EEG PATTERN CLASSIFICATION FOR THE BRAIN-COMPUTER MUSICAL INTERFACE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
ELECTRONICS AND ELECTRICAL ENGINEERING
OF GLASGOW UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Alexander A Duncan
31st January 2001

Abstract

Building on the work of Rosenboom, who created numerous *interactive musical environments* that utilised human brainwaves [Ros90], this thesis extends the concept of the *brain-computer interface* by defining the concept of a brain-computer musical interface (BCMI), described concisely as:

> "A musical synthesis device that uses the knowledge of the *presence* or *absence* of certain musical thoughts or experiences, by means of a *brain-computer interface and EEG analysis system*, so as to allow thought-control of the music that is subsequently created."

Developing BCMI systems requires a fusion of genres including the arts, neurosciences, and engineering. This thesis makes a practical contribution towards the development of such *thought-controlled musical devices* by evaluating a number of *EEG pattern classification* techniques. In particular, it is concerned with the critical issue of identifying patterns in the EEG that correspond to the kind of musical tasks or experiences of relevance to the hypothetical BCMI. In this respect, the degree of success achieved acts to confirm that the BCMI is, in principle, no longer an aspect of fiction, rather an opportunity waiting to be realised.

An iterative procedure of *hypothetical BCMI application design, experimental design and implementation*, and *data analysis*, is the means by which this research has been evolved. To this end, three novel experiments are designed and implemented, each of which contributes to the making of a working BCMI development environment. The first experiment, based on the classification of event-related potentials (ERPs) resulting from the auditory stimulus of simple tones heard over silence, demonstrates that successful classification of single segments of pre and post-stimulus onset EEG is possible. This is achieved by means of a novel correlation-based feature extraction technique in combination with a multilayer perceptron neural network classifier. Three subjects are tested, yielding average classification accuracies of 84.7%, 80.8%, 91.8% respectively. Most importantly, the experiment shows that the EEG contains

information concerning the experience of music - a pivotal requirement for the realisation of any BCMI. Two further experiments, involving the mental tasks *musical imagery* and *musical focusing*, boast positive results in favour of BCMI systems that utilise discrete musically relevant mental tasks. Classification results in the order of 70% to 80% are achieved for most subjects.

A structured classification framework is adopted that incorporates the following sub-stages, within which a number of options are varied:

- *Pre-processing:* Raw representation (i.e. no pre-processing), average referencing, Hjorth's Laplacian spatial filter, low pass filtering.

- *Feature extraction:* Linear autoregressive model coefficients, autoregressive model order estimate, binned fast Fourier transform and estimated power spectral density coefficients, and a novel correlation-based detector.

- *Feature selection:* Varied number of EEG channels used. Of the 128 channels available, subsets of 4, 18 and 92 channels are tested.

- *Nonlinear Classification*: Generalised linear models and single hidden-layer static multilayer perceptron neural networks. Compared with the (linear) Fisher discriminant, these are shown to offer a higher performance.

Optimal strategies correlate with findings from BCI research, in particular, the success of Laplace spatial filtering for pre-processing raw EEG data, linear autoregressive modelling for feature extraction, and static feedforward multilayer perceptron neural networks for classification.

Due to the novelty of this research, further experimentation involving tasks such as musical imagery and focusing could be useful for validation purposes, and as a means of testing new methods for future BCMI applications. New experiments need to be implemented that attempt to mimic real-world, on-line environments. Efforts should be made towards reducing the number of channels required to achieve suitable classification accuracies, techniques such as *committee networks* and *multiple segment averaging* are worth evaluating for this purpose. Adaptive classification techniques such as *hidden Markov models* might be necessary as experiments show that the underlying statistical properties of EEG data may change during the operational time frame of a prospective BCMI system, and thus may contaminate the data if treated with off-line or non-adaptive learning mechanisms.

Acknowledgements

Special thanks to my joint supervisors, Dr. Eduardo Reck Miranda, for his early idea-forming efforts as joint supervisor to this Ph.D. project, and Dr. Kerry Kilborn for his generous help in all matters psychological, experimental and motivational. Thanks to Dr. Ken Sharman, my original supervisor, for authorising such an unusual Ph.D. project. Finally, an extra special thanks to Dr. Yun Li, who has acted wonderfully as primary supervisor since Dr. Sharman's departure.

Various members of staff have at one point or another come into contact with this Ph.D. project and in some way offered help. These people include Professor John Barker, Dr. H. Morgan, Professor S. Beaumont, Professor John Cooper, Tom O'Hara, and no doubt others.

I would also like to thank various colleagues who have assisted me in one way or another. These people include Craig Slorach, Richard McAleer and Steve Fulton.

From further afield, I wish to thank those people who have engaged in discussions with myself via emails or by phone. Especially: Dr. Will Penny, Oxford University. Dr Colwyn Trevarthen, Edinburgh University. Mireille Besson, University of Marseille. Dick Dafters, University of Glasgow. Petr Janata, University of Oregon. Jim Kay, University of Glasgow.

Thanks also to Dave Peterson and Chuck Anderson of Colorado State University for making me welcome during my overseas visit.

Finally, thanks to Sharon and Skye for just being themselves, and to the rest of my family who have 'endured' me during the ups and downs.

Glossary

AR Autoregressive Model. Representing the linear AR method of feature extraction. 6th order model coefficients used implying 6-valued feature set.

ARMO Autoregressive Model Order. Representing the optimal (estimated) AR model order feature extraction method. Single valued feature set.

AVR Average Reference. Representing the average reference pre-processing method.

BCI Brain Computer Interface.

BCMI Brain Computer Musical Interface. Representing the concept of a thought-controlled musical device as explained in this thesis.

CF Classifier. Represents the parameter for the particular classifier options used in the classification strategy. FISHER (Fisher discriminant). GLM (General linear model). MLP (Multilayer perceptron).

CND Condition Combination. Represents the parameter for the particular combination of conditions being classified. I (Imagery). F (Focusing). FL (Focusing left). FR (Focusing right). R (Relaxing - passive listening). C (Counting).

COR Representing the correlation detector feature extraction method used in the auditory stimulus experiment. Single valued feature set..

DFT Discrete Fourier Transform.

DSP Digital Signal Processing.

EEG Electroencephalogram.

Epoch A period of time, in this case, relating to a segment of EEG.

ERP Event Related Potential.

FFT Fast Fourier Transform. Representing the fast Fourier transform feature extraction method. 5 binned frequency ranges resulting in a 5-valued feature set.

FISHER Fisher Discriminant. Representing the linear Fisher Discriminant classifier method

FS Feature Selection. Represents the parameter for the particular feature selection options used in the classification strategy. FS1 (4 Temporal channels). FS2 (International 10-20 montage). FS3 (128 channels less perimeter electrodes).

FX Feature Extraction. Represents the parameter for the particular feature extraction method used in the classification strategy. AR (6th order linear autoregressive model coefficients). ARMO (Estimated optimal AR model order). FFT (Binned absolute fast Fourier transform coefficients). PSD (FFT squared). COR (correlation detector method).

GLM Generalised Linear Model. Representing the generalised linear model neural network classifier method.

ISI Inter-Stimulus-Interval.

LPF Low Pass Filter. Representing the low pass filter pre-processing method.

MLP Multilayer Perceptron. Representing the multilayer perceptron neural network classifier method.

Montage A specific set of EEG electrode locations on the head.

NE Number of Epochs. Representing the number of epochs used whilst training an MLP.

NES Number of Training Set Patterns. Determined by the Split Ration (SR) and the size of the complete available data set (NDS). Also used: N_{ES}.

NH Number of Hidden Units. Representing the number of hidden units in an MLP. Also used N_{HU}.

NI Number of Input Units. Representing the number of input units in an MLP. Also used N_{IU}.

NONE Representing the no-pre-processing method, I.e. raw EEG data.

NP Number of (Bootstrap) Permutations. Representing the number of times that a strategy is initialised (new ES and TS and new classifier weights) and trained in order to obtain an average strategy fitness with a greater degree of confidence.

NTS Number of Test Set Patterns. Determined by the NES and the size of the complete available data set (NDS). NTS = NDS - NES.

P99 Probability that a chance classifier could have performed as well as or better than the given strategy. The 99 indicates that the goal performance used to calculate the probability of chance was the average strategy fitness (for each pattern in the test set) less the 99

PP Pre-processing. Represents the parameter for the particular pre-processing method used in the classification strategy. NONE (Raw mean corrected EEG). AVR (Average reference filter). SPF (Laplace spatial filter). LPF (Low pass filter).

PS Power Spectral Density. Representing the power spectral density feature extraction method. 5 binned frequency ranges (same as FFT) resulting in a 5-valued feature set.

SPF Spatial Filter. Representing the Laplace filter pre-processing method.

SR Split Ratio. Patterns (cases) are picked at random from the complete data set (DS) according to a split ratio SR = NES:NTS. This is performed for every cycle of the bootstrapping process.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the key features that distinguish humans from other animals is the fact that we are intrinsically musical. Music is generally associated with the expression of emotions [HH77], but it is also common sense that the intellect plays an important role in musical activities [Deu77]. All the same, music appreciation requires the ability to recognise and imagine patterns of sounds, it requires sophisticated memory mechanisms, involving both conscious manipulation of concepts and subconscious access to millions of networked neurological bonds [Mir97]. Countless studies have been undertaken which address these ideas, some of which have already been mentioned.

In the late 1960's, inspired by a fusion of ideas from the fields of brain science and biofeedback[1], one person, Rosenboom[2], began a life work which, in the author's opinion, has been the most comprehensive attempt to date at harnessing the musical potential of the EEG in a creative and artistic way [Ros90]. He developed a variety of EEG based musical interfaces and associated compositional and performance environments that utilised the latest EEG analysis and interpretation techniques. In particular, use was made of the fact that certain aspects of a person's musical experience, such as their level of surprise related to the perception of a rare musical

[1]The term 'biofeedback' refers to the act of self-regulation, that one may be able to achieve a degree of conscious, wilful control of particular body functions formerly thought only to be regulated by unconscious, autonomic processes. Biofeedback is most popularly associated with its therapeutic application, where a patient learns to control some body function, or biosignal, such as their EEG, in order to alleviate symptoms of illness. Here, however, it is used in its broadest sense, corresponding to the 'type (2)' BCI described below.

[2]David Rosenboom, School of Music, Center for Experiments in Art, Information and Technology, California Institute of the Arts, 24700 McBean Parkway, Valencia, CA 91355, U.S.A. E-mail: david@music.calarts.edu. Web site: http://music.calarts.edu/~david.

event, can be inferred from a transient EEG component known as the event-related-potential (ERP). He concludes his report by describing an EEG based musical interface/synthesis system that would be able to make inferences about certain aspects of one's musical experience from the EEG, then put this information to use in a live musical-biofeedback environment. Rosenboom's enthusiasm was evident, as he writes (of this system) [Ros90],

> *"...this goal is immediately achievable with existing and affordable technology. All that is required is the vision, support and time to realise it."*

Inspired by Rosenboom's work, this thesis is concerned with exploring new areas of research and developing the concept of a *brain computer musical interface* (BCMI) - a hypothetical *thought-controlled musical device* that would infer knowledge about a performer's musical experience, by analysing his electroencephalogram (EEG), then use the knowledge to control or influence the music he subsequently hears. An embellished description of this concept follows some background information on *EEG pattern classification*, which, according to the author, is the critical area in need of investigation [DMS98a].

1.1 EEG pattern classification

In this section a succinct introduction to the electroencephalogram (EEG) is followed by an overview of those aspects of its analysis that are relevant to this thesis. For a full treatment of the EEG in today's rich and varied field, refer to a good text book, such as [NLDS98].

1.1.1 The electroencephalogram

> *"The human brain produces a complex, multidimensional, pulsating, electromagnetic field resulting from the electrochemical behaviour of masses of neurons acting in small to very large groups"* [Ros90].

According to Rosenboom [Ros90], EEG data can been categorised into four main components: a random-seeming background signal, long-term coherent waves, short-term transient waves, and complex ongoing waves. The random-seeming background signal is the residue observed after all known methods of waveform decomposition are exhausted, very little is known about this signal. Long-term coherent waves are the well-known alpha, beta, delta, and theta rhythms, which range from approximately 1 to 30 Hertz. They are often associated with certain states of consciousness, such as alertness and sleep. Short-term transient waves reflect the 'singular experience' associated with an external stimulus and up to now they have been accessible only by event-related-potential (ERP) analysis (discussed in Chapter 2). Finally, it is suggested that a non-random complex component exists, whose ever changing pattern comes from the build-up of baseline activations from the vast neuronal masses within the brain. This pattern is expected to be the result of the ongoing, self-organisation of information during a person's own life experience. If these patterns could be successfully measured, and sense made of them, one might witness the mechanisms of higher level thought processes.

From an engineering perspective there are two distinct areas of EEG analysis:

- *Event related potentials (ERP)* (or evoked potentials) which focus on short lived components within the ongoing EEG, specific to some event, usually the result of sensory stimulus.

- *spontaneous-EEG* which looks at the ongoing EEG for patterns or trends that correspond to certain 'brain states' of interest.

These two ways of treating the EEG are discussed in some detail in Chapter 2. In both cases however, there is the need to be able to discriminant between complex (often multi-channel) sets of EEG data belonging to a number of different *classes* [CDA94a].

1.1.2 Classification methods

In this thesis, a number of EEG experiments are designed and implemented, all of which supply segments of multi-channel EEG data belonging to a number of classes. These classes correspond to specific *conditions* or *tasks* relating to an aspect of the subjects' experience of music. These experiments have been designed with one aim in mind, that is, to find a number of classifiable tasks or conditions relating to aspects of musical experience that could be utilised in a BCMI system. Since the success of such a search is both dependent on the tasks or conditions *and* the methods used to classify the EEG, the search also incorporates a selective evaluation of state-of-the-art classification methods. Some of these are discussed below.

Power spectral analysis

The reason for analysing the EEG in the frequency domain is the hope that certain patterns will emerge from the features extracted from its power spectral density (PSD) and that these patterns correlate to the conditions that are being tested for.

Typical features that are extracted from the PSD include the dominant frequency, mean power, and of course individual frequency powers. These values are often analysed visually by plotting them on a map of the head, or by coherence measures - the correlation between values at different locations on the head.

Some authors have opted to split the PSD into several bands that correspond to the popular EEG frequency bands. Janata [JP93], for example, chose the following bands: delta (1.5-3.5 Hz), theta (4.0-7.5 Hz), alpha (8.0-12.5 Hz), beta1 (13.0-18.0 Hz), beta2 (18.5-24.0 Hz), and beta3 (24.5-31.5 Hz). Other people have opted to consider a greater number of frequency bands, and then make a selection on which bands (or other features) are based on some statistical measure, such as principal components analysis (PCA) [JMS97].

The most common method of generating a PSD is by using the Fast Fourier Transform (FFT) algorithm, which is based on the premise that any signal can be broken down into a number of sinusoids.

Despite its popularity, the FFT has two limitations that are of consequence to the analysis of EEG signals:

- The FFT has a poor spectral resolution for signals of finite length (i.e. digital signals), especially when the number of samples is small [Roa96].
- It deals poorly with signals that are of a short transient nature, and surrounded by noise [Roa96].

However, for many of the studies encountered in this literature survey, the FFT has been adequate since the duration of the epochs of EEG have been sufficient to yield a reasonable resolution.

It has only been in cases where the epoch duration has been required to be considerably shorter, that alternative methods have been employed. For instance, Saiwaki [SKI97] used an Auto-regressive (AR) model to produce the PSD of short EEG epochs, while Jung [JMS97] used the moving average technique, ARMA. It is probable that these methods might be useful in this work as the aim is to work towards the instantaneous detection and classification of musical thought patterns, which will require the EEG to be broken into reasonably short epochs (tens of milliseconds).

Correlation and coherence analysis

Coherence analysis and general measures of correlation between measures of PSD, or ERP amplitude at different sites on the scalp, are common.

Petsche performed a series of experiments where individuals were given a variety of creative mental tasks to complete [Pet96]. 5-minute epochs of EEG were obtained for the tasks, as well as epochs for a resting state, where the subjects just relaxed. These epochs were then converted to PSDs and split into the six bands. Statistically significant differences between the correlations of the resting EEG and task EEGs were plotted on maps of the scalp. He concludes that EEG patterns do change in comparison to resting EEG whilst performing creative tasks, and that the upper alpha band (12Hz) seems to reflect individual features apart from the group.

Using a coherence technique similar to Petsche, Janata [JP93] shows that it can be applied to subjects who perform listening tests where they hear a variety of musical resolutions (cadences) of varying dissonance. He identified that the electrodes placed near the auditory sites and right frontal and parietal regions were most likely to show significant differences in coherence.

Artificial neural networks

In their description of biosignal classification methods, Ciaccio *et al.,* [CDA94a] indicate that artificial neural networks (ANNs) are particularly suitable when little is known about the signals to be classified.

An ANN may be described as a statistical model of a real world problem that has a network structure built around several layers of interconnected processing units, commonly referred to as *neurons* [Gur97]. The tuneable parameters of the model, *weights*, represent the strength of the connections between neurons. These weights are adjusted during a training period over which a sequence of known input-output vectors are presented to the network until the error between the actual output and the desired output reaches an acceptable level.

In this way, a neural network could be tuned to perform pattern classification [Swi96], where the inputs are the features that have been chosen to represent the patterns extracted from the EEG data, and the outputs are a set of classes that correspond to the patterns under classification. Referring to all the possible values of a set of parameters as its space, the neural network makes a mapping from feature space to classification space. Some examples of EEG pattern classification using ANNs are given below.

- William Penny and Steve Roberts (Oxford University) are working on a brain-computer-interface related problem where the aim is to recognise imagined hand movements [PRCS00, RP00]. They claim to be able to achieve classification accuracies in the order of 80% in on-line trials. Among the techniques explored are Bayesian neural networks and a linear discriminant classifier [RP00].
- Jung et al., [JMS97] used a feed-forward ANN to estimate the level of alertness of subjects during target detection experiments. The ANN was trained using a set of features extracted from two channels of EEG. Principal component analysis (PCA) was used as a means of reducing the pattern space, which was an 81 point PSD derived from ongoing EEG.
- In the clinical field, Weng et al., [WK96] confirmed the effectiveness of ANNs as a quantitative EEG analysis tool by performing pattern recognition of epileptic seizures. 16 channels of EEG, recorded at 200 Hz, were segmented into 2.56 second epochs, then FFT'd to give the PSD. The mean power and dominant power were used as the main features for the ANN. Galicki et al., [GWD+97] employed a similar strategy for pattern recognition of burst-interburst cycles of neonatal children using only two channels of EEG.

1.2 Brain-computer musical interface

1.2.1 Brain-computer interfaces

One particular area where EEG pattern classification plays a central part is the field of brain computer interfaces (BCI). Brain computer interfaces (BCIs) provide a new way for people to interact with the world, via computer systems that interpret their EEG. BCI research began in the early 1970's [Mul73], yet only recently has the field become established, as witnessed by the first international meeting of "Brain-Computer Interface Technology: Theory and Practice, New York State Department of Health, June 1999." For a review of the state-of-the-art in today's fast-moving BCI field, refer to the work of Peters et al. [PPF97].

Broadly speaking, BCI systems can be divided into three operational categories relating to adaptation:

1. *Computer adapts to the user.* Metaphorically speaking, these systems attempt to read the mind of the user. For example, Anderson *et al.* [AS96] classify EEG patterns associated with specific mental tasks, such as mental arithmetic, letter writing and object rotation. The idea being that a disabled person could control a wheel chair by performing the appropriate mental task. These systems can be developed off-line.

2. *User adapts to the computer.* These systems utilise the users' capacity to learn to control certain aspects of their EEG, affording them the ability to exert some control of events in their environment. Examples are cursor movement control [WMNF91], or simple selection devices such as Birbaumer's letter selection system for the disabled [BGH+99].

3. *Both the computer and user adapt.* Combining the functionality of (1) and (2), these are systems that allow biofeedback and adaptation. For example, the combined use of mental task pattern classification and biofeedback assisted online learning allow the computer and the user to adapt. Prototype cursor control systems have been developed in this fashion [PPF97, PRCS00].

To date, efforts have been aimed largely at developing ways to help severely disabled people communicate via computer systems. However, little has been undertaken that combines state-of-the-art BCI technology with experimental musical applications, an area that in principle should be possible by current standards [Ros90]. This is the topic matter of this thesis, as will be explained in the remainder of this chapter.

1.2.2 New BCMIs

Saiwaki et al. [SKI97] introduce the concept that a human brain listening to music can be thought of as a system where the input is a sound, and that the recognition of music is dependent on the many subsystems of the brain operating in co-operation. The output, the EEG, is hoped to represent this internal functioning. This general theory – that aspects of our musical experience might be reflected in our EEG – is the underlying thread that supports the creative idea of this thesis.

Imagine if it were possible to make music burst forth around you by merely imagining a tune in your head. Crazy though it may seem, this could be the way people perform music in the 21st century – with a BCI for musical applications, or *brain computer musical interface* (BCMI). However, this idea is not unique, in fact, it has been clearly hinted at by the impressive work of Rosenboom [Ros90], who developed several experimental musical environments for composition and performance that incorporate EEG-biofeedback. In these cases, music is controlled, or steered, by the user who learns to 'will' their EEG in a certain way (a category (2) BCI). Other examples of similar approaches can be found in [LKL94, Oki95, Roa96, FH98].

Rosenboom [Ros90] concludes by describing an expert biofeedback system that is capable of inferring, from a performer's EEG, certain things about their experience whilst immersed in a musical environment. In doing this, it could direct the manner in which the music evolves in a way that reflects the performer's response to the previous parts of the performance. Such a system, as depicted in Figure 1.1, can be conceptualised as the interaction between performer / participant, and three functional sub-systems: a *music engine*, *EEG analysis engine* and *co-ordinator*. These are described below:

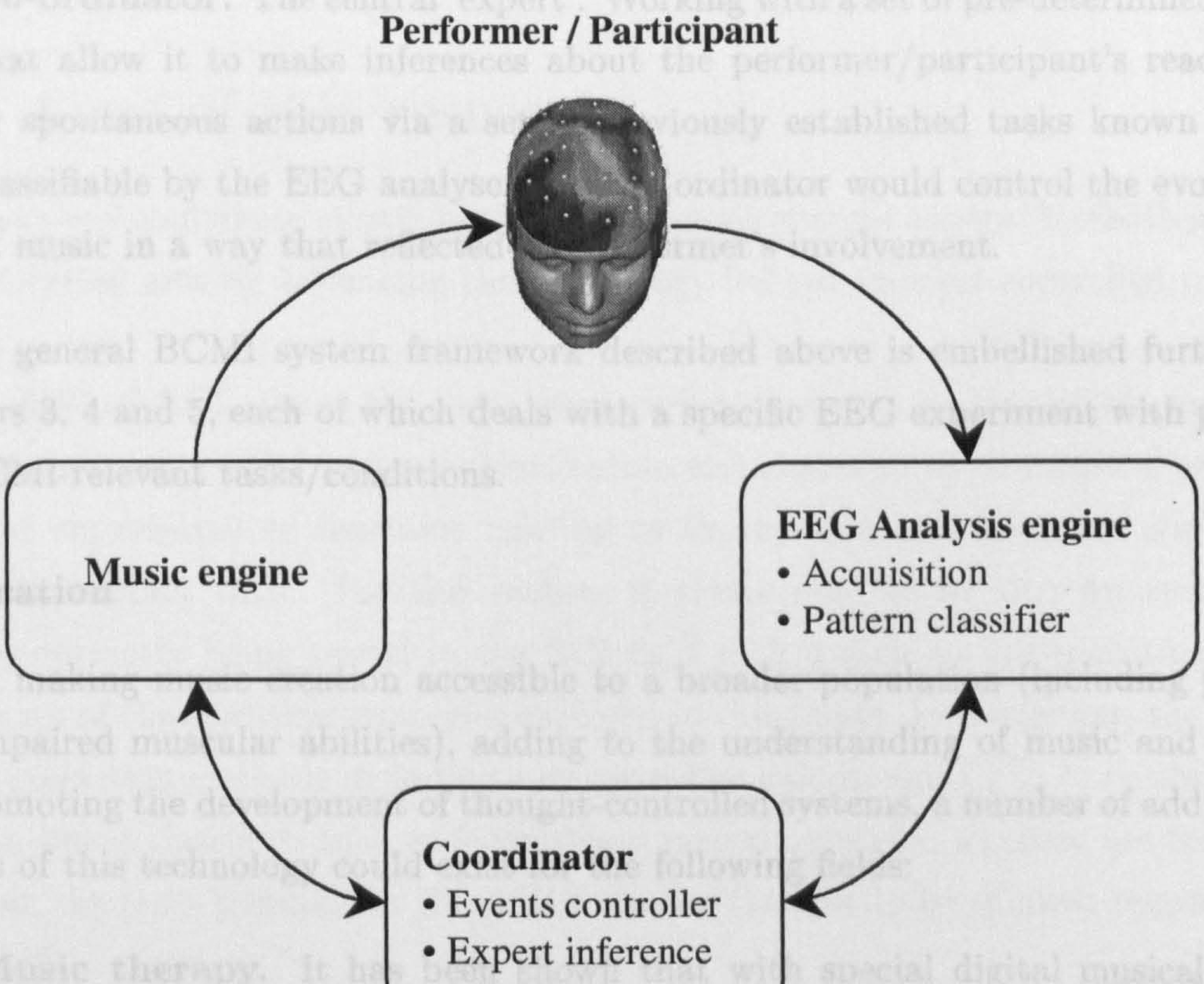

Figure 1.1: Illustration of the general BCMI concept comprising a *performer*, *EEG analysis engine*, *co-ordinator* and *music engine*.

- **Music engine.** Responsible for generating the music according to the instructions coming from the co-ordinator. This could be realised as an ongoing score generation and music synthesis system, or more simply, a sequencer / sampler that had a data base of loopable parts, the auditioning of which would be determined by the co-ordinator.

- **EEG analysis engine.** Incorporating EEG acquisition and pattern classification. Capable of classifying numerous mental tasks and musical experiences or states of consciousness, the nature of which has been developed and tested in works such as this thesis. This sub-system, working closely with the co-ordinator, would be selectively directed as to what type of classification to perform, depending on the current state of the music engine, and on previous inferences made by the co-ordinator.

- **Co-ordinator.** The central 'expert'. Working with a set of pre-determined rules that allow it to make inferences about the performer/participant's reactions, or spontaneous actions via a set of previously established tasks known to be classifiable by the EEG analyser. The co-ordinator would control the evolution of music in a way that reflected the performer's involvement.

The general BCMI system framework described above is embellished further in Chapters 3, 4 and 5, each of which deals with a specific EEG experiment with particular BCMI-relevant tasks/conditions.

Justification

Besides making music creation accessible to a broader population (including people with impaired muscular abilities), adding to the understanding of music and mind, and promoting the development of thought-controlled systems, a number of additional benefits of this technology could exist for the following fields:

- **Music therapy.** It has been shown that with special digital musical input devices, the well-being of people with special needs, especially handicapped children, can be cultivated. For example, Soundbeam, a system that allows the slightest movement of the body to play a synthesiser via MIDI is one such device in use today [Sou]. As mentioned above, this work could lead to numerous BCMI systems which could be tailored for therapeutic purposes.

- **Biofeedback therapy.** An area of therapy claiming to assist people with various chronic illnesses. It has been receiving more attention as technology has become more sophisticated and affordable [RL97]. Any progress in the fields of EEG analysis and BCI will be of benefit to this field. As for music therapy, BCMI systems tailored to therapeutic purposes could be of merit in this setting also.

- **Clinical diagnosis.** The EEG analysis techniques explored in this thesis could profit the field of clinical diagnosis where quantitative EEG analysis is becoming increasingly popular as a diagnostic tool. Examples include: epileptic seizure detection [WK96], diagnosis of Alzheimer's disease [HPGM95], serotonin deficiency research [HJ93], MDMA ("Ecstasy") use [DDO98], and memory impairment in multiple sclerosis [PGH97]. For further details, see the article by Eric Fottorino published in Le Monde [Fot98].

1.3 Aims of this research

This thesis will contribute mostly to the fields of experimental musical biofeedback and the performing arts by advancing the technology behind thought-controlled musical devices.

The field of BCI shares the same engineering challenge as the BCMI, namely, the need to develop EEG pattern classification techniques so as to utilise a person's thoughts, experiences or reactions relating to the environment they are aiming to control or interact with. For this reason, it seems sensible to look for successful methods currently being tested in the BCI field with a view to incorporating them into the set of classification strategies explored in this thesis. In particular, the use of feature extraction methods including autoregressive models and FFT, in combination with non-linear classifiers such as feedforward neural networks, as these are found to be among the most popular for those BCI studies thought to be of most relevance to the BCMI.

The work described in this thesis addresses what the author believes are the key problem areas inherent to the development of new BCMI systems:

1. Expanding the concept of the BCMI - a thought-controlled musical instrument/environment - by combining the ideas of computer music research with the state-of-the-art classification and experimental methods of the BCI field.

2. Finding pattern classification methods that estimate, to a reasonable degree of accuracy[3], the probability that a segment of EEG belongs to one of a number of classes, where each class relates to a particular BCMI musical task or condition.

3. Developing experiments that both provide EEG data of a suitable nature so as to allow for the off-line evaluation of a number of EEG classification methods, and attempt to account for the fact that, ultimately, the BCMI system for which they are intended would be required to operate in real-time.

4. Developing a complete 'BCMI evaluation system' that provides a systematic way to achieve the above.

The efforts made towards addressing the above problems form the main contributions of this thesis, which are fully detailed in Chapter 6. Note that, where possible, efforts are made to backup the decisions made during this research with relevant references. However, due to the novelty of this work, as well as the scarcity of similar research, there are many cases where referencing is not possible. Much of the ground work, including the development of the BCMI concept, the experimental paradigms, and the choice of EEG pattern classification methods, are the result of numerous communications (and subsequent assessments and refinements made by the author) with experts in the fields of psychology, musical-psychology, statistics, computer music, and DSP-engineering. These people have been mentioned in the Acknowledgements section.

1.4 Outline

Chapter 2 deals with an account of the EEG pattern classification methods developed and evaluated in this thesis. These methods, drawn mostly from the BCI field, are arranged and applied in a standard off-line pattern classification strategy [CDA93a], namely: *data-acquisition, pre-processing, feature extraction, feature selection, data*

[3]This is a somewhat ambiguous definition as it makes no explicit mention of the classification accuracy sought after. Reasons for this are that there are no previous studies of this exact nature, and therefore no benchmark exists for comparison. This work, unlike many engineering theses, is not an attempt to improve on what has gone before. That is not to say that similar problems have not been tackled - on the contrary - work in the BCI field supports this work and offers the next best thing to a working benchmark. It has been decided to compare classification results to the worst case, that is the 'chance classifier' (described in Chapter 2). Emphasis is directed at finding the best combination of the methods investigated, keeping in mind the restraints imposed by a potential BCMI device.

preparation and classification. This collation and organisation of methods forms the second contribution of this thesis. Chapters 3 - 5 describe three experiments, all of which have been designed with the BCMI concept in mind. The first of these experiments (Chapter 3) involves the musical experience of perception, and is based on the analysis of *ERP* data using a novel correlation-based feature extraction technique, the third contribution of this thesis. Chapters 4 and 5 present a different type of classification problem based on the discrimination of different mental tasks imbedded in semi-spontaneous-EEG data. Both experiments have required the design of new paradigms for assessing specific mental tasks for the novel BCMI systems under development. Jointly, these chapters represent a considerable contribution towards the exciting topic of this thesis. Finally, Chapter 6 draws together conclusions, and makes recommendations for future work in this area.

Chapter 2

Classification Methodology

2.1 Introduction

The EEG, often referred to as *brainwaves*, is a measurement of the voltage-difference between two or more electrodes on the surface of the scalp [Hug95]. This electrical activity is thought to be the result of large numbers of neurons, within the cortex, depolarising in synchronisation with each other. Unfortunately, these signals are naturally filtered by the fact that they must first conduct through the cerebral fluid, the bone of the skull, and the skin of the scalp, before reaching the electrodes.

Modern digital EEG recording systems have head nets that hold as many as 128 electrodes, and can sample a signal at 250Hz or more. This is more than adequate, as the literature supports the idea that all or most of the important EEG activity lies between 0 and 50 Hz [Hug95].

This section introduces some top-down considerations relating to the classification methodology. This is followed by an overview of the main EEG pattern classification methodology which is employed as part of the *systemised BCMI evaluation protocol*, to data from three BCMI-relevant experiments (see Chapters 3 - 6). The remainder of the chapter presents a detailed description of the DSP methods used in this thesis, along with reasoning as to why they are used.

2.1.1 Off-line approach

Although the concept of the BCMI necessitates real-time analysis of EEG data, the approach taken here is to evaluate DSP methods off-line. The main reasons for this are as follows:

1. Many DSP methods can be evaluated in parallel using pre-recorded EEG data that represents the kind of data expected from an on-line setting. This is a particular advantage since very little is known about which DSP methods, if any, are capable of performing EEG pattern classification of this nature.

2. Methodology comparisons are based on a constant data set, and are therefore open to future comparisons using improved methods.

3. Off-line analysis requires considerably less computational power and a simpler implementation, allowing more emphasis to be placed on the investigation of DSP methods.

However, it is important to keep in mind that the BCMI concept is on-line by nature. Therefore, effort has been made to design EEG experiments that take this into account. For example, experiments and analysis methods have been designed to work with the classification of up to 2-second EEG segments. This time frame, although slow by some BCI standards [ADS95], is felt to be reasonable in the musical context of the BCMI. This is because the melodic aspects of a piece of musical often last several seconds before changing or stopping.

Secondly, although less important in the early stages of experimentation, some thought is given to the fact that the complexity of the methods might impose practical constraints in a real-time environment. In particular, the 128-channel clinical EEG system used to gather data for this thesis, would be unsuitable for a mass produced, portable, or widely accessible BCMI system, as these clinical devices are relatively expensive and cumbersome. As mentioned in Chapter 1, the ideal would be to use a system that only requires a handful of EEG channels, in conjunction with a portable or desktop PC. Such systems exist in the field of BCI, for example, Birbaumer *et all's* spelling device for the paralysed [BGH+99]. To accommodate for this, the issue of minimal complexity is kept in mind throughout the thesis by way of including an evaluation of the DSP methods that use subsets of the 128 channels available, so as to mimic a smaller EEG acquisition system. Nevertheless, the complex (or impractical to implement in real-time by current standards) approach is not ignored since the focus of this thesis is mainly concerned with proof of the BCMI concept, as opposed to finding direct engineering solutions.

2.1.2 Single subjects

It is often necessary to engineer an EEG pattern classifier that is able to perform reliably on unseen data from new individuals, having 'learned' how to classify from a sample population of representative individuals [Swi96]. This is particularly relevant in the field of clinical diagnosis where the problem is often to discriminate between the EEG of a healthy person and that of someone who is unwell [HJ93]. In this case, the very nature of the problem requires that the system is trained on a known data set (derived from a sample of 'normal' and unhealthy individuals), and is capable of generalisation across the population. With a BCI based problem, there is not the same need to generalise across a population of individual people - although it would be an admirable system that could do so. Rather, it must be able to learn to classify new instances of EEG from a single subject, within certain operational constraints. Such constraints might include the environment the device is to be used in, or the mental tasks it has been trained to operate with.

The pattern classification methods described in this thesis are only applied to single, as opposed to multiple subjects[1]. This means that the data used for training the classification systems (since they incorporate neural networks), and for evaluation, is always limited to that of one subject at a time.

In the field of EEG pattern classification, there are two types, or rather, two ways of treating the EEG, namely, *event-related potentials* (ERPs) - also known as *evoked potentials*, and *spontaneous-EEG*. These are described below.

