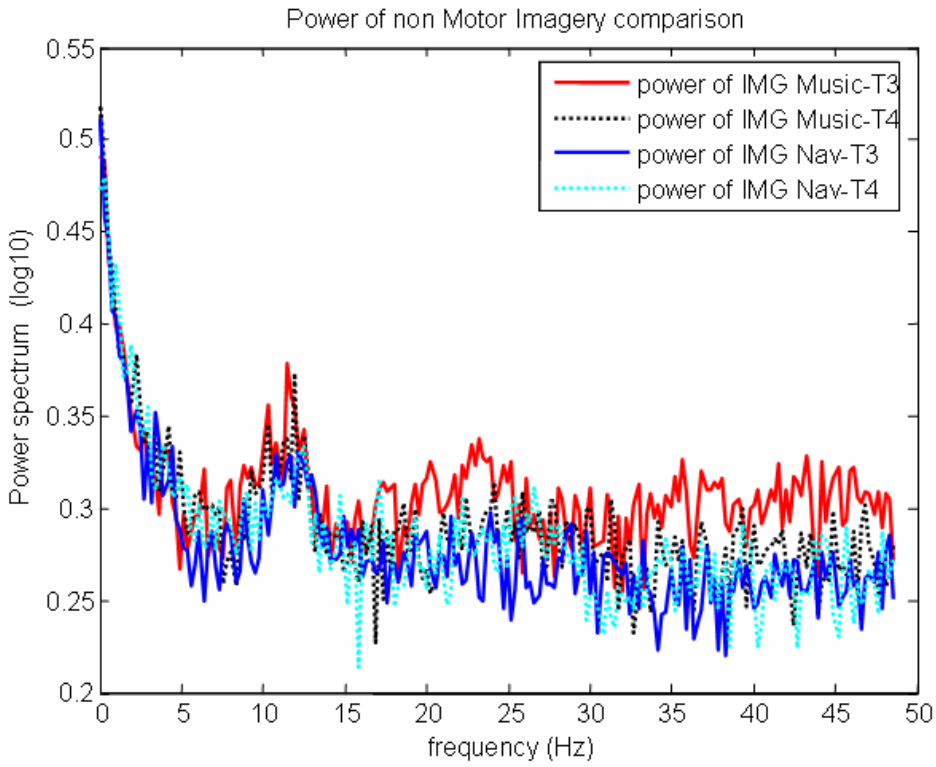

Figure C.49: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 4.*

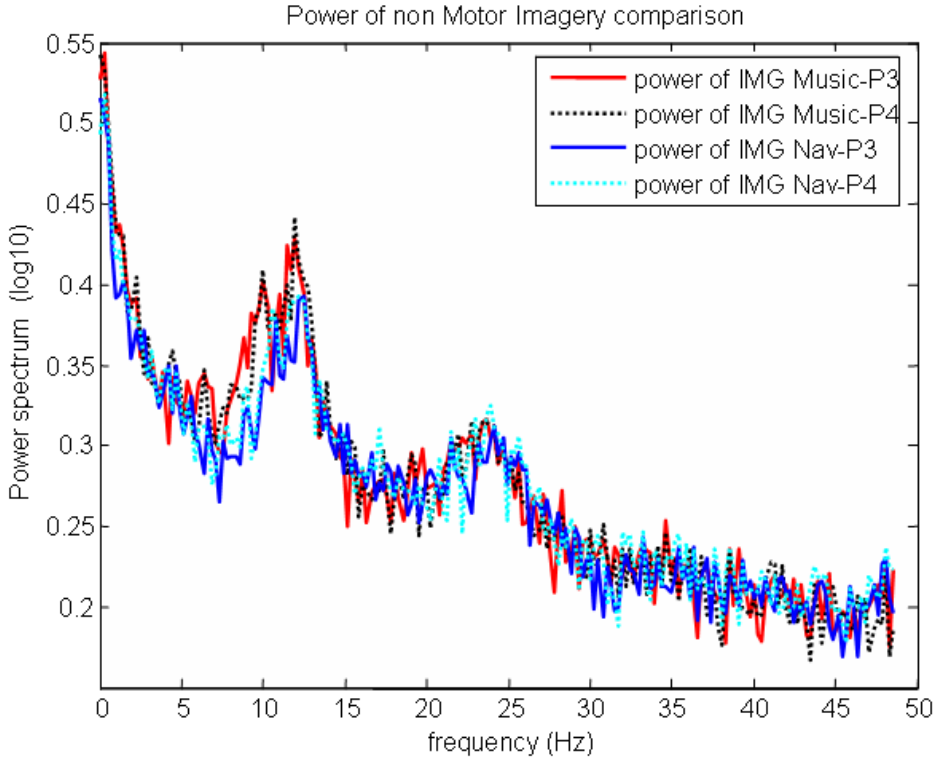

Figure C.50: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 4.*

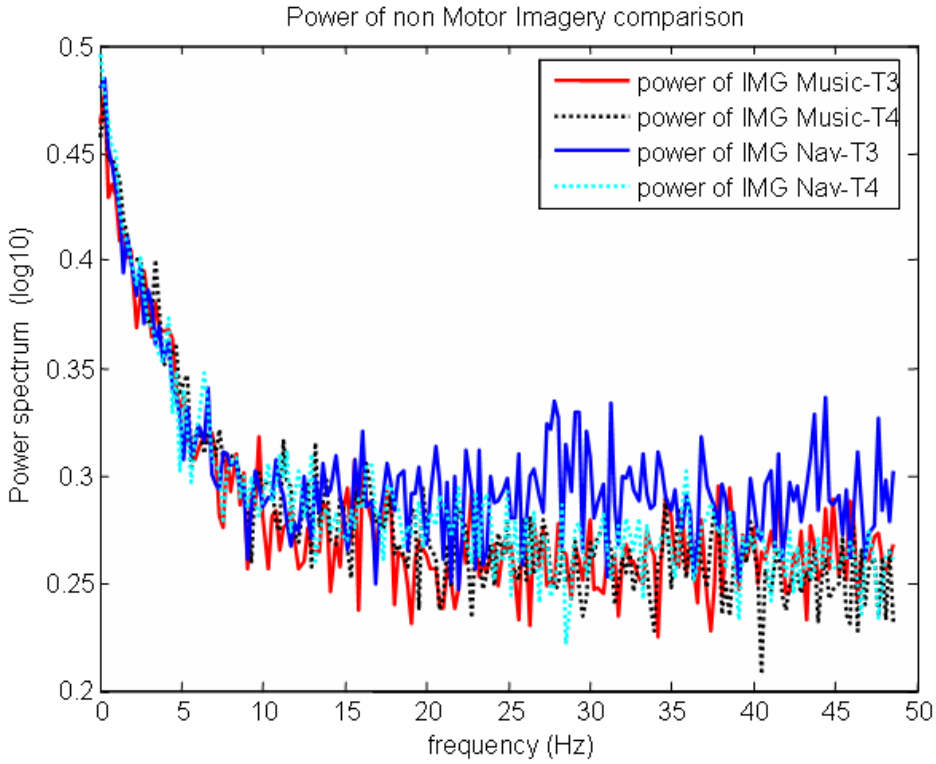

Figure C.51: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 5.*

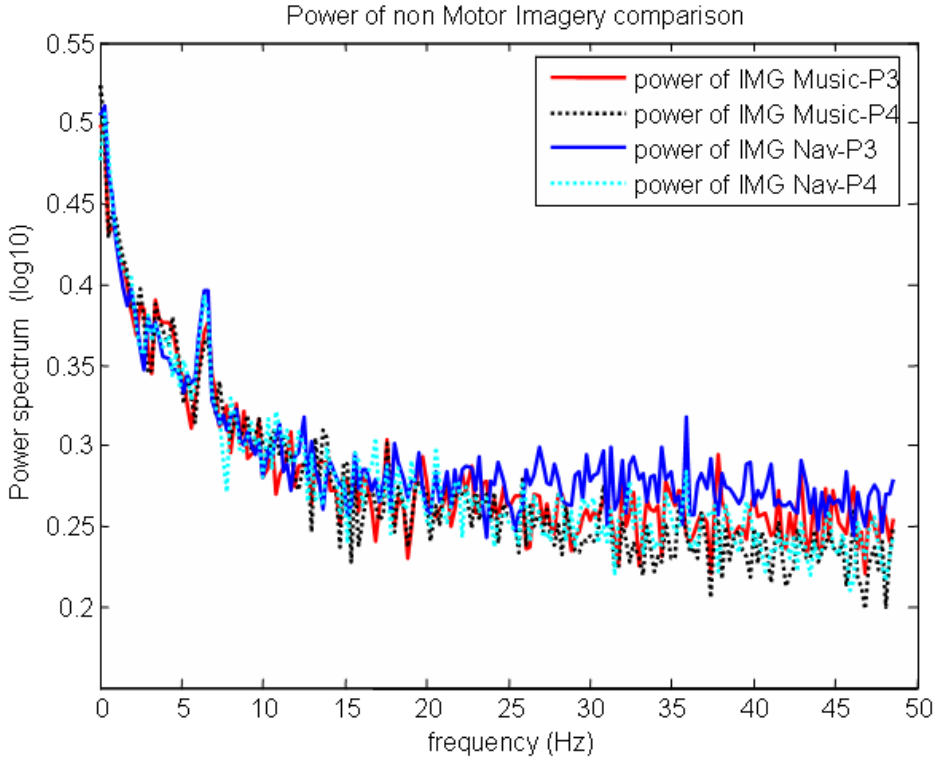

Figure C.52: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 5.*

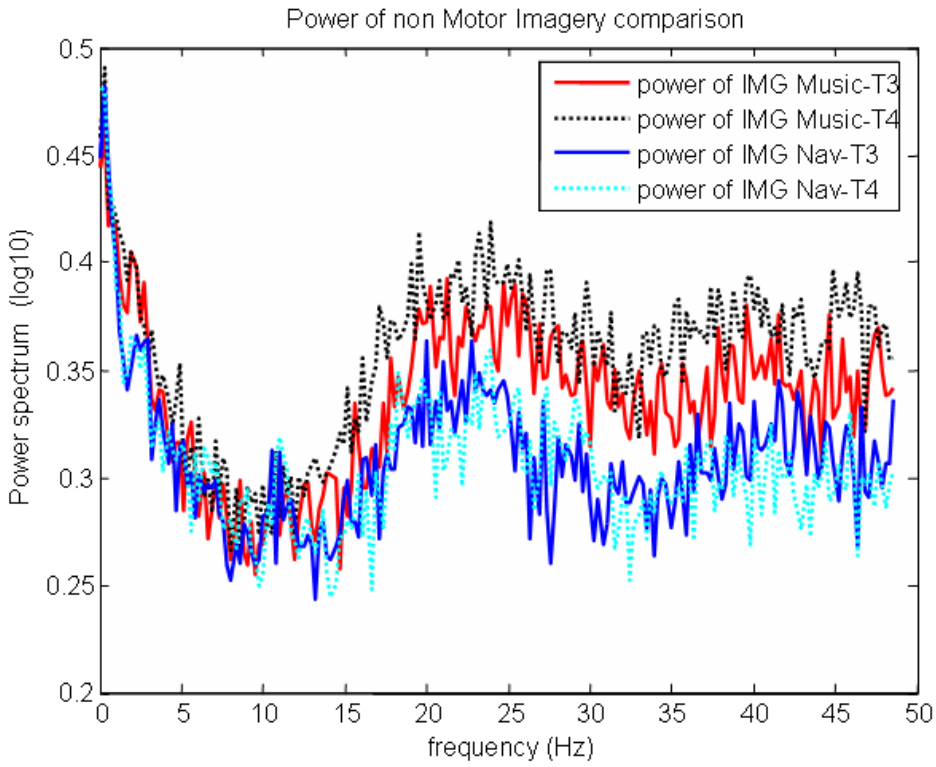

Figure C.53: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 6.*

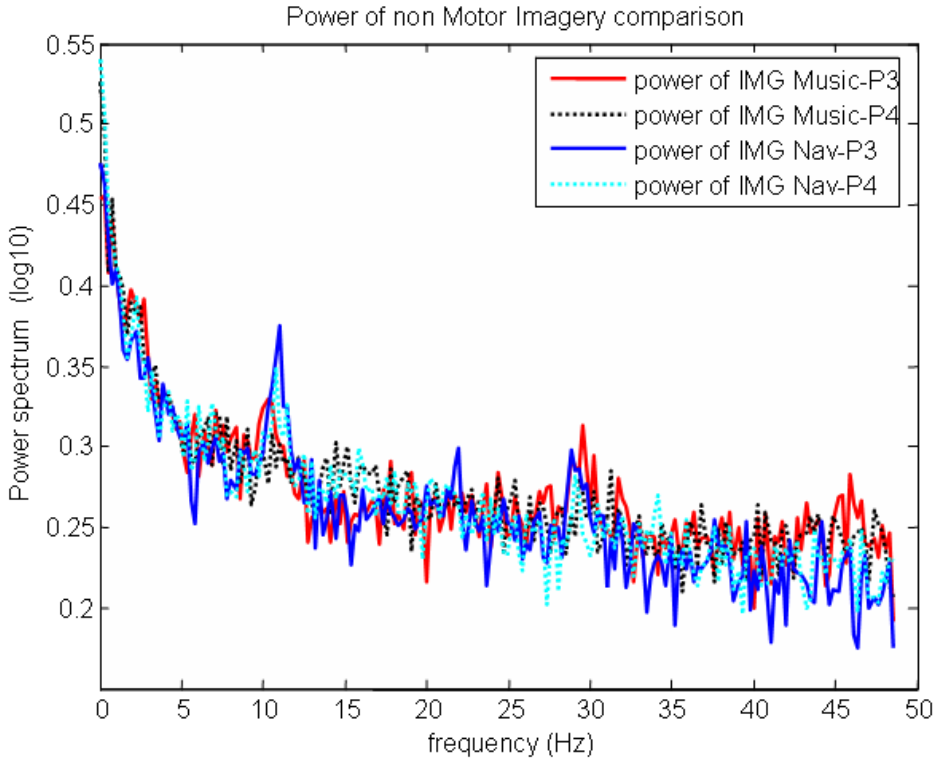

Figure C.54: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 6.*

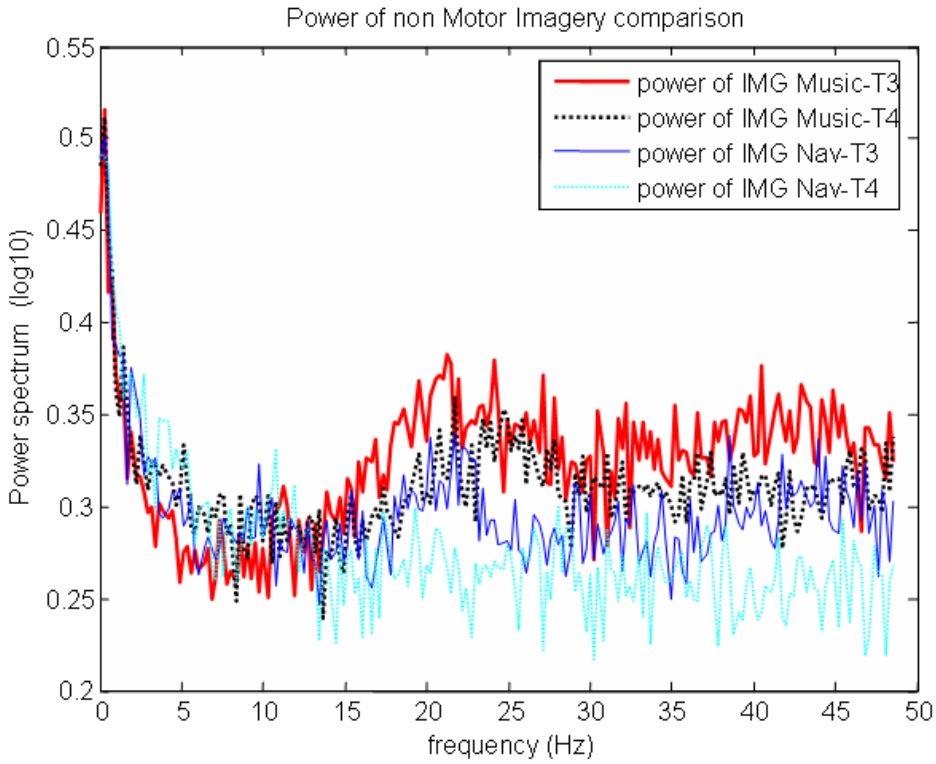

Figure C.55: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the temporal area on the data from Subject 7.*

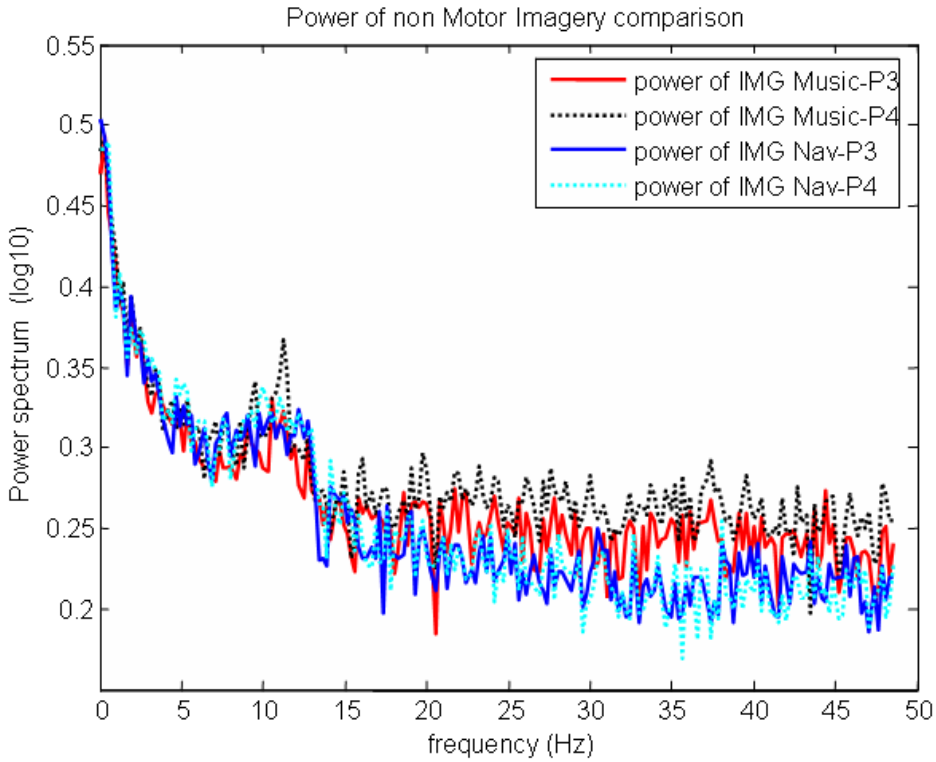

Figure C.56: *The averaged power spectra over ten trials for the spatial navigation imagery and music imagery at the parietal area on the data from Subject 7.*

Appendix D

Publications

Journal Papers

S. Wang, and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis", *Computational Intelligence and Neuroscience*, 2007(ID41468), 9pp, 2007.

Conference Papers

C. J. James and S. Wang, "Single Channel ICA on P300 based BCI", *Proceedings of the IET Medical Signal and Information Processing Conference MEDSIP 2008*, Italy, 14-16 July, CD-ROM, 2008