2.1.3 Evoked potentials

Defined as "an electrical peak [in the EEG] related to a particular stimulus." [Car98]. More generally, it is a transient signal that forms part of the ongoing EEG that relates to the brain's activity whilst processing some stimulus or other discrete event. If the occurrence of the event is clearly defined, such as the onset of a sound, then the ERP is assumed to be time and phase-locked to the event onset [MEJS00]. Extracting ERPs from the EEG is desirable, as their characteristics can lead to insights about how the brain functions, or even the nature of our immediate experience, perceptual or otherwise [NLDS98]. However, as well as the ERP signal, there are other non related signal components, treated as noise, which make observing the ERP in the

[1] A single subject approach was adopted for 2 reasons. Firstly, limited human and computational resources being especially significant as the project was the first of its kind in the University. Secondly, it was assumed that inter-subject variation would add to the complexity of the classification problem.

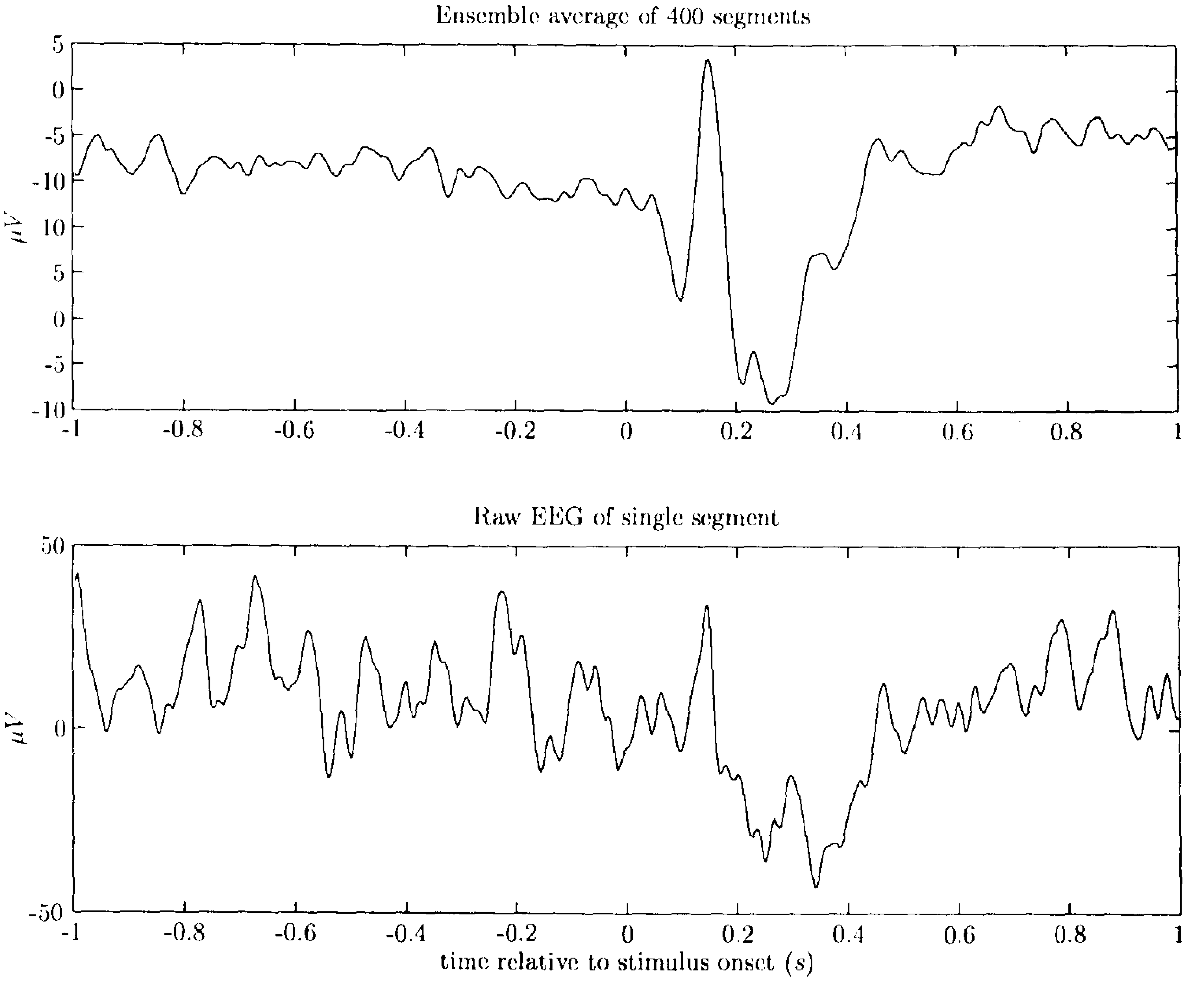

Figure 2.1: Example plots of a raw EEG segment containing an ERP signal and the ensemble average signal of 100 time-locked segments.

raw EEG very difficult. The most common method for extracting ERPs, used mostly in the fields of psychology and neuroscience, is to record the EEG, whilst presenting the same event again and again. Afterwards, the EEG is segmented around the stimulus onset, then, an ensemble average of the time-locked segments (a.k.a *trials*) is made. This has the effect of emphasising event-related activity, whilst attenuating the noise. The result is an estimate of the ERP. Figure 2.1 gives an example of raw EEG containing an ERP (i.e. a single trial), and an ensemble average of multiple trials. The weakness of this approach is that it assumes the ERPs don't vary between trials, and that the correlation between the noise components is zero. Despite recent attempts at improved ERP detection, that is, estimating ERPs from single or very few trials [LPI95], this basic technique remains the standard approach in the majority of cases.

2.1.4 spontaneous-EEG

spontaneous-EEG analysis is concerned with longer-term qualities of the ongoing EEG, as opposed to short-term transient effects. Research fields include *clinical* *diagnosis, psychology, neuroscience, biofeedback,* and *human-computer interfacing.* Quantitative measures transform raw EEG data so that meaningful patterns, that would otherwise be missed, can be identified. Traditional techniques include *Fourier*-based spectral analysis [JMS97] and *cross-channel coherence* [Pet96], whilst more recently, methods such as *parametric modelling* [SKI97], *independent components analysis* (ICA) [MEJS00], and *chaos* [Fre98] are being explored. Techniques such as these are utilised as part of automatic pattern classification systems, where known (or more often unknown) patterns in the EEG are detected without any human, i.e. *qualitative* efforts. A brief review of some typical methods used in such systems now follows.

The general engineering challenge common to all experiments described in this thesis is to successfully predict, from a person's EEG, whether they are engaged in one of a number of music-related mental tasks, as opposed to a passive baseline task. This requires a systematic approach involving the design of novel EEG experiments, and the evaluation of suitable DSP methods, both of which must be framed in the context of the BCMI concept introduced in Chapter 1. To this end, a number of pattern classification strategies based on successful and closely related methodologies of BCI research, such as [PRCS00, PPF97, AS96, JMS97], are explored.

2.1.5 Top-down approach

The raw multi-channel-EEG data acquired for off-line analysis in this thesis represents a very large *input-space*. This is because each segment of data that is to be classified is represented by a huge number of variables. For example, a 128-channel segment lasting for 1-second, and sampled at 250 Hz, would constitute a total of $128 \times 250 = 32000$ values. The main task of pattern classification is to take such a set of variables, and map them onto a *classification-space*, thus performing a classification. Effectively, this process involves a drastic kind of feature reduction, whereby as little information is lost as possible, yet, sufficient spurious information, called *noise*, is discarded. In this way, the system is able to generalise, that is, to classify new, unseen segments of data, without being sensitive to the inevitable changes in the 'noise.' [Bis95]. The general procedure for achieving this goal is well documented in the literature. In the field of biosignal classification (which is the case here), a series of papers by Ciaccio et al., provides a good overview of this procedure, including a variety of methods and applications [CDA93a, CDA93b, CDA94a, CDA94b].

As stated above, the basic aim is to perform a mapping from input-space to classification-space. It is often the case, as is chosen here, to split this task into a number of sub-stages, namely, *pre-processing, feature extraction & selection, classification*, and *post-processing* [CDA93a]. The following sections provide a description of the classification methods evaluated as part of the *systemised BCMI evaluation protocol* successfully developed and tested in this thesis.

2.2 Pre-processing

A number of alterations including: scaling, artefact analysis, and filtering, are applied to raw multi-channel EEG data before attempting feature extraction.

2.2.1 Scaling

Consider a single-channel segment of EEG data, notated

$$\{x_c^{i,j}(t), t = 1, \dots N_t\}$$

where c indexes the complete set of channels, $C = \{1, 2, \dots, 128\}$, i indexes the complete set of segments $I = \{1, 2, \dots, N_i\}$, j indexes the complete set of classes (i.e. experimental conditions) $J = \{1, 2, \dots, N_j\}$, and t is the discrete time index. Written as a time series vector, one has, $\mathbf{x}_c^{i,j}$.

Raw data from the EEG recording equipment is stored in binary files, and once imported to MATLAB, has to be scaled to adjust for channel gain and D.C. offset. This involves a subtraction of a zero calibration constant $zero_c$, then a multiplication by a gain calibration constant $gain_c$, for each EEG sample $x_c(t)$, in other words

$$x'_c(t) = (x_c(t) - zero_c)gain_c, \tag{2.1}$$

where c denotes the channel and t the discrete time index.

2.2.2 Artefact analysis

The EEG is comprised of cortex born components (the signals of interest) and non-cortex born components, known in the literature as artefacts. EEG artefacts can be divided into two categories: (i) those derived from muscle activity of the individual subject and (ii) those due to measurement noise.

Muscle artefacts

Artefacts due to eye movement, blinking, swallowing and other spurious limb movements generate large EEG components which cover up the weaker cortex-born signals that one is interested in capturing. Figures 2.2 and 2.3 show examples of artefact contaminated multi-channel EEG. Research in the EEG pattern classification field acknowledges the fact that artefacts are problematic and must be considered. Commonly, muscle artefacts are tackled in one of three ways [NLDS98]:

1. Discard contaminated segments, detected manually or automatically.

2. Include contaminated segments in analysis, i.e. ignore artefacts.

3. Model and subtract artefacts from EEG.

In this thesis, method (2) is mostly employed, since all the experiments were 'eyes-closed' designs that resulted in a low rate of eye related artefacts. However, a brief look at artefact detection with the view to excluding contaminated data (1) is attempted, but gives poor results. In this case, automatic detection of eye-blink and eye-movement artefacts is performed by algorithms that compare fast and slow running averages of the difference between eye channel signals (based on those used by EGI's Averager software[2]). These are explained in Appendix A.

Measurement artefacts

It is inevitable that a measurement system is going to introduce some noise to the signal under observation. With EEG, this is a real concern, as it is quite a small signal (microvolt range) that requires sensitive electrical measurement equipment.

When using a commercial EEG system, the main cause of noise is *mains hum* (a 50 to 60 Hz signal due to the carrier frequency of mains electricity). In order to minimise the effect of this undesirable artefact, electrode impedances have to be kept very low, which requires diligent placement of electrodes.

With dense sensor arrays (such as the one used here) it is often the case that some sensors may be faulty, become misplaced during the experiment, or simply dry out[3].

---

[2] Visit www.egi.com for full documentation, included i which are details of the artefact detection algorithms utilised by the Averager software.

[3] The sensors in EGI's geodesic net are made from silver-silver chloride electrodes housed in small plastic casings. Contact between the scalp and electrode is made via a sponge (also in the casing). This requires that the net is soaked in an electrolyte solution (salt water) prior to use. Often, due to the quality of the subjects hair, some electrodes dry out before the experiment is complete, causing the impedance and hence measurement noise to increase.

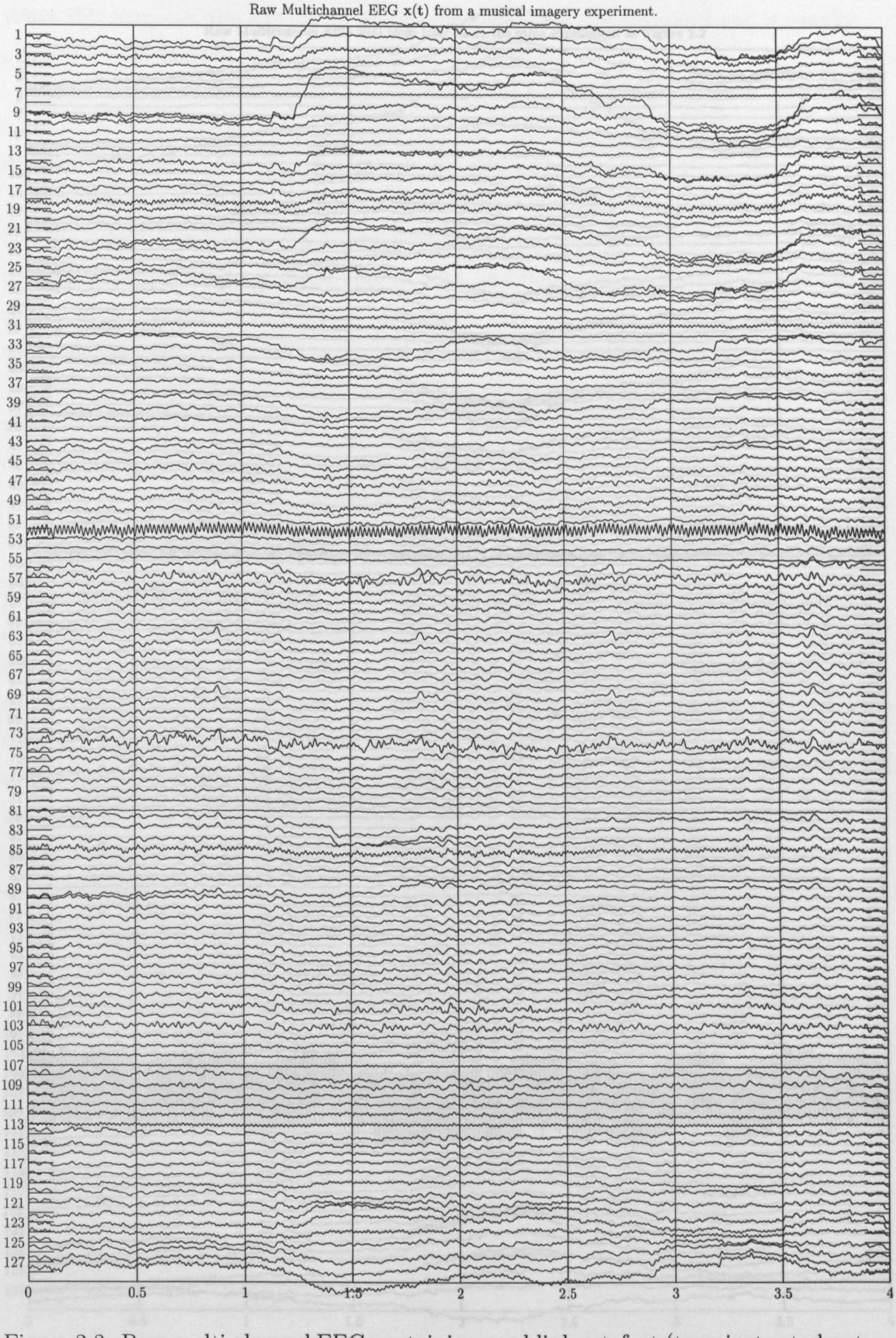

Figure 2.2: Raw multi-channel EEG containing eye-blink artefact (transients at about 1.25 seconds in channels near eyes, i.e. low channel numbers and high numbers), and a possible bad channel (52) probably due to poor electrode placement.

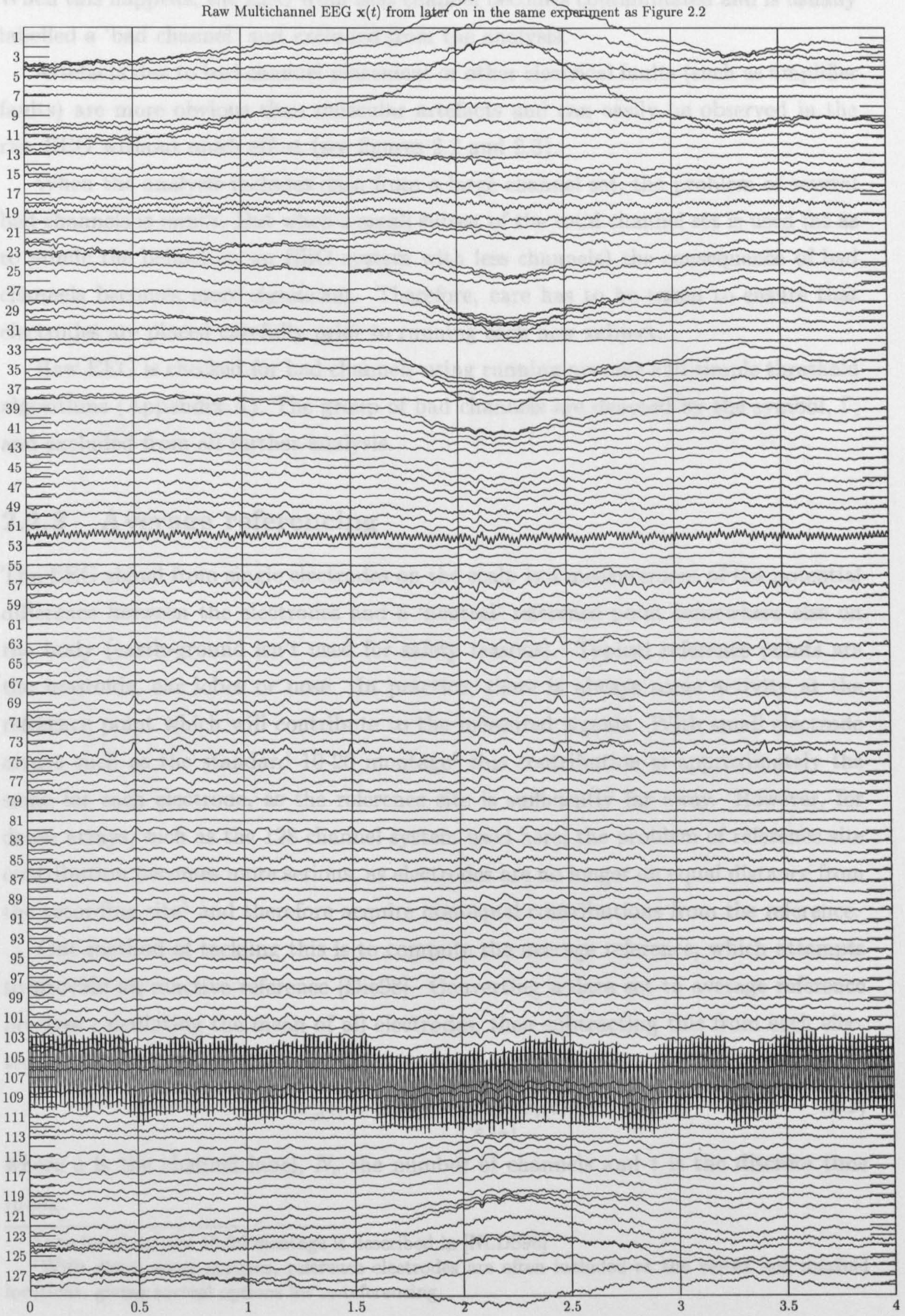

Figure 2.3: Raw multi-channel EEG containing eye-movement artefact (peaking at 2.125 secs. channels near eyes), and a bad channel (107), perhaps due to electrode becoming misplaced during experiment.

When this happens, the EEG from that channel becomes contaminated and is usually labelled a 'bad channel' and excluded from the analysis.

Artefacts due to bad channel placement or other electrical faults (such as amplifier faults) are more obvious than muscular artefacts and can easily be observed in the raw EEG without much effort (see figures 2.2 and 2.3).

When the analysis includes data from a large channel set, the problem of several bad channels is minor. But when a small subset of the total channel set is used (so as to mimic the behaviour an EEG system with less channels) the consequence of bad channels becomes more significant. Therefore, care has to be taken to ensure that electrodes are placed carefully prior to running each new subject.

Raw EEG is checked for bad channels using running-average and simple threshold algorithms (Appendix A). The group of bad channels are denoted by the symbol, Γ, and excluded from all further analysis.

2.2.3 Average referencing

The EEG signal from single electrodes on the scalp is a measurement of the potential difference between the electrodes and a 'neutral' reference point somewhere else on the body (earth-ground isn't used for safety reasons). Typical reference points are the mastoids, ear lobes or nose. In practice, there is always some activity at the reference point which will contribute to the measured signals. With small electrode arrays such as the standard 10-20 montage[4] this contribution is approximately the same for each electrode, as the reference site is sufficiently far away. However, for dense arrays, such as the 128 channel system used here, the problem of reference site contribution becomes more serious, as electrodes are no longer an equal distance from the recording site[5] and therefore acquire non-equal contributions from the reference.

One method of tackling this is to compute the average reference, which attempts to provide an inactive reference [Die98]. Converting a data set to average reference involves calculating the mean of all electrodes, then subtracting this from each electrode, in other words

$$x'_c(t) = x_c(t) - \frac{1}{N_c}\sum_{i=1}^{N_c} x_i(t), \tag{2.2}$$

where c is the channel label, N_c the number of channels and t is the discrete time index.

[4] The international 10-20 montage is described in [NLDS98]

[5] With dense array systems, reference electrodes are often included at the vertex and mastoid locations, giving several options for re-referencing.

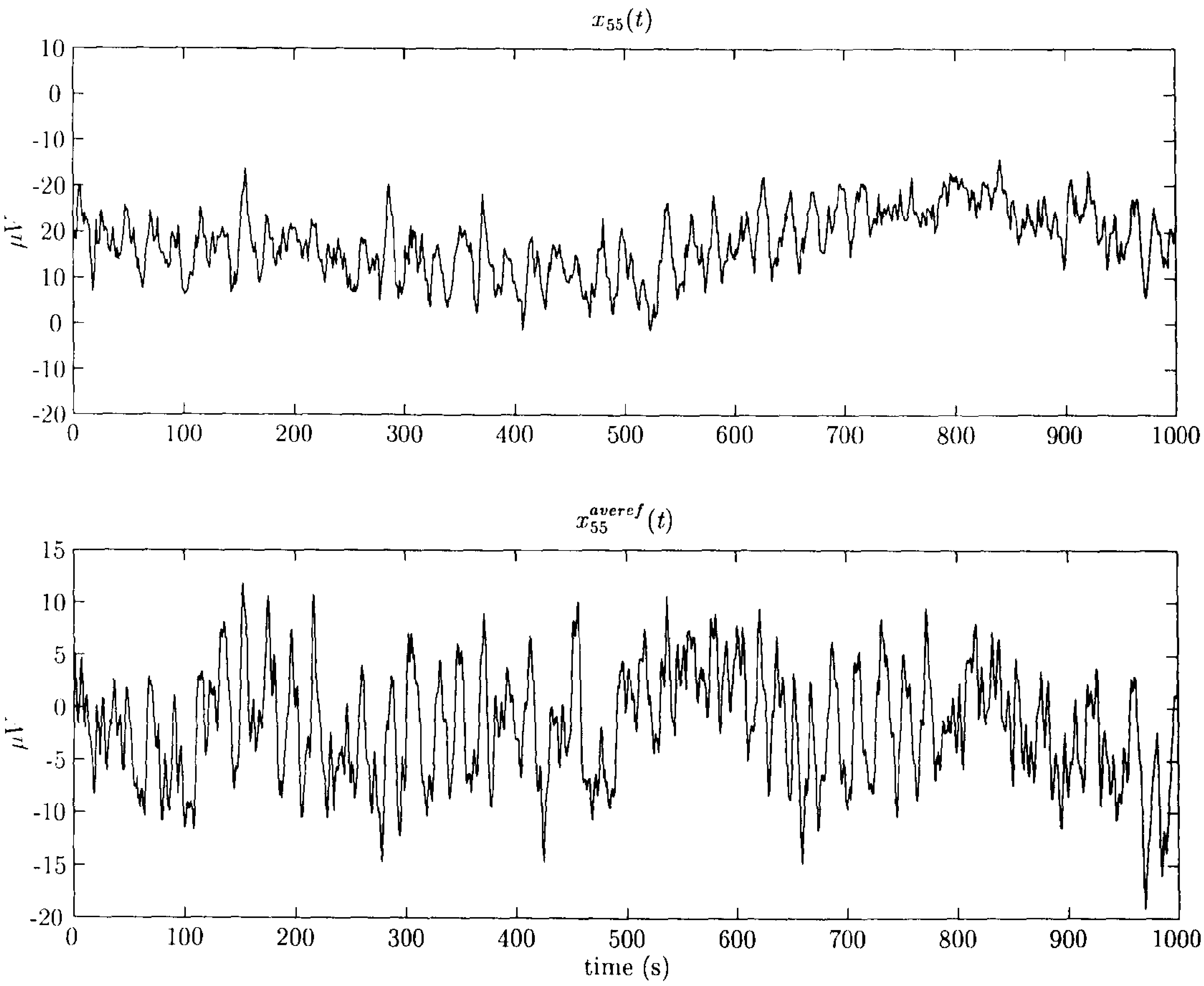

Figure 2.4: Plots showing effects of average-referencing versus raw (i.e. no re-referencing) on a channel close to the reference site (vertex).

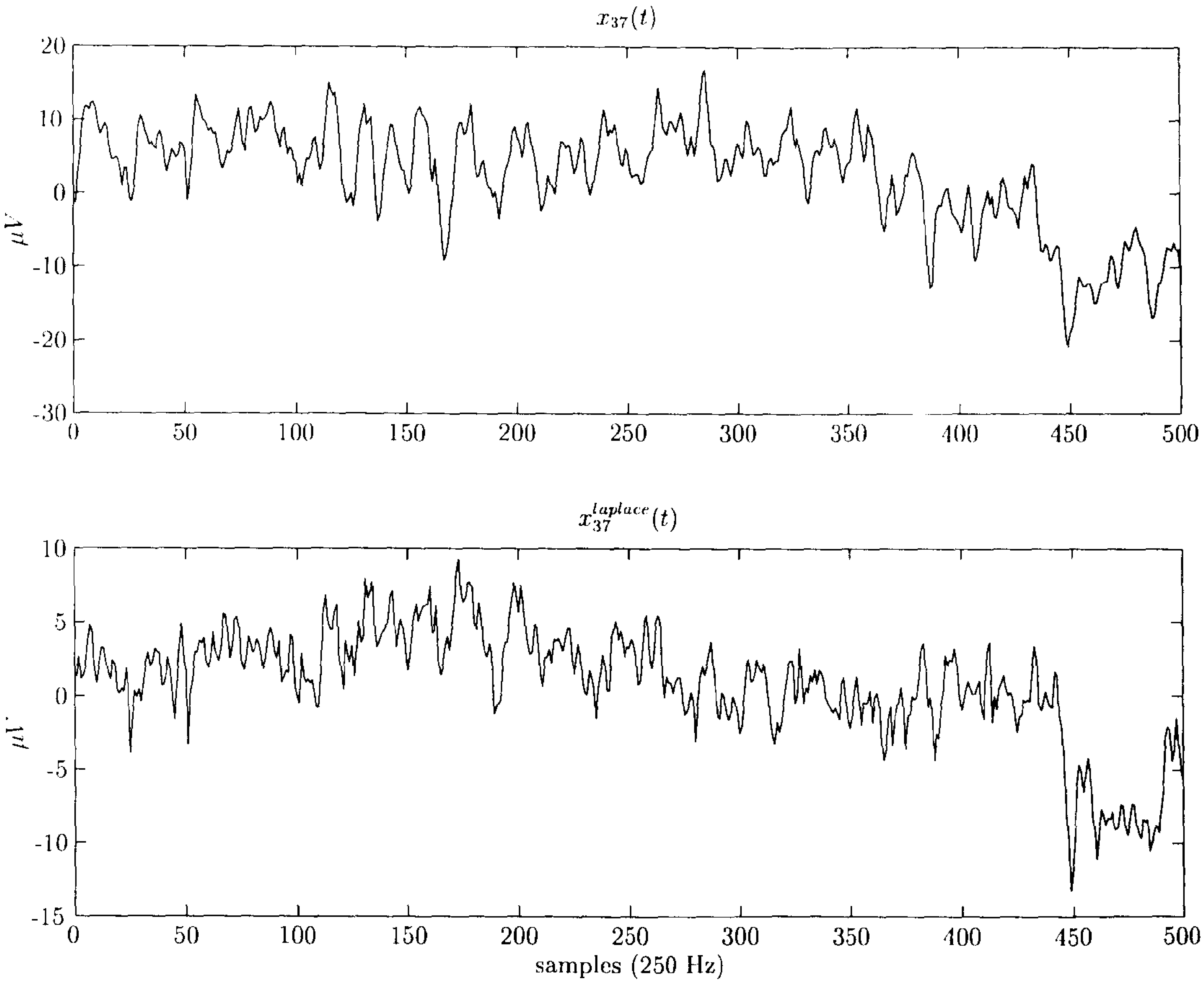

Figure 2.5: Plots comparing raw versus Laplace filtered EEG.

2.2.4 Spatial filtering

A spatial filter, based on the Laplace filter described in [Hjo75] is employed as a measure to separate local EEG from larger global effects, a technique used by [RP00] and [PPF97] (See Figure 2.5). It involves subtracting from each electrode's signal, the average of its nearest neighbours' signals, in other words

$$x_c^{'}(t) = x_c(t) - \frac{1}{|\Omega_c|} \sum_{i \in \Omega_c} x_i(t), \tag{2.3}$$

where Ω_c is the neighbourhood of channel c, $c \in \Lambda$ (all used channels), and $|\Omega_c|$ is the cardinality of Ω_c. In practice, bad channels are removed from the neighbourhood when computing (2.3), i.e. $\Omega_c \cap \Gamma = \emptyset$, where Γ is the set of bad channels.

Due to the irregular layout of the geodesic net, Ω_c varies according to the electrodes' location. See Table 2.1 for details.

Table 2.1: Nearest neighbour channel sets used when calculating the Laplace filter transform. See Figure 2.6 for channel locations.

c	Ω_c	c	Ω_c	c	Ω_c
1	2 8 125 122	2	1 122 123 3 9 8	3	2 123 124 4 10 9
4	3 124 119 5 11 10	5	4 119 113 6 12 11	6	5 113 107 7 13 12
7	6 107 32 31 13	8	1 2 9 14 126	9	2 3 10 15 14 8
10	3 4 11 16 15 9	11	4 5 12 20 19 16 10	12	5 6 13 21 20 11
13	6 7 31 30 21 12	14	8 9 15 17 126	15	9 10 16 18 14
16	10 11 19 18 15	17	14 22	18	15 16 19 23 22
19	11 20 24 23 18 16	20	11 12 21 25 24 19	21	12 13 30 29 25 20
22	17 18 23 26 127	23	18 19 24 27 26 22	24	19 20 25 28 27 23
25	20 21 29 28 24	26	23 27 127	27	23 24 28 34 33 26
28	24 25 29 35 34 27	29	21 31 36 35 28 25	30	13 31 37 36 29 21
31	7 32 38 37 30 13	32	7 55 54 38 31	33	27 34 128 26
34	27 28 35 40 39 33	35	28 29 36 41 40 34	36	29 30 37 42 41 35
37	30 31 38 43 42 46	38	31 32 54 53 43 37	39	33 34 40 45 44 128
40	34 35 41 46 45 39	41	35 36 42 47 46 40	42	36 37 43 48 47 41
43	37 38 53 52 48 42	44	39 45 49 128	45	39 40 46 49 44
46	40 41 47 50 56 49 45	47	41 42 48 51 50 46	48	42 43 52 51 47
49	44 45 46 56	50	46 47 51 58 57 56	51	47 48 52 59 58 50
52	43 53 60 59 51 48	53	38 54 61 60 52 43	54	32 55 62 61 53 38
55	32 81 80 62 54	56	40 50 57 63 49	57	50 58 64 63 56
58	50 51 59 65 64 57	59	51 52 60 66 65 58	60	52 53 61 67 66 59
61	53 54 62 68 67 60	62	54 55 80 79 68 61	63	56 57 64 69
64	57 58 65 70 69 63	65	58 59 66 71 70 64	66	59 60 67 72 71 65
67	60 61 68 73 72 66	68	61 62 79 78 73 67	69	63 64 70 74
70	64 65 71 75 74 69	71	65 66 72 76 75 70	72	66 67 73 77 76 71
73	67 68 78 77 72	74	69 70 75 82	75	70 71 76 83 82 74
76	71 72 77 84 83 75	77	72 73 78 85 84 76	78	73 68 79 86 85 77
79	62 80 87 86 78 68	80	55 81 88 87 79 62	81	50 107 106 88 80
82	74 75 83 89	83	75 76 84 90 89 82	84	76 77 85 91 90 89 83
85	77 78 86 92 91 90 84	86	79 87 93 92 85 78	87	78 79 87 93 92 85
88	80 81 106 105 94 87	89	82 83 90 95	90	83 84 91 96 95 89
91	84 85 92 97 96 90	92	85 86 93 98 97 91	93	86 87 94 99 98 92
94	87 88 105 104 99 93	95	89 90 96 100	96	90 91 97 101 100 95
97	91 92 98 102 100 101 96	98	92 93 99 103 102 97	99	93 94 104 103 98
100	95 96 101 108	101	96 97 102 108 100	102	97 98 103 109 114 108 101
103	98 99 104 110 109 102	104	94 105 111 110 103 99	105	88 106 112 111 104 94
106	88 81 107 113 112 105	107	6 113 106 81 7	108	100 101 102 109 114
109	102 103 110 116 115 114 108	110	103 104 111 117 116 109	111	104 105 112 118 117 110
112	105 106 113 119 118 111	113	5 119 112 106 107 6	114	108 102 109 115 120
115	114 109 116 121 120	116	122 121 115 109 110 117	117	123 122 116 110 111 118
118	124 123 117 111 112 119	119	5 4 124 118 112 113	120	114 115 116 121 125
121	120 115 116 122 1 125	122	1 125 121 116 117 123 2	123	1 112 117 118 124 3 2
124	4 3 123 118 119	125	1 122 121 120	126	14 8 9
127	26 23 22	128	33 34 39		

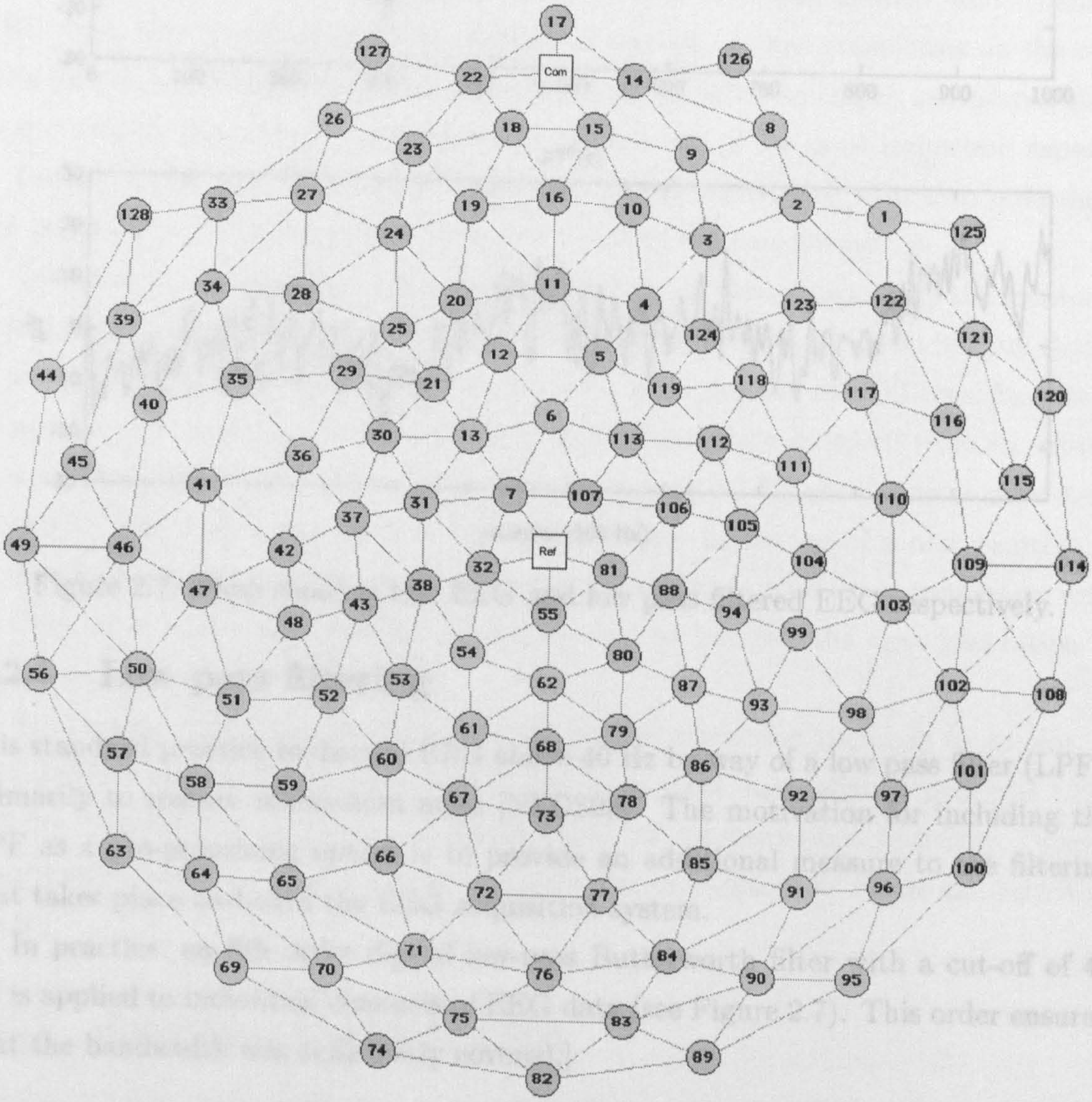

Figure 2.6: Relative channel locations of 128-channel Geodesic Net.

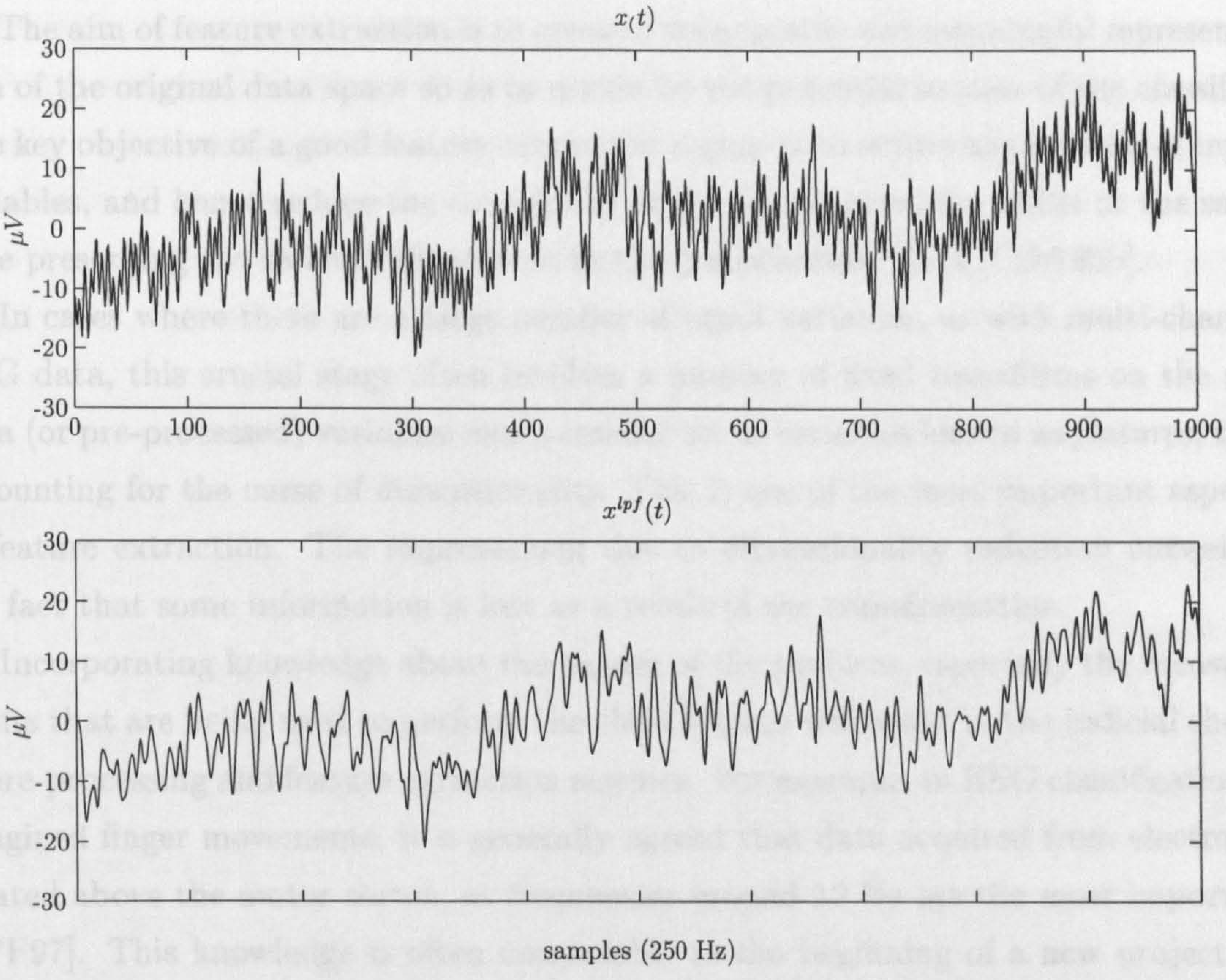

Figure 2.7: Plots showing raw EEG and low pass filtered EEG respectively.

2.2.5 Low pass filtering

It is standard practice to discard EEG above 40 Hz by way of a low pass filter (LPF), primarily to remove mains-hum noise [NLDS98]. The motivation for including the LPF as a pre-processing option is to provide an additional measure to the filtering that takes place on-board the EEG acquisition system.

In practice, an 8th order digital low-pass Butterworth filter with a cut-off of 40 Hz is applied to individual channels of EEG data (see Figure 2.7). This order ensured that the bandwidth was sufficiently covered.]

2.3 Feature extraction

In practical situations, the overall performance of a neural network based classification system can be improved by breaking the 'mapping' into two stages where an initial (often fixed) transformation of the input variables precedes the neural network [Bis95]. This stage is often one of the most important.

The aim of feature extraction is to create a manageable and meaningful representation of the original data space so as to maximise the potential success of the classifier. The key objective of a good feature extraction regime is to reduce the number of input variables, and hence reduce the complexity of the neural network, whilst at the same time preserving the salient information in the pre-processed data [CDA93b].

In cases where there are a large number of input variables, as with multi-channel EEG data, this crucial stage often involves a number of fixed transforms on the raw data (or pre-processed) variables into a smaller set of variables known as *features*, thus accounting for the curse of dimensionality. This is one of the most important aspects of feature extraction. The improvement due to dimensionality reduction outweighs the fact that some information is lost as a result of the transformation.

Incorporating knowledge about the nature of the problem, especially the measurements that are being used to perform the classification will assist in the judicial choice of pre-processing and feature extraction regimes. For example, in EEG classification of imagined finger movements, it is generally agreed that data acquired from electrodes located above the motor cortex, at frequencies around 12 Hz are the most important [PPF97]. This knowledge is often unavailable at the beginning of a new project. It is therefore important to evaluate and compare methods used in a systematic way, so that, at a later stage, knowledge can be used to improve the next generation of solutions.

In choosing which of the many possible feature extraction methods to employ, it was decided to investigate those which had been most successful in previous EEG pattern classification studies. Techniques based on *autoregressive modelling, Fourier transforms, coherence*, and *a correlation detector* are employed. These are described below.

2.3.1 Autoregressive modelling

Autoregressive (AR) modelling, a popular method for signal identification is also proving to be successful in the field of EEG pattern classification, especially as an efficient feature extraction method. AR modelling of EEG data for BCI systems has been employed successfully by many groups including [AS96, ADS95, PPF97, PRCS00]. The AR model can be used to estimate the spectral density of the signal it is modelling, which is desirable for the EEG since its changing spectral density is thought to be an important feature [NK81]. Here, two AR based feature extraction methods are used: *AR model coefficients* and an estimate of the *optimal AR model* *order*. These are described below.

AR model coefficients contain information about the statistical nature of the signal being modelled, which is desirable for the EEG since its changing spectral density is thought to be an important feature [NK81].

AR models are a class of parametric model where the future samples of a sequence $u(t), t = 1, \ldots, T$ are modelled as a linear combination of past samples, i.e.

$$\hat{u}(t) = \sum_{i=1}^{N_{AR}} \alpha(i)u(t-i), \tag{2.4}$$

where $\alpha(i), i = 1, \ldots, N_{AR}$ are the coefficients of the AR model of order N_{AR}.

Using autoregressive model parameters to represent segments of single channel EEG time series data has the advantage of greatly reducing the dimensionality of the data. For example, one might wish to classify 2-second segments of EEG data which, if sampled at 250 Hz, would result in 500-sample segments of multi-channel EEG data. By using a 5th order AR process to estimate the AR model coefficients, one achieves a compression ratio of 100:1.

AR model parameters can be estimated using a number of methods such as the Yule-Walker and Burg methods. In this case, a *stepwise least squares* algorithm [NS] is employed[6].

All features derived from EEG time series data, including AR-derived features, are calculated from non-overlapping segments of EEG between 1 and 2 seconds (250 - 500 samples) duration. The choice of a non-overlapping scheme minimises computational costs. The choice of window *size* is also influenced by computational costs, as well as temporal factors inherent in the paradigms. Further reasoning for window length is presented on a per-paradigm basis.

A well known affliction inherent to all sorts of optimisation problems is the trade-off between generalisation and accuracy. In the case of AR modelling, this is governed by the model order (N_{AR}). If the order is low, it is unlikely to sufficiently model the detail. On the other hand, if the order is too high, it will over-fit, that is, model the noise, and generalise poorly. One approach to finding the optimal order is to try a variety of models on a sub-set of the available data (a kind of exploratory stage) then choose the model order that performed the best using the rest of the data. However, this can be time consuming, and in some cases impossible (for example, where the AR model order is just one of many parameters in a larger classification system). More efficient methods for estimating the optimal model order have been

[6] Realised in the ARFIT toolbox for MATLAB [SN].

devised. For example: Akaike's Information Criterion (AIC) or Schwarz' Bayesian Criterion (SBC). In practice, a combination of the above methods are utilised, as well as indications from other EEG studies.

Optimal (AR) model order

A second AR based feature: the *optimal AR model order* (ARMO) provides a measure of signal complexity, something that has been suggested as a potentially useful measure of the EEG during different mental activities [RR98]. Although other methods exist for estimating the complexity of a time series, such as the Lyapanov dimension and correlation dimension, these all require a large data set (ten-of-thousands of values as opposed to hundreds, as in EEG) [ABST93]. Therefore, an AR based optimal model order estimation criterion is considered.

A modified Schwarz' Bayesian Criterion (MSC) was used to estimate the optimal model order of non-overlapping single channel segments of EEG, thus producing single integer valued features[7].

2.3.2 Fourier analysis

The fast Fourier transform (FFT) algorithm is the most established method for performing spectral analysis of the EEG [NLDS98]. It enables an efficient computation that yields an estimate of the spectral content of a finite length time series.

Here, the FFT algorithm is used to compute the *discrete Fourier transform* (DFT) and the *power spectral density*[8] (PSD) of single-channel EEG segments. The *magnitude squared coherence*, or simple coherence, between pairs of EEG channels is also computed. These transforms are described below.

Discrete Fourier transform and power spectral density

The discrete Fourier transform[9] (DFT), denoted, $X(k)$, of a discrete time series, $x(n)$, of finite length $n = 0, 2, \ldots, N-1$ is defined as the complex series

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk}, \qquad k = 0, 1, \ldots N-1, \tag{2.5}$$

where $W_N = \exp^{-j2\pi/N}$ is the tweak factor, k is the discrete frequency index and n the discrete time index. The first half of the DFT terms represent the positive

---

[7]The MSC was computed using the ARFIT toolbox, described in [SN].

[8]also referred to as a *spectrogram*.

[9]Calculated using the FFT algorithm in MATLAB

frequency values where the kth value corresponds to a frequency of $\frac{k}{NT}$ Hz, where T is the sampling interval in seconds.

Another popular Fourier based measure used in EEG analysis is the power spectral density (PSD) which is closely related to the FFT. There are several ways of estimating the power spectral density (PSD), $K_x(k)$, of a discrete time series $x(n)$. These are described below:

Periodogram. The periodogram is defined as the magnitude square of the DFT of $x(n)$,

$$\hat{K}_x(k) = \frac{1}{N}|X(k)|^2. \tag{2.6}$$

Averaged periodogram. To obtain a smoother spectral estimate, the averaged periodogram can be computed by dividing the signal $x(n)$ into a number of equal length segments, then computing the average of the peroidograms of each segment. Consider a signal $x(n)$ of length NL, split into L consecutive N-length segments, the averaged periodogram is defined as

$$\hat{K}_x(k) = \frac{1}{L}\sum_{l=0}^{L-1}\left\{\frac{1}{N}\left|\sum_{n=0}^{N-1} x(n+lN)W_N^{nk}\right|^2\right\}. \tag{2.7}$$

This method produces a smoother estimate than the first method, at the expense of the frequency resolution. Furthermore, both methods assume that the finite length signal is stationary for its duration.

Windowed averaged periodogram. To further smooth the randomness of $\hat{K}_x(k)$, it is helpful to window $x(n)$ before computing the DFTs, i.e.

$$\hat{K}_x(k) = \frac{1}{L}\sum_{l=0}^{L-1}\left\{\frac{1}{N}\left|\sum_{n=0}^{N-1} w(n)x(n+lN)W_N^{nk}\right|^2\right\}, \tag{2.8}$$

where $w(n)$ is a window function of which many exist. For example, the Hann window

$$w(n) = 0.5\left\{1 - \cos\left(\frac{2\pi(n+1)}{N+1}\right)\right\}. \tag{2.9}$$

For a practical coverage of the topic, see [Por97].

When working in the frequency domain and analysing EEG data, it is common practice to represent the spectrum in terms of a number of frequency bands believed to reflect certain functional aspects of the brain. This is achieved by summing the appropriate coefficients of the DFT or PSD into the following standard bands[10]: *Delta* (1 - 4 Hz), *Theta* (4 - 8 Hz), *Alpha* (8 - 13 Hz), *Beta-1* (18 - 24 Hz) and *Beta-2* (24 - 32 Hz). Thus producing 5-valued feature vectors for the DFT and PSD representations.

2.3.3 Correlation detector method

The problem of detecting whether or not a single channel segment of EEG, $x(t)$, contains a particular ERP signal, where the ERP is thought of as a target signal, $y(t)$, and the ongoing EEG $(x(t) - y(t))$ is treated as stationary noise, $n(t)$, can be tackled using a *correlation detector*[11]. The correlation detector computes the cross-correlation, r_{xy}, between a segment of the received signal (in our case the EEG), $x(t)$, and the target signal template $y(t)$ (the ERP), where $t = 1, 2, \ldots, N$ is the discrete time index. In other words

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xy} s_{xy}}}, \tag{2.10}$$

where s_{xy} is the unbiased sample covariance between the finite length time series $x(t)$ and $y(t)$, defined as

$$s_{xy} = \frac{1}{N-1} \sum_{t=1}^{N} x(t)y(t) - \overline{x}\,\overline{y}, \tag{2.11}$$

where $\overline{x}$ and $\overline{y}$ are the sample means of $\{x(t)\}$ and $\{y(t)\}$ respectively. If $x(t)$ is comprised of either (i) $n(t)$, or (ii) $y(t)+n(t)$, then r_{xy} will be maximum for case (ii). In communication detectors, the target signal (which is used directly to form the optimal matched filter response) is usually known, however, with the ERP detection problem, it has to be estimated. The simplest method for ERP estimation is to compute the ensemble average of a number of pre-recorded, time-locked EEG segments, each of which contains the ERP signal, that is

$$\hat{y}(t) = \sum_{i=1}^{N_{ES}} x^i(t), \tag{2.12}$$

where i is the index into the set of segments $\{x^i(t) : i = 1, 2, \ldots, N_{ES}\}$ used for estimation. Hence, the correlation detector method converts single-channel segments of EEG, $x(t)$, into single valued features, $r_{x\hat{y}}$, which are the cross-correlation coefficients calculated using (2.10).

[10]Note that these ranges are approximate and vary slightly in the literature.

[11]The correlation detector method is based on the theory of matched filters used in communications systems. For general information on matched filters, see [Str90].

The complete set of features obtained as a result of whatever pre-processing and feature extraction has been performed is written

$$\{f_c^{i,j}(m), m = 1, \ldots N_m\}$$

where m indexes the feature number of the set of features, $M = \{1, 2, \ldots, N_m\}$. For example, if the linear autoregressive model coefficients feature extraction was used, $N_m = N_{AR}$, where N_{AR} is the model order. Written as a vector of features, one has, $\mathbf{f}_c^{i,j}$.

2.4 Nonlinear classification

In this thesis, two nonlinear classifiers are evaluated, the static-multilayer perceptron (MLP), and a generalised linear modal (GLM). These are both types of artificial neural networks. As a comparison, Fisher's linear discriminant is also used.

An artificial neural network (ANN) may be described as a statistical model of a real world problem that has a network structure built around several layers of interconnected processing units, commonly referred to as *neurons* [Gur97]. The tuneable parameters of the model, *weights*, represent the strength of the connections between neurons. These weights are adjusted during a training period over which a sequence of known input-output vectors are presented to the network until the error between the actual output and the desired output reaches an acceptable level. In this way, an ANN can be tuned to perform pattern classification [Swi96] where the inputs are the features that have been chosen to represent the patterns extracted from the EEG data, and the outputs are a set of classes that correspond to the patterns under classification.

There are numerous reasons why the EEG pattern classification problem under investigation is difficult, and hence warrants the use of neural networks, some of these reasons are given below:

- Very little is known about the discerning qualities of normal EEG and the EEG that manifests as a result these types of mental activity. Ciaccio *et al.*, [CDA94a] indicate that neural networks are particularly suitable when little is known about the signals to be classified [AS96].

- EEG is generally accepted as a complex time varying signal that comprises of a large noise component, requiring a non-trivial approach to its analysis [NLDS98].

- Multi-channel EEG adds another dimension of complexity to the already challenging case of single channel EEG analysis [NLDS98].
- Although the field off EEG analysis is growing steadily, little work is being undertaken that addresses the multidisciplinary needs of the BCMI concepts discussed herein [Ros90].

Furthermore, neural networks are of particular interest because:

- They are capable of modelling complex linear and/or non-linear structures from real data.
- They employ a processes where a set of representative data (the training set), consisting of input vectors and desired output vectors, is presented to the network which uses an iterative training process to adjust the network's free parameters until a successful classification fitness has been achieved, thus modelling the structure within the data.
- They have proved successful when applied to similar problems. See for example the work of [PRCS00, PPF97, AS96, JMS97].

The processes by which a computer based pattern classification system achieves its goal can be broken down into a number of sub-processes each fulfilling a certain role. With neural network based classification systems it is convenient to group these sub-processes into the following stages [Bis95]:

1. Pre-processing. Raw data is manipulated in a number of ways, making it more suitable (reduced dimensionality) for presenting to a neural network. Typical processes include:

   (a) *Pattern localisation*, such as scaling to adjust for gains and bias in the measurement system, filtering to remove noise etc.

   (b) *Feature extraction*, where characteristic features based on transforms or other measures are assembled.

   (c) *Feature selection* where the best features are chosen.

(b) and (c) are often treated as separate stages from pre-processing, as is the case in this thesis.

2. Neural network classification. Presented with the pre-processed data, the neural network's adaptive parameters (weights) are adjusted during an iterative training process which minimises the network error.

3. Post-processing. Enables the network output to be interpreted as a classification, and is often achieved by simple decision rule acting directly on the network's output.

Off-line implementation of a neural network as part of a pattern classification system can be broken down into four stages:

1. *Data preparation.* Where the set of feature vectors derived from the pre-processed raw data is massaged into a form suitable for presentation to the network. This stage may involve arranging the data into suitable training and test sets, scaling the data so as to optimise the performance of the networks, and so on.

2. *Architecture.* Where applicable, decide how many hidden layer and output units the network will have, what transfer functions will be used etc.

3. *Training.* Decide which training algorithm (and subsequent parameters) will be used to recursively tune the network's weights, where relevant, determine the *stopping criterion* so that the network reaches a suitable compromise between generalisation and accuracy.

4. *Testing.* Evaluate the network's ability to generalise by presenting it with a set of unseen data.

The following sub-sections deal with these steps in turn.

2.4.1 Data preparation

Concatenation

Before presentation to a neural network, feature vectors are organised by concatenating the channels into pattern vectors,

$$\mathbf{p}^{i,j} = \begin{bmatrix} \mathbf{f}^{i,j}_{FS(1)} \\ \mathbf{f}^{i,j}_{FS(2)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}^{i,j}_{FS(N_{FS})} \end{bmatrix}$$

of length $N_{FS} \times N_m$, where FS is an N_{FS} sized indexed list representing a subset of the total channel set (C), $i \in \xi$, where ξ is the subset of classes. The choice of which channels to include in the analysis, (FS) is considered part of the feature selection stage.

Target vectors

For each pattern vector, a target vector $\mathbf{t}^{i,j}$ must also be constructed which represents the desired output of the network for that particular pattern. The form of the target vectors depends on the number of output units in the network. The way that target vectors are constructed depends on how many outputs the classifier has, the number of classes being classified (ξ), and the number of output units being used. Table 2.2 illustrates the target vector arrangements for the different classifiers.

Data set partitioning

In order to train and test a network, the complete set of pattern vectors and their associated target vectors (DS) are split into a training set, $ES = \{[\mathbf{p}^{i,j}, \mathbf{t}^{i,j}], i \in \Psi, j \in \xi\}$, and test set, $TS = \{[\mathbf{p}^{i,j}, \mathbf{t}^{i,j}], i \in \Upsilon, j \in \xi\}$, where Ψ is an N_Ψ sized random list of segments from the set of total set of segments I, where $\Upsilon \cap \Psi = \emptyset$ and $\Psi \cup \Upsilon = I$, and ξ is the set of classes (also referred to as conditions) used in the particular strategy. The size of the complete data set is, $N_{DS} = N_I N_\xi$ [12], where N_ξ is the size of ξ. Thus, the size of the training set can be calculated $N_{ES} = N_\Psi N \xi$, and the test set $N_{TS} = N_{DS} - N_{ES}$.

[12] In all cases, the number of segments, N_I, is equal for each class / condition.

Table 2.2: Target vector formats for various 2-way and 3-way classification problems for each classifier. The single-output MLP was used in the Auditory Stimulus experiment, whereas, the 2 and 3 output networks were used elsewhere.

	FISHER		MLP	MLP / GLM	
Number of outputs	1		1	2	3
Class	2-way	3-way	2-way	2-way	3-way
1	[1]	[1]	[0]	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\\0\end{bmatrix}$
2	[2]	[2]	[1]	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\\0\end{bmatrix}$
3	n/a	[3]	n/a	n/a	$\begin{bmatrix}0\\0\\1\end{bmatrix}$

Scaling

Consider the set of pattern vectors which make up the training set, $ES = \{[\mathbf{p}^{i,j}], i \in \Psi, j \in \xi\}$ which can be represented as a pattern matrix

$$\mathbf{P} = \left[\mathbf{p}^{1,1}, \mathbf{p}^{1,2}, \mathbf{p}^{1,N_\xi - 1}, \mathbf{p}^{1,N_\xi}, \mathbf{p}^{2,1}, \ldots, \mathbf{p}^{N_\Psi - 1, N_\xi - 1}, \mathbf{p}^{N_\Psi, N_\xi}\right].$$

The rows $\mathbf{P}$, which correspond to the set of values presented to a single network input, are linearly scaled within the range $[-1, 1]$ prior to presentation to the networks. This is common practice when using neural networks, as it limits the range of values presented to any one network input, which can reduce the number of training iterations required before the network converges [Swi96]. In the same way, the set of pattern vectors making up the test set are also scaled in the same way. Sets of target vectors need not be scaled as they are naturally matched to the range of the network output units, i.e $[0, 1]$.

2.4.2 Architecture and training

Multilayer perceptron

Single hidden-layer static MLP networks with a variety of hidden unit numbers, and between 1 and 3 output units (depending on the task) are trained in batch mode using a scaled conjugate gradient algorithm [Bis95].

Over-fitting is tackled by restricting the number of training epochs to a fixed number (stopping criterion), determined empirically from initial exploratory studies, where the performance of networks doesn't improve significantly for a greater number of training epochs. All network units use the *logistic sigmoid* transfer function:

$$\varphi(v) = \frac{1}{1 + \exp(-v)}. \tag{2.13}$$

which is a good general choice according to various neural network texts and numerous EEG studies.

Generalised linear model

The GLM used here has a single layer of sigmoid units (see 2.13) and is trained using the iterative re-weighted least squares (IRLS) algorithm [MN83]. Figure 2.8 illustrates a GLM network with a single output unit. This network has d inputs (plus a bias unit) feeding into a single output unit which computes a weighted linear combination of inputs plus bias. This is then fed through a logistic sigmoid to produce the output. This is linear regression with a non-linear activation function, which is precisely generalised linear regression. Both multilayer perceptron (MLP) and generalised linear model (GLM) networks are realised using the Netlab toolbox for MATLAB[13] developed by Bishop and described in his book [Bis95].

2.5 Post-processing

Patterns in the test set are forward propagated through a trained network to assess its classification fitness. Patterns presented in this manner are awarded '1' for a correct classification and '0' for an incorrect classification, in other words

$$award(\mathbf{p}) = \begin{cases} 1 : f(\mathbf{y}) = \mathbf{t} \\ 0 : \text{otherwise} \end{cases} \tag{2.14}$$

[13] http://www.ncrg.aston.ac.uk/netlab/.

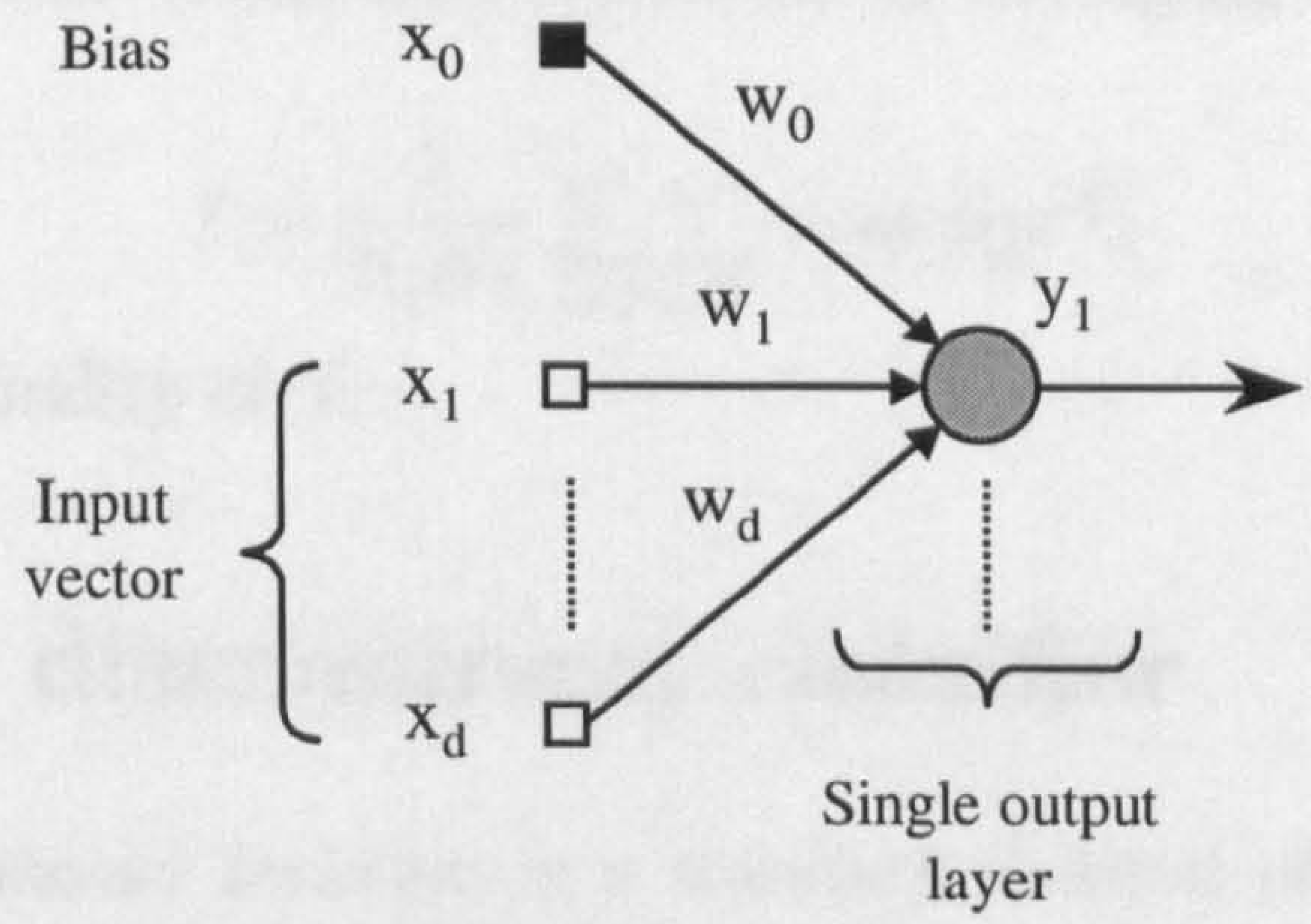

Figure 2.8: Representation of a single output unit GLM network.

where $f(\mathbf{y})$ is the post-processing function which converts the continuous outputs of the sigmoidal output units, $\mathbf{y}(\mathbf{p})$ into a binary form comparable to the target vectors. The exact nature of $f(\mathbf{y})$ depends on the number of output units in the network, and how the target vectors were constructed. Two basic post-processing functions are employed for the two types of output vector / target vector regimes (see Table 2.2):

Case 1: Single output unit. Used with the MLP in the auditory stimulus experiment (Chapter 3). In this case, two-way classifications are encoded as single bit target vectors: a '0' representing class-1, and a '1' representing class-2. The trained network, when presented with a test pattern will produce a single valued output within the range [0,1]. The post-processing function used here is a simple threshold rule, such that

$$f(\mathbf{y}) = \begin{cases} 1 : \mathbf{y} > 0.5 \\ 0 : \mathbf{y} \leq 0.5 \end{cases} \tag{2.15}$$

Case 2: Multiple output units. Used with the MLP and GLM in the musical imagery and focusing experiments (Chapters 4 and 5). Here, two and three-way classifications are encoded as two and three-bit target vectors. For example, the 3-class target vectors: [1 0 0], [0 1 0] and [0 0 1] representing class-1, class-2 and class 3 respectively. The trained network, when presented with a test pattern will produce three continuous (real valued) outputs within the range [0,1]. The post-processing function used here is a competitive transfer function which returns a vector in the form of the target vectors, such that the bit with the highest value is allocated '1' and the rest a '0'.

The fitness of the network (f) is then calculated by averaging the awards for the test set patterns (Υ),

$$f = \frac{1}{N_j N_\Upsilon} \sum_{j \in J} \sum_{i \in \Upsilon} award(\mathbf{p}^{i,j}) \tag{2.16}$$

where N_Υ is the cardinality of Υ.

2.6 Fisher's discriminant classifier

Fisher's linear discriminant function is a standard method of discriminant analysis which assumes nothing about the probability densities of the individual class populations [MKB97]. Its aim here is to achieve an optimal linear dimensionality reduction from the multi-dimensional feature-space to a single dimension classification-space. It can then be employed as a classifier by way of a simple decision rule[Bis95].

The method applied in this thesis uses the MATLAB function *fisherp.m*[14] which is based on a modified perceptron algorithm [SH99]. The function parameter *maxIts* which determines the maximum number of training iterations, is set to 200 for all strategies employed in this thesis.

2.7 Statistical analysis of results

2.7.1 Bootstrapping

In order to assess the confidence of a particular classification strategy more reliably, classifiers are re-initialised, trained and tested for a number of permutations of training sets (ES) and test sets (TS). Then the average fitness $\overline{f}$ is taken to represent that particular combination of pre-processing, feature extraction and network particulars. The bootstrapping procedure is described by the following pseudo code:

repeat NP times:
  select randomised training and test sets (sizes depend on strategy).
  initialise, train and test classifier.
  record fitness, f.
calculate average fitness, $\overline{f}$, standard deviation, confidence intervals etc.

The greater the number of permutations, (NP), the lower the confidence limits for the average fitness, which provides a more confident assessment of the strategy.

[14] Of the Statistical Pattern Recognition Toolbox. Written by Vojtech Franc, Vaclav Hlavac, Czech Technical University Prague. Available from http://cmp.felk.cvut.cz.

2.7.2 Random classifier

The expected performance of a random classifier - i.e. one which simply chooses the classes for each trial at random - is used as a means of gauging the performance of a particular classification strategy. The comparison helps address the problem of whether or not the result has arisen by chance.

Suppose this is a problem with c classes. For each trial a random classifier chooses one of these classes at random. If it is assumed that the underlying classes in the observation data are equiprobable then it can be stated that the probability of a success is $p = 1/c$, and of failure (misclassification), $1 - p$.

One is interested in the probability of k successes over n independent trials which is given by,

$$\text{Prob (k successes in n trials)} = p^k(1-p)^{n-k} \tag{2.17}$$

This is actually the probability of *one particular sequence* of successes and failures. To find out how many ways one can have k successes in n trials, one computes the following,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \tag{2.18}$$

Thus, the total probability of any k successes in n trials is,

$$f(n,k,p) = \binom{n}{k} \tag{2.19}$$

This is the *binomial probability mass function.*

More often one is interested in the *distribution function* computed from the following cumulative sum,

$$F(n,k,p) = \sum_{q=0}^{k} f(n,q,p) \tag{2.20}$$

This tells us the probability that the random classifier will have k *or less* successes in n trials, which can be turned around to yield the probability of getting more than k successes,

$$\text{Prob (} k \text{ or more successes)} = 1 - F(n, k-1, p) \tag{2.21}$$

For example, if the mean fitness of a 2-class strategy gives a result of 40% fitness for a batch of 100 test patterns. The probability of the random classifier getting 40 or more hits in 100 is 0.0966, or nearly 10%. It could therefore be said that the classifier strategy under test was doing nothing better than random choices. Only when this probability drops to around 5%, or even 1% does one conclude that the classification strategy is outperforming a random classifier.

In cases where the performance of a classification strategy is high (say greater than 70%), and/or when there are many trials in the test set, then probability of the random classifier matching the performance of the strategy under test becomes minuscule. The comparison is only really useful in the context of medium performance results. For example, suppose a three-class strategy yields an average fitness of 50% (having subtracted the lower confidence limit[15]). The performance of the random classifier depends entirely on the number of times it must guess. If there were only 10 test set patterns ($N_\Upsilon = 10$), the probability of guessing 5 trials correctly would be about 21% (Using equation 2.21, where $n = 10$, $k = 5$, and $p = 0.333$). However, if, as is the case in the analysis in the following chapters, N_Υ is in the order of 50 or more, then the performance of a random classifier becomes insignificant (less than 1%). In this case, the result of 50%, even though it is poor in an engineering setting, is nevertheless statistically superior to random guessing.