M. Davies, C. J. James and S. Wang, "Space-Time ICA and EM Brain Signals", *Proceeding of 7th International Conference, ICA 2007*, 577-584, 2007.

S. Wang, and C. J. James, "On the independent component analysis of evoked potentials through single or few recording channels", *29th International Conference of IEEE Engineering in Medicine and Biology Society (EMBC2007)*, Lyon, France, 23-26 August 2007, 5433-5436, 2007.

C. J. James and S. Wang, "Blind source separation in single-channel EEG analysis: An application to BCI", *Proceedings 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 31 August - 2 September 2006, CD-ROM, 2006,

S. Wang, and C. J. James, "Enhancing evoked responses for BCI through advanced ICA techniques", *Proceedings of the IET Medical Signal and Information Processing Conference MEDSIP 2006*, Glasgow, Scotland, 17 - 19 July 2006, CD-ROM, 2006.

Conference Abstracts

S. Wang, C. J. James and M. Stokes, "The Southampton BCI Research Programme", *Abstract book of Satellite symposium of IEEE EMBS 27th Annual Conference*, Beijing, China, 2005.

S. Wang, and C. J. James, "Preprocessing the P300 word speller with ICA for Brain-Computer Interfacing", *Abstract book of ICA Research Network Workshop on Applied Blind Source Separation and Independent Component Analysis*, Southampton, 2005.

References