2.8 Summary

The analytical emphasis in this thesis is placed on the evaluation of a selection of pre-processing methods in conjunction with two popular neural networks, namely, multilayer perceptrons (MLPs) and later, generalised linear models (GLMs). As a way of gauging the success of the nonlinear classifiers, a linear method, the Fisher discriminant is employed (as recommended by communications with Will Penny from Oxford University's pattern analysis group). It is the application of these methods to the novel EEG pattern classification problems of the BCMI, rather than an analysis of the methods themselves which constitute the main contribution of this thesis. This chapter describes the DSP methods which are later tested on two types of EEG pattern classification problems, one *ERP*-based, the others *spontaneous*.

[15] In practice, k is obtained by subtracting the lower confidence limit from the average fitness, $\overline{f}$, then rounding down to the nearest integer.

Chapter 3

Auditory Evoked EEG Experiment

3.1 Introduction

3.1.1 Motivation

This experiment has been designed with the following hypothesis in mind: *there exists information in the EEG which allows one to distinguish between segments recorded immediately preceding and immediately following a simple auditory stimulus heard over silence.*

This may be a comparatively simple problem to address, however, it is an interesting problem nonetheless. As an initial exploration into music related EEG pattern classification, it offers a number of benefits:

- Simple problem.
- Ideal for testing overall methods including: experimental skills, EEG data management, systematic off-line EEG analysis etc.
- Potential for development into more extravagant problems with greater BCMI relevance.

3.1.2 Hypothetical BCMI system utilising ERP detection

Consider a BCMI system, such as the one depicted in chapter 1. The ability to be able to reliably differentiate between pre and post-stimulus-onset EEG alone does not lend itself to many exciting applications. However, it could lay the foundation for an ERP based system based on the idea of classifying between the EEG immediately following a self elected target stimulus, and other, non target stimuli. For example. Consider the composition of a simple piece of dance music[1], repetitive in nature - it often builds on a theme, adding and removing layers of parts which weave in and out of the piece. Suppose it is possible to classify between the EEG which immediately follows a the re-introduction of a part of music (to the ongoing dance tune) that the subject has heard before, and really likes, and is looking out for, as opposed to hearing other less favourite parts. Knowing what this response would look like in the EEG would allow the *co-ordinator* to embellish this part, in favour of the other parts. This processes could continue, until the subject no longer choose to look out for favourite parts, at which point, the music would stop.

3.1.3 Classification problem

The experiment described below allows the following classification problem to be tackled: *Determine, on a segment by segment basis, which class (pre or post-stimulus-onset) a 1-second multi-channel EEG segment belongs to, where the stimulus is any one of four tones heard over silence.* Note that the question of which tone is heard is not addressed here, merely that a tone is or isn't heard.

3.1.4 Objectives

The main objective of the experiment is to evaluate a number of pattern classification strategies (outlined in Chapter 2) on the EEG data resulting from auditory stimulus perception. What's more, it serves as a test bed for the author to become accustomed to the interdisciplinary nature of EEG experiment design, implementation, and analysis.

3.2 Paradigm

3.2.1 Overview

Subjects perform a single recognition task whilst listening to a continues sequence of auditory stimulus trials, each consisting of one of four pure tones. Upon hearing the tone, the subject decides which of the four tones they heard. The experiment

[1]The modern variety, also known as club music.

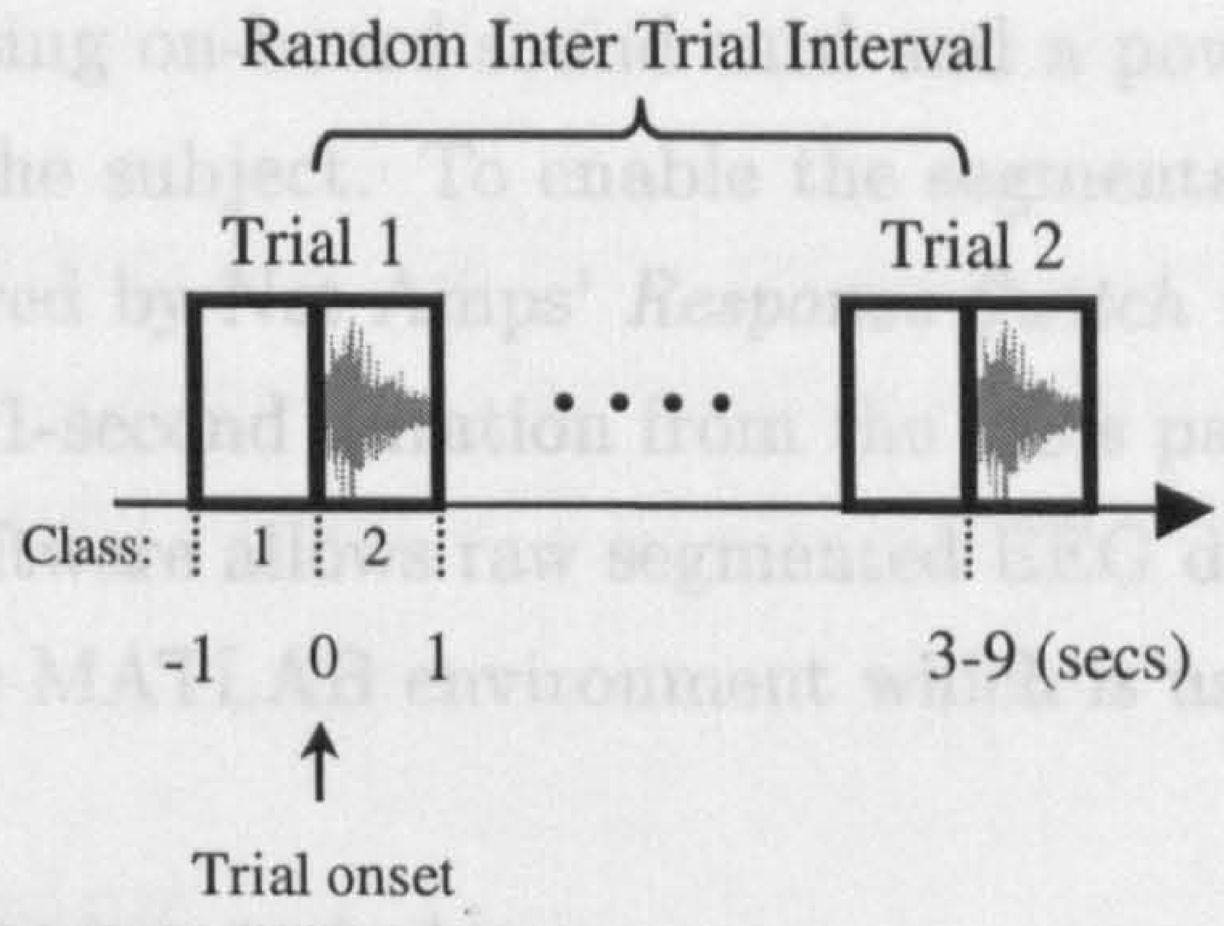

Figure 3.1: Illustration of auditory stimulus experiment trial format.

is divided into 4 blocks of trials, giving the subject a chance to relax. In total, the experiment lasts for approximately 1 hour, including set-up time.

3.2.2 Trial details

The experiment consists of four blocks of 100 1-second trials, each with a random inter-stimulus interval (ISI) between three and nine seconds (Figure 3.1). Each trial auditions one of four 1-second tones (300Hz, 400Hz, 420Hz and 600Hz) from a pseudo random play-list, such that there are 25 trials of each tone per block. Subjects are asked to listen to the tones and think to themselves which of the four they have just heard. The reason for having four tones instead of one, a varying ISI, and the random order is to maintain the interest of the subjects (for further discussion see [Ros90]). To minimise artefacts due to muscular activity and sensor displacement, subjects are asked to sit still and keep their eyes closed during the experiment. A rest period of a minute or so is allowed between blocks. Three adult male subjects are used in the experiment.

3.2.3 Data acquisition

A 128-channel (plus Cz reference[2]) Electrical Geodesics Incorporated (EGI) EEG system is employed [Tuc93]. (See Figure 2.6 for channel locations.) The system consists of a state-of-the-art *geodesic sensor net*, purpose built amplifier (*Net Amps*), acquisition control and analysis software (*Net Station* and *EGIS*). EEG data is digitised at a sample rate of 250Hz and an A/D resolution of 12 bits, then band-pass filtered between 0.1 and 40 Hz. For detailed technical notes, see EGI's web-site: http://www.egi.com.

[2]From the international 10-20 electrode placement standard.

An IBM compatible PC running in-house developed software manages the presentation of the tones using on-board sound card and a powered loudspeaker, which is placed 1 meter from the subject. To enable the segmentation of EEG data, trial-onset markers are captured by Net Amps' *Response-Switch Connector* which receives negative going pulses of 1-second duration from the PC's parallel port.

EGI's Net Station software allows raw segmented EEG data to be saved in binary files compatible with the MATLAB environment which is used for all EEG analysis.

3.2.4 Raw data segmentation

For each trial, the two seconds of EEG centred around the time of stimulus onset are split into two non-overlapping 1-second segments. Thus, each trial yields a pre-stimulus-onset and post-stimulus-onset segment. These are labelled *class 1* and it class 2 respectively. Subjects 1, 2 and 3 yield 400, 395 and 190 trials[3] resulting in $N_I = 800, 790$ and 380 segments in total.

3.3 Classification methodology

3.3.1 Overview

The problem addressed by the classification methods described below is to distinguish between pre-stimulus-onset (class 1) and post-stimulus-onset (class 2) multi-channel EEG segments, on a subject by subject basis. A number of classification strategies are evaluated where the following variations are implemented[4] (See figure 3.3.1:

- *Pre-processing.* Raw representation, i.e. no pre-processing (NONE) and Hjorth's Laplacian spatial filter (SPF). Exclusion of bad channels is standard.
- *Feature extraction.* 6th order linear autoregressive model coefficients (AR) and a correlation template method (COR) based on matched filter theory.
- *Data preparation.* Following feature extraction, pattern vectors are scaled, then divided into a training and test set. Numerous training set sizes (N_{ES}) are investigated, ranging from 100 to 700 (from a maximum of 800 segments).

[3]Subject 1 yielded 400 trials. However, due to technical mistakes, subjects 2 and 3 only yielded 195 and 396 trials respectively.

[4]For a detailed explanation of these methods, refer to Chapter 2.

- *Classifiers.* Single hidden-layer static multilayer perceptron (MLP) neural networks, and a Fisher discriminant (FISHER) classifier are compared.

The following sub sections describe these variations, and where appropriate, give justification for their inclusion into the classification system under investigation. Where justification is not given, refer to chapter 2 which explains the various methods in more detail.

3.3.2 Pre-processing

For each subject's data, bad channels are identified and removed from all further analysis. This is achieved using bad channel detection algorithm A (Appendix A). Data is then either Laplace filtered (SPF) according to equation 2.3, or left in its raw form (NONE).

3.3.3 Feature extraction

Two feature extraction methods are compared, namely AR coefficients (6th order) and correlation coefficients, as described in Chapter 2. These are computed for both raw and Laplace filtered EEG.

3.3.4 Data preparation

Pattern vectors are scaled within the range $[-1, 1]$ so as to limit the range of values presented at each network input. Scaled pattern vectors and single value target vectors are then randomly organised (see Chapter 2) into a training set (ES) and test set (TS). The number of patterns in the training set, N_{ES}, is varied between 100 and 700, depending on the number of available segments (N_I). This is done to see what effect the training set size has on classification fitness.

3.3.5 Classification

Classifications are made using static multilayer-perceptron (MLP) neural networks. The number of hidden layer units (N_{HU}) is varied between 1 and 16[5]. All networks have a single unit output layer.

[5] For computational restrictions, larger networks are not evaluated.

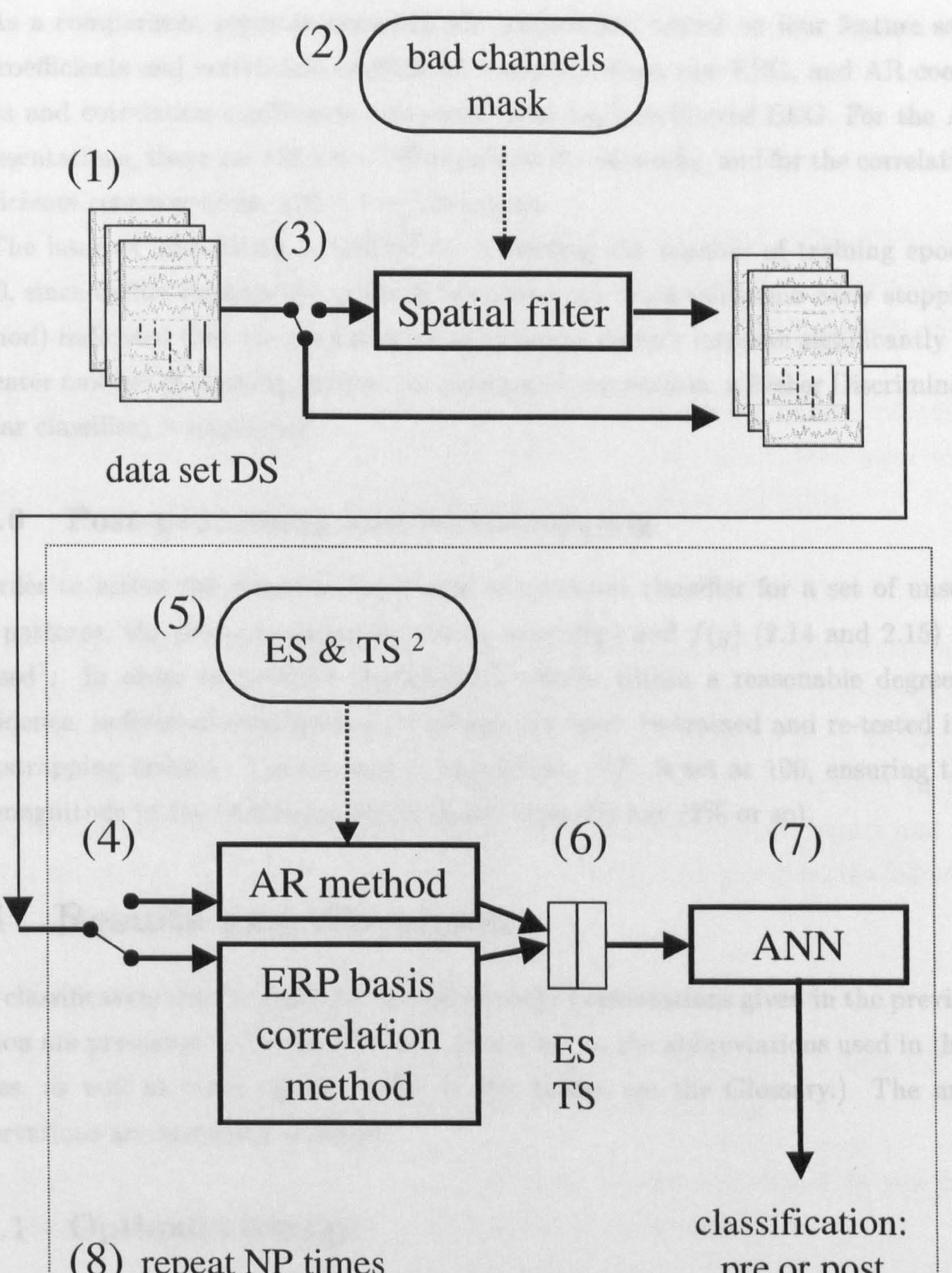

Figure 3.2: Flowchart representation of classification methodology. (1) Raw data set (DS), comprised of an equal number of pre and post-stimulus-onset multi-channel EEG segments, is filtered (or not) by Laplace method (3) which takes into account bad channels (2). The filtered data sets are then randomly divided into training (ES) and test (TS) sets according to parameter NES (5). Feature extraction (4) is performed, after which feature sets are massaged (in case of AR, scaled) into a suitable form (6) for batch training with MLP neural networks (7). Steps 4 to 7 are repeated NP times (8) to validate the results of the classifier.

As a comparison, separate networks are trained and tested on four feature sets: AR coefficients and correlation coefficients computed from raw EEG, and AR coefficients and correlation coefficients computed from Laplace filtered EEG. For the AR representations, there are 128 x 6 = 768 inputs to the networks, and for the correlation coefficients representation, 128 x 1 = 128 inputs.

The issue of over-fitting is tackled by restricting the number of training epochs to 20, since earlier exploratory analysis (involving the cross-validation early stopping method) indicated that the performance of networks doesn't improve significantly for a greater number of training epochs. As a means of comparison, a Fisher Discriminant (linear classifier) is employed[6].

3.3.6 Post-processing and bootstrapping

In order to assess the classification fitness of a trained classifier for a set of unseen test patterns, the post-processing functions, $award(\mathbf{p})$ and $f(y)$ (2.14 and 2.15) are utilised[7]. In order to produce classification results within a reasonable degree of confidence, individual classification strategies are reset, re-trained and re-tested in a bootstrapping fashion. The number of repetitions, NP, is set at 100, ensuring that the magnitude of the confidence intervals are typically low (2% or so).

3.4 Results and discussion

The classification results from the various strategy permutations given in the previous section are presented in Tables 3.1 – 3.8. (For a key to the abbreviations used in these tables, as well as other results tables in this thesis, see the Glossary.) The main observations are summarised below.

3.4.1 Optimal strategy

The optimal strategy common to all three subjects utilises the raw pre-processing option (NONE), the correlation detector feature extraction method (COH), the largest training sets (N_{ES} = 700 and 360 for subjects 1,3, and subject 2 respectively), and

[6]The reason for evaluating the Fisher discriminant classifier is due to the fact that certain BCI studies, such as [PRCS00], indicate that simple linear classifiers can perform fairly well in real-world situations.

[7]For the Fisher discriminant, its outputs are massaged into a form that can be processed in the same way, as if it were a neural network output.

the largest MLP (N_{HU}=16). This strategy yields 84.7%, 80.8%, 91.8% for subjects 1, 2 and 3 respectively.

3.4.2 Classifier

Strategies which use the MLP classifier perform marginally better than those that employ the FISHER classifier, especially when combined with the COR feature extraction method. Average classification fitnesses for these optimal strategies range from 80 to 90% across subjects.

Of the range of MLP network sizes tested, those with the greater number of hidden units only perform marginally better than those with only one hidden layer unit. There could be several reasons for this, two of which are described below.

Insufficient number of training examples

The dimensionality of the input space (to the neural networks) is comparatively large compared to the number of free parameters in the network. Although it is recognised that finding the optimal settings for a neural network can be more of an art than a science, there are some theoretical guidelines which can be used as 'rules of thumb' in choosing good ratios between, for example, the number network inputs and the number of patterns in the training set. In his book [Swi96], Swingler gives the following suggestions:

- Never choose N_{HU} to be greater than twice the number of hidden units.
- You can load p patterns of i elements into $ilog_2p$ hidden units. So never use more. If you want good generalisation, use considerably less.
- Ensure you have at least $1/\epsilon$ times as many training examples as you have weights (w) in your network, where ϵ is the network target error.

The final guideline is probably most relevant to the results obtained in this experiment. Consider the smallest and largest networks. The smallest network, belonging to $FX = COR$ and $N_{HU} = 1$ has $w = 128 \times 1 + 1 = 129$ weights, implying that $N_{ES} \geq \frac{129}{\epsilon}$. If $\epsilon = 0.1$, then $N_{ES} \geq 1290$. This lower limit is only just approached by the largest training set size, $N_{ES} = 700$. However, as N_{HU} increases, this lower limit for N_{ES} is considerably missed. For example, taking the same strategy as before, but increasing N_{HU} to 16. Thus $w = 128 \times 16 + 16 = 2064$, which implies $N_{ES} \geq \frac{2064}{\epsilon}$.

Using the same error level, this implies a lower limit for N_{ES} of 20640, a figure which is nearly 30 times greater than the maximum N_{ES} used in this experiment. This effect is further inflated when considering strategies that utilise the AR method of feature extraction, which results in approximately 6 times more network weights. This might explain the relatively poor performance of the AR representation compared to the COR method.

Sparse data set

If the data set presented to the network is sparse, that is, only a small proportion of multidimensional space is covered, then the smaller network may have an adequate capacity to model a suitable decision boundary. Increasing the network size would only increase the potential accuracy to which the network could operate to. However, if the underlying statistics of the input data are simple, then this greater modelling capacity will be useless.

One area of research that offers effective solutions to these kind of neural network application problems is the field of Bayesian statistics. Bishop [Bis95] gives a detailed account of this area, describing a variety of Bayesian techniques for neural networks with the advantages of providing an analytical method of determining confidence intervals, and the lack of dependence upon data sets other than a training set.

3.4.3 Pre-processing

When using the MLP classifier, the raw data (NONE) method performs slightly better than SPF. In the case of the FISHER classifier, a similar result is found, though the effect is less clear. This result is somewhat surprising as other studies (referenced in chapter 2) indicate that the Laplace spatial filter is beneficial. In fact, this is found to be the case in the other experiments presented in the following chapters. One could speculate that the reason for this finding is related to the nature of the classification problem. In this case, the task is to classify between stimulus-absent versus stimulus-present data, whereas in other studies, the task is often of the form: classify between stimulus-A and stimulus-B, or common-stimulus versus rare-stimulus. It is possible that, for the latter type of problem, a more localised (high frequency spatial resolution) representation of the EEG is beneficial. Whereas, for the stimulus on/off case, a more globalised (low frequency spatial resolution) representation could be important. This tallies with the effect of the Laplace filter which attenuates global

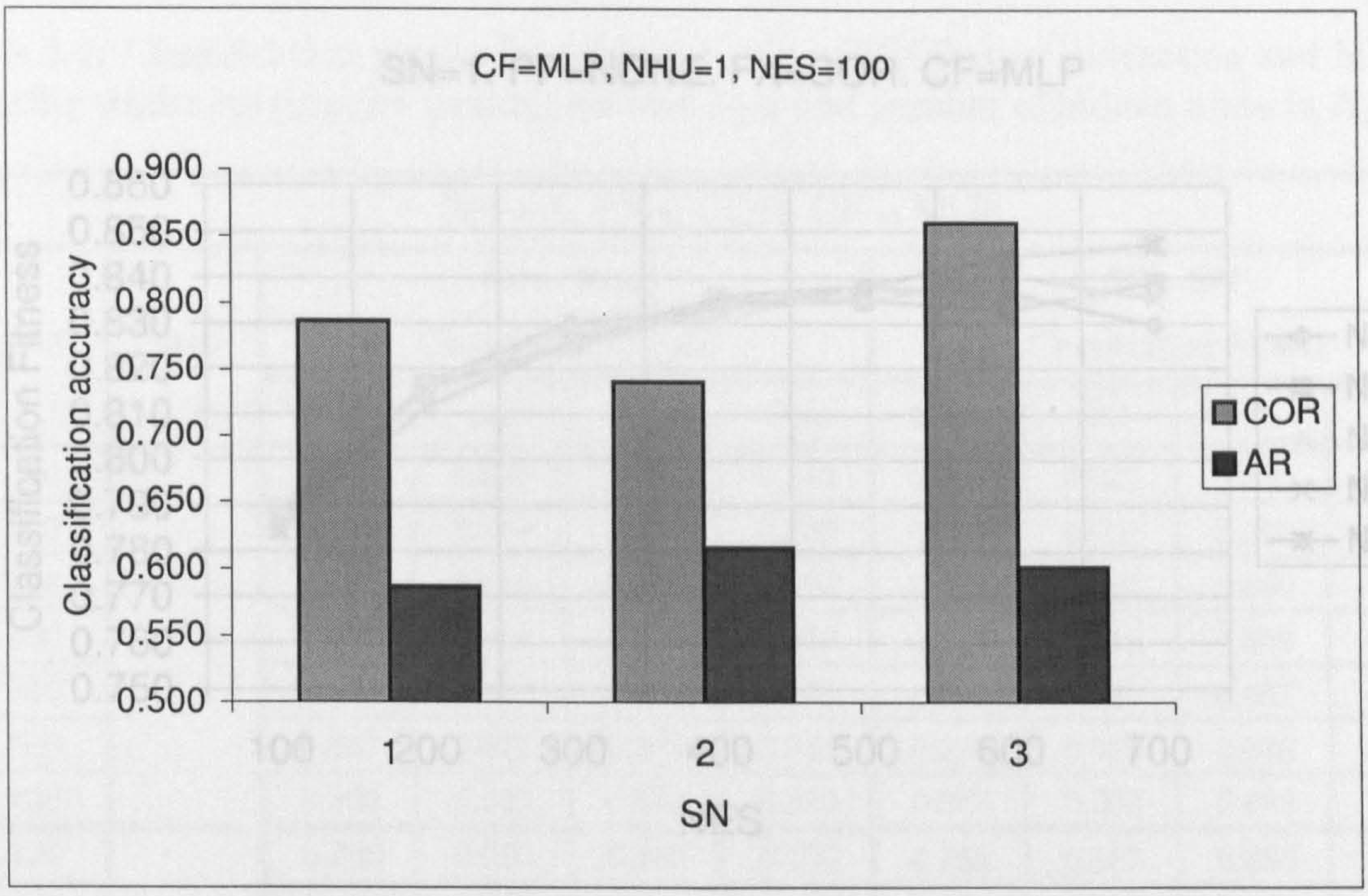

Figure 3.3: Bar plot of classification fitness for all four condition comparisons as a function of feature extraction. SN=1

effects whilst emphasising local ones [PPF97].

3.4.4 Feature extraction

Strategies employing the COR method (and the MLP) perform about 20% better than those using the AR method. As mentioned above, this could be a function of the network size problem mentioned above. However, if AR and COH are equally effective FX methods, one might at least expect (according to the network size / training set size trade-off) to find that the strategy $\{FX = AR, N_{HU} = 1\} \Longrightarrow w = 6 \times 128 + 1 = 769$ with the largest training set sizes would produce better results than the strategy $\{FX = COH, N_{HU} = 16\} \Longrightarrow w = 1 \times 128 \times 16 + 16 = 2064$ with the smaller training set sizes. However, this is not the case. Clearly, for this particular choice of strategy variables, COH is the better method of feature extraction. Figure 3.3 shows a comparison of the two feature extraction methods for subject 1.

3.4.5 Data partitioning (training set size)

The best results are obtained when using larger training set sizes (N_{ES}). However, as N_{ES} increases, the improvement becomes less pronounced. Figure 3.4 shows a typical

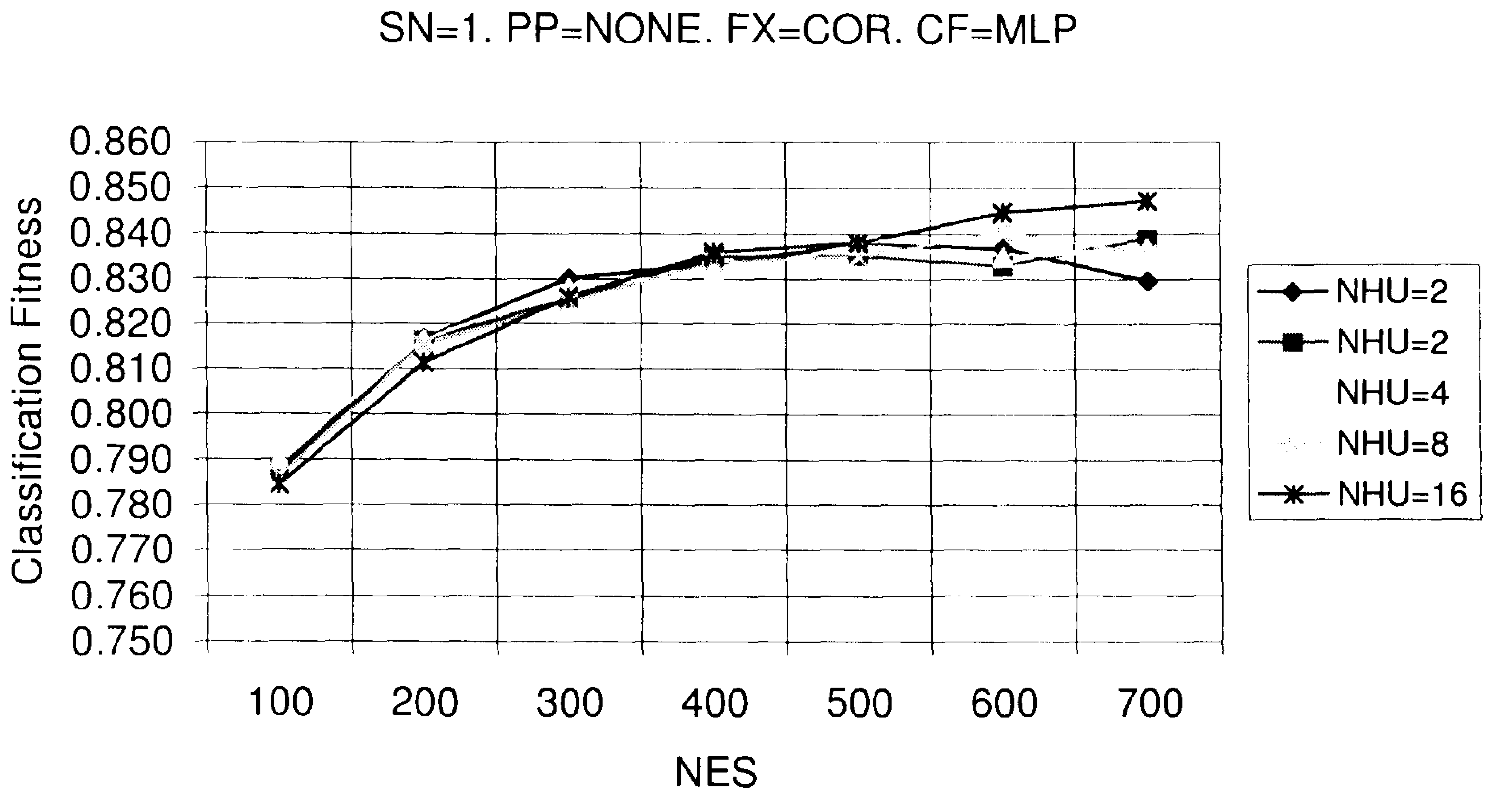

Figure 3.4: Line plot of classification fitness versus N_{ES} and N_{HU} of pre and post-stimulus onset EEG for subject 1 using the following variations: PP = NONE, FX = COR, CF = MLP.

example of this.

3.4.6 Subject variation

All subjects yielded successful results. In order of highest to lowest, subject 3 came first, followed by subject 1 and finally subject 2.

3.4.7 Performance as a function of class

When using the MLP classifier, the difference between pre and post-stimulus-onset classification scores depends on the method of feature extraction[8]. In the case of COH, pre-stimulus classification fitnesses are some 6 to 18% higher than post-stimulus segments. Whereas, in the case of AR, the pre-stimulus segments are classified some 6% lower than post-stimulus segments. See Table 3.8.

[8]The same trend is observed in the FISHER data.

Table 3.1: Classification results for subject 1 using COR feature extraction and MLP classifier whilst varying the training set size N_{ES} and number of hidden units in N_{HU}.

SN = 1. FX = COR. CF = MLP									
NES	NHU	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.787	0.024	0.831	0.743	0.754	0.025	0.855	0.652
200		0.817	0.019	0.845	0.789	0.789	0.022	0.856	0.723
300		0.830	0.018	0.847	0.814	0.802	0.020	0.850	0.753
400		0.833	0.016	0.847	0.819	0.810	0.019	0.848	0.772
500		0.838	0.020	0.850	0.825	0.822	0.021	0.857	0.787
600		0.837	0.027	0.848	0.826	0.820	0.028	0.845	0.794
700		0.830	0.033	0.840	0.819	0.827	0.033	0.844	0.810
100	2	0.788	0.020	0.846	0.730	0.753	0.027	0.865	0.641
200		0.816	0.017	0.845	0.788	0.787	0.019	0.862	0.711
300		0.826	0.017	0.845	0.807	0.803	0.017	0.857	0.748
400		0.835	0.018	0.849	0.821	0.814	0.019	0.856	0.772
500		0.835	0.018	0.850	0.820	0.820	0.021	0.853	0.788
600		0.833	0.023	0.846	0.820	0.827	0.027	0.859	0.795
700		0.839	0.034	0.856	0.822	0.831	0.035	0.863	0.799
100	4	0.790	0.023	0.843	0.737	0.757	0.023	0.862	0.651
200		0.817	0.019	0.840	0.793	0.784	0.021	0.862	0.707
300		0.828	0.017	0.845	0.811	0.801	0.019	0.854	0.748
400		0.831	0.018	0.842	0.821	0.810	0.020	0.853	0.767
500		0.837	0.020	0.841	0.834	0.822	0.021	0.848	0.795
600		0.835	0.021	0.847	0.823	0.820	0.027	0.844	0.797
700		0.837	0.036	0.852	0.821	0.832	0.036	0.856	0.808
100	8	0.787	0.021	0.846	0.727	0.754	0.022	0.866	0.643
200		0.815	0.019	0.842	0.789	0.783	0.019	0.857	0.709
300		0.825	0.017	0.837	0.813	0.802	0.018	0.850	0.753
400		0.834	0.018	0.842	0.825	0.808	0.019	0.851	0.766
500		0.835	0.020	0.838	0.832	0.818	0.019	0.849	0.787
600		0.840	0.023	0.846	0.835	0.826	0.026	0.846	0.805
700		0.836	0.038	0.843	0.828	0.827	0.038	0.840	0.813
100	16	0.784	0.021	0.847	0.722	0.751	0.023	0.876	0.625
200		0.811	0.019	0.833	0.790	0.784	0.022	0.864	0.704
300		0.826	0.016	0.836	0.816	0.798	0.020	0.851	0.744
400		0.836	0.018	0.850	0.821	0.811	0.015	0.846	0.777
500		0.838	0.020	0.841	0.834	0.817	0.020	0.847	0.788
600		0.845	0.023	0.852	0.837	0.822	0.026	0.844	0.799
700		0.847	0.032	0.848	0.846	0.838	0.037	0.857	0.819

Table 3.2: Classification results for subject 2 using COR feature extraction and MLP classifier whilst varying the training set size N_{ES} and number of hidden units in N_{HU}.

SN = 2. FX = COR. CF = MLP									
NES	NHU	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.740	0.032	0.830	0.651	0.639	0.031	0.879	0.400
200		0.765	0.031	0.835	0.695	0.706	0.031	0.873	0.538
300		0.781	0.044	0.835	0.727	0.750	0.045	0.880	0.621
360		0.791	0.117	0.862	0.720	0.745	0.130	0.812	0.678
100	2	0.741	0.032	0.841	0.640	0.644	0.029	0.884	0.404
200		0.769	0.029	0.827	0.710	0.712	0.032	0.878	0.546
300		0.786	0.047	0.833	0.739	0.753	0.047	0.876	0.630
360		0.791	0.132	0.836	0.746	0.777	0.127	0.874	0.680
100	4	0.740	0.029	0.826	0.653	0.779	0.129	0.894	0.664
200		0.784	0.045	0.832	0.735	0.714	0.030	0.880	0.547
300		0.800	0.115	0.840	0.760	0.759	0.049	0.881	0.637
360		0.774	0.029	0.831	0.717	0.654	0.027	0.884	0.424
100	8	0.741	0.029	0.840	0.642	0.646	0.024	0.893	0.399
200		0.770	0.028	0.829	0.712	0.711	0.029	0.884	0.539
300		0.788	0.043	0.827	0.749	0.757	0.042	0.873	0.641
360		0.777	0.129	0.818	0.736	0.764	0.124	0.864	0.664
100	16	0.739	0.030	0.844	0.635	0.642	0.027	0.892	0.391
200		0.768	0.030	0.826	0.711	0.709	0.036	0.877	0.540
300		0.771	0.044	0.809	0.733	0.757	0.051	0.876	0.639
360		0.808	0.100	0.836	0.780	0.770	0.128	0.868	0.672

Table 3.3: Classification results for subject 3 using COR feature extraction and MLP classifier whilst varying the training set size N_{ES} and number of hidden units in N_{HU}.

SN = 3. FX = COR. CF = MLP									
NES	NHU	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.861	0.023	0.902	0.819	0.719	0.029	0.892	0.546
200		0.887	0.016	0.909	0.866	0.817	0.021	0.912	0.722
300		0.899	0.017	0.915	0.883	0.856	0.021	0.924	0.789
400		0.903	0.017	0.913	0.893	0.877	0.019	0.933	0.820
500		0.906	0.019	0.915	0.897	0.889	0.025	0.926	0.852
600		0.913	0.020	0.923	0.902	0.899	0.022	0.924	0.873
700		0.911	0.033	0.920	0.901	Accidental data loss			
100	2	0.860	0.023	0.892	0.828	0.725	0.028	0.895	0.554
200		0.888	0.016	0.909	0.867	0.819	0.023	0.916	0.722
300		0.899	0.015	0.911	0.887	0.858	0.020	0.928	0.788
400		0.904	0.016	0.915	0.893	0.880	0.018	0.934	0.827
500		0.909	0.016	0.918	0.899	0.892	0.020	0.935	0.849
600		0.912	0.021	0.916	0.907	0.899	0.018	0.931	0.867
700		0.915	0.032	0.923	0.907	0.911	0.033	0.939	0.883
100	4	0.863	0.021	0.905	0.821	0.725	0.027	0.894	0.556
200		0.888	0.015	0.907	0.870	0.819	0.024	0.917	0.720
300		0.900	0.015	0.909	0.891	0.856	0.018	0.920	0.793
400		0.907	0.015	0.918	0.896	0.881	0.019	0.929	0.833
500		0.908	0.018	0.923	0.894	0.893	0.019	0.927	0.860
600		0.912	0.021	0.918	0.906	0.901	0.025	0.931	0.871
700		0.913	0.030	0.920	0.906	0.910	0.027	0.934	0.886
100	8	0.863	0.021	0.909	0.818	0.882	0.019	0.928	0.836
200		0.887	0.016	0.902	0.873	0.893	0.019	0.930	0.857
300		0.898	0.016	0.913	0.883	0.901	0.022	0.932	0.870
400		0.905	0.016	0.913	0.897	0.907	0.032	0.933	0.882
500		0.907	0.017	0.910	0.904	0.732	0.027	0.901	0.564
600		0.912	0.021	0.921	0.904	0.821	0.021	0.917	0.725
700		0.911	0.033	0.917	0.904	0.859	0.018	0.922	0.795
100	16	0.865	0.018	0.901	0.829	0.728	0.033	0.903	0.553
200		0.889	0.017	0.908	0.870	0.818	0.022	0.914	0.721
300		0.899	0.016	0.916	0.882	0.858	0.019	0.925	0.791
400		0.905	0.014	0.914	0.895	0.878	0.017	0.924	0.833
500		0.913	0.018	0.920	0.906	0.893	0.020	0.926	0.860
600		0.910	0.022	0.918	0.903	0.899	0.021	0.929	0.869
700		0.918	0.027	0.928	0.908	0.908	0.032	0.928	0.888

Table 3.4: Classification results for subject 1 using AR feature extraction and MLP classifier whilst varying the training set size N_{ES} and number of hidden units in N_{HU}.

SN = 1. FX = AR. CF = MLP									
NES	NHU	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.586	0.048	0.577	0.595	0.549	0.033	0.514	0.584
200		0.584	0.050	0.533	0.636	0.559	0.039	0.588	0.529
300		0.600	0.052	0.585	0.615	0.562	0.044	0.528	0.597
400		0.604	0.054	0.582	0.625	0.571	0.048	0.573	0.568
500		0.594	0.063	0.542	0.647	0.574	0.043	0.571	0.577
600		0.594	0.065	0.547	0.641	0.570	0.056	0.574	0.566
700		0.596	0.073	0.555	0.636	0.571	0.056	0.565	0.577
100	2	0.596	0.041	0.590	0.601	0.557	0.029	0.577	0.537
200		0.605	0.042	0.596	0.615	0.570	0.032	0.604	0.535
300		0.613	0.043	0.606	0.620	0.577	0.033	0.557	0.597
400		0.619	0.041	0.607	0.631	0.587	0.035	0.548	0.626
500		0.617	0.049	0.610	0.624	0.580	0.036	0.576	0.585
600		0.633	0.046	0.595	0.670	0.588	0.041	0.578	0.599
700		0.619	0.055	0.590	0.648	0.578	0.051	0.605	0.552
100	4	0.589	0.037	0.603	0.575	0.560	0.025	0.568	0.553
200		0.614	0.029	0.594	0.634	0.577	0.025	0.562	0.592
300		0.619	0.033	0.595	0.642	0.578	0.031	0.577	0.579
400		0.623	0.035	0.585	0.661	0.586	0.027	0.605	0.568
500		0.628	0.037	0.630	0.627	0.589	0.033	0.571	0.608
600		0.628	0.039	0.618	0.638	0.598	0.037	0.563	0.633
700		0.636	0.045	0.645	0.626	0.598	0.044	0.596	0.599

Table 3.5: Classification results for subject 2 using AR feature extraction and MLP classifier whilst varying the training set size N_{ES} and number of hidden units in N_{HU}.

SN = 2. FX = AR. CF = MLP									
NES	NHU	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.616	0.061	0.540	0.692	0.567	0.041	0.558	0.576
200		0.634	0.051	0.576	0.692	0.577	0.049	0.622	0.533
300		0.628	0.067	0.555	0.701	0.583	0.059	0.591	0.576
360		0.627	0.126	0.498	0.756	0.576	0.119	0.584	0.568
100	2	0.641	0.033	0.560	0.723	0.581	0.034	0.586	0.576
200		0.644	0.043	0.564	0.724	0.596	0.045	0.612	0.580
300		0.644	0.053	0.561	0.727	0.587	0.058	0.582	0.593
360		0.628	0.128	0.508	0.748	0.607	0.124	0.572	0.642
100	4	0.643	0.030	0.581	0.704	0.582	0.030	0.583	0.582
200		0.648	0.030	0.583	0.713	0.596	0.036	0.612	0.579
300		0.648	0.046	0.576	0.721	0.594	0.048	0.577	0.612
360		0.648	0.113	0.564	0.732	0.595	0.117	0.570	0.620

Table 3.6: Classification results for subjects 1, 2 and 3 using COR feature extraction and FISHER classifier whilst varying the training set size N_{ES}.

FX = COR. CF = FISHER									
NES	SN	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
100	1	0.777	0.024	0.779	0.775	0.769	0.021	0.775	0.763
200		0.795	0.022	0.796	0.795	0.792	0.020	0.795	0.789
300		0.795	0.046	0.794	0.795	0.786	0.022	0.794	0.779
400		0.776	0.052	0.776	0.776	0.801	0.022	0.811	0.791
500		0.791	0.048	0.799	0.784	0.795	0.024	0.807	0.782
600		0.783	0.057	0.781	0.784	0.803	0.034	0.810	0.796
700		0.806	0.051	0.809	0.804	0.792	0.038	0.802	0.782
100	2	0.729	0.036	0.743	0.715	0.674	0.030	0.681	0.667
200		0.756	0.047	0.769	0.744	0.734	0.037	0.755	0.713
300		0.743	0.070	0.753	0.733	0.755	0.043	0.769	0.742
360		0.740	0.135	0.752	0.728	0.776	0.120	0.776	0.776
100	3	0.863	0.024	0.879	0.847	0.762	0.024	0.769	0.756
200		0.881	0.018	0.900	0.862	0.831	0.016	0.845	0.817
300		0.867	0.033	0.883	0.851	0.861	0.018	0.879	0.843
400		0.880	0.019	0.900	0.861	0.879	0.016	0.896	0.861
500		0.878	0.018	0.904	0.852	0.872	0.013	0.902	0.856
600		0.877	0.031	0.902	0.851	0.879	0.022	0.913	0.846
700		0.875	0.043	0.897	0.852	0.882	0.024	0.920	0.836

Table 3.7: Classification results for subjects 1, 2 and 3 using AR feature extraction and FISHER classifier whilst varying the training set size N_{ES}.

FX = AR. CF = FISHER									
NES	SN	PP = NONE Classification Fitness				PP = SPF Classification Fitness			
		Mean	Std	Pre	Post	Mean	Std	Pre	Post
50	1	0.586	0.031	0.615	0.557	0.572	0.024	0.578	0.567
100		0.580	0.040	0.608	0.553	0.578	0.019	0.586	0.569
150		0.578	0.036	0.608	0.549	0.577	0.023	0.580	0.574
200		0.576	0.023	0.591	0.560	0.573	0.028	0.573	0.572
250		0.581	0.042	0.608	0.554	0.563	0.029	0.571	0.554
300		0.560	0.031	0.572	0.548	0.582	0.030	0.588	0.576
350		0.586	0.044	0.590	0.582	0.554	0.039	0.565	0.544
50	2	0.616	0.049	0.617	0.615	0.577	0.031	0.579	0.575
100		0.595	0.050	0.599	0.590	0.588	0.031	0.594	0.582
150		0.623	0.057	0.629	0.617	0.595	0.050	0.582	0.609
180		0.608	0.191	0.592	0.624	0.584	0.165	0.568	0.600
50	3	0.626	0.033	0.628	0.624	0.531	0.022	0.525	0.536
100		0.612	0.042	0.609	0.615	0.544	0.022	0.539	0.549
150		0.620	0.033	0.621	0.618	0.530	0.023	0.520	0.540
200		0.614	0.040	0.612	0.616	0.544	0.019	0.537	0.550
250		0.617	0.033	0.615	0.619	0.541	0.033	0.525	0.557
300		0.606	0.052	0.605	0.607	0.527	0.031	0.522	0.532
350		0.610	0.055	0.609	0.612	0.530	0.063	0.530	0.530

Table 3.8: Combined (grand average) classification results for MLP classifier. All subjects, training set sizes and network sizes are combined in order to summarise trends as a function of pre-processing and feature extraction.

PP	FX	Mean	Std	Pre	Post
NONE	COH	0.832	0.032	0.864	0.799
SPF		0.792	0.035	0.884	0.700
NONE	AR	0.626	0.052	0.594	0.659
SPF		0.601	0.043	0.605	0.596

3.5 Conclusions

This chapter describes an evoked auditory stimulus EEG experiment which has been designed as a first measure to assessing the viability of the BCMI concept. Classifications are performed on single 1-second segments of multi-channel EEG belonging to one of two classes: pre or post-stimulus EEG, where the stimulus is a pure tone heard over silence. Firstly, it has been shown that by combining a correlation detector based feature extraction method and MLP neural networks, one can classify between pre stimulus-onset and post stimulus-onset auditory evoked EEG, a problem with relevance to the BCMI concept [DMS98b], and any application involving single trial ERP detection. Secondly, it is found that a simple non-linear classifer -the Fisher discriminant -only performs marginally less successful than the MLP, a significant result from the point of finding methods that minimise computational costs. Thirdly, a number of interesting questions relating to specific aspects of the classification methodology have been raised. In particular

- Is the information in multi-channel EEG sparsely distributed, in which case, can standard data reduction techniques be employed to simplify the task of the classifier? For example, would a reduced channel provide successful classification results?
- Is the Laplace spatial filter suited to specific types of EEG classification problem, or can it be safely employed as a general pre-processing tool?
- The correlation detector, simple yet effective in this situation: could it be useful in more challenging situations, such as differentiating, on a trial by trial basis, between a number of different classes of ERP relating to different aspects of musical experience?

Overall, the experiment re-confirms what countless other research indicates - that the EEG contains information about our experience, and that techniques exist which enable aspects of that experience to be witnessed in near real time (since classifications are based on 1-second segments).

3.6 Summary

It is shown that autoregressive (AR) modelling, a correlation detector technique (a type of matched filter), and a multilayer perceptron (MLP) neural network enable the classification of single-trial, pre and post stimulus-onset auditory evoked EEG, a problem which has relevance in the design of musical BCI systems. It is found that the correlation detector method of feature extraction significantly out-performs the AR method. It is also found that pre-processing with a Laplace filter does not improve classification fitness. This may suggest that auditory perception in the EEG is distributed over the entire scalp in a more gross manner, since the Laplace filter aims to pick out local detail in favour of global detail [PPF97]. A linear classifier, the Fisher discriminant, is also tested and found to operate almost as successfully as the MLP.

Chapter 4

Musical Imagery Experiment

4.1 Introduction

4.1.1 Overview

This chapter presents details of a novel EEG experiment involving three BCMI related mental tasks: *musical imagery, passive listening* and *counting*. Numerous classification strategies (based on the classification framework described in Chapter 2) are employed in an off-line analysis of the three classes of EEG data relating to the above three tasks. The results suggest that a BCMI system based on these tasks alone is viable. This section gives some motivational issues and a description of the engineering problem intrinsic to the design of the experiment. The remaining sections give a detailed explanation of the experimental paradigm, the analysis methodology, results and discussion and some brief conclusions.

4.1.2 Motivation

One of the key engineering problems upon which the concept of a new-era BCMI heavily depends is being able to classify between segments of EEG belonging to a variety of musical tasks. When considering which music related mental tasks to test as possible BCMI candidate tasks, imagery appears to be a good option. Brain imaging studies back this up. For example Zatorre [ZHP96], by gives very strong evidence that both listening to music and imagining music use the same parts of the brain. Zatorre finds that there are many common areas of activation for perception and imagery, including the frontal and temporal-lobe areas, the hippocampal and thalamic areas, and more. A previous study by Zatorre [ZEM94] concludes that activity in the right superior temporal cortex is representative of perceptual analysis of melodies, pitch comparisons are likely to involve the right prefrontal cortex, and active pitch memory recruiting both the right temporal and frontal regions. These studies are important, since they re-enforce the hypothesis that the human brain uses the same cortical areas and processes for the perception and imagery of music. Moreover, with clues as to which parts of the brain are involved in these tasks, one might be able to focus on the EEG sites which correspond to these areas.

Related studies in the field of BCI technology report the success of comparing (non-musical) imagery tasks against a baseline task, such as relaxing or some other passive activity. For example, Anderson et al [AS96] shows how half-second segments of EEG data recorded whilst subjects perform imagined letter writing, object rotation and relaxation tasks can be classified using a combination of AR modelling and standard feedforward neural networks.

The success of this study, and the brain imagery work of Zatorre readily leads to the following hypothesis, upon which the musical imagery experiment described in this chapter is based:

> *There exists information in the EEG that allows one to identify whether a person is engaged in one of two mental tasks: musical imagery or passive listening.*

In this context, the mental task of musical imagery means to re-play the experience of hearing some music, or a part of that music in the 'mind's ear'. This activity should be familiar to anyone who enjoys listening to and humming along with a favourite tune. Notice that once the song has finished, one finds oneself humming the tune in one's head without actually making any sound. When a composer 'hears' an entire tune (or part of it) in his head, he is using the 'mind's ear'. It is just like an artist who uses their 'mind's eye' to visualise how a piece of art would look after some changes, except with the auditory faculty.

To listen to a piece of music without making any special effort, such as imagining hearing parts (during or after they have finished), or actively focusing on one aspect or part of the music, is passive listening. In day to day life one is likely to be listening passively if they are relaxing to music or engaged in some other task whilst listening to music at the same time.

In most contexts one chooses to perform musical imagery, invoking it as a tool to aid composition or assist performance, or simply as a means of enhancing the overall musical experience. Musical imagery is, for the following reasons, a prime candidate mental-task for the BCMI investigation:

- It is a well defined aspect common to the experience of music.
- It is creative and enjoyable by nature.
- It requires a deliberate mental effort.
- It is the closest thing to singing or performing, yet is without any muscular activity.

The passive listening task is an activity requiring no effort, and is thus a suitable candidate for the baseline task, the 'neutral state' upon which a set of music related mental tasks (such as imagery) can be classified from.

4.1.3 Objectives

The remainder of this chapter concerns the successful implementation of the following two objectives believed to be necessary for assessing the validity of the above hypothesis:

1. Design and implement an experiment incorporating musical imagery and passive listening tasks that allows the following classification problem to be tackled:

   Successfully determine, on a segment by segment basis, which class (musical imagery or passive listening) a 2-second multi-channel EEG segment belongs to, where the class is named after the mental task the subject is performing while the segment is being recorded.

2. Seek a solution to the above problem by employing a variety of digital signal processing methods based on existing methods found to be of success in the EEG pattern classification field (such as those described in Chapter 1), thus gaining insight into whether the methods explored are likely to be suitable for a working BCMI system.

4.2 Paradigm

4.2.1 Overview

Subjects perform one of three mental tasks: imagery (I), passive listening (relaxing) (R) and counting (C) whilst listening to a continuous sequence of trials. Each trial consists of two parts: a rhythm part, lasting for the entire trial, and a riff[1] part, lasting for the first half of the trial. It is during the second half of each trial that the mental task is performed. The experiment is divided into 6 blocks of trials, giving the subject a chance to relax. The experiment lasts for approximately 1 and a half hours, including set-up time.

Due to a lack of research in this area, the paradigm described herein is necessarily unique, and therefore certain particularities (such as the blocking arrangement - see below) are also treated as an experimental variables. In terms of style, a balance has been sought between existing standard practice (use of blocks, randomisation of trials, inclusion of a control condition etc.), and the anticipated needs of hypothetical BCMI system. For example, the choice to use life-like musical stimulus, as opposed to more simple 'repeatable' options.

4.2.2 Trial format

Each trial lasts for 8 seconds and consists of 4 repetitions of a 1-bar rhythm loop[2]. Superimposed onto this rhythm part are 2 repetitions of a 1-bar riff loop which starts at the beginning of the trial and are finished halfway through the trial. Altogether, there are 15 unique riff loops: 5 piano, synth and guitar loops respectively. All these musical parts were created using a multimedia PC computer using a MIDI-audio sequencer package, an electric guitar with amplifier simulator, and an audio i/o sound card. The music composed is in the style of a moderate tempo popular dance / club tune (120 beats per minute, 4 beats per bar). An inter-trial interval of 8 seconds means that there are no gaps between trials. In this way, the background part loops seamlessly for the entire duration of each block of trials. See Figure 4.1.

[1] Riff: popular music jargon for a short catchy melody which is usually repeated many times in the course of a song. Classical music calls it an 'ostinato'.

[2] Loop: popular music jargon for a segment of music which, when played in a continuous 'loop' sounds seamless, as if it were being played as a continuous part with no obvious beginning and end. The term loop refers to the segment which is to be looped.

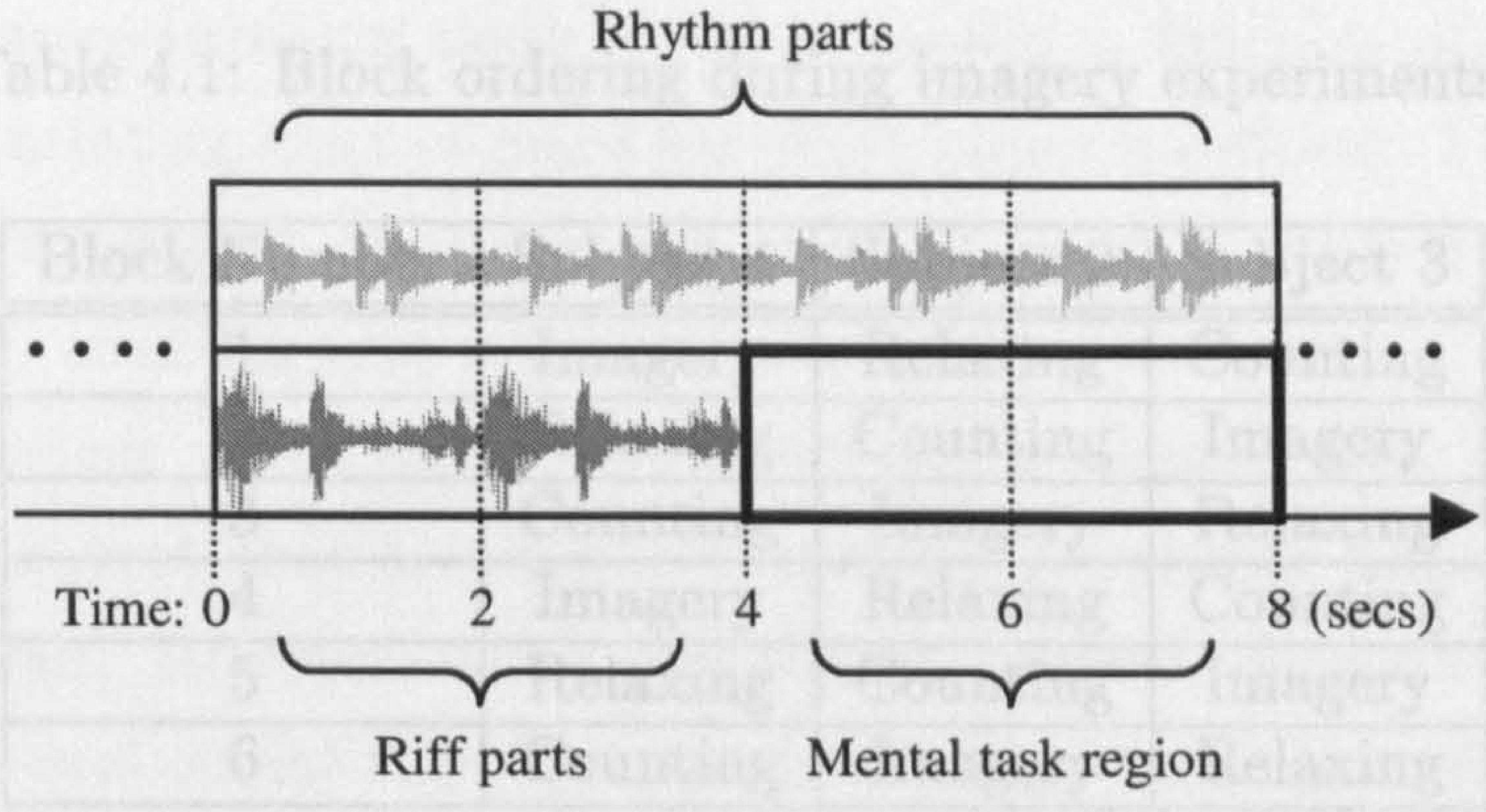

Figure 4.1: Diagrammatic representation of an imagery experiment trial.

4.2.3 Conditions

There are three conditions: Imagery, Passive Listening / Relaxing and Counting. A blocked system is adopted where the subject is instructed before each block of trials to perform one of three tasks, named after the conditions. These tasks are described below:

- Imagery task (I): listen to the looped riff, which lasts for 2 bars, then immediately after it finishes, imagine in your 'minds ear' that the riff continues for another 2 bars until the next trial begins.
- Passive listening task (relaxing - R): listen to the entire 4 bar trial with no effort, just relax. If you like, focus on the background part but in a relaxed way.
- Counting task (C): once the riff part has stopped, begin counting the following self repeating sequence: 1, 10, 3, 8, 5, 6, 7, 4, 9, 2, 1, 10 and so on. Do this at your own pace.

4.2.4 Blocks

The experiment is divided into 6 blocks each consisting of 60 trials. Blocks are named after the task that the subject is instructed to perform in that block, and are ordered according to Table 4.1. The reason for having 2 blocks for each condition, and for changing the ordering of these blocks for each subject is to minimise the effects of time dependent features (such as increased alpha due to tiredness) which could confuse the results [Ros90]. Each of the 15 riff parts are auditioned 4 times in each block, but in a pseudo random order (so as to alleviate boredom).

Table 4.1: Block ordering during imagery experiments.

Block Number	Subject 1	Subject 2	Subject 3
1	Imagery	Relaxing	Counting
2	Relaxing	Counting	Imagery
3	Counting	Imagery	Relaxing
4	Imagery	Relaxing	Counting
5	Relaxing	Counting	Imagery
6	Counting	Imagery	Relaxing

4.2.5 Subjects

Four male subjects with musical experience were run. However, due to poor net placement, subject number 4's data has been excluded from the analysis.

4.2.6 Rational for including counting task

Suppose, hypothetically, that it is possible to perform a two way classification between the EEG of musical imagery and passive listening tasks respectively. An important line of questioning that begs to be answered goes something like this:

> *What is the phenomenon, in this case, which leads to the classifiable characteristics contained within the EEG? Is it that musical imagery requires more effort than the passive task, and this effort appears as a gross discriminating characteristic? Or are there other, subtler task dependent characteristics, such as the fact that it is a musical task, or that it requires the skill of visualisation?*

The problem highlighted here is central to the field of EEG pattern classification. As mentioned early, it is only recently that science has began to study the central nervous system, the brain, and mind in any depth, and it is a long way off its goal [Car98]. Consequently, scientific insight into the inner workings of the brain, and of conscious experience and how this is reflected in the EEG, falls short of being able to answer the above question. Even with the most cleverly designed experiments which take into account many cognitive factors, there is still the uncertainty that it is the analysis methods which are not sensitive enough, or suitably attuned to the details hidden within the EEG.

Nevertheless, the rational for including the counting task is that it provides a way to check that the characteristics which might allow one to classify between the imagery and passive listening tasks, are not merely a function of a concentrating versus not-concentrating state of mind (or effort versus no effort). The counting task is a control for the mental-effort component of the musical imagery task, since, like the imagery task, counting also requires the subject to concentrate. If it was the act of concentration alone which lead to the classification imagery from passive listening, then one might expect that a two-way classification between musical imagery and counting might fail. However, if one can classify between these three tasks (two of which require mental effort) then it supports the hypothesis that effort related tasks involving different faculties produce different EEG based characteristics.

Bearing this in mind, one should note that there have been successful attempts at controlling or interacting with a 'musical environment' by the human EEG which have utilised non-musical tasks. These are typically based on the subjects' ability to self regulate the power of their alpha wave frequency component (8-12Hz) by some relaxation or mental stilling technique [Ros90]. For this reason, it is useful to consider counting as a candidate task for the BCMI concept.

4.2.7 Acquisition details

EEG data is acquired using the same system as described in the previous chapter. However, EGI's *EGIS* system is used to control the presentation of musical media and to manage the acquisition of data.

4.2.8 Raw data segmentation

Only the last four seconds (the later half) of each trial are used in the analysis. These 4-second segments are further divided into two 2-second segments. Thus each trial yields 2 segments. With 120 trials for each condition, each subject produces 720 segments (240 segments for each of the three conditions).

4.3 Classification methodology

4.3.1 Overview

EEG data is analysed on a subject by subject basis. Classifications are made between 2-second multi-channel segments belonging to pairs of conditions (2-way classifications) and to all three conditions (3-way classifications). A number of classification strategies are evaluated where the following variations are implemented[3]:

- *Pre-processing (PP):* raw representation, i.e. no pre-processing (NONE), average referencing (AVR), Hjorth's Laplacian spatial filter (SPF), low pass filtering (LPF). Exclusion of bad channels is standard. Automatic detection and exclusion of 'bad' segments (muscle artefact contamination) is attempted, but is found to be unsuccessful.

- *Feature extraction (FX):* 6th order linear autoregressive model coefficients (AR), autoregressive model order estimate (ARMO), binned fast Fourier transform (FFT) and estimated power spectral density coefficients (PSD).

- *Feature selection (FS):* various channel groupings consisting of subsets of the full 128 possible channels. (See Table 4.3.)

- *Data encoding - training set size (SR):* two basic training set sizes are investigated, the exact size is determined by the split ratio (SR) which is either ninety-ten (9:1) or fifty-fifty (1:1).

- *Classifier:* generalised linear model (GLM) and single hidden-layer static multilayer perceptron (MLP) neural networks, and a Fisher discriminant (FISHER).

The following sub sections describe these variations, and where appropriate, give justification for their inclusion into the classification system under investigation.

4.3.2 Pre-processing

Filtering

As stated above, in addition to the raw representation, three filtering variations are employed, namely: average referencing (AVR), Hjorth's Laplacian spatial filter (SPF) and a low pass filter (LPF).

Bad channels

Removal of bad channels is standard procedure when using dense array EEG systems. In this case, a bad channel detection algorithm is employed as a means to automatically detecting bad channels. See Appendix A - 'Bad channel detection algorithm B' for details. These channels are removed from all further analysis.

[3]For a detailed explanation of these methods, refer to Chapter 2.

Bad segments

EEG identified as being contaminated by muscle artefacts is often removed from analysis [NLDS98]. In this case, two artefact detection algorithms: one for eye-blinks, the other for eye-movements are employed in an attempt to identify segments of EEG which contain these types of artefacts (Appendix A). Both algorithms are applied to all 720 2-second segments from all 3 conditions. If either of the algorithms identify an artefact, the segment is labelled as a 'bad', allowing it to be excluded from the analysis.

As can be seen from Table 4.2, the number of bad channels for each subject varies significantly. Of the three subjects' data, subject 1's was the worst with more than half the segments identified as being bad. This could be due to the calibration parameters of the detection algorithms, or, more likely, a problem with one or more of the four eye channel electrodes. However, visual inspection at the time of recording confirmed that the overall appearance of the data was 'clean'.

Table 4.2: Percentage of bad segments identified by artefact detection algorithms. Total number of segments per subject is 720.

Subject number	Percentage bad segments
1	58.9%
2	9.4%
3	18.6%

4.3.3 Feature extraction

The following transformations are performed on all the four variations of pre-processed segments, resulting in a number of sets of feature vectors.

AR coefficients

Single channel 2-second (500 samples) segments of pre-processed EEG are converted into 6-valued feature vectors, which are the 6th order AR model coefficients obtained using a *stepwise least squares* algorithm [NS] which is realised in the ARFIT toolbox for MATLAB [SN]. The reasons for using a 6th order model are as follows:

- A preliminary exploratory analysis indicated that the 6th order AR model is the better choice in comparison to 12 and 24 ordered models, which are found to offer insignificant improvements to the classification accuracy.
- Existing research supports this finding, especially [AS96, PPF97, PRCS00].

Autoregressive model order estimation

Single channel 2-second (500 samples) segments of pre-processed EEG are converted into single-valued feature vectors, where each value is the result of the optimal model order estimation using a modified Schwarz' Bayesian Criterion (MSC).

Binned DFT and PSD

Single channel 2-second (500 samples) segments of pre-processed EEG are converted using MATLAB's built in FFT function (fft.m) to 500-valued FFT coefficients vectors. The first 250 values are kept and their absolute values calculated (using the abs.m MATLAB function). This gives the DFT coefficients ranging from 0 to 125 Hz, since the EEG is sampled at 250 Hz. To compute the PSD coefficients, the simplest method is employed, that is, the squared DFT coefficients. DFT and PSD coefficients are then arranged into bins and summed according to the 5 popular EEG frequency ranges (*Delta* (1 - 4 Hz), *Theta* (4 - 8 Hz), *Alpha* (8 - 13 Hz), *Beta-1* (18 - 24 Hz) and *Beta-2* (24 - 32 Hz)). Hence, both representations reduce a 500-valued segment of EEG time series into a 5-valued binned frequency representation.

4.3.4 Feature selection

Having transformed raw segments of EEG data into a number of reduced dimensionality representations, the total number of features representing a single multi-channel segment of EEG is still great. This is due to the number of channels available, i.e. 128. For example, with the AR representation, the total number of values in the feature vector is 128 x 6 = 768 variables. Whereas, with the ARMO representation, there are only 128 x 1 = 128 variables. It is a well known malady in the field of pattern classification that there is a price to pay for high dimensionality data sets. This is discussed in Chapter 2. Suffice to say that, reducing the dimensionality on the one hand throws away information, yet on the hand improves the likelihood that the network will perform well. In cases where there are many input variables presented to the classifier, there must also be significantly more exemplar patterns making up the

Table 4.3: Channel sets used in feature selection along with code names used in tables throughout the remainder of this thesis.

Channel sets used in feature selection stage			
Channel sets	Code name used in tables	Number of channels	Geodesic Nets Electrode Numbers
Temporal channels	FS1	4	46 58 97 109
International 10-20 montage	FS2	18	9 11 23 25 34 37 46 58 60 62 71 84 86 97 105 109 122 124
Full 128 EGI net minus the peripheral electrodes[1]	FS3	92	1 8 14 17 22 26 33 39 44 45 49 56 57 63 64 69 70 74 75 82 83 89 90 95 96 100 101 108 114 115 120 121 125 126 127 128
1. Peripheral electrodes - the ones that were removed - are given here.			

training set [Bis95]. For this reason, a number of reduced channel set representations (FS1, FS2 and FS3) are presented to the networks, rather than using the full 128 channels. These are defined in Table 4.3.

An additional reason for evaluating smaller channel sets is that from the practical engineering point of view, fewer channels are better (cheaper and more portable - not to mention simpler).

4.3.5 Data set partitioning

It is unusual that the data analyst has an unlimited data set representative of the problem at hand. In this case, the finite-sized data set (DS) is partitioned into a training set (ES) and test set (TS). (For details see Chapter 2, 'Data set partitioning'.) Two partitionings are considered here, a fifty-fifty split (SR=1:1) and a ninety-ten split (SR=9:1), where SR denotes the split ratio.

These two split ratios are evaluated to identify whether the classification system is sensitive to the size of ES. This is a very important question which is worthy of an entire branch of investigation because the engineer wishes to know how little training data will sufficiently allow the classification system to learn to classify unseen data. In the case of EEG, the less time an individual has to spend in a training phase, the better.

4.3.6 Classification

In the field of EEG pattern classification, nonlinear classification techniques have been shown to perform well in comparison to linear methods [NLDS98]. Here, nonlinear classifications are performed by static-MLP and GLM neural networks. A linear method - the Fisher discriminant classifier - is also evaluated[4]. A range of MLPs are evaluated where the number of hidden units is varied between 2 and 32. Rather than attempting validation early stopping, or pruning, MLPs were trained in batch mode[5] for a fixed number of epochs - in this case 50 - which was found to be a suitable value in previous exploratory studies.

4.3.7 Bootstrapping and statistics

Every classification strategy is re-tested between 10 and 25 times[6] according to the bootstrapping method explained in Chapter 2. This yields an average fitness, $\overline{f}$. In addition to this, the following statistical measures are calculated: 99% confidence limits, standard deviation, and p_{chance} - the probability that the result average fitness $\overline{f}$ could have arisen by chance (using the random classifier comparison described in Chapter 2).

4.4 Results and discussion

In this section, results are presented in a way which focuses on the overall mean classification fitnesses (resulting from the bootstrap analysis) as a function of the numerous strategy variations (described above)[7]. The main purpose for doing this is to identify which strategies are best suited to the problem at hand, as well as gaining some insight into the relative contribution of each sub-stage (with respect

[4] As stated before, the reason for employing the Fisher discriminant is to add to the current body of research which compares linear and nonlinear methods in the EEG pattern classification field.

[5] Batch mode is preferred over incremental mode since (1) all data is available off-line, and (2) batch mode gives better convergence [Swi96].

[6] This figure was set arbitrarily so as to give a suitable number of repetitions, and hence, a good confidence statistic. The trade-off between the time taken to re-compute was the main factor in making the choice of how many repetitions to perform.

[7] In this context, the term ***strategy*** refers to a set of classification parameters, one parameter for each of the five sub-stages of the classification procedure: pre-processing (PP), feature extraction (FX), feature selection (FS), data-encoding (SR) and classifier (CF). Additional 'variations' which are not particular to the classification methodology itself are the experimental conditions, or classes (CND), and the subject number (SN). These abbreviations are used throughout the following two chapters as a shorthand was of describing the particular strategy components.

to its variations) to the overall classification performance. Note that, as expressed in Chapter 1, the overall aim is to develop a *systematic BCMI evaluation protocol*, which has been achieved, rather than dwelling on the specific ramifications of the individual stages.