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, "Brain-computer interface technology: a review of the first international meeting", *IEEE Transactions. on Neural Systems and Rehabilitation,* Vol. 8, 2,164-173, 2000.

[2] P. R. Kennedy, R. A. Bakay, "Restoration of neural output from a paralysed patient by a direct brain connection". *Neuroreport*. 1;9(8):1707-11, 1998.

[3] S. Seymour. "The Brain*", HarperTrophy*, 1999.

[4] H. Berger, "Über das Elektroenkephalogram des Menschen", *Archives of Psychiatry.* 87: 527-70, 1929.

[5] M. Le Van Quyen, V. Navarro, M. Baulac, B. Renault, J. Martinerie, "Anticipation of epileptic seizures from standard EEG recordings". *The Lancet*, Vol. 361, Issue 9361, pp 970-971, 2000.

[6] B. Pohlmann-Eden, M. Newton, "First seizure: EEG and neuroimaging following an epileptic seizure". *Epilepsia*, Vol. 49 Suppl 1:19-25, 2008.

[7] I. Giuseppina, F. Fabio, M, Nadia and M, Francesco Carlo, "Brain Activity Investigation by EEG Processing: Wavelet Analysis, Kurtosis and Renyi's Entropy for Artifact Detection", *International Conference on Information Acquisition*. pp:195-200, 2007.

[8] C. Liu**,** P. Lo, "Investigation of spatial characteristics of meditation EEG using wavelet analysis and fuzzy classifier", *Proceedings of the fifth IASTED International Conference: biomedical engineering*, pp: 91 – 96, 2007

[9] www.sce.com/kidsscience/body/nervous.html.

[10] D. Amit, “Modelling brain function”, *Cambridge University Press*, 1989.

[11] S. Rush and D. A. Driscoll, “EEG-electrode sensitivity - An application of reciprocity”. *IEEE Transactions on Biomedical Engineering,* BME-16:(1) 15-22, 1969.

[12] H.H. Jasper, “The ten–twenty electrode system of the International Federation”, *Electroencephalogram. Clinical. Neurophysiology.* Vol. 10, pp: 367–380, 1958.

[13] F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Lüders, M. Nuwer, and T. W. Picton, “American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature”. *Journal of Clinical Neurophysiology,* 8: 200-2, 1991.

[14] T. W. Picton, P. van Roon, M. L. Armilio, et al, “The correction of ocular artifacts: a topographic perspective”. *Journal of Clinical Neurophysiology,* 111(1): 53-65, 2000.

[15] W. T. Roth, J. M. Ford, A. Pfefferbaum and T. R. Elbert, “Methodological issues in event-related potential and magnetic field studies,” *Psychopharmacology, The Fourth Generation of Progress, F.E. Bloom and D.J. Kupfer (Eds.)*, pp. 895-910, 1995.

[16] S. R. Benbadis, D. Rielo, “EEG Atlas: EEG Artifacts”, *Neurology in eMedicine*, 2006

[17] J. Sijbers. 1. , J. Van Audekerke, et al, "Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data", *Magnetic Resonance Imaging*, Vol. 18, Nr. 7, p. 881-886, 2000.

[18] T. Jung, “Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects”, *Clinical Neurophysiology*, Vol. 111, Issue 10, pp: 1745-1758, 2000.

[19] J. J. Vidal. “Towards direct brain–computer communication”, *Annual Review of Biophysics and Bioengineering*, 2:157–180, 1973.

[20] J. Phukan, N. P. Pender, O. Hardiman. "Cognitive impairment in amyotrophic lateral sclerosis". *Lancet Neurol* Vol. 6, Issue. 11. pp:994–1003. 2007.

[21] T.M. Miller, B.K. Kaspar, G.J. Kops, *et al*. "Virus-delivered small RNA silencing sustains strength in amyotrophic lateral sclerosis", *Annals of Neurology,* Vol. 57, pp: 773-6, 2005.

[22] D. Reed, D. Labarthe, K.M. Chen, R. Stallones, "A cohort study of amyotrophic lateral sclerosis and parkinsonism-dementia on Guam and Rota". *American Journal of Epidemiology*, Vol. 125, pp: 92–100, 1987.

[23] H. Hayashi and E. A. Oppenheimer, "LS patients on TPPV Totally locked-in state, neurologic findings and ethical implications", *Neurology*, Vol. 61 pp:135-137. 2001.

[24] H. Moravec, "When will computer hardware match the human brain"? *Journal of Evolution and Technology*. Vol. 1, 1998.

[25] *http://ida.first.fraunhofer.de/projects/bci/bbci_official/index_en.html.*

[26] *http://www.mp.uni-tuebingen.de/mp/index.php?id=137.*

[27] J. R. Wolpaw, D. J. McFarland, T. M. Vaughan, G. Schalk, "The Wadsworth Center brain-computer interface (BCI) research and development program", *IEEE Transactions on Neural Systems and Rehabilitation,* 11(2):204-7, 2003.

[28] *http://www.bci-info.tugraz.at/.*

[29] B. Graimann and G. Townsend, G. Pfurtscheller, "Brain-computer communication - A brief introduction". *Available from <http://www.bci-info.tugraz.at/Research_Info/documents/articles/briefintro >*

[30] T. N. Lal, M. Schröder, J. Hill, H. Preissl, T. Hinterberger, et al. "A Brain Computer Interface with Online Feedback based on Magnetoencephalography". *Proceedings of the 22nd International Conference on Machine Learning*, pp: 465 - 472. 2005.

[31] S. Coyle, T. Ward, C. Markham, and G. McDarby, "On the suitability of Near-Infrared Systems for Next Generation Brain Computer Interfaces." *Physiological Measurement Special Issue: World Congress on Medical Physics and Biomedical Engineering 2003,* 25: 815-822, 2003.

[32] N. Weiskopf, K. Mathiak, S. W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N. Birbaumer, "Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI)," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 966-70, 2004.

[33] B. K. Siesjo, "Brain Energy Metabolism", *New York: Wiley*, 1978.

[34] S. P. Levine, J. E. Huggins, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, M. M. Rohde, E. A. Passaro, D. A. Ross, K. V. Elisevich, and B. J. Smith, "A direct brain interface based on event-related potentials," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 180-5, 2000.

[35] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 954-62, 2004.

[36] G. J. Feldman, R. D. Cousins, "Unified approach to the classical statistical analysis of small signals" *Physical Review*, 1998.

[37] R. A, Fisher, "The Use of Multiple Measurements in Taxonomic Problems", *Annals of Eugenics*, Vol. 7, pp: 179-188, 1936.

[38] M. Aizerman, E. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning". *Automation and Remote Control*, Vol. 25, pp: 821–837, 1964.

[39] K. Hornik, M. Stinchcombe, H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks archive*, Vol. 2, Issue 5, 1989.

[40] B. E. Boser, I. M. Guyon, and V. N. Vapnik. "A training algorithm for optimal margin classifiers" *5th Annual ACM Workshop on COLT*, pp: 44-152, 1992.

[41] M. Rychetsky, S. Ortmann, M. Glesner, "Support Vector Approaches for Engine Knock Detection", *International Joint Conference on Neural Networks (IJCNN 99)*, 1999.

[42] R. T. Lauer, P. H. Peckham and K. L. Kilgore, "EEG-based control of a hand grasp neuroprothesis" *NeuroReport,* 10:1767–1771, (1999)

[43] P. Boord, Y. Tran, and J. Middelton, etc. “Alpha Band activity during eye-closure in people with spinal cord injury”, *Proceeding of International Functional Electrical Stimulation Society Conference*, 2003.

[44] J. A. Turner, J. S. Lee and O. Martinez, etc. “Somatotopy of the motor cortex after long-term spinal cord injuryor amputation”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 9, Issue 2, pp:154 – 160, 2001.

[45] J. Wessberg, C. R. Stambaugh and J. D. Kralik, etc. “Real-time prediction of hand trajectory by ensembles of cortical neurons in primates”. *Nature*, Vol.16, pp: 361-365, 2000.

[46] I. Wickelgren, “Neuroscience: Tapping the Mind”, *Science,* Vol. 299. no. 5606, pp: 496 – 499, 2003.

[47] E. E. Sutter, D. Tran, “The field topography of ERG components in man”. *The photopic luminance response*;32:433–446, 1992.

[48] L. A. Farwell LA and E. Donchin, “Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials”. *Electroencephalography and Clinical Neurophysiology*, 70:510–523, 1988.

[49] C. Ming, and G. Shangkai. “An EEG-Based Cursor Control System”, *Proceedings of the First Joint BMES/EMBS Conference: Serving Humanity, Advancing Technology*. p. 669, 1999.

[50] E. E. Sutter. “The brain response interface: communication through visually induced electrical brain responses”, *Journal of Microcomputer Applications*, 15:31–45, 1992.

[51] M. Middendorf, G. McMillan, G. Calhoun and K. S. Jones. “Brain–computer interfaces based on steady-state visual evoked response”, *IEEE Transactions on Rehabilitation Engineering,* 8:211–213, 2000.