4.4.1 Optimal strategies

Classification results for the optimal strategies are presented for the each of the three classifiers: Fisher discriminant (FISHER), generalised linear model (GLM) and static-multilayer perceptron (MLP) in Tables[8] 4.4, 4.5 and 4.6 respectively. The following attributes are common among all three classifiers:

- Pre-processing: Laplace spatial filter (SPF).
- Feature extraction: 6th order linear autoregressive model coefficients (AR).
- Split ratio (determines training set size): 9:1.

Furthermore, strategies which utilise the FISHER and MLP classifiers perform best when using the largest number of electrodes (FS3). When using the GLM the two way classifications (IR, IC, RC) perform best with the international 10-20 electrode montage (FS2) whereas the three-way classification problem (IRC) does best when using the larger channel set (FS3).

4.4.2 Best classifier

Taking the best strategies for the three classifiers, the MLP is marginally better than the FISHER and GLM which both achieve similar classification fitnesses. This trend can be seen in Figures 5.3 – 5.5.

To compare the overall performance of each classifier the average performance across a number of strategies is calculated (see Table 4.7). Clearly, the MLP achieves the best overall classification fitness. The GLM appears to be slightly better than the FISHER method. However, when restricting this comparison to the best strategies for each condition and subject, it is found that the FISHER method performs marginally better (96.2%) than the GLM (95.4%)[9]. These differences in performance fall within the 99% confidence limits and the one (typical) standard deviation which for these

[8] Refer to Table the Glossary for a key of abbreviated code-names used in tables.

[9] Computed from the average of the 12 optimal strategies in Tables 4.4, 4.5.

Table 4.4: Optimal classification strategies using the FISHER classifier.

Optimum Strategies for the FISHER classifier[1]											
SN	Classification fittnes				Confidence Limits (+/-)	CND			NI	NES	NTS
	Mean	Std.	Min.	Max.		Class 1	Class 2	Class 3			
	0.977	0.018	0.958	1.000	0.019	I	R				
1	0.994	0.010	0.979	1.000	0.010	I	C	none	474	432	48
	0.971	0.020	0.938	1.000	0.021	R	C				
	0.961	0.022	0.931	1.000	0.023	I	R	C	474	648	72
	0.979	0.017	0.958	1.000	0.017	I	R				
2	0.913	0.040	0.833	0.958	0.041	I	C	none	474	432	48
	0.958	0.035	0.896	1.000	0.036	R	C				
	0.893	0.049	0.792	0.958	0.051	I	R	C	474	648	72
	0.960	0.039	0.875	1.000	0.040	I	R				
3	0.973	0.030	0.917	1.000	0.030	I	C	none	432	432	48
	0.996	0.009	0.979	1.000	0.009	R	C				
	0.965	0.029	0.903	1.000	0.030	I	R	C	432	648	72

1. These parameters were common to all the above variations: PP=SPF, FX=AR, FS=FS3, SR=9:1, CF=FISHER., NP=10.

Table 4.5: Optimal classification strategies using the GLM classifier.

Optimum Strategies for GLM classifier[1]												
SN	Classification fittnes				Confidence Limits (+/-)	CND			FS	NI	NES	NTS
	Mean	Std.	Min.	Max.		Class 1	Class 2	Class 3				
	0.922	0.031	0.854	0.979	0.018	I	R					
1	0.965	0.022	0.917	1.000	0.012	I	C	none	FS2	78	432	48
	0.928	0.038	0.813	0.979	0.021	R	C					
	0.945	0.023	0.889	0.986	0.013	I	R	C	FS3	474	648	72
	0.989	0.014	0.958	1.000	0.008	I	R					
2	0.932	0.034	0.854	0.979	0.019	I	C	none	FS2	90	432	48
	0.927	0.035	0.833	0.979	0.019	R	C					
	0.938	0.029	0.889	1.000	0.016	I	R	C	FS3	474	648	72
	0.963	0.032	0.896	1.000	0.018	I	R					
3	0.986	0.017	0.938	1.000	0.009	I	C	none	FS2	90	432	48
	0.996	0.010	0.958	1.000	0.006	R	C					
	0.963	0.020	0.917	1.000	0.011	I	R	C	FS3	432	648	72

1. These parameters were common to all the above variations: PP=SPF, FX=AR, SR=9:1, CF=GLM, NP=25.

Table 4.6: Optimal classification strategies using the MLP classifier.

Optimum Strategies for MLP classifier[1]

SN	Classification fittnes				Confidence Limits (+/-)	CND			NI	NES	NTS
	Mean	Std.	Min.	Max.		Class 1	Class 2	Class 3			
1	0.998	0.007	0.979	1.000	0.007	I	R				
	0.996	0.009	0.979	1.000	0.009	I	C	none	474	432	48
	0.994	0.010	0.979	1.000	0.010	R	C				
	0.988	0.015	0.958	1.000	0.016	I	R	C		648	72
2	0.994	0.010	0.979	1.000	0.010	I	R				
	0.973	0.031	0.896	1.000	0.032	I	C	none	474	432	48
	0.954	0.038	0.896	1.000	0.039	R	C				
	0.951	0.023	0.903	0.986	0.024	I	R	C		648	72
3	0.973	0.014	0.958	1.000	0.014	I	R				
	0.992	0.011	0.979	1.000	0.011	I	C	none	432	432	48
	0.994	0.014	0.958	1.000	0.014	R	C				
	0.985	0.015	0.958	1.000	0.016	I	R	C	432	648	72

1. These parameters were common to all the above variations: PP=SPF. FX=AR. FS=FS3. SR=9:1. CF=MLP. NP=10. NH=8. NE=50.

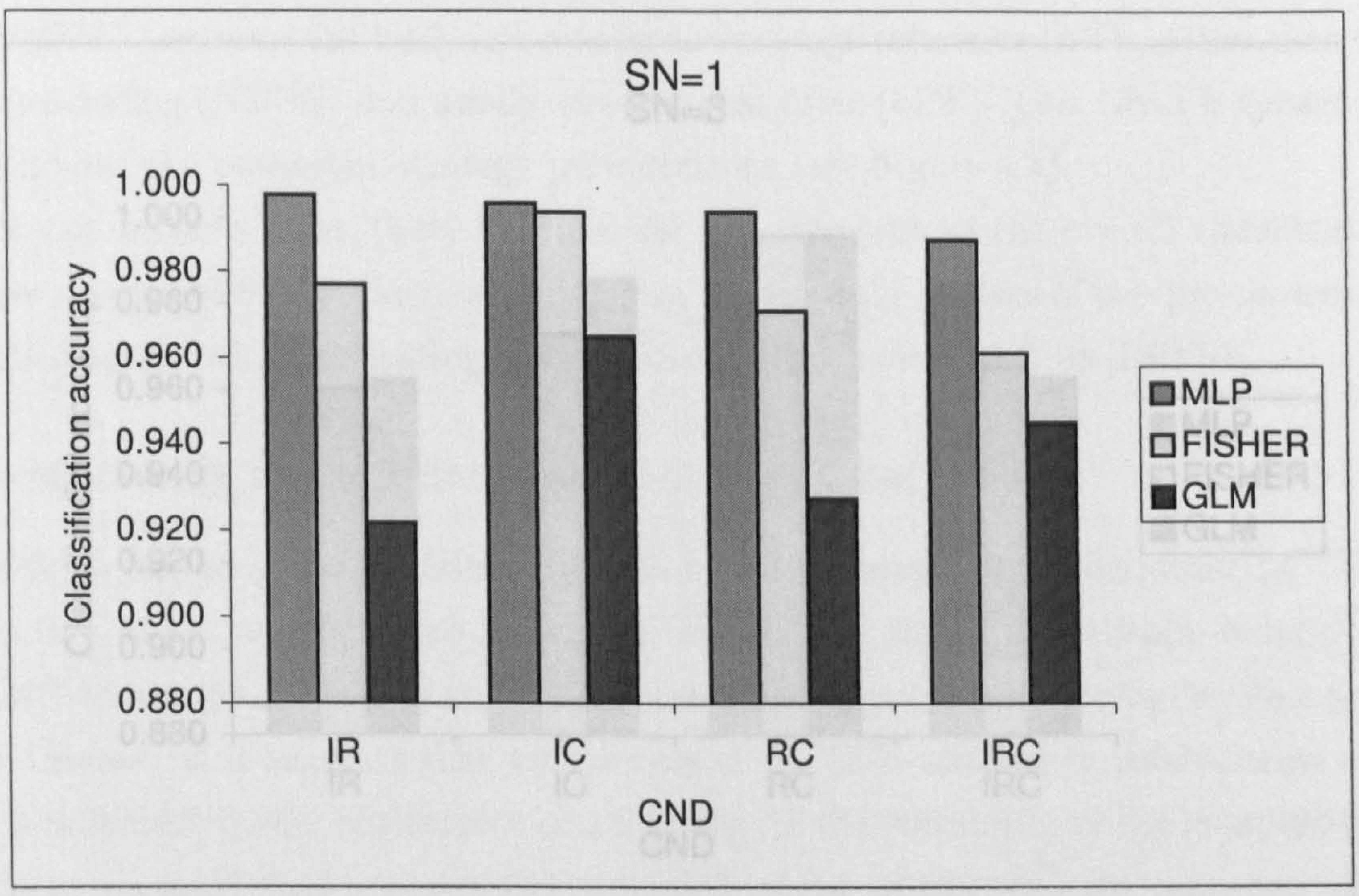

Figure 4.2: Bar plot of classification fitness for all four condition comparisons as a function of classifier. SN=1

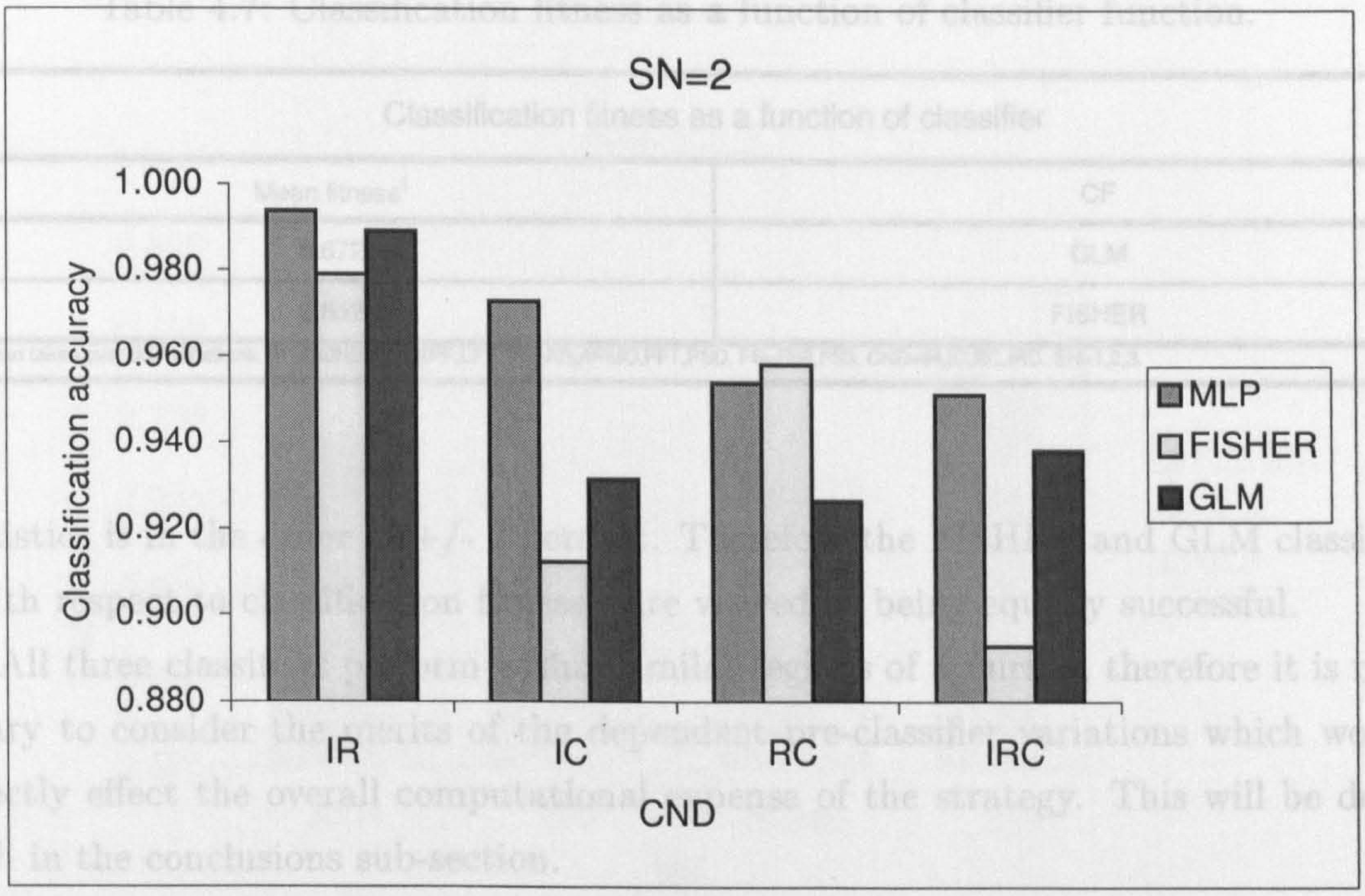

Figure 4.3: Bar plot of classification fitness for all four condition comparisons as a function of classifier. SN=2

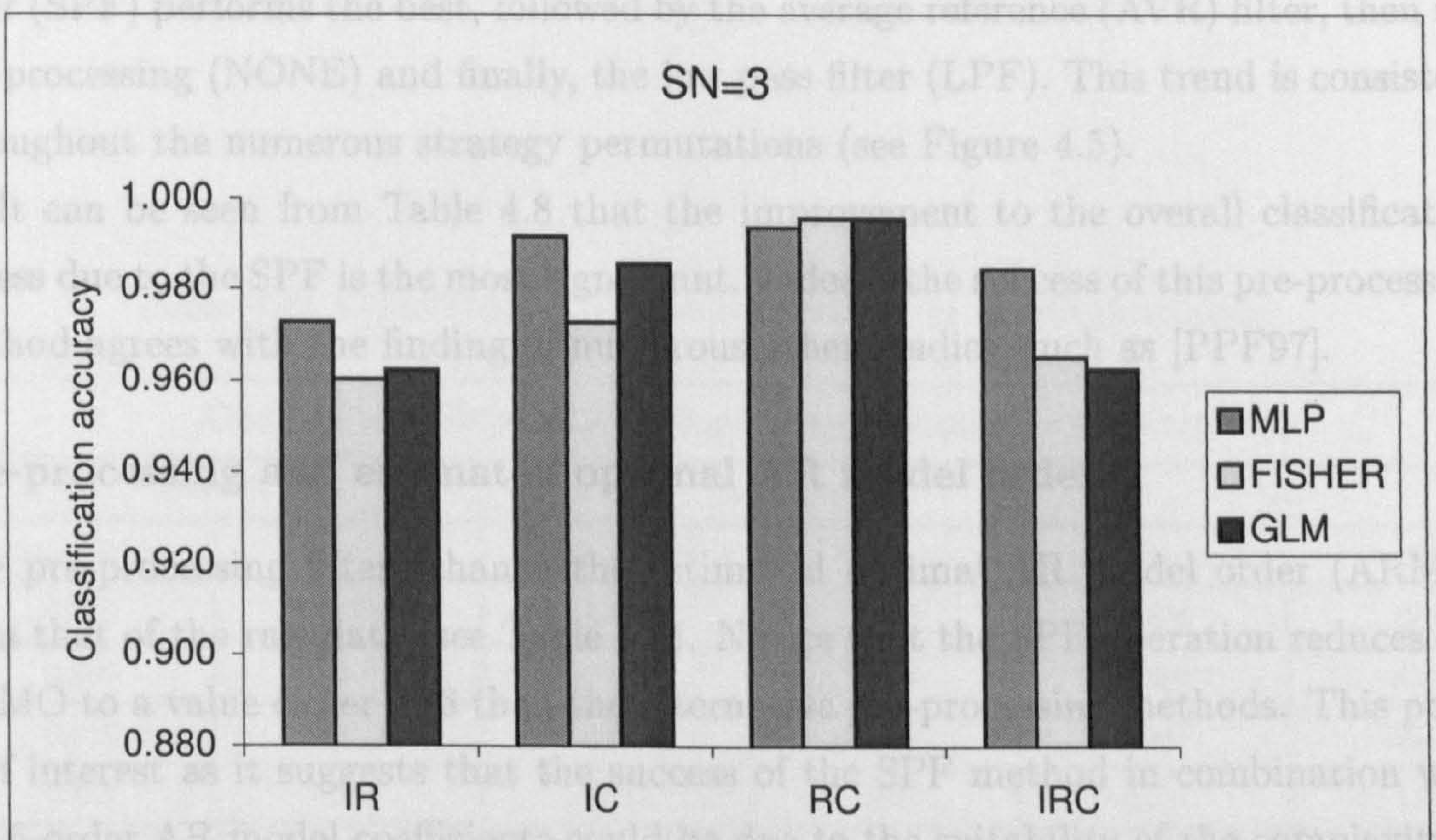

Figure 4.4: Bar plot of classification fitness for all four condition comparisons as a function of classifier. SN=3

Table 4.7: Classification fitness as a function of classifier function.

Classification fitness as a function of classifier	
Mean fitness[1]	CF
0.672	GLM
0.652	FISHER

1. Mean taken over 384 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. CND=IR,IC,RC,IRC. SN=1,2,3.

statistics is in the order of +/- 2 percent. Therefore the FISHER and GLM classifier - with respect to classification fitness - are viewed as being equally successful.

All three classifiers perform within similar regions of accuracy, therefore it is necessary to consider the merits of the dependent pre-classifier variations which would directly effect the overall computational expense of the strategy. This will be dealt with in the conclusions sub-section.

4.4.3 Pre-processing

Of the four pre-processing variations (NONE, AVR, SPF, LPF) the Laplacian spatial filter (SPF) performs the best, followed by the average reference (AVR) filter, then no-pre-processing (NONE) and finally, the low pass filter (LPF). This trend is consistent throughout the numerous strategy permutations (see Figure 4.5).

It can be seen from Table 4.8 that the improvement to the overall classification fitness due to the SPF is the most significant. Indeed, the success of this pre-processing method agrees with the finding of numerous other studies, such as [PPF97].

Pre-processing and estimated optimal AR model order

The pre-processing filters change the estimated optimal AR model order (ARMO) from that of the raw data (see Table 4.9). Notice that the SPF operation reduces the ARMO to a value closer to 6 than the alternative pre-processing methods. This point is of interest as it suggests that the success of the SPF method in combination with the 6-order AR model coefficients could be due to the suitability of the complexity of the pre-processed data with the 6th order AR model. If this were the case, one might expect that the classification fitness for the AR feature extraction method might be a function of the optimal model order estimate. However, when looking at the data more carefully, it appears that there is no trend indicating the above nature. Table 4.10

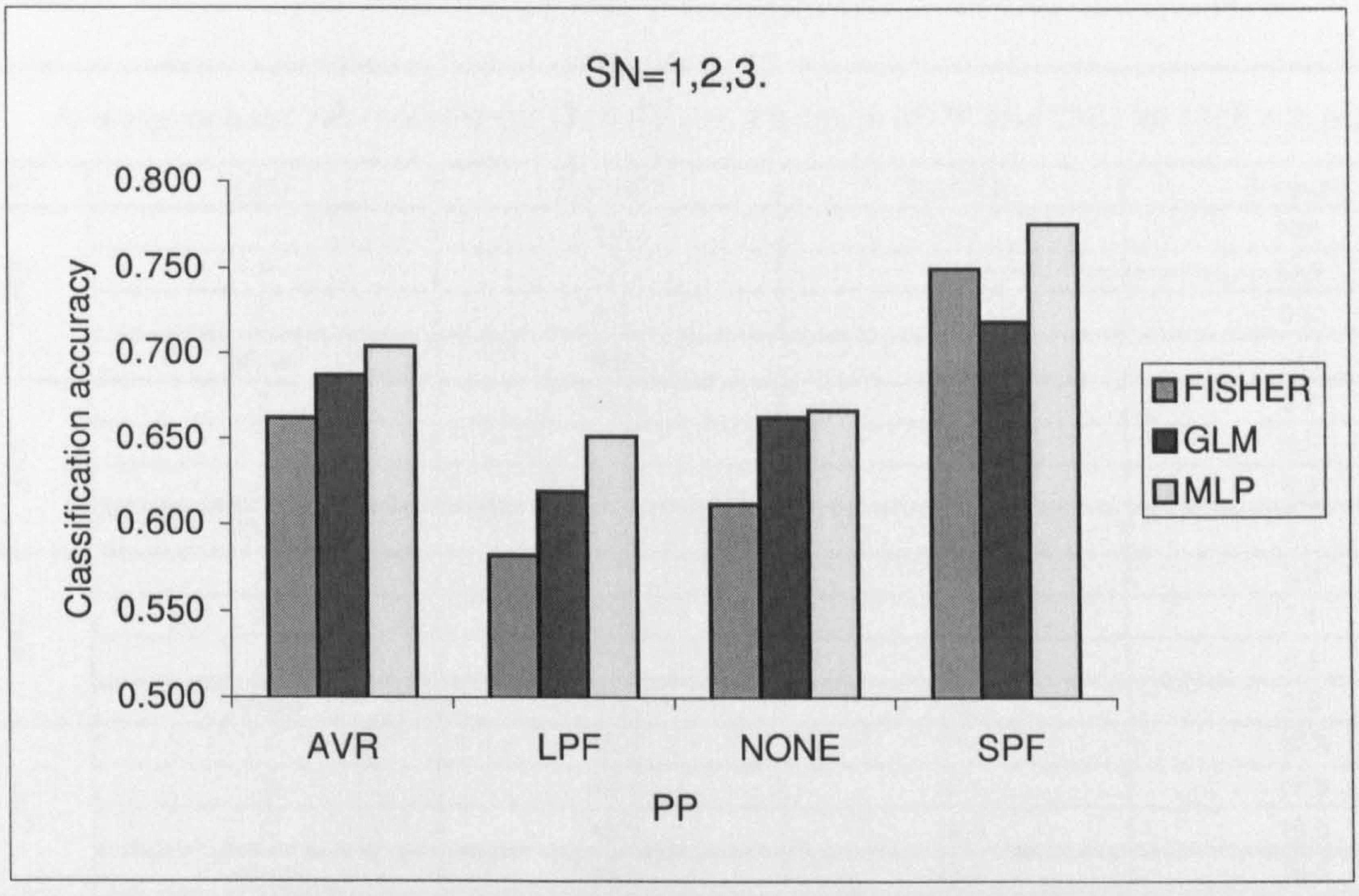

Figure 4.5: Bar plot of classification fitness for all four condition comparisons as a function of pre-processing. SN=1,2,3.

Table 4.8: Classification fitness as a function of pre-processing.

Classification fitness as a function of pre-processing for each classifier					
FISHER		GLM		MLP	
Mean fitness[1]	PP	Mean fitness[2]	PP	Mean fitness[3]	PP
0.750	SPF	0.719	SPF	0.776	SPF
0.662	AVR	0.687	AVR	0.704	AVR
0.612	NONE	0.663	NONE	0.667	NONE
0.582	LPF	0.619	LPF	0.651	LPF
1. Mean taken over 96 variations: FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=FISHER, CND=IR,IC,RC,IRC. SN=1,2,3.		2. Mean taken over 96 variations: FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=FISHER, CND=IR,IC,RC,IRC. SN=1,2,3.		3. Mean taken over 12 variations: FX=AR,ARMO,FFT,PSD. FS=FS1. SR=9:1, CF=MLP, CND=IR. SN=1,2,3.	

Table 4.9: The effects of pre-processing on the estimated optimal AR model order (ARMO).

Average optimal AR model order (ARMO[1]) as a function of PP and CND for each subject				
PP	CND	Subject 1	Subject 2	Subject 3
NONE	I	9.4	14.0	10.6
	R	9.1	14.0	10.1
	C	8.5	14.0	9.6
	Mean	9.0	14.0	10.1
AVR	I	10.9	10.4	9.7
	R	11.5	10.7	9.6
	C	10.8	10.8	8.9
	Mean	11.1	10.6	9.4
SPF	I	7.7	9.2	6.9
	R	7.5	9.4	7.1
	C	7.1	9.3	6.9
	Mean	7.4	9.3	7.0
LPF	I	18.3	18.4	18.6
	R	18.4	18.4	18.5
	C	18.2	18.5	18.6
	Mean	18.3	18.4	18.5
	Grand Mean	11.4	13.1	11.3

1. ARMO feature set used to calculate above values. All 128 channels excluding bad channels were used.

illustrates this by showing the classification fitnesses for the AR and FFT methods for the four pre-processing methods along with the corresponding average ARMO values. Instead, one finds that it is the pre-processing method and not the ARMO value which governs the success of the strategies.

Artefact removal

A brief look at the impact on overall classification performance with respect to artefact contamination (see Table 4.11) indicates that the removal of bad trials from the analysis makes no impact on the performance of the above strategies. This could be due to several factors, namely:

- Poor or poorly calibrated artefact detection algorithm.
- Low proportion of genuine artefacts.
- Overall classification system is insensitive to artefacts.

Table 4.10: Possible relationship between pre-processing, mean ARMO and classification fitness.

Possible relationship between PP, Mean ARMO and Classification fitness				
SN	PP[3]	Mean ARMO	Mean fitness (AR)[1]	Mean fitness (FFT)[2]
1	SPF	7	0.945	0.815
	AVR	11	0.851	0.596
	NONE	9	0.838	0.565
	LPF	18	0.776	0.587
2	SPF	9	0.938	0.762
	AVR	11	0.906	0.636
	NONE	14	0.900	0.559
	LPF	18	0.846	0.587
3	SPF	7	0.963	0.583
	AVR	9	0.892	0.484
	NONE	10	0.906	0.476
	LPF	19	0.912	0.512
ARMO feature set used to calculate Mean ARMO. All 128 channels excluding bad channels were used. Mean fitnesses are calculated by averaging all the permutations of the strategieswhere: CF=GLM. FS=FS2. CND=IRC. FX= (1) AR (2) FFT. (3) PP column ordered according to the best-to-worst findings of the PP analysis.				

The issue of how to handle artefacts in EEG and general signal analysis is an entire topic in its own right. If nothing else, this brief exploration into the matter confirms this.

4.4.4 Feature extraction

Irrespective of the classifier used, the 6th order autoregressive model coefficients (AR) representation performs significantly better than the alternatives (see Figure 4.6). The AR model order estimation (ARMO) representation fairs the worst, with the Fourier methods (FFT, PSD) somewhere in-between (see Table 4.12).

Most noteworthy is the finding that the AR representation significantly out-performs (10% + improvement) the Fourier methods. This finding is generally confirmed by other spontaneous-EEG pattern classification studies, for example [AS96, PPF97, PRCS00].

4.4.5 Feature selection

In general, both MLP and FISHER classifiers perform best when using the largest number of channels (FS3), whereas the GLM perform best with the smaller 10-20 grouping (FS2) (see Table 4.13). When only the optimal strategies for each classifier

Table 4.11: Effect on classification fitness after removing bad trials. Bad trials referring to segments of EEG that were identified by the muscle artefact detection algorithms.

Effect on classification fitness after removing bad trials			
SN	Mean fitness[1]		PP
	Bad trials removed	All trials used	
2	0.993	0.989	SPF
	0.964	0.965	AVR
	0.929	0.939	NONE
	0.881	0.876	LPF
3	0.896	0.963	SPF
	0.853	0.893	AVR
	0.861	0.857	NONE
	0.832	0.829	LPF

1. Mean fitness is the strategy fitness where: FX=AR. FS=FS2. CND=IR. CF=GLM.

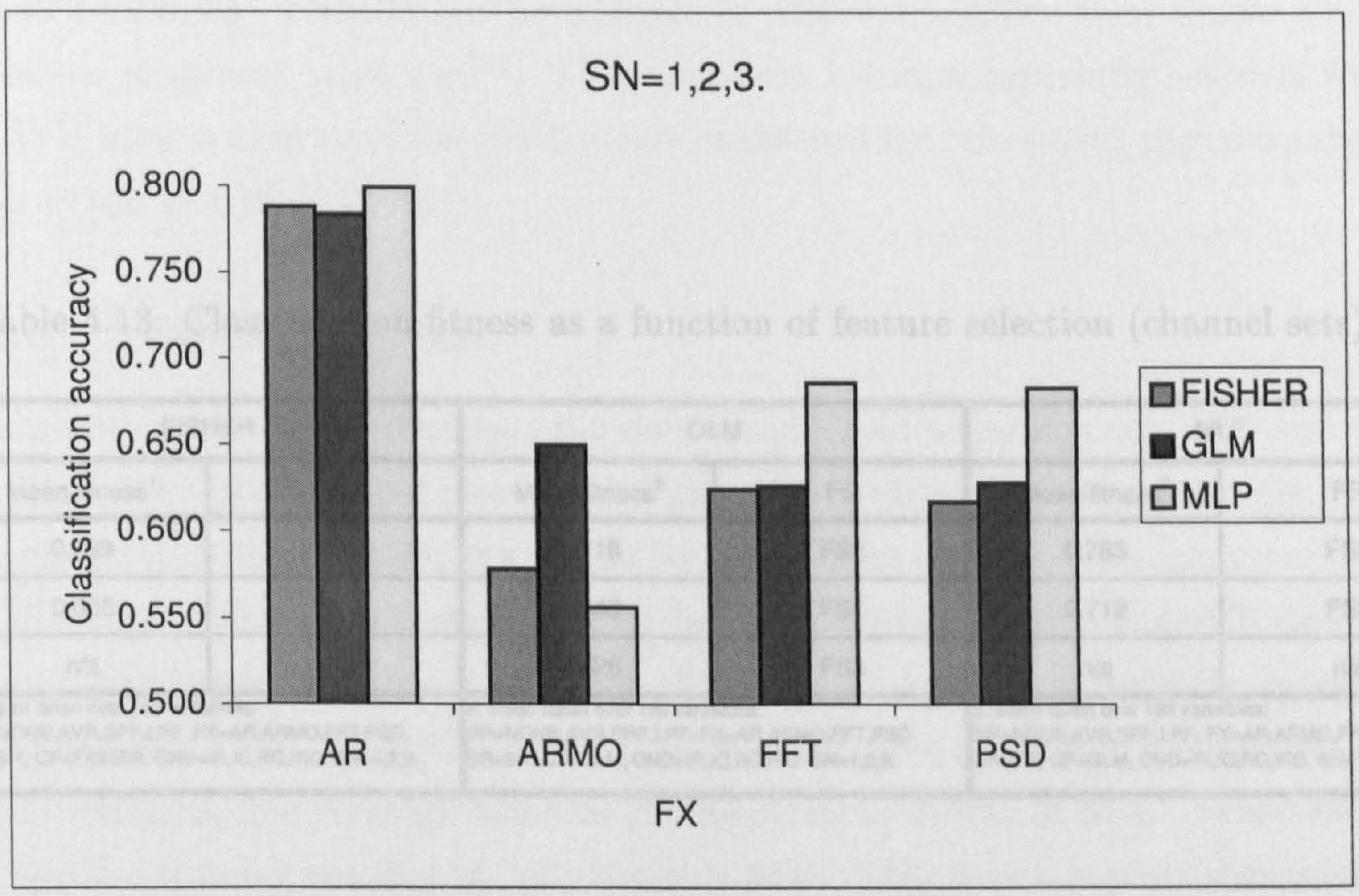

Figure 4.6: Bar plot of classification fitness for all four condition comparisons as a function of feature extraction. SN=1,2,3.

Table 4.12: Classification fitness as a function of feature extraction variations (and classifier variations).

Classification fitness as a function of feature extraction for each classifier					
FISHER		GLM		MLP	
Mean fitness[1]	FX	Mean fitness[2]	FX	Mean fitness[3]	FX
0.788	AR	0.784	AR	0.799	AR
0.624	FFT	0.650	ARMO	0.686	FFT
0.617	PSD	0.628	PSD	0.683	PSD
0.578	ARMO	0.626	FFT	0.556	ARMO
1. Mean taken over 96 variations: PP=NONE,AVR,SPF,LPF. FS=FS2,FS3. SR=9:1, CF=FISHER, CND=IR,IC,RC,IRC. SN=1,2,3.		2. Mean taken over 96 variations: PP=NONE,AVR,SPF,LPF. FS=FS2,FS3. SR=9:1, CF=GLM, CND=IR,IC,RC,IRC. SN=1,2,3.		3. Mean taken over 4 variations: PP=NONE,AVR,SPF,LPF. FS=FS1. SR=9:1, CF=MLP, CND=IR. SN=1,2,3.	

Table 4.13: Classification fitness as a function of feature selection (channel sets).

FISHER		GLM		MLP	
Mean fitness[1]	FS	Mean fitness[2]	FS	Mean fitness[2]	FS
0.699	FS3	0.718	FS2	0.783	FS3
0.605	FS2	0.640	FS1	0.712	FS2
n/a	n/a	0.626	FS3	n/a	n/a
1. Mean taken over 192 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. SR=9:1, CF=FISHER, CND=IR,IC,RC,IRC. SN=1,2,3.		2. Mean taken over 192 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. SR=9:1, CF=GLM, CND=IR,IC,RC,IRC. SN=1,2,3.		2. Mean taken over 192 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. SR=9:1, CF=GLM, CND=IR,IC,RC,IRC. SN=1,2,3.	

are considered, i.e. PP=SPF, FX=AR, it is found that the improvement due to optimal FS choice is in the order of a 5%. For example, for condition pair IR, the MLP classifier yields 99.2% when using FS3, compared to only 95.8% when using the smaller FS2 channel set. Of the three classifiers, the GLM seems to yield best fitness for the smaller channel sets, a result which sets it apart from the other classifiers in terms of computational and practical complexity.

Note that when comparing the results of strategies utilising the smaller channels sets (FS1 and FS2) in combination with either the Laplace filter (SPF) or the average reference filter (AVR) with strategies which use the full channel set (FS3), it must be remembered that both of these pre-processing options require additional electrodes (all the electrodes are used for the AVR calculation, whereas all the nearest neighbours are used for the SPF filter). This becomes relevant only when the merits of reduced channel sets over dense arrays are being pronounced. For example, consider the results from subject 1 and the strategies employing the 4 temporal channels (i.e. FS=FS1) as well as the following settings: FX=AR, SR=9:1, CF=GLM. One finds that the SPF strategy yields a significantly higher classification fitness (82%) over the no pre-processing strategy (77%). However, because the SPF filter actually uses more than 4 channels in its derivation, the engineer must bear in mind that, should a 4-channel system be chosen, it would not be possible to utilise a Laplace filter unless special Laplacian electrodes were used[10]. This argument becomes especially relevant when the AVR filter is used since the effectiveness of the average referencing requires a dense array in the first place [Die98].

4.4.6 Training set size

Before evaluation by the classifier, the data is split into a training set (ES) and test set (TS). The proportional size of these sets is determined by the split ratio (SR). It is found that the SR resulting in the larger training set size, $9:1$, only performs slightly better than the alternative, $1:1$ (see Table 4.14). This is generally true for comparisons between pairs of strategies. For example, subject 1, with the following attributes: PP=SPF, FX=AR, FS=FS2, CND=IRC. For the larger training set size of 648 patterns, the strategy achieves a classification fitness of 87%. Whereas, for the smaller training set size of 360, it yields 84%. This is not a great difference in performance, despite the big difference in the size of ES. Obviously, the smaller number of training-patterns required, the better, since a BCMI which minimises the

[10]Such Laplacian electrodes are available commercially as self contained sensors.

Table 4.14: Classification fitness as a function of training set size.