[52] X. Gao, D. Xu, etc. “A BCI-based environmental controller for the motion-disabled” *IEEE Transactions on Rehabilitation Engineering.* V11-2, 2003.

[53] S. G. Mason and G. E. Birch, "A General Framework for Brain-Computer Interface Design," *IEEE Transactions on Rehabilitation Engineering, 11, 70-85,* 2000.

[54] S. G. Mason, M. M. Moore, and G. E. Birch, "Designing pointing devices using brain-computer interface technology," *Proceedings of IEEE 1st International Conference on Neural Engineering,* 2003.

[55] T. Hinterberger, G. Baier, J. Mellinger, et al, "Auditory Feedback of the human EEG for direct brain-computer communication", *Proceedings of the International Conference on Auditory Displays*, 2004.

[56] N. Neumann and A. Kübler, "Training locked-in patients: A challenge for the use of brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, pp. 169-172, 2003.

[57] N. Birbaumer, T. Hinterberger, A. Kubler, N. Neumann, "The thought-translation-device (TTD): Neurobehavioral Mechanism and Clinical Outcome". *IEEE Transaction on neural systems and Rehabilitation Engineering*, v. 11, p. 120-123, 2003.

[58] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, "Brain-computer interface research at the Wadsworth Center," *IEEE Transactions on Rehabilitation Engineering,* vol. 8, pp. 222-226, 2000.

[59] S. Sutton, M. Braren, J. Zublin, and E. John, "Evoked potential correlates of stimulus uncertainty" *Science*, 150, 1187–1188, 1965.

[60] E. W. Sellers, D. J. Krusienski, D. J. McFarland, et al, "A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance", *Biological Psychology*, Vol.73, Issue 3, 2006.

[61] B. Z. Allison and J. A. Pineda, "ERP's evoked by different matrix sizes: implications for a brain computer interface (BCI) system," *IEEE Transaction on neural systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 110-113, 2003.

[62] R. C. Wu, S. F. Liang, C. T. Lin, and C. F. Hsu, "Applications of Event-Related-Potential-Based Brain Computer Interface to Intelligent Transportation

Systems," *The 2004 IEEE International Conference on Networking, Sensing, and Control*, 2004.

[63] A. Rakotomamonjy, V. Guigue, G. Mallet and V. Alvarado, "Ensemble of SVMs for improving Brain Computer Interface P300 speller performances", *Proceedings of International Conference on Artificial Neural Networks,* 2005..

[64] I. Jentzsch, "Independent Component Analysis Separates Sequence-Sensitive ERP Components", *International Journal of Bifurcation and Chaos*, 14(2), 667-678, 2004.

[65] J. W. Kozelka and T. A. Pedley, "Beta and mu rhythms," *Journal of Clinical Neurophysiology*, vol. 7, no. 2, pp. 191–207, 1990.

[66] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)," *Electroencephalography and Clinical Neurophysiology*, vol. 43, no. 5, pp. 757–760, 1977.

[67] S. Makeig. "A natural basis for efficient brain-actuated control", *IEEE transaction on rehabilitation engineering*, Vol. 8, pp: 208-211, 2000.

[68] M. Pregenzer, G. Pfurtscheller, "Frequency component selection for and EEG-based brain computer interface", *IEEE Transactions. on Neural Systems and Rehabilitation,*7:413–419.1999.

[69] J. A. Pineda, D. S. Silverman, A. Vankov, J. Hestenes, "Learning to control brain rhythms: making a brain-computer interface possible", *IEEE Transaction on neural systems and Rehabilitation Engineering,* 11(2):181-4, 2003.

[70] C. Guger, A. Schlögl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller, "Rapid prototyping of an EEG-based brain–computer interface (BCI)" *IEEE Transaction on Rehabilitation Engineering*, vol. 9, no. 1, pp. 49-58, 2001.

[71] T. Wang and B. He, "An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in brain computer interface," *Journal of Neural Engineering,* 1(1): 1-7, 2004.

[72] A. Delorme and S. Makeig. “EEG changes accompanying learned regulation of 12-Hz EEG Activity”, *IEEE Transaction on neural systems and Rehabilitation Engineering*, 2(2):133-136, 2003.

[73] N. J. Huan and R. Palaniappan, “Classification of mental tasks using fixed and adaptive autoregressive models of EEG signals,” *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC ’04)*, pp. 507–510, 2004.

[74] H. Ramoser, J. M¨uller-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.

[75] Z. A. Keirn and J. I. Aunon, “A new mode of communication between man and his surroundings”. *IEEE Transactions on Biomedical Engineering*, Vol. 37, Issue 2, pp: 1209–1214, 1990.

[76] E. Curran, P. Sykacek, M. Stokes, et al. “Cognitive tasks for driving a brain-computer interfacing system: a pilot study,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 12(1):48-54, 2004.

[77] http://www.cyberkineticsinc.com/content/researchproducts/index.jsp.

[78] P. R. Kennedy, R. A. Bakay, “Restoration of neural output from a paralyzed patient by a direct brain connection”, *NeuroReport,* 9:1707–1711, 1998.

[79] J. K. Chapin, K. A. Moxon, R. S. Markowitz and M. A. Nicolelis, “Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex”, *Nature Neuroscience,* 2:664–670, 1999.

[80] J. Wessberg, C. R. Stambaugh, J. D. Kralik, et al. “Real-time prediction of hand trajectory by ensemble of cortical neurons in primates”. *Nature,* 408:361–365, 2000.

[81] Y. Prut and S. I. Perlmutter, “Firing Properties of Spinal Interneurons during Voluntary Movement. I. State-Dependent Regularity of Firing”, *Journal of Neuroscience*, Vol. 23, Issue 29, pp: 9600-9610, 2003.

[82] R. Leigh. Hochberg, D. Mijail. M. Gerhard, et al. "Neuronal ensemble control of prosthetic devices by a human with tetraplegia". *Nature* 442: 164-171, 2006.

[83] L. Bolinger, R. D. Arcy, Y. Marchand, et al, "Expert Assessment of Neuroimaging and Brain-Computer Interface", *Defence R&D Canada contractor report*, 2005.

[84] T.N. Lal, T. Hinterberger, G. Widmann, et al, "Methods towards invasive human brain–computer-interfaces", *Advanced Neural Information Process System,* 17:737–44, 2005.

[85] N. Birbaumer, "Brain–computer-interface research: Coming of age", *Clinical Neurophysiology*, Vol. 117, Issue 3, Pages 479-483, 2005.

[86] R. G. Lyons, "Understanding Digital Signal Processing, Second Edition", *Prentice Hall,* Chapter 11, 2004.

[87] O. Rompelman, H. H. Ros, "Coherent averaging technique: a tutorial review." *Journal of Biomedical Engineering* . Vol. 8(1):30-5, 1986.

[88] D. C. Wood, "Habituation in Stentor: produced by mechanoreceptor channel modification", *Journal of Neuroscience*, Vol. 8, 2254, 1998.

[89] F. Beverina, G. Palmas and S. Silvoni, etc., "User adaptive BCIs: SSVEP and P300 based interfaces", *Psychology Journal*, Vol.1, Num. 4, pp: 331 – 354, 2003.

[90] N. Srinivasan, "Cognitive neuroscience of creativity: EEG based approaches", *Methods*, Vol. 42, Issue 1, pp: 109-116, 2007.

[91] N. A. de Beer, M. van de Velde and P. J. Cluitmans, "Clinical evaluation of a method for automatic detection and removal of artifacts in auditory evoked potential monitoring", *Journal of clinical monitoring and computing*. Vol. 11, pp: 381–391, 1995.

[92] P. B. C. Fenwick, P. Mitchie, J. Dollimore and G.W Fenton, "Application of the autoregressive model to EEG analysis," *Agressoiogia*, Vol. 10, pp: 553-564, 1969.

[93] L. H. Zetterberg, "Estimation of parameters for a Linear difference equation with application to EEG analysis," *Mathematics Bioscience*, Vol. 5, pp: 227-275, 1969.

[94] A. Isaksson and A. Wennberg, "Visual evaluation and computer analysis of the EEG," *Electroencephalography and Clinical Neurophysiology*, Vol. 38, pp: 79-86, 1975.

[95] G. Bodenstein and H.M. Praetorius, "Feature extraction from the EEG by adaptive segmentation," *Proceeding of IEEE*, Vol. 65, pp: 642-652, 1977.

[96] A.C. Sanderson and J. Segen, " Hierarchical Modelling of EEG Signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-2, pp: 405-415, 1980.

[97] F. H. Lopes da Silva, "Analysis of EEG non-stationarities," *Electroencephalography and Clinical Neurophysiology*, Vol. 34, pp: 163-179, 1978.

[98] H. Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control,* Vol. 19, pp: 716-723, 1974.

[99] G. Schwarz, "Estimating the dimension of a model", *The Annals of Statistics* Vol. 6, pp: 461-464, 1978

[100] D. J. McFarland and J.R. Wolpaw, "Sensorimotor rhythm-based brain–computer interface (BCI): model order selection for autoregressive spectral analysis", *Journal of Neural Engineering*, Vol. 5. PP: 155-162, 2008.

[101] R. J. Croft, R. J. Barry. "Removal of ocular artifact from the EEG: a review", *Neurophysiology Clinic*, Vol. 30, pp: 5–19, 2000.

[102] J. S. Barlow, "Artifact processing (rejection and minimization) in EEG data processing". *Electroencephalograph Clinical Neurophysiology,* Vol. 2, pp:15–62, 1986.

[103] H. Dym and H. McKean, "Fourier Series and Integrals", *Academic Press*, 1985.

[104] K. Yosida, "Functional Analysis", *Springer-Verlag*, 1968.

[105] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, "Discrete-time signal processing", *Upper Saddle River, N.J.: Prentice Hall*, 1999.

[106] E. O. Bringham, "The Fast Fourier Transform", *Prentice-Hall, New Jersey*, 1974.

[107] R. P. Michael, "Time-frequency representation of digital signals and systems based on Short-time Fourier analysis", *IEEE Transactions on Acoustics, speech, and signal processing*, Vol. ASSP-28, No. 1, 1980.

[108] J. W. Cooley, and W. T. John, "An algorithm for the machine calculation of complex Fourier series", *Mathematic Computing*, Vol. 19, pp: 297–301, 1965.

[109] F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proceedings of the IEEE,* Vol. 66, No. 1, 1978.

[110] C. Alloway, R. Ogilvie and C. Shapiro, "EEG power spectral analysis of the sleep onset period in narcolepsy". *Journal of Sleep Research*, Vol. 25: 185, 1996.

[111] H. Olkkonen, "Running discrete Fourier transform for time-frequency analysis of biomedical signals", *Medical Engineering & Physics,* Vol. 17, pp: 455–458, 1995.

[112] S. M. Hosni, M. E. Gadallah, S. F. Bahgat, M. S. AbdelWahab, "Classification of EEG signals using different feature extraction techniques for mental-task BCI", Proceeding of International conference on Computer Engineering & Systems, 2007.

[113] E. T. Craig, K. Jürgen, "Reference-free quantification of EEG spectra: Combining current source density (CSD) and frequency principal components analysis (fPCA)", Clinical Neurophysiology Vol.116, Issue 12, pp: 2826-2846, 2005.

[114] S. Choi, A. Cichocki, H. M. Park, S. Y. Lee. "Blind source separation and independent component analysis: a review". *Neural Information Processing - Letters and Reviews,* 2005.

[115] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture", *Signal Processing*, Vol. 24, pp:1-10, 1991.

[116] P. Comon, "Independent Component Analysis: a new concept", *Signal Processing, Elsevier*, Vol. 36, Issue. 3, pp: 287-314, 1994.

[117] T. P. Jung, S. Makeig and M. Westerfield, etc. "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects", Clinical Neurophysiology, Vol. 111, pp: 1745–58, 2000.

[118] T. P. Jung, S. Makeig and M. Westerfield, etc. "Analysis and visualization of single-trial event-related potentials", *Human Brain Mapping,* Vol. 14, pp: 166–85, 2001.

[119] R. Vigario, J. Sarela and V. Jousmiki V, etc. “Independent component approach to the analysis of EEG and MEG recordings”, *IEEE Transactions on Biomedical Engineering*, Vol. 47, pp: 589–93, 2000.

[120] S. M. Le, B. W. Jervis, N. R. Porter and A. R. Maye, “Study of ERPs on an individual trial basis using independent component analysis”, *IEE Colloquium on Medical Applications of Signal Processing*, pp: 3/1 - 3/10, 1999.

[121] S. Makeig, S. Enghoff, T. P. Jung and T. J. Sejnowski. “A natural basis for efficient brain-actuated control”, *IEEE Transactions on Rehabilitation Engineering*, Vol.8, Issue, 208–11, 2000.

[122] R. J. Croft, R. J. Barry, “Issues relating to the subtraction phase in EOG artefact correction of the EEG”, *International Journal of Psychophysiology*. Vol. 44, pp: 187–95, 2002.

[123] A. Cichocki, S. Shishkin, T. Musha, et al, “EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease”, *Clinical Neurophysiology*, Vol: 116, pp: 729-737, 2000.

[124] A. D. Sanchez, “Frontiers of research in BSS/ICA”, *Neurocomputing*, Vol. 49, No. 1, pp: 7-23, 2002.

[125] J. Qin, Y. Li and A Cichocki, “ICA and Committee Machine-Based Algorithm for Cursor Control in a BCI System”, *Lecture Notes in Computer Science*, Volume 3496, pp: 973-978, 2005.

[126] A. Cichocki , S. Shishkin and T. Musha, etc. “EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease”, *Clinical Neurophysiology*, Vol. 116, Issue 3, pp: 729 – 737, 2004.

[127] J. Patricia, J. C. DeCoursey, J. L. Dunlap, “Chronobiology: Biological Timekeeping”, *Sinauer Associates Inc. ISBN-10: 087893149*, 2003.

[128] A. Hyvarinen, J. Karhunen and E. Oja, “Independent component analysis”. *John Wiley & Sons, New York, NY*, 2001.

[129] J. Herault and C. Jutten, “Space or time processing by neural network models”, *Proceeding of AIP Conference: Neural Networks for Computing*, Vol. 151, 1986.

[130] C. Jutten and J. Herault, "*Blind separation of sources. I, An adaptive algorithm based on neuromimetic architecture*", Signal processing, Vol. 24, pp: 1-10, 1991.

[131] P. Comon "*Independent component analysis, a new concept?*", Signal Process. Vol. 36, pp: 287–314, 1994.

[132] A. Hyvärinen and P. Pajunen, "*Nonlinear independent component analysis: Existence and uniqueness results*", Neural Networks, Vol. 12, Issue 3, pp: 429-439, 1999.