Classification fitness as a function of split ratio (training set size)	
Mean fitness[1]	SR
0.661	9:1
0.635	1:1

1. Mean taken over 576 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS1,FS2,FS3. CF=GLM. CND=IR,IC,RC,IRC. SN=1,2,3.

amount of training sessions will be favourable. If successful strategies could be found, it would necessary to determine a suitable lower-limit for the training set size.

4.4.7 Conditions

Of the 4 possible variations of condition comparisons (IR, IC, RC and IRC), the former (IR) is the most important. Musical imagery is the primary mental task which this experiment is based on, and relaxing (or passive listening) is the baseline task. If successful classifications can be made between unseen EEG segments recorded whilst the subject is engaged in either of these two tasks, then this will help the strengthen case for the BCMI concept. Table 4.15 demonstrates that all the four condition pairs are classified to similar fitnesses.

The fact that strategies tested on the three 2-way condition pairs yield similar classifications is slightly concerning. Consider tasks I and C. Both require a the subject to make concerted mental effort, whereas task R being passive requires little mental effort. This point was discussed at length earlier in the chapter. The point here is that on a hierarchical scale of concentration (or mental effort), tasks I and C should rank comparatively higher than R. The fact that segments belonging to conditions I and C can be successfully distinguished from each other - and just as well as with the active-passive combinations (IR and CR) - suggests that the finer aspects of the brains' functioning (i.e. musical imagery versus numerical counting) are being picked up. Following this line of thinking, one would expect the active-passive combinations to yield better results than the more challenging IC combination. However, this is not the case, therefore a closer look at the reasons behind this behaviour are given below.

Table 4.15: Classification fitness as a function of condition variations (and classifier variations). 2-way classifications are of similar order of classification fitness, whereas, the 3-way problem is somewhat less successful.

Classification fitness as a function of condition-pairs for each classifier			
FISHER		GLM	
Mean fitness[1]	CND	Mean fitness[2]	CND
0.715	RC	0.684	RC
0.688	IC	0.678	IC
0.683	IR	0.676	IR
0.520	IRC	0.650	IRC
1. Mean taken over 96 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=FISHER. SN=1,2,3.		2. Mean taken over 96 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=MLP. SN=1,2,3.	

Effects of blocking conditions

During the EEG acquisition stage of the experiment, trials are arranged in a single-condition-per-block fashion. This block-wise ordering of trials reveals an interesting result which suggests there is a relationship between the within-subject classification fitness of the 3 possible 2-way classifications - IR, IC and RC - and the order in which the blocks are presented during the experiment. Table 4.16 demonstrates that the condition pair which performs best corresponds to the order (position in time) of the blocks. This relationship could be explained by the idea that throughout the recording phase of the experiment there is a time dependent component to the EEG which grows as the experiment goes on.

Suppose that there was a time dependent component of the EEG which increased steadily whilst the experiment was underway. For example, as the experiment progresses the subject's attention might become weaker leading to an increasing magnitude of the alpha band component (related to the state of relaxation / sleepiness. If this were the case one would expect to find that trials recorded further into the experiment would contain a greater prevalence of this component. A quantitative pattern classification system operates on the statistical properties of the data choosing the salient features in order to perform the classification [CDA93a]. It will always converge to the simplest solution which is why the data must be free from hidden decoys - features which would lead to a classification, but for the wrong reasons. In cases where the engineer is unaware of the underlying nature of the data, they cannot know which features to exclude or account for. This lack of prior knowledge must

Table 4.16: Classification fitness as a function of block ordering.

Relationship between block ordering and 2-way classification fitness							
SN	Ordering	Relative Condition Pair Ranking			Mean classification fitness[1]		
		Best	Medium	Worst	IR	IC	RC
1	I R C I R C	IC	RC	IR	0.922	0.965	0.928
2	R C I R C I	IR	IC	RC	0.989	0.932	0.927
3	C I R C I R	RC	IR	IC	0.963	0.986	0.996
1. PP=SPF. FX=AR. FS=FS2. CF=GLM. SR=9:1.							

Table 4.17: Classification fitness as a function of subject.

Classification fitness as a function of subject for each classifier					
FISHER		GLM		MLP	
Mean fitness[1]	SN	Mean fitness[2]	SN	Mean fitness[3]	SN
0.674	1	0.682	1	0.764	1
0.649	3	0.672	3	0.730	3
0.632	2	0.661	2	0.715	2
1. Mean taken over 128 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=FISHER, CND=IR,IC,RC,IRC.		2. Mean taken over 128 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=9:1, CF=GLM, CND=IR,IC,RC,IRC.		3. Mean taken over 256 variations: PP=NONE,AVR,SPF,LPF. FX=AR,ARMO,FFT,PSD. FS=FS2,FS3. SR=1:1,9:1, CF=MLP, CND=IR,IC,RC,IRC.	

be countered for by rigorous experimental design, so as to ensure that, as best as possible, no masking factors creep into the data and caused a false positive result.

This finding highlights a possible flaw in the experimental design. It is possible that the success of the classification strategies described above are due to something other than the EEG components related to the mental tasks of interest. If this is the case, then the results obtained from this experiment must be treated as tentative.

4.4.8 Subjects

Table 4.17 demonstrates that all three subjects' data are classifiable to a similar extent.

4.5 Conclusions

A novel musical imagery experiment has been developed targeting BCMI systems as future applications. This experiment is also designed to validate whether state-of-the-art DSP might be able to perform computerised pattern classification of a subject's EEG whilst he/she is engaged in active musical imagery tasks, as opposed to passive listening or counting tasks. The EEG pattern classification system described in Chapter 2 is applied and tested on the data from 3 subjects. Results are mixed.

The data recorded from 3 subjects has been analysed with a variety of classification strategies, incorporating numerous pre-processing, feature extraction, channel selection, and classifier variations. The general behaviour of the classification system acts accordingly with respect to the results of similar research into spontaneous-EEG pattern classification. The following list summarises these findings:

- The Laplacian spatial filter as a pre-processing step significantly outperforms both average reference filtering (also known as common referencing) and raw un-filtered data.
- Linear autoregressive model coefficients significantly outperform Fourier transform based representations as a feature extraction method.
- Classification of mental tasks might be possible with a handful of EEG sensors, as opposed to dense arrays of electrodes.
- Nonlinear classifiers, such as the MLP, may not be significantly better than simpler linear alternatives, such as the Fisher discriminant.

At first glance, the classification figures seem outstanding, ranging from 90 to 99%. Whilst the overall classification performances are very high (in the order of 95% +), there is some doubt as to whether this success is due to in fact miss-classification due to time-related experimental artefacts born from a weakness in the design of the paradigm. The results can be contaminated (due to the way that trials are blocked according to condition, as opposed to mixing them up so that each block contained all three tasks) and this could be remedied by paying careful attention to the design of off-line experiments which are aimed at testing tasks for BCI technology. The closer the experiment can be made to the envisaged real world situation, the better. Clearly, this experiment demonstrates that the *systemised BCMI evaluation protocol* works.

Chapter 5

Musical Focusing Experiment

5.1 Introduction

This chapter presents another novel experiment, similar to the previous one, which has been designed with the end use of a BCMI system in mind. The EEG pattern classification methodology employed is the same as that used for the musical imagery experiment. However, the mental tasks and paradigm are different.

In this section, motivational issues are given, as well as a description of the engineering problem intrinsic to the design of the experiment. The remaining sections give a detailed explanation of the experimental paradigm, results and discussion, and conclusions. Since the classification methodology used is practically identical to that used in the imagery experiment - there is no need to re-present it in this chapter.

5.1.1 Motivation

The musical focusing experiment has been designed with the following hypothesis in mind:

> *firstly, there exists information in the EEG which allows one to identify whether a person is engaged in one of two mental tasks: musical focusing or passive listening. Secondly, one can establish from the EEG alone, which ear the music the person is focusing on is being heard through the headphones.*

In the context of this experiment, musical focusing requires the subject to pay special attention to a particular part of the music they are listening to. For example, suppose you are listening to a piece of pop music. There will be many instruments, all playing something different. The instruments may even be placed apart in the stereo mix. For example, the vocal part might be centre panned[1], the guitar panned hard-left, and the keyboards hard right. The drums could be panned according to the actual position of the parts making up the kit. Suppose you notice that there is flute part hiding somewhere amongst all the other parts. If you listened carefully and tried to pick out this part from the others, then you would be focusing - deliberately steering you awareness towards that particular part in favour over the others. This is just one example of musical focusing of which there are many others.

Suppose that it is possible to tell from a persons' EEG whether they are performing the mental task of focusing, as opposed to just listening to the music in a normal relaxed manner. Furthermore, suppose that it is possible to tell whether they are focusing on a part of the music which is panned hard-left, centre, or hard-right. The ability to do this would open up numerous possible BCMI applications. One such system is described below.

5.1.2 Hypothetical BCMI system utilising musical focusing

Figure 5.1 shows how musical focusing could be utilised in a BCMI context. Taking the BCMI system described in Chapter 1 (Figure 1.1) as a basic framework, the *EEG analysis engine*, capable of identifying when the person is focusing and whether they are focusing on a part which is placed left or right in the stereo mix, would be communicating with the *co-ordinator*. The co-ordinator would instruct the *music engine* in a way which related to which part the person had just been focusing on. For example, suppose the person hears a guitar part appear[2] in the right hand side of the mix, and liking it, they focus on it. The co-ordinator (which is responsible for telling the music engine what to play, then waiting for the EEG analyser's response) would interpret this as 'person is focusing on guitar part', since it knows that the guitar part is right in the current mix. Then, depending on a set of rules, the co-ordinator would take some action, such as incorporating the guitar as a regular continuous part in the overall mix, or adding new variations for the performer to 'choose'. The possibilities are endless. This framework would allow the performer / participant to

[1] 'Panning' refers to the relative volume (loudness) that a sound makes in each of a stereo pair of speakers. The three extremes are hard-left (sound only heard in left speaker), center (sound heard in both speakers at an equal level), and hard-right (sound in right speaker only).

[2] Assume that there are some initial conditions, or rules, that as well as starting the music off at the beginning, would also introduce new parts as the music evolved. The integration of these parts could then be influenced by the performer.

steer their own unfolding musical performance without touching a button or making any other gesture. They would simply involve themselves in the experience, using musical focusing like a conductor uses a baton.

Musical focusing is, for the following reasons, a prime candidate mental-task for the BCMI concept:

- It is a natural activity in the context of musical experience.
- It is a simple task which requires only an experiential understanding of music. In other words, everyone who has grown up with music should be able to do it.
- It requires a deliberate mental effort of a kind which can be set apart from other types of musical tasks, such as musical imagery.
- It would readily find use in a BCMI system such as the one described above.

5.1.3 Classification problem

The experiment described below allows the following classification problems to be tackled:

1. successfully determine, on a segment by segment basis, which class - musical focusing or passive listening - a 2-second multi-channel EEG segment belongs to, where the class is named after the mental task the subject is performing while the segment is being recorded.
2. Within the focusing set of data, successfully determine which headphone speaker the target part was heard in. In other words, ascertain from the EEG whether the subject was focusing on music heard through the left ear or right ear.

5.1.4 Objectives

Seek a solution to the above problem by employing a variety of digital signal processing methods based on existing methods found to be of success in the EEG pattern classification field, thus gaining insight into the validity of the hypothesis stated above, and whether the methods explored are likely to be suitable for a working BCMI system[3].

[3]This experiment was designed and performed at the same time as the imagery experiment.

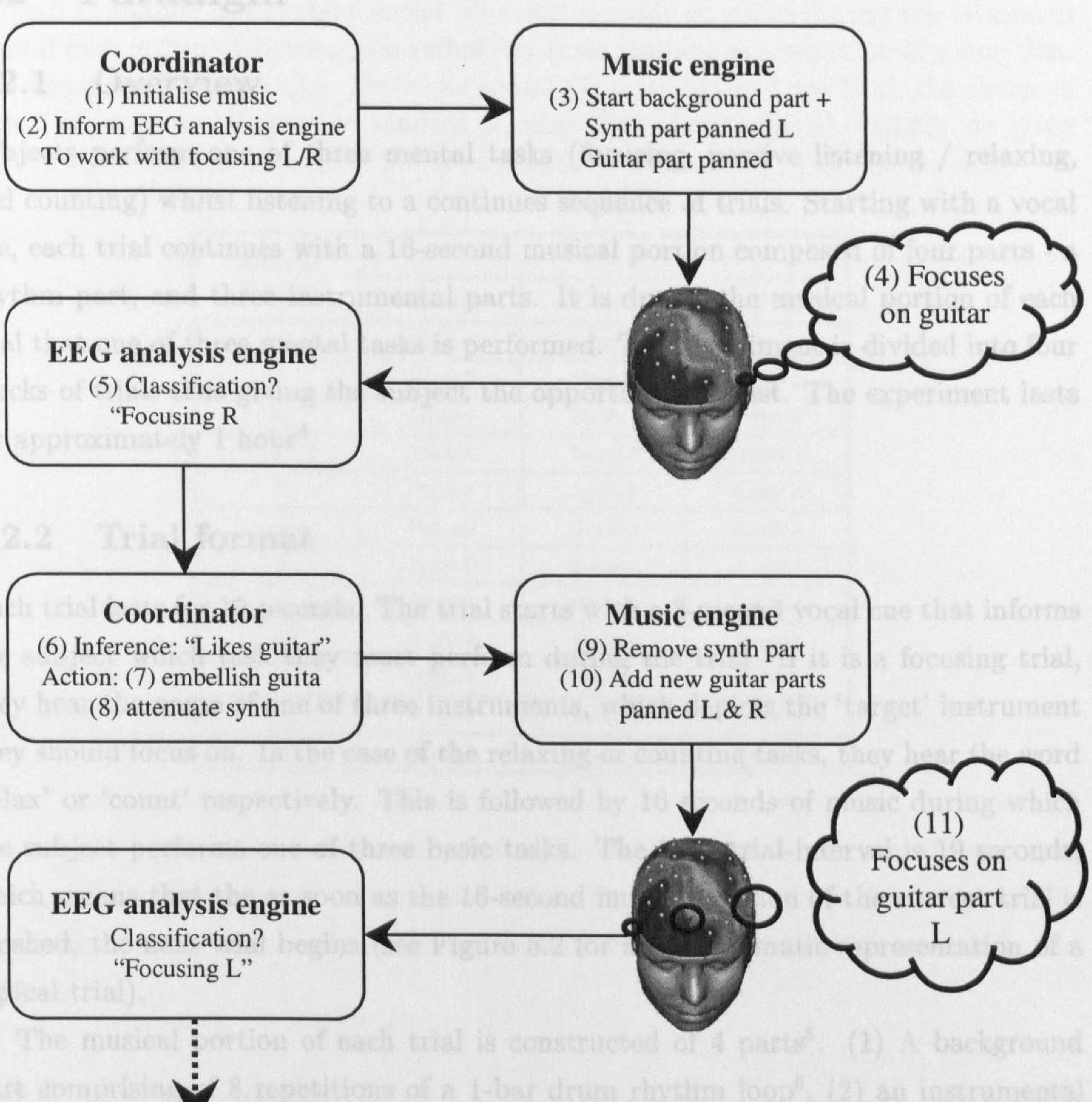

Figure 5.1: Hypothetical BCMI system using focusing. (1) *Co-ordinator* informs *music engine* to start music, and (2) tells *EEG analysis engine* to classify data as *focusing-left* or *focusing-right*. (3) Music engine starts background part along with a synth part (panned left) and guitar part (panned right). (4) The performer, preferring the guitar part, focuses on it. (5) The *EEG analysis engine* detects this and relays it to the *co-ordinator* which, in turn (6) makes the inference that the performer likes the guitar part. At this stage, action is taken (which would be based on a pre-defined rule) in the form of (7) instructing the *music engine* to embellish the guitar and (8) attenuate the synth. (9) The *music engine* thus removes the synth part and (10) adds two new guitar parts (panned L and R) based on the previous one. (11) The performer, preferring the guitar part in the left of the mix, focuses on it. (12) Finally, the *EEG analysis engine* classifies the EEG data as *focusing-left*, and so on.

5.2 Paradigm

5.2.1 Overview

Subjects perform one of three mental tasks (focusing, passive listening / relaxing, and counting) whilst listening to a continues sequence of trials. Starting with a vocal cue, each trial continues with a 16-second musical portion composed of four parts - a rhythm part, and three instrumental parts. It is during the musical portion of each trial that one of three mental tasks is performed. The experiment is divided into four blocks of trials thus giving the subject the opportunity to rest. The experiment lasts for approximately 1 hour[4].

5.2.2 Trial format

Each trial lasts for 19 seconds. The trial starts with a 3-second vocal cue that informs the subject which task they must perform during the trial. If it is a focusing trial, they hear the name of one of three instruments, which depicts the 'target' instrument they should focus on. In the case of the relaxing or counting tasks, they hear the word 'relax' or 'count' respectively. This is followed by 16 seconds of music during which the subject performs one of three basic tasks. The inter-trial-interval is 19 seconds, which means that the as soon as the 16-second musical portion of the current trial is finished, the next trial begins (see Figure 5.2 for a diagrammatic representation of a typical trial).

The musical portion of each trial is constructed of 4 parts[5]. (1) A background part comprising of 8 repetitions of a 1-bar drum rhythm loop[6], (2) an instrumental part, panned hard-left, (3) a second instrumental part, center panned, (4) and a third instrumental part, panned hard-right. Each instrumental part is comprised of eight repetitions of a 1-bar riff[7] played on one of three instruments: a synthesiser, electric guitar and piano. Altogether there are 24 unique riffs which are split among the three instruments resulting in eight synth, electric guitar and piano parts respectively. The

[4]The focusing experiment was run in conjunction with the imagery experiment. Consequently, subjects were involved in the experiments for approximately 2 and a half hours.

[5]All musical parts were created using a multimedia PC computer using a MIDI-audio sequencer package, an electric guitar with amplifier simulator, and an audio i/o sound card. The music composed is in the style of a moderate tempo popular dance / club tune (120 beats per minute, 4 beats per bar). These musical parts are the same as those used in the imagery experiment.

[6]Loop: popular music jargon for a segment of music which, when played in a continuous 'loop' sounds seamless, as if it were being played as a continuous part with no obvious beginning and end.

[7]Riff: popular music jargon for a short catchy melody which is usually repeated many times in the course of a song.

Table 5.1: Details of the eight sound files that provide stimulus during the 16 second mental task portion following the verbal cue (also realised as a selection of sound files. See Appendix B for details). Only one sound file is auditioned per trial, the choice of which depends on the pseudo random audition list (Appendix B). Details are given as to which instrumental parts are used and their positioning in the stereo field.

Sound file number	Panning		
	Hard-left	Centre	Hard-right
1	Synth-1	Guitar-1	Piano-3
2	Synth-2	Piano-1	Guitar-3
3	Guitar-1	Synth-1	Piano-4
4	Guitar-2	Piano-2	Synth-3
5	Piano-1	Synth-2	Guitar-4
6	Piano-2	Guitar-2	Synth-4
7	Synth-5	Guitar-3	Piano-5
8	Synth-6	Piano-3	Guitar-5
9	Guitar-6	Synth-3	Piano-6
10	Guitar-7	Piano-4	Synth-7
11	Piano-7	Synth-4	Guitar-8
12	Piano-8	Guitar-4	Synth-8

instrumental parts are 'auditioned' (see Table 5.1) so that there are always three instruments playing, each one panned either hard-left, centre, or hard-right. The reason for having eight unique riffs for each instrument is to provide variety for the subject.

5.2.3 Mental tasks

There are three main mental tasks: focusing (F), passive listening / relaxing (R) and counting (C). Data gathered whilst subjects perform these tasks are allocated to one of three conditions who's names are the same as the tasks. The focusing condition is further divisible into two sub-conditions: focusing-left (FL) and focusing-right (FR), dependent on which headphone speaker - left or right - the target instrument is heard in[8]. Before the experiment begins, subjects are given the following instructions as to the nature of the tasks:

- Focusing task. Listen to the looped riffs whilst focusing especially hard on the target part, which belongs to the instrument that is defined during the cue at

[8]Note that for the sake of simplicity, it was decided not to include a third sub-condition for focusing on center panned instrumental parts.

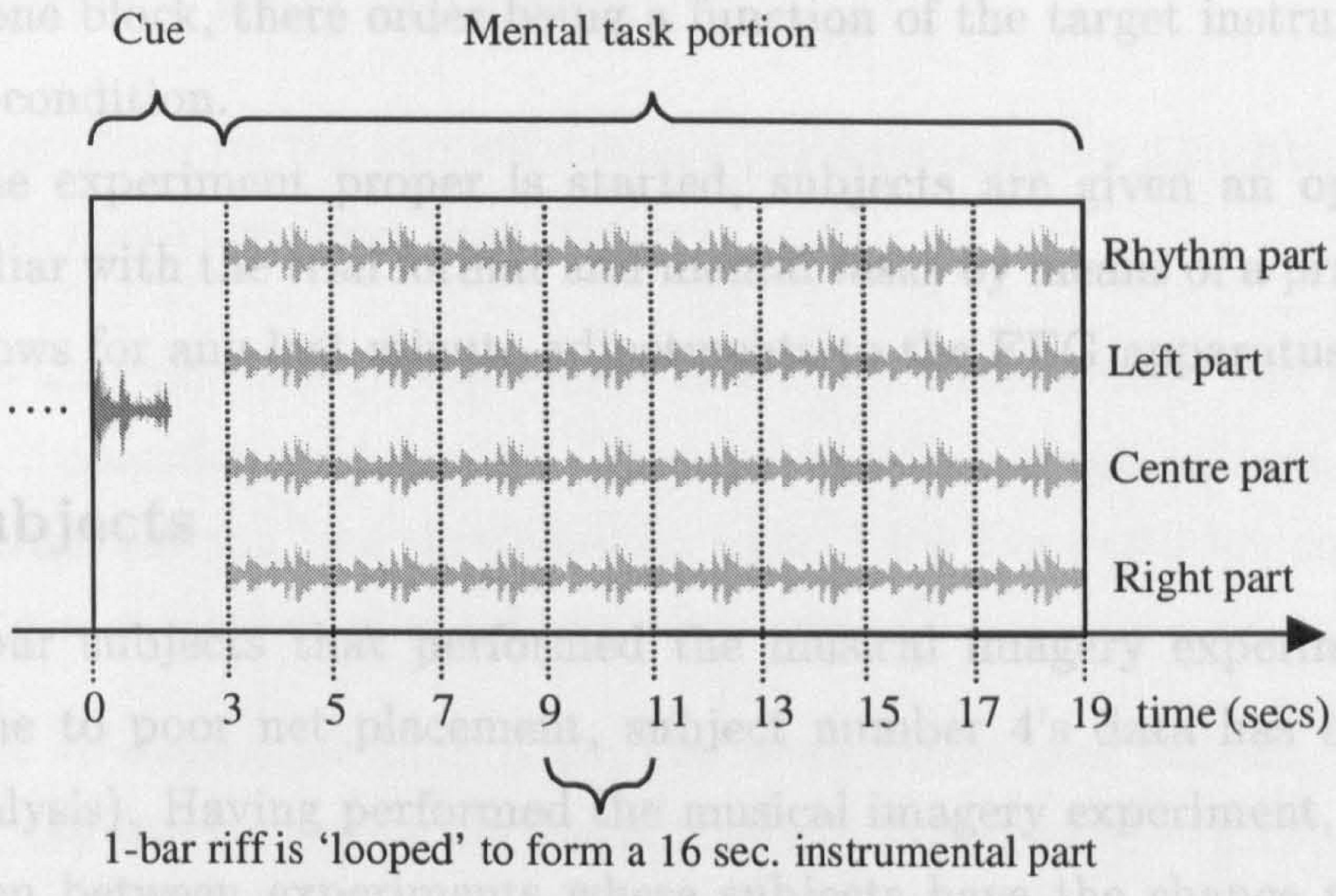

Figure 5.2: Diagrammatic representation of a focusing experiment trial.

the beginning of the trial.

- Passive listening task. Listen to the entire 4 bar trial with no effort, just relax.
- Counting task. Count the following self repeating sequence - 1, 10, 3, 8, 5, 6, 7, 4, 9, 2, 1, 10 and so on - until the trial is finished[9].

Subjects are informed that the instrument they are cued to focus on in each trial will either be panned hard-left or hard-right. However, they are not informed as to the experimental relevance for doing this.

5.2.4 Blocks

A blocked system is adopted whereby an equal number of trials from each of the three main conditions are presented in a pseudo random order (Appendix B). Altogether there are four blocks, each consisting of 12 focusing trials (six with left target and six with right target parts), 12 relaxing trials and 12 counting trials. This makes a total of 36 trials lasting 19 seconds each, hence the experiment lasts for about 45 minutes (not including rests in-between blocks).

Within each block the trials are auditioned in a pseudo random order, so that the subject does not learn to predict which type of trial is coming next. Each of the 12 sound files (see Table 5.1) which are comprised of the 4 musical parts, are played once

[9] The rational for including the counting task is the same as for the imagery experiment (see Chapter 4 for discussion).

only in any one block, there order being a function of the target instrument and the panning sub-condition.

Before the experiment proper is started, subjects are given an opportunity to become familiar with the trial format and mental tasks by means of a practice session. This also allows for any last minute adjustments to the EEG apparatus to be made.

5.2.5 Subjects

The same four subjects that performed the musical imagery experiment are used (however, due to poor net placement, subject number 4's data has been excluded from the analysis). Having performed the musical imagery experiment, a five-minute break is given between experiments where subjects have the chance to get up and walk around.

5.2.6 Data acquisition and segmentation

The same protocol is used for EEG acquisition as described in the previous chapter. Each trial is segmented into eight non-overlapping 2-second segments. In this way, each subject yields 1152 segments comprising of 384 focusing (which can be further divided into an equal number of focusing-left and focusing-right segments), 384 relaxing and 384 counting segments.

5.3 Classification methodology

The classification methodology used here is essentially the same as for the imagery experiment, hence a full explanation of the methodology is not necessary. Instead, the differences are listed below:

- *Amount of data.* The focusing experiment yields a greater number of data segments for each class, and therefore the sizes of training and test sets are different from those in the imagery experiment.

- *MLP classifier.* Only one static multilayer perceptron neural network classifier (MLP) is evaluated. It has the following properties. 8 units in the hidden layer, two or three units in the output layer, for two-way and three-way classifications respectively, training lasts for 50 epochs in batch mode.

5.4 Results and discussion

The analysis results from the various strategy permutations are presented in numerous tables of which there are three basic types:

1. *Top-ten optimal strategies for each condition combination*: one table for each subject / classifier combination. These tables take a single subject and classifier, and present the ten best strategies for each of the five condition combinations[10], resulting in a total of nine tables: 5.2 – 5.10.

2. *Optimal strategies for each condition combination and subject*: one table for each of the three classifier methods, hence 3 Tables (5.11 – 5.13). These tables condense the results giving only the highest scoring strategy for each subject, classifier and condition combination.

3. *Strategy averages*: one table for each subject (5.14 – 5.16) plus a fourth Table (5.17) for grand strategy averages. These tables present the average classification statistics for each of the main methodology variations namely: PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1, 1:1. CF=FISHER, GLM, MLP. For example, to compute the average classification fitness for the SPF pre-processing method for subject 1, all the strategies for each condition combination, feature extraction method, feature selection method, data split ratio and classifier are taken into the average. The grand strategy averages table further reduces this by combining the data from each subject. The reason for doing these averaging procedures is to reveal general trends for the individual methodological variations.

5.4.1 Pre-processing

The optimal strategies of subject 1 point at SPF as being the best pre-processing method, followed by AVR. However, for subjects 2 and 3, AVR appears to lead, followed by NONE. These findings are particularly evident in strategies which utilise the FISHER and MLP classifier. A similar trend is apparent when looking at the strategy averages and grand average tables.

[10]Note that in the case of the GLM classifier, the fifth condition combination F-R-C has was not computed due to the comparatively poor performance of the GLM classifier compared with the FISHER and MLP.

Table 5.2: Ten optimal strategies of subject 1 using the Fisher discriminant classifier.

Top-ten optimum strategies for each condition combination for subject 1 using the FISHER classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.741	0.036	0.703	0.781	0.074	SPF	FFT	FS3	1 1	486	192	192
			0.722	0.025	0.693	0.760	0.052	SPF	PSD	FS3	1 1	486	192	192
			0.716	0.057	0.658	0.789	0.117	SPF	PSD	FS3	9 1	486	342	38
			0.711	0.037	0.658	0.763	0.077	SPF	AR	FS3	9 1	486	342	38
			0.710	0.024	0.682	0.734	0.049	SPF	AR	FS3	1 1	486	192	192
			0.684	0.049	0.605	0.737	0.101	AVR	PSD	FS3	9 1	486	342	38
			0.679	0.086	0.579	0.816	0.177	AVR	AR	FS3	9 1	486	342	38
			0.674	0.044	0.605	0.711	0.091	SPF	FFT	FS3	9 1	486	342	38
			0.665	0.048	0.615	0.740	0.098	AVR	AR	FS3	1 1	342	192	192
			0.658	0.087	0.553	0.789	0.180	NONE	PSD	FS3	9 1	486	342	38
F	R	NONE	0.636	0.022	0.609	0.664	0.045	SPF	FFT	FS3	1 1	486	384	384
			0.621	0.037	0.579	0.671	0.075	SPF	FFT	FS3	9 1	486	684	76
			0.620	0.031	0.578	0.661	0.064	SPF	PSD	FS3	1 1	486	384	384
			0.613	0.038	0.566	0.671	0.078	NONE	AR	FS3	9 1	486	684	76
			0.613	0.053	0.526	0.671	0.109	AVR	AR	FS3	9 1	486	684	76
			0.613	0.038	0.566	0.671	0.078	AVR	PSD	FS3	9 1	486	684	76
			0.609	0.023	0.573	0.635	0.048	SPF	AR	FS3	1 1	486	384	384
			0.608	0.087	0.474	0.711	0.179	SPF	PSD	FS3	9 1	486	684	76
			0.603	0.055	0.566	0.697	0.112	AVR	AR	FS3	9 1	456	684	76
			0.600	0.039	0.570	0.664	0.081	AVR	PSD	FS3	1 1	486	384	384
F	C	NONE	0.816	0.021	0.789	0.842	0.043	SPF	AR	FS3	9 1	486	684	76
			0.805	0.024	0.776	0.829	0.048	SPF	FFT	FS3	9 1	486	684	76
			0.801	0.037	0.766	0.859	0.076	SPF	FFT	FS3	1 1	486	384	384
			0.789	0.053	0.724	0.842	0.110	SPF	PSD	FS3	9 1	486	684	76
			0.781	0.010	0.768	0.792	0.020	SPF	PSD	FS3	1 1	486	384	384
			0.774	0.054	0.697	0.842	0.111	AVR	PSD	FS3	9 1	486	684	76
			0.773	0.023	0.750	0.802	0.047	AVR	FFT	FS3	1 1	486	384	384
			0.769	0.029	0.734	0.802	0.061	AVR	PSD	FS2	1 1	78	384	384
			0.768	0.075	0.645	0.829	0.155	AVR	PSD	FS2	9 1	78	684	76
			0.763	0.101	0.645	0.868	0.207	LPF	PSD	FS3	9 1	486	684	76
R	C	NONE	0.834	0.048	0.776	0.908	0.099	SPF	FFT	FS3	9 1	486	684	76
			0.807	0.019	0.781	0.831	0.039	SPF	FFT	FS3	1 1	486	384	384
			0.801	0.029	0.781	0.852	0.060	SPF	PSD	FS3	1 1	486	384	384
			0.789	0.052	0.711	0.829	0.107	SPF	PSD	FS3	9 1	486	684	76
			0.784	0.036	0.750	0.842	0.073	AVR	FFT	FS3	9 1	486	684	76
			0.782	0.020	0.763	0.807	0.041	AVR	FFT	FS3	1 1	486	384	384
			0.770	0.023	0.742	0.805	0.048	SPF	AR	FS3	1 1	486	384	384
			0.766	0.075	0.711	0.882	0.155	SPF	AR	FS3	9 1	486	684	76
			0.765	0.024	0.727	0.792	0.050	AVR	PSD	FS3	1 1	486	384	384
			0.758	0.044	0.697	0.789	0.091	NONE	FFT	FS3	9 1	486	684	76
F	R	C	0.636	0.022	0.609	0.664	0.045	SPF	FFT	FS3	1 1	486	384	384
			0.621	0.037	0.579	0.671	0.075	SPF	FFT	FS3	9 1	486	684	76
			0.620	0.031	0.578	0.661	0.064	SPF	PSD	FS3	1 1	486	384	384
			0.613	0.038	0.566	0.671	0.078	NONE	AR	FS3	9 1	486	684	76
			0.613	0.053	0.526	0.671	0.109	AVR	AR	FS3	9 1	486	684	76
			0.613	0.038	0.566	0.671	0.078	AVR	PSD	FS3	9 1	486	684	76
			0.609	0.023	0.573	0.635	0.048	SPF	AR	FS3	1 1	486	384	384
			0.608	0.087	0.474	0.711	0.179	SPF	PSD	FS3	9 1	486	684	76
			0.603	0.055	0.566	0.697	0.112	AVR	AR	FS3	9 1	456	684	76
			0.600	0.039	0.570	0.664	0.081	AVR	PSD	FS3	1 1	486	384	384
From total of 480 Strategies: CND=F-R, F-C, F-C-R, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=FISHER. SN=1.														

Table 5.3: Ten optimal strategies of subject 2 using the Fisher discriminant classifier.