[133] P. Pajunen, J. Karhunen. "*A Maximum Likelihood Approach to Nonlinear Blind Source Separation*". ICANN. Pp: 541-546, 1997.

[134] L. Parra, G. Deco and S. Miesbach, "*Statistical Independence and Novelty Detection with Information Preserving Nonlinear Maps*", Neural Compution, Vol. 8, No. 2, pp: 260-269, 1996.

[135] J. Lin, D. Grier, and J. Cowan. "*Feature extraction approach to blind source separation*". IEEE Workshop on Neural Networks for Signal Processing, pp: 398--405, 1997.

[136] K. Torkkola. "Blind separation of convolved sources based on information maximization". *Proceeding of IEEE Workshop on Neural Networks and Signal Processing*, pp: 423-432, 1996.

[137] P. Smaragdis. "Blind separation of convolved mixtures in frequency domain". *Neurocomputing*, Vol. 22, pp: 21-34, 1998.

[138] N. Mitianoudis and M. Davies. "New fixed-point ICA algorithm for convolved mixtures". *Proceeding of 3rd International Workshop on Independent Component Analysis and Blind Source Separation*, pp: 633-638, 2001.

[139] A. Dapena, M.F. Bugallo and L. Castedo. "Separation of convolutive mixtures of temporally-white signals: a novel Frequency-Domain Approach". *Proceeding of 3rd International Workshop on Independent Component Analysis and Blind Source Separation*, pp: 179-184, 2001.

[140] R. K. Prasad, H. Saruwatari, and K. Shikano, "Enhancement of speech signals separated from their convolutive mixture by FDICA algorithm", *Digital Signal Processing*, Vol. 19, Issue 1, pp: 127-133, 2009.

[141] R.K. Prasad, H. Saruwatari, A. Lee and K. Shikano, "A fixed point ICA algorithm for convoluted speech separation", *Proceeding of International Symposium on ICA and BSS*, pp. 579–584, 2003.

[142] M. Abramowitz and I. A. Stegun, "Probability Functions". – Chapter 26 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, *9th printing. New York: Dover, ISBN-10: 0486612724*, pp: 925-964, 1972.

[143] A. Papoulis, "Probability, Random Variables, and Stochastic Processes", *2nd edition. New York: McGraw-Hill*, pp. 145-149, 1984.

[144] J. J. A. Moors, "The Meaning of Kurtosis: Darlington Reexamined". *The American Statistician*, Vol. 40, pp: 283-284, 1986.

[145] M.G. Kendall, A. Stuart, "The Advanced Theory of Statistics", Chapter 3 in Volume 1, *3rd Edition. Griffin, London, ISBN-10: 0340614307*. 1969.

[146] J. Cohen, P. Cohen, S. G. West and L. S. Aiken, "Applied multiple regression/correlation analysis for the behavioral sciences". *3rd edition. Hillsdale, NJ: Lawrence Erlbaum Associates, ISBN-10: 0805822232*, 2003.

[147] M. Kac, "Statistical Independence in Probability, Analysis and Number Theory (Carus Mathematical Monographs, No. 12)", *Mathematical Association of America (MAA), ISBN 10: 0883850257*, 1959.

[148] A. Hyvärinen, E. Oja, "Independent component analysis: algorithms and applications", *Neural Network,* Vol. 13, pp:411-30, 2000.

[149] A. Hyvarinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis". *Neural Computation*, 9(7), pp.1483-1492, 1997.

[150] http://www.cis.hut.fi/projects/ica/fastica/;

[151] T. M. Cover and J. A. Thomas, "*Elements of Information Theory*", John Wiley & Sons, 1991.

[152] A. Hyvärinen, "*Survey on independent component analysis*", Neural Computing Surveys, Vol. 2, pp: 94-128, 1999.

[153] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", Neural Computation, Vol. 7, 1129-59, 1995.

[154] C. J. James and C. W. Hesse, "Independent Component Analysis for Biomedical Signals", *Physiological Measurement*, 26, R15-R39, 2005.

[155] A. Ziehe and K.-R. Müller, "*TDSEP - an efficient algorithm for blind separation using time structure*", Proceeding of international conference on Artificial Neural Networks, pp: 675-80, 1998.

[156] A. Belouchrani and M.G. Amin, "*Blind source separation based on time-frequency signal representations*", IEEE Transaction on Signal Processing, Vol. 46, Issue. 11, pp: 2888-2897, 1998.

[157] A. Ziehe and K. R. M¨uller, "TDSEP—an efficient algorithm for blind separation using time structure," *Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN '98)*, pp. 675–680, 1998.

[158] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, 1998.

[159] A. Ziehe, P. Laskov, K.-R. M¨ueller, and G. Nolte, "A linear least-squares algorithm for joint diagonalization," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pp. 469–474, 2003.

[160] C. J. James, C. W. Hesse, "*Mapping scalp topographies of rhythmic EEG activity using temporal decorrelation based constrained ICA"*, Proceeding of International Conference of IEEE Engineering in Medicine and Biology Society Vol.1, Issue.1, pp: 994 – 997, 2004.

[161] J. Karvanen, T. Tanaka, "*Temporal decorrelation as preprocessing for linear and post-nonlinear ICA*", Proceeding of International conference on independent component analysis and blind signal separation, Vol. 3195, pp: 774-781, 2004.

[162] T. Joliffe, "Principal Component Analysis", *Springer-Verlag*, 1986.

[163] R. A. Horn and C. R. Johnson, "Matrix Analysis", *Cambridge University Press*, Section 7.3, 1985.

[164] M. Vlachos, C. Domeniconi and D. Gunopulos, etc. "Non-Linear Dimensionality Reduction Techniques for Classification and Visualization" *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002.

[165] A. Hyvärinen, E. Oja, "Independent component analysis: algorithms and applications", *Neural Network,* Vol. 13, pp: 411-30, 2000.

[166] N. Castaeda-Villa, C. J. James, "Objective source selection in Blind Source Separation of AEPs in children with Cochlear Implants", *29th Annual International Conference of the IEEE on EMBS*, Issue, 22-26, pp: 6223 – 6226, 2007.

[167] C. W. Hesse and C. J. James, "On Semi-Blind Source Separation using Spatial Constraints with Applications in EEG Analysis", *IEEE Transactions on Biomedical Engineering,* 2005.

[168] A. Cichocki, S. Amari, and K. Siwek, "ICALAB toolbox for signal processing," http://www.bsp.brain.riken.jp/ICALAB/.

[169] C. W. Hesse and C. J. James, "Stepwise model order estimation in blind source separation applied to ictal EEG", *International Conference of IEEE EMBS*, 2004.

[170] C. W. Hesse and C. J. James, "The FastICA algorithm with spatial constraints", *Signal Processing Letters*, IEEE Volume 12, Issue 11, Nov, pp: 792 – 795, 2005.

[171] C. J. James and C. W. Hesse, "On Semi-Blind Source Separation in the analysis of Electromagnetic Brain Signals", *Journal of the Franklin Institute*, 2006.

[172] W. Lu and J. C. Rajapakse, "ICA with reference" Proceeding of 3rd International Conference on Independent Component Analysis and Blind Signal Separation, pp: 120–5, 2001.

[173] C. W. Hesse and C. J. James, "Blind source separation through multi-spectral signal decorrelation applied to multi-channel ictal EEG", *Proceeding of Medical conference on Medical and Biological Engineering*, 2004.

[174] M. E. Davis and C. J. James, "Source-separation using single-channel ICA," *Signal Processing*, 2007.

[175] F. Takens, "Detecting strange attractors in turbulence". D. A. Rand and L.-S. Young *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898: 366–381, 1981.

[176] C. J. James and D. Lowe, "Extracting multisource brain activity from a single electromagnetic channel," *Artificial Intelligence in Medicine*, vol. 28(1), pp. 89-104, 2003.

[177] A. C. Fisher, R. P. Hagan, M. C. Brown, W. El-Deredy, P. J. G. Lisboa, "ICA-Based Blind Source Separation (BSS) Recovery of the Pattern Electroretinogram (PERG) from Single Channel Records with Poor SNR", *Processing of Advances in Medical Signal and Information. MEDSIP,* 2006.

[178] C. J. James and C. W. Hesse, "On the use of spectrally constrained ICA applied to single-channel ictal EEG recordings within a dynamical embedding framework", *Processing of 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3pp, 956-959, 2005.

[179] C. J. James, D. Lowe and O. J. Gibson, "An Objective Study into Single vs Multiple Channel Brain Signal Analysis using Realistic Ictal EEG", *Proceedings of 5th International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, 2003.

[180] T. W. Lee. S. E. Lewicki, "ICA mixture models for image processing", *Proceedings of 6th Joint Synposiurn on Neural Computation*, 1999.

[181] Technical specification group radio access network: physical layer – general specification, 3GPP, Release 6, www.3gpp.org, 2003.

[182] T. Ristaniemi and J. Joutsensalo, "*On the performance of blind source separation in CDMA downlink*". Proceeding of international workshop on Independent Component Analysis and Signal Separation (ICA'99), pp: 437–441, 1999.

[183] A. K. Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ECG signals using independent components analysis," *Neurocomputing,* vol. 22, pp. 173–186, 1998.

[184] J. F. Cardoso, "Multidimensional independent component analysis," *Processing of IEEE International Conference on Acoustics, Speech and Signal,* vol. 4, pp. 1941–1944, 1998.

[185] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, S. Makeig and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, pp. 163–178, 2000.

[186] M. J. McKeown, T. P. Jung, S. Makeig, G. G. Brown, G. Kindermann, T. W. Lee and T. J. Sejnowski, "Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop colornaming task," in *Proceedings of the National Academy of Sciences, vol. 95*, pp. 803–810, 1998.

[187] C. J. James, and S. Wang, "Blind source separation in single-channel EEG analysis: an application to BCI", *Proceedings 28th Annual International Conference IEEE on EMBS*. 2006.

[188] www.gtec.at

[189] www.biopac.com

[190] www.neuroscan.com

[191] *http://www.neurobs.com/*

[192] S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. C. Gunter and A. D. Friederici, "Music, language and meaning: brain signatures of semantic processing," *Nature Neuroscience*, 7, 302 – 307, 2004.

[193] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interface—A new communication device for handicapped people," *Journal of Microcomputer Application*, vol. 16, pp. 293–299, 1993.

[194] E. A. Maguire, R. S. Frackowiak, and C. D. Frith, "Recalling routes around London: Activation of the right hippocampus in taxi drivers," *Journal of Neuroscience*, vol. 17, no. 18, pp. 7103–7110, 1997.

[195] A. Graps, "An introduction to wavelets", *IEEE Computational Science and Engineering,* Vol. 2, (2), 50–61, 1995.

[196] Y. Juliana, K. Vasil and P. John, "P300 and alpha event-related desynchronization (ERD)", *Psychophysiology*, 38: 143-152, 2001.

[197] T. Hinterberger, B. Wilhelm, J. Mellinger, B. Kotchoubey, N. Birbaumer, "A device for the detection of cognitive brain functions in completely paralyzed or unresponsive patients". *IEEE Transactions on Biomedical Engineering*, Vol. 52-2, pp: 211-220, 2005.

[198] M. Hansenne, "The P300 event-related potential. II. Interindividual variability and clinical application in psychopathology." *Clinical Neurophysiology*, Vol. 30 (4): pp: 211-231, 2000.

[199] T. Frodl-Bauch, R. Bottlender, U. Hegerl, "Neurochemical substrates and neuroanatomical generators of the event-related P300," *Neuropsychobiology*, Vol. 40 (2): 86-94 1999.

[200] E. Curran and M. Stokes, "Learning to control brain activity: A review of the productuon and control of EEG components for driving brain-computer interface (BCI) systems," *Brain Cognitive*, Vol. 51, pp. 326–335, 2003.

[201] B. Blankertz, K. R. Muller and G. Curio, etc, "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials", *IEEE Transactions on Biomedical Engineering*, Vol. 51, Issue 6, pp: 1044-1051, 2004.

[202] (*ida.first.fraunhofer.de/projects/bci/competition_ii*)

[203] D. G. Altman, J. M. Bland JM, "Diagnostic tests. 1: Sensitivity and specificity". *BMJ,* Vol. 308 Issue 6943, (1994).

[204] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner T and H. Ritter, "BCI competition 2003—Data set IIb: Support vector machines for the P300 speller paradigm", *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 1073–1076, 2004.

[205] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI competition 2003—data set IIb: enhancing P300 wave detection using ICA-based subspace

projections for BCI applications," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1067–1072, 2004.

[206] S. Wang and C. J. James, "Enhancing evoked responses for BCI through advanced ICA techniques," in *Proceeding of IET 3rd International Conference on Advances in Medical, Signal and Information Processing (MEDSIP '06)*, p. 38,Glasgow,UK, 2006.

[207] A. Erfani and A. Erfanian, "The effects of mental practice and concentration skills on EEG brain dynamics during motor imagery using independent component analysis," in *Proceedings of the 26th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC '04)*, vol. 1, pp. 239– 242, 2004.

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

[209] C. J. James and C. W. Hesse, "Observing and tracking brain dynamics in electromagnetic brain signals through blind source separation," in *Proceedings of the IEE Medical Signal and Information Processing Conference (MEDSIP '04)*, pp. 145–152, 2004.

[210] N. Nagelkerke, "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, Vol. 78, no. 3, pp: 691-692, 1991.,

[211] *www.isis.ecs.soton.ac.uk/resources.*

[212] L. Song and J. Epps, "Classifying EEG for brain-computer interfaces: learning optimal filters for dynamical system features," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 857–864, 2006.

[213] P. Tichavsky, Z. Koldovsky and E. Oja, "Performance analysis of the FastICA algorithm and Crame/spl acute/r-rao bounds for linear independent component analysis", *IEEE Transactions on Signal Processing*, Vol. 54, Issue 4, pp: 1189 – 1203, 2006.

[214] Z. Shi, L. P. H. Tang, and Y. Tang, "An improved gradient learning algorithm for spatial independent component analysis of functional MRI data", Neural Information Processing -- Letters and Reviews, Vol. 5 (1), pp:1-8, 2004.

[215] M. E. Wall, R. Andreas and M. R. Luis, "Singular value decomposition and principal component analysis", *A Practical Approach to Microarray Data Analysis*, pp: 91-109, 2003.

[216] P. Grassberger, R. Hegger and H. Kantz, etc., "On noise reduction methods for chaotic data". *Chaos*, Vol. 3, pp: 127-141, 1993.

[217] N. Nemoto, K. Momose, M. Kiyosawa, et al, "Characteristics of first and second order kernels of Visually Evoked Potentials elicited by pseudorandom stimulation", *Documenta Ophthalmologica,* 108: 157-163, 2004.

[218] K. Momose, "Evaluation of an eye gaze point detection method using VEP elicited by mult-pseudorandom stimulation for brain computer interface", *Proceedings of the 29th annual International conference of the IEEE EMBS*, 2007.