Top-ten optimum strategies for each condition combination for subject 2 using the FISHER classifier

CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.641	0.027	0.599	0.667	0.056	AVR	AR	FS3	1 1	456	192	192
			0.605	0.119	0.395	0.684	0.245	NONE	AR	FS2	9 1	72	342	38
			0.601	0.037	0.563	0.641	0.076	AVR	FFT	FS3	1 1	456	192	192
			0.600	0.078	0.474	0.658	0.160	AVR	AR	FS3	9 1	456	342	38
			0.595	0.100	0.526	0.763	0.205	NONE	PSD	FS3	9 1	456	342	38
			0.592	0.039	0.542	0.630	0.080	NONE	AR	FS3	1 1	456	192	192
			0.589	0.105	0.421	0.711	0.215	AVR	FFT	FS3	9 1	456	342	38
			0.584	0.114	0.447	0.711	0.234	NONE	FFT	FS3	9 1	456	342	38
			0.563	0.082	0.474	0.684	0.170	SPF	AR	FS3	9 1	456	342	38
			0.558	0.032	0.521	0.609	0.067	NONE	PSD	FS3	1 1	456	192	192
F	R	NONE	0.603	0.055	0.566	0.697	0.112	AVR	AR	FS3	9 1	456	684	76
			0.592	0.084	0.513	0.711	0.172	AVR	FFT	FS2	9 1	72	684	76
			0.567	0.010	0.557	0.583	0.021	AVR	AR	FS3	1 1	456	384	384
			0.556	0.033	0.505	0.591	0.068	NONE	FFT	FS2	1 1	72	384	384
			0.555	0.025	0.526	0.592	0.052	NONE	AR	FS1	9 1	12	684	76
			0.555	0.054	0.487	0.632	0.111	AVR	PSD	FS2	9 1	72	684	76
			0.554	0.021	0.521	0.576	0.042	NONE	FFT	FS3	1 1	456	384	384
			0.553	0.023	0.521	0.578	0.048	AVR	FFT	FS3	1 1	456	384	384
			0.552	0.030	0.521	0.589	0.062	AVR	PSD	FS2	1 1	72	384	384
			0.547	0.027	0.526	0.592	0.056	AVR	ARMO	FS3	9 1	76	684	76
F	C	NONE	0.671	0.076	0.592	0.789	0.157	AVR	PSD	FS2	9 1	72	684	76
			0.661	0.099	0.500	0.763	0.204	AVR	FFT	FS3	9 1	456	684	76
			0.650	0.044	0.618	0.724	0.091	NONE	FFT	FS3	9 1	456	684	76
			0.629	0.042	0.565	0.680	0.087	AVR	PSD	FS3	1 1	456	384	384
			0.627	0.029	0.589	0.659	0.060	NONE	FFT	FS3	1 1	456	384	384
			0.626	0.051	0.536	0.661	0.106	AVR	FFT	FS2	1 1	72	384	384
			0.624	0.050	0.566	0.671	0.102	AVR	FFT	FS2	9 1	72	684	76
			0.620	0.026	0.583	0.648	0.053	AVR	FFT	FS3	1 1	456	384	384
			0.605	0.048	0.549	0.654	0.099	NONE	PSD	FS3	1 1	456	384	384
			0.595	0.045	0.526	0.645	0.093	AVR	PSD	FS3	9 1	456	684	76
R	C	NONE	0.684	0.053	0.618	0.750	0.108	NONE	FFT	FS3	9 1	456	684	76
			0.662	0.024	0.630	0.693	0.049	AVR	FFT	FS3	1 1	456	384	384
			0.655	0.078	0.553	0.737	0.161	NONE	PSD	FS3	9 1	456	684	76
			0.654	0.040	0.617	0.719	0.082	NONE	FFT	FS3	1 1	456	384	384
			0.653	0.065	0.566	0.711	0.134	AVR	PSD	FS2	9 1	72	684	76
			0.647	0.073	0.526	0.711	0.151	AVR	FFT	FS2	9 1	72	684	76
			0.645	0.019	0.622	0.669	0.039	AVR	PSD	FS3	1 1	456	384	384
			0.639	0.058	0.579	0.724	0.119	AVR	FFT	FS3	9 1	456	684	76
			0.639	0.044	0.594	0.688	0.090	AVR	PSD	FS2	1 1	72	384	384
			0.631	0.045	0.583	0.685	0.093	NONE	PSD	FS3	1 1	456	384	384
F	R	C	0.458	0.052	0.412	0.526	0.106	NONE	FFT	FS3	9 1	456	1026	114
			0.453	0.031	0.410	0.479	0.065	AVR	FFT	FS3	1 1	456	576	576
			0.433	0.043	0.368	0.474	0.088	AVR	FFT	FS2	9 1	72	1026	114
			0.425	0.036	0.368	0.456	0.075	AVR	PSD	FS2	9 1	72	1026	114
			0.423	0.052	0.342	0.474	0.107	AVR	FFT	FS3	9 1	456	1026	114
			0.411	0.042	0.344	0.455	0.086	AVR	PSD	FS2	1 1	72	576	576
			0.409	0.031	0.359	0.441	0.064	AVR	FFT	FS2	1 1	72	576	576
			0.409	0.026	0.373	0.436	0.053	AVR	PSD	FS3	1 1	456	576	576
			0.407	0.019	0.389	0.432	0.038	NONE	FFT	FS3	1 1	456	576	576
			0.405	0.045	0.325	0.432	0.093	AVR	AR	FS3	1 1	456	576	576

From total of 480 Strategies: CND=F-R, F-C, F-C-R, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=FISHER. SN=2.

Table 5.4: Ten optimal strategies of subject 3 using the Fisher discriminant classifier.

Top-ten optimum strategies for each condition combination for subject 3 using the FISHER classifier

CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.665	0.048	0.615	0.740	0.098	AVR	AR	FS3	1 1	342	192	192
			0.658	0.026	0.632	0.684	0.054	NONE	AR	FS3	9 1	342	342	38
			0.658	0.081	0.579	0.789	0.167	NONE	PSD	FS3	9 1	342	342	38
			0.658	0.104	0.500	0.763	0.213	AVR	AR	FS3	9 1	342	342	38
			0.656	0.026	0.625	0.688	0.053	AVR	FFT	FS3	1 1	342	192	192
			0.653	0.066	0.605	0.763	0.135	AVR	FFT	FS3	9 1	342	342	38
			0.647	0.055	0.579	0.711	0.112	NONE	PSD	FS2	9 1	72	342	38
			0.631	0.018	0.609	0.656	0.037	AVR	ARMO	FS3	1 1	57	192	192
			0.621	0.044	0.553	0.658	0.091	AVR	ARMO	FS3	9 1	57	342	38
			0.616	0.101	0.474	0.737	0.208	LPF	FFT	FS3	9 1	342	342	38
F	R	NONE	0.592	0.057	0.526	0.658	0.117	AVR	FFT	FS3	9 1	342	684	76
			0.584	0.056	0.539	0.658	0.116	SPF	AR	FS3	9 1	342	684	76
			0.582	0.072	0.487	0.658	0.148	NONE	FFT	FS3	9 1	342	684	76
			0.574	0.067	0.513	0.684	0.139	NONE	PSD	FS2	9 1	72	684	76
			0.571	0.033	0.513	0.592	0.068	AVR	PSD	FS3	9 1	342	684	76
			0.563	0.050	0.500	0.632	0.104	AVR	AR	FS3	9 1	342	684	76
			0.561	0.082	0.461	0.658	0.169	NONE	FFT	FS1	9 1	24	684	76
			0.558	0.032	0.518	0.596	0.066	NONE	PSD	FS3	1 1	342	384	384
			0.558	0.071	0.474	0.632	0.145	SPF	PSD	FS2	9 1	72	684	76
			0.558	0.026	0.531	0.586	0.053	AVR	AR	FS3	1 1	342	384	384
F	C	NONE	0.666	0.083	0.526	0.737	0.171	NONE	FFT	FS3	9 1	342	684	76
			0.653	0.034	0.605	0.684	0.071	AVR	FFT	FS3	9 1	342	684	76
			0.642	0.027	0.609	0.669	0.056	AVR	FFT	FS3	1 1	342	384	384
			0.583	0.015	0.565	0.607	0.031	AVR	PSD	FS3	1 1	342	384	384
			0.582	0.034	0.526	0.612	0.069	NONE	FFT	FS3	1 1	342	384	384
			0.577	0.039	0.529	0.633	0.081	NONE	AR	FS3	1 1	342	384	384
			0.576	0.025	0.539	0.605	0.052	AVR	PSD	FS2	9 1	72	684	76
			0.574	0.021	0.544	0.602	0.043	NONE	FFT	FS1	1 1	24	384	384
			0.574	0.036	0.526	0.618	0.073	NONE	PSD	FS1	9 1	24	684	76
			0.571	0.090	0.447	0.671	0.184	SPF	PSD	FS3	9 1	342	684	76
R	C	NONE	0.679	0.026	0.645	0.711	0.053	NONE	PSD	FS3	9 1	342	684	76
			0.668	0.081	0.579	0.763	0.166	NONE	PSD	FS2	9 1	72	684	76
			0.655	0.014	0.632	0.671	0.030	AVR	FFT	FS3	9 1	342	684	76
			0.650	0.015	0.635	0.672	0.030	AVR	FFT	FS3	1 1	342	384	384
			0.646	0.020	0.620	0.664	0.041	AVR	PSD	FS3	1 1	342	384	384
			0.644	0.036	0.594	0.680	0.075	NONE	FFT	FS3	1 1	342	384	384
			0.644	0.035	0.586	0.677	0.071	NONE	PSD	FS3	1 1	342	384	384
			0.642	0.030	0.592	0.671	0.062	NONE	FFT	FS3	9 1	342	684	76
			0.632	0.016	0.618	0.658	0.033	SPF	FFT	FS3	9 1	342	684	76
			0.624	0.048	0.566	0.697	0.099	SPF	PSD	FS3	9 1	342	684	76
F	R	C	0.430	0.036	0.377	0.465	0.073	NONE	PSD	FS3	9 1	342	1026	114
			0.425	0.050	0.386	0.509	0.103	NONE	FFT	FS3	9 1	342	1026	114
			0.411	0.034	0.375	0.458	0.071	AVR	FFT	FS3	1 1	342	576	576
			0.410	0.020	0.382	0.434	0.040	NONE	PSD	FS3	1 1	342	576	576
			0.407	0.061	0.325	0.491	0.125	AVR	FFT	FS3	9 1	342	1026	114
			0.406	0.026	0.375	0.429	0.054	NONE	FFT	FS3	1 1	342	576	576
			0.398	0.049	0.316	0.439	0.100	NONE	FFT	FS2	9 1	72	1026	114
			0.393	0.048	0.333	0.465	0.098	AVR	FFT	FS2	9 1	72	1026	114
			0.392	0.030	0.345	0.427	0.061	NONE	AR	FS3	1 1	342	576	576
			0.391	0.040	0.333	0.439	0.081	NONE	AR	FS3	9 1	342	1026	114

From total of 480 Strategies: CND=F-R, F-C, F-C-R, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=FISHER. SN=3.

Table 5.5: Ten optimal strategies of subject 1 using the generalised linear model classifier.

Top-ten optimum strategies for each condition combination for subject 1 using the GLM classifier

CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.688	0.078	0.553	0.868	0.044	SPF	PSD	FS2	9 1	78	342	38
			0.685	0.092	0.474	0.816	0.051	NONE	AR	FS2	9 1	78	342	38
			0.664	0.083	0.500	0.842	0.046	SPF	AR	FS2	9 1	78	342	38
			0.660	0.059	0.526	0.763	0.033	SPF	FFT	FS2	9 1	78	342	38
			0.654	0.062	0.526	0.789	0.035	LPF	AR	FS2	9 1	78	342	38
			0.654	0.095	0.421	0.816	0.053	LPF	FFT	FS2	9 1	78	342	38
			0.649	0.073	0.500	0.763	0.041	AVR	AR	FS2	9 1	78	342	38
			0.643	0.047	0.547	0.729	0.026	SPF	PSD	FS2	1 1	78	192	192
			0.636	0.066	0.500	0.737	0.037	AVR	AR	FS1	9 1	24	342	38
			0.633	0.078	0.474	0.842	0.044	NONE	PSD	FS1	9 1	24	342	38
F	R	NONE	0.672	0.050	0.553	0.803	0.028	AVR	AR	FS2	9 1	78	684	76
			0.653	0.050	0.553	0.750	0.028	SPF	PSD	FS2	9 1	78	684	76
			0.649	0.043	0.579	0.737	0.024	LPF	AR	FS2	9 1	78	684	76
			0.642	0.053	0.553	0.776	0.030	AVR	FFT	FS2	9 1	78	684	76
			0.638	0.065	0.513	0.737	0.036	AVR	AR	FS3	9 1	486	684	76
			0.634	0.020	0.596	0.667	0.011	LPF	AR	FS2	1 1	78	384	384
			0.627	0.028	0.557	0.680	0.016	AVR	AR	FS2	1 1	78	384	384
			0.626	0.068	0.474	0.737	0.038	NONE	PSD	FS1	9 1	24	684	76
			0.625	0.049	0.526	0.737	0.028	SPF	FFT	FS2	9 1	78	684	76
			0.621	0.051	0.500	0.711	0.028	AVR	PSD	FS2	9 1	78	684	76
F	C	NONE	0.819	0.047	0.711	0.882	0.026	AVR	FFT	FS2	9 1	78	684	76
			0.812	0.041	0.737	0.895	0.023	SPF	FFT	FS2	9 1	78	684	76
			0.807	0.037	0.737	0.882	0.021	NONE	PSD	FS2	9 1	78	684	76
			0.803	0.030	0.711	0.842	0.017	NONE	FFT	FS2	9 1	78	684	76
			0.802	0.044	0.711	0.882	0.025	LPF	PSD	FS2	9 1	78	684	76
			0.799	0.032	0.750	0.855	0.018	LPF	FFT	FS2	9 1	78	684	76
			0.797	0.053	0.684	0.882	0.030	NONE	AR	FS2	9 1	78	684	76
			0.788	0.051	0.697	0.855	0.028	AVR	PSD	FS2	9 1	78	684	76
			0.782	0.043	0.711	0.855	0.024	AVR	AR	FS2	9 1	78	684	76
			0.778	0.055	0.671	0.882	0.031	SPF	PSD	FS2	9 1	78	684	76
R	C	NONE	0.816	0.029	0.763	0.882	0.016	LPF	AR	FS2	9 1	78	684	76
			0.810	0.048	0.697	0.908	0.027	AVR	AR	FS2	9 1	78	684	76
			0.807	0.037	0.711	0.868	0.021	AVR	PSD	FS2	9 1	78	684	76
			0.803	0.044	0.737	0.882	0.025	NONE	FFT	FS2	9 1	78	684	76
			0.800	0.036	0.724	0.868	0.020	LPF	FFT	FS2	9 1	78	684	76
			0.793	0.049	0.684	0.882	0.027	NONE	AR	FS2	9 1	78	684	76
			0.789	0.044	0.711	0.868	0.025	AVR	FFT	FS2	9 1	78	684	76
			0.789	0.048	0.711	0.868	0.027	NONE	PSD	FS2	9 1	78	684	76
			0.786	0.055	0.671	0.895	0.031	LPF	PSD	FS2	9 1	78	684	76
			0.783	0.014	0.758	0.813	0.008	LPF	AR	FS2	1 1	78	384	384

From total of 384 Strategies: CND=F-R, F-C, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=GLM. SN=1.

Table 5.6: Ten optimal strategies of subject 2 using the generalised linear model classifier.

Top-ten optimum strategies for each condition combination for subject 2 using the GLM classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.608	0.076	0.447	0.763	0.043	AVR	AR	FS2	9 1	72	342	38
			0.606	0.065	0.500	0.763	0.036	SPF	AR	FS2	9 1	72	342	38
			0.606	0.075	0.474	0.737	0.042	NONE	PSD	FS2	9 1	72	342	38
			0.604	0.071	0.421	0.711	0.040	NONE	AR	FS2	9 1	72	342	38
			0.604	0.073	0.500	0.816	0.041	NONE	AR	FS1	9 1	12	342	38
			0.588	0.076	0.447	0.737	0.042	NONE	FFT	FS2	9 1	72	342	38
			0.585	0.033	0.510	0.630	0.019	SPF	AR	FS2	1 1	72	192	192
			0.583	0.087	0.395	0.763	0.049	AVR	FFT	FS2	9 1	72	342	38
			0.583	0.031	0.510	0.630	0.017	NONE	AR	FS1	1 1	12	192	192
			0.582	0.030	0.521	0.641	0.017	AVR	AR	FS2	1 1	72	192	192
F	R	NONE	0.623	0.047	0.526	0.684	0.026	AVR	AR	FS3	9 1	456	684	76
			0.618	0.045	0.513	0.697	0.025	NONE	AR	FS2	9 1	72	684	76
			0.615	0.045	0.513	0.724	0.025	AVR	AR	FS2	9 1	72	684	76
			0.612	0.032	0.553	0.671	0.018	NONE	FFT	FS2	9 1	72	684	76
			0.605	0.052	0.500	0.684	0.029	AVR	FFT	FS2	9 1	72	684	76
			0.602	0.023	0.552	0.635	0.013	AVR	AR	FS2	1 1	72	384	384
			0.602	0.021	0.568	0.646	0.012	AVR	PSD	FS2	1 1	72	384	384
			0.597	0.056	0.461	0.711	0.031	AVR	PSD	FS2	9 1	72	684	76
			0.593	0.023	0.547	0.633	0.013	AVR	FFT	FS2	1 1	72	384	384
			0.584	0.020	0.542	0.617	0.011	NONE	FFT	FS2	1 1	72	384	384
F	C	NONE	0.752	0.041	0.671	0.829	0.023	AVR	AR	FS2	9 1	72	684	76
			0.723	0.049	0.632	0.816	0.028	NONE	PSD	FS2	9 1	72	684	76
			0.720	0.018	0.680	0.758	0.010	AVR	PSD	FS2	1 1	72	384	384
			0.718	0.045	0.632	0.816	0.025	AVR	FFT	FS2	9 1	72	684	76
			0.718	0.021	0.685	0.755	0.012	AVR	AR	FS2	1 1	72	384	384
			0.717	0.044	0.618	0.789	0.024	AVR	PSD	FS2	9 1	72	684	76
			0.709	0.063	0.579	0.868	0.035	NONE	FFT	FS2	9 1	72	684	76
			0.697	0.023	0.633	0.745	0.013	AVR	FFT	FS2	1 1	72	384	384
			0.685	0.034	0.632	0.750	0.019	NONE	AR	FS2	9 1	72	684	76
			0.674	0.021	0.641	0.716	0.012	NONE	PSD	FS2	1 1	72	384	384
R	C	NONE	0.735	0.044	0.658	0.829	0.025	AVR	PSD	FS2	9 1	72	684	76
			0.728	0.042	0.658	0.816	0.024	AVR	FFT	FS2	9 1	72	684	76
			0.721	0.045	0.632	0.803	0.025	NONE	AR	FS2	9 1	72	684	76
			0.707	0.022	0.654	0.740	0.012	AVR	FFT	FS2	1 1	72	384	384
			0.704	0.043	0.618	0.776	0.024	AVR	AR	FS2	9 1	72	684	76
			0.699	0.047	0.566	0.776	0.026	NONE	PSD	FS2	9 1	72	684	76
			0.696	0.018	0.669	0.740	0.010	AVR	PSD	FS2	1 1	72	384	384
			0.686	0.017	0.641	0.721	0.009	NONE	AR	FS2	1 1	72	384	384
			0.684	0.054	0.579	0.776	0.030	AVR	FFT	FS3	9 1	456	684	76
			0.683	0.020	0.651	0.714	0.011	AVR	AR	FS2	1 1	72	384	384
From total of 384 Strategies: CND=F-R, F-C, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=GLM. SN=2.														

Table 5.7: Ten optimal strategies of subject 3 using the generalised linear model classifier.

Top-ten optimum strategies for each condition combination for subject 3 using the GLM classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.707	0.070	0.553	0.842	0.039	AVR	ARMO	FS3	9 1	57	342	38
			0.700	0.074	0.526	0.816	0.042	AVR	AR	FS1	9 1	24	342	38
			0.695	0.087	0.526	0.842	0.048	AVR	AR	FS2	9 1	72	342	38
			0.681	0.029	0.635	0.729	0.016	AVR	AR	FS1	1 1	24	192	192
			0.671	0.068	0.526	0.763	0.038	AVR	PSD	FS2	9 1	72	342	38
			0.665	0.077	0.474	0.789	0.043	AVR	ARMO	FS2	9 1	12	342	38
			0.665	0.028	0.609	0.719	0.015	AVR	AR	FS2	1 1	72	192	192
			0.661	0.067	0.500	0.763	0.038	AVR	FFT	FS2	9 1	72	342	38
			0.650	0.033	0.583	0.719	0.018	AVR	ARMO	FS3	1 1	57	192	192
			0.649	0.079	0.474	0.789	0.044	SPF	FFT	FS2	9 1	72	342	38
F	R	NONE	0.621	0.068	0.487	0.789	0.038	NONE	AR	FS3	9 1	342	684	76
			0.616	0.055	0.487	0.711	0.031	NONE	FFT	FS3	9 1	342	684	76
			0.614	0.059	0.487	0.724	0.033	NONE	PSD	FS3	9 1	342	684	76
			0.612	0.053	0.513	0.737	0.030	NONE	AR	FS2	9 1	72	684	76
			0.606	0.061	0.500	0.724	0.034	AVR	AR	FS3	9 1	342	684	76
			0.603	0.065	0.487	0.737	0.037	AVR	PSD	FS3	9 1	342	684	76
			0.595	0.046	0.500	0.684	0.026	NONE	PSD	FS1	9 1	24	684	76
			0.592	0.054	0.474	0.671	0.030	NONE	FFT	FS2	9 1	72	684	76
			0.591	0.048	0.487	0.658	0.027	NONE	FFT	FS1	9 1	24	684	76
			0.591	0.052	0.487	0.697	0.029	NONE	PSD	FS2	9 1	72	684	76
F	C	NONE	0.684	0.051	0.592	0.789	0.029	NONE	AR	FS2	9 1	72	684	76
			0.679	0.045	0.566	0.750	0.025	AVR	AR	FS2	9 1	72	684	76
			0.677	0.054	0.592	0.803	0.030	AVR	AR	FS3	9 1	342	684	76
			0.671	0.067	0.539	0.803	0.037	AVR	PSD	FS3	9 1	342	684	76
			0.666	0.058	0.526	0.750	0.032	SPF	AR	FS2	9 1	72	684	76
			0.663	0.048	0.579	0.750	0.027	AVR	PSD	FS2	9 1	72	684	76
			0.662	0.070	0.526	0.776	0.039	NONE	PSD	FS2	9 1	72	684	76
			0.661	0.021	0.615	0.698	0.012	NONE	AR	FS2	1 1	72	384	384
			0.659	0.045	0.539	0.724	0.025	NONE	AR	FS3	9 1	342	684	76
			0.650	0.024	0.607	0.690	0.013	AVR	AR	FS2	1 1	72	384	384
R	C	NONE	0.717	0.045	0.658	0.829	0.025	AVR	PSD	FS3	9 1	342	684	76
			0.716	0.051	0.605	0.803	0.028	NONE	FFT	FS2	9 1	72	684	76
			0.707	0.049	0.618	0.803	0.028	NONE	PSD	FS2	9 1	72	684	76
			0.707	0.052	0.632	0.789	0.029	AVR	FFT	FS3	9 1	342	684	76
			0.701	0.045	0.592	0.776	0.025	AVR	AR	FS2	9 1	72	684	76
			0.694	0.040	0.632	0.763	0.022	NONE	AR	FS2	9 1	72	684	76
			0.689	0.067	0.566	0.816	0.038	NONE	FFT	FS3	9 1	342	684	76
			0.685	0.046	0.618	0.789	0.026	NONE	FFT	FS1	9 1	24	684	76
			0.684	0.047	0.592	0.776	0.026	NONE	AR	FS3	9 1	342	684	76
			0.683	0.051	0.553	0.776	0.028	AVR	AR	FS3	9 1	342	684	76
From total of 384 Strategies: CND=F-R, F-C, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=GLM. SN=3														

Table 5.8: Ten optimal strategies of subject 1 using the multilayer perceptron classifier.

Top-ten optimum strategies for each condition combination for subject 1 using the MLP classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.805	0.056	0.684	0.868	0.057	SPF	FFT	FS3	9 1	486	342	38
			0.771	0.070	0.658	0.895	0.072	SPF	PSD	FS3	9 1	486	342	38
			0.755	0.058	0.684	0.842	0.060	SPF	FFT	FS2	9 1	78	342	38
			0.754	0.022	0.734	0.792	0.023	SPF	FFT	FS3	1 1	486	192	192
			0.747	0.047	0.658	0.816	0.048	AVR	FFT	FS3	9 1	486	342	38
			0.737	0.063	0.658	0.842	0.065	SPF	AR	FS2	9 1	78	342	38
			0.731	0.034	0.656	0.766	0.035	SPF	PSD	FS3	1 1	486	192	192
			0.729	0.021	0.698	0.771	0.022	SPF	AR	FS3	1 1	486	192	192
			0.724	0.031	0.684	0.789	0.032	SPF	AR	FS3	9 1	486	342	38
			0.718	0.078	0.579	0.789	0.080	SPF	PSD	FS2	9 1	78	342	38
F	R	NONE	0.709	0.039	0.658	0.776	0.041	SPF	FFT	FS3	9 1	486	684	76
			0.701	0.029	0.643	0.742	0.030	SPF	FFT	FS3	1 1	486	384	384
			0.691	0.020	0.648	0.711	0.020	SPF	PSD	FS3	1 1	486	384	384
			0.683	0.064	0.605	0.789	0.065	SPF	AR	FS3	9 1	486	684	76
			0.682	0.050	0.618	0.776	0.052	AVR	PSD	FS3	9 1	486	684	76
			0.676	0.072	0.592	0.803	0.074	SPF	PSD	FS3	9 1	486	684	76
			0.671	0.019	0.633	0.703	0.019	SPF	AR	FS3	1 1	486	384	384
			0.671	0.038	0.579	0.697	0.039	AVR	AR	FS3	9 1	486	684	76
			0.670	0.060	0.579	0.750	0.062	NONE	FFT	FS3	9 1	486	684	76
			0.663	0.041	0.592	0.711	0.042	NONE	AR	FS3	9 1	486	684	76
F	C	NONE	0.805	0.056	0.684	0.868	0.057	SPF	FFT	FS3	9 1	486	342	38
			0.771	0.070	0.658	0.895	0.072	SPF	PSD	FS3	9 1	486	342	38
			0.755	0.058	0.684	0.842	0.060	SPF	FFT	FS2	9 1	78	342	38
			0.754	0.022	0.734	0.792	0.023	SPF	FFT	FS3	1 1	486	192	192
			0.747	0.047	0.658	0.816	0.048	AVR	FFT	FS3	9 1	486	342	38
			0.737	0.063	0.658	0.842	0.065	SPF	AR	FS2	9 1	78	342	38
			0.731	0.034	0.656	0.766	0.035	SPF	PSD	FS3	1 1	486	192	192
			0.729	0.021	0.698	0.771	0.022	SPF	AR	FS3	1 1	486	192	192
			0.724	0.031	0.684	0.789	0.032	SPF	AR	FS3	9 1	486	342	38
			0.718	0.078	0.579	0.789	0.080	SPF	PSD	FS2	9 1	78	342	38
R	C	NONE	0.883	0.035	0.842	0.934	0.036	SPF	FFT	FS3	9 1	486	684	76
			0.868	0.023	0.829	0.895	0.024	SPF	PSD	FS3	9 1	486	684	76
			0.857	0.010	0.846	0.875	0.010	SPF	FFT	FS3	1 1	486	384	384
			0.842	0.026	0.789	0.882	0.026	AVR	FFT	FS3	9 1	486	684	76
			0.838	0.019	0.805	0.870	0.020	SPF	PSD	FS3	1 1	486	384	384
			0.826	0.028	0.776	0.855	0.028	AVR	PSD	FS3	9 1	486	684	76
			0.825	0.030	0.789	0.882	0.031	NONE	FFT	FS3	9 1	486	684	76
			0.817	0.023	0.776	0.842	0.023	NONE	PSD	FS2	9 1	78	684	76
			0.817	0.040	0.750	0.855	0.042	AVR	PSD	FS2	9 1	78	684	76
			0.813	0.015	0.797	0.839	0.015	NONE	FFT	FS3	1 1	486	384	384
F	R	C	0.669	0.024	0.625	0.714	0.025	SPF	FFT	FS3	1 1	486	576	576
			0.658	0.020	0.632	0.693	0.021	SPF	FFT	FS3	9 1	486	1026	114
			0.651	0.048	0.579	0.719	0.049	SPF	AR	FS3	9 1	486	1026	114
			0.637	0.019	0.602	0.660	0.019	AVR	FFT	FS3	1 1	486	576	576
			0.633	0.043	0.561	0.684	0.044	NONE	FFT	FS2	9 1	78	1026	114
			0.632	0.013	0.615	0.653	0.013	AVR	PSD	FS3	1 1	486	576	576
			0.632	0.014	0.613	0.658	0.014	SPF	AR	FS3	1 1	486	576	576
			0.632	0.056	0.535	0.702	0.058	SPF	PSD	FS3	9 1	486	1026	114
			0.625	0.037	0.570	0.675	0.038	AVR	PSD	FS2	9 1	78	1026	114
			0.625	0.051	0.518	0.702	0.053	AVR	PSD	FS3	9 1	486	1026	114
From total of 256 Strategies: CND=F-R, F-C, FL-FR, R-C, F-R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS2, FS3. SR=9:1,1:1. NP=5. CF=MLP. SN=1.														

Table 5.9: Ten optimal strategies of subject 2 using the multilayer perceptron classifier.

Top-ten optimum strategies for each condition combination for subject 2 using the MLP classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.703	0.068	0.579	0.816	0.070	AVR	AR	FS3	9 1	456	342	38
			0.679	0.033	0.646	0.740	0.033	AVR	AR	FS3	1 1	456	192	192
			0.668	0.071	0.579	0.789	0.073	NONE	AR	FS3	9 1	456	342	38
			0.663	0.059	0.579	0.763	0.061	SPF	AR	FS3	9 1	456	342	38
			0.639	0.053	0.553	0.711	0.054	AVR	AR	FS2	9 1	72	342	38
			0.631	0.025	0.599	0.682	0.025	SPF	AR	FS3	1 1	456	192	192
			0.627	0.023	0.589	0.672	0.024	NONE	AR	FS3	1 1	456	192	192
			0.621	0.054	0.526	0.684	0.056	AVR	FFT	FS3	9 1	456	342	38
			0.616	0.027	0.589	0.682	0.028	AVR	AR	FS3	1 1	72	192	192
			0.616	0.085	0.526	0.789	0.088	NONE	AR	FS3	9 1	72	342	38
F	R	NONE	0.639	0.056	0.513	0.697	0.057	AVR	FFT	FS3	9 1	456	684	76
			0.629	0.041	0.566	0.684	0.042	SPF	FFT	FS3	9 1	456	684	76
			0.621	0.018	0.602	0.659	0.018	AVR	FFT	FS3	1 1	456	384	384
			0.621	0.041	0.553	0.684	0.042	AVR	PSD	FS2	9 1	72	684	76
			0.620	0.054	0.513	0.684	0.055	AVR	AR	FS2	9 1	72	684	76
			0.618	0.039	0.566	0.671	0.040	NONE	FFT	FS2	9 1	72	684	76
			0.614	0.061	0.487	0.697	0.062	AVR	PSD	FS3	9 1	456	684	76
			0.614	0.081	0.500	0.750	0.083	AVR	FFT	FS2	9 1	72	684	76
			0.610	0.021	0.576	0.646	0.022	AVR	FFT	FS2	1 1	72	384	384
			0.609	0.041	0.539	0.671	0.042	AVR	AR	FS3	9 1	456	684	76
F	C	NONE	0.768	0.031	0.724	0.829	0.032	AVR	FFT	FS3	9 1	456	684	76
			0.727	0.014	0.708	0.747	0.014	AVR	FFT	FS3	1 1	456	384	384
			0.722	0.053	0.618	0.789	0.054	AVR	PSD	FS2	9 1	72	684	76
			0.721	0.024	0.698	0.758	0.024	AVR	PSD	FS3	1 1	456	384	384
			0.716	0.053	0.645	0.803	0.055	NONE	FFT	FS3	9 1	456	684	76
			0.711	0.056	0.618	0.789	0.058	AVR	AR	FS3	9 1	456	684	76
			0.709	0.047	0.658	0.776	0.048	SPF	FFT	FS3	9 1	456	684	76
			0.709	0.053	0.658	0.803	0.054	AVR	AR	FS2	9 1	72	684	76
			0.706	0.021	0.677	0.737	0.021	AVR	PSD	FS2	1 1	72	384	384
			0.700	0.056	0.605	0.789	0.058	AVR	FFT	FS2	9 1	72	684	76
R	C	NONE	0.751	0.060	0.658	0.842	0.062	AVR	PSD	FS3	9 1	456	684	76
			0.749	0.049	0.658	0.816	0.050	NONE	FFT	FS3	9 1	456	684	76
			0.743	0.014	0.708	0.755	0.015	AVR	FFT	FS3	1 1	456	384	384
			0.742	0.046	0.671	0.803	0.047	AVR	FFT	FS2	9 1	72	684	76
			0.728	0.065	0.605	0.829	0.067	AVR	FFT	FS3	9 1	456	684	76
			0.712	0.042	0.671	0.789	0.043	NONE	FFT	FS2	9 1	72	684	76
			0.711	0.024	0.674	0.750	0.024	NONE	PSD	FS3	1 1	456	384	384
			0.710	0.018	0.693	0.747	0.019	AVR	FFT	FS2	1 1	72	384	384
			0.709	0.028	0.664	0.747	0.029	AVR	PSD	FS3	1 1	456	384	384
			0.707	0.048	0.645	0.803	0.050	AVR	AR	FS3	9 1	456	684	76
F	R	C	0.561	0.050	0.456	0.623	0.052	AVR	FFT	FS3	9 1	456	1026	114
			0.546	0.031	0.509	0.605	0.032	NONE	FFT	FS3	9 1	456	1026	114
			0.544	0.027	0.491	0.578	0.028	AVR	FFT	FS3	1 1	456	576	576
			0.531	0.018	0.505	0.563	0.019	AVR	FFT	FS2	1 1	72	576	576
			0.531	0.051	0.439	0.605	0.052	AVR	PSD	FS2	9 1	72	1026	114
			0.526	0.031	0.482	0.553	0.032	AVR	AR	FS3	9 1	456	1026	114
			0.523	0.018	0.491	0.543	0.018	NONE	FFT	FS3	1 1	456	576	576
			0.519	0.038	0.465	0.588	0.039	AVR	FFT	FS2	9 1	72	1026	114
			0.515	0.052	0.412	0.623	0.054	NONE	FFT	FS2	9 1	72	1026	114
			0.504	0.015	0.490	0.533	0.016	AVR	PSD	FS2	1 1	72	576	576
From total of 256 Strategies: CND=F-R, F-C, FL-FR, R-C, F-R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS2, FS3. SR=9:1,1:1. NP=5. CF=MLP. SN=2.														

Table 5.10: Ten optimal strategies of subject 3 using the multilayer perceptron classifier.

Top-ten optimum strategies for each condition combination for subject 3 using the MLP classifier														
CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
FL	FR	NONE	0.774	0.053	0.684	0.842	0.054	AVR	AR	FS3	9 1	342	342	38
			0.726	0.061	0.632	0.816	0.063	AVR	AR	FS2	9 1	72	342	38
			0.719	0.014	0.682	0.734	0.015	AVR	AR	FS3	1 1	342	192	192
			0.700	0.073	0.579	0.816	0.075	SPF	AR	FS3	9 1	342	342	38
			0.692	0.053	0.579	0.763	0.054	AVR	FFT	FS3	9 1	342	342	38
			0.689	0.072	0.579	0.789	0.074	NONE	FFT	FS3	9 1	342	342	38
			0.688	0.033	0.615	0.729	0.034	AVR	AR	FS2	1 1	72	192	192
			0.687	0.067	0.579	0.789	0.069	SPF	FFT	FS3	9 1	342	342	38
			0.686	0.041	0.615	0.750	0.042	AVR	PSD	FS3	1 1	342	192	192
			0.683	0.030	0.641	0.724	0.031	NONE	AR	FS3	1 1	342	192	192
F	R	NONE	0.662	0.059	0.579	0.763	0.060	AVR	AR	FS3	9 1	342	684	76
			0.659	0.058	0.579	0.737	0.059	NONE	PSD	FS3	9 1	342	684	76
			0.654	0.061	0.579	0.763	0.063	NONE	FFT	FS3	9 1	342	684	76
			0.638	0.028	0.602	0.685	0.029	NONE	FFT	FS3	1 1	342	384	384
			0.634	0.015	0.607	0.648	0.016	NONE	PSD	FS3	1 1	342	384	384
			0.618	0.065	0.539	0.711	0.067	AVR	PSD	FS3	9 1	342	684	76
			0.617	0.022	0.573	0.643	0.023	NONE	AR	FS3	1 1	342	384	384
			0.616	0.021	0.589	0.646	0.022	AVR	AR	FS3	1 1	342	384	384
			0.616	0.045	0.539	0.684	0.046	NONE	FFT	FS2	9 1	72	684	76
			0.613	0.060	0.526	0.724	0.062	NONE	AR	FS2	9 1	72	684	76
F	C	NONE	0.730	0.044	0.645	0.789	0.046	AVR	FFT	FS3	9 1	342	684	76
			0.687	0.026	0.646	0.727	0.027	NONE	FFT	FS3	1 1	342	384	384
			0.680	0.010	0.661	0.693	0.010	AVR	FFT	FS3	1 1	342	384	384
			0.676	0.044	0.579	0.737	0.045	NONE	FFT	FS3	9 1	342	684	76
			0.674	0.051	0.592	0.750	0.053	AVR	FFT	FS2	9 1	72	684	76
			0.674	0.056	0.579	0.750	0.058	LPF	FFT	FS3	9 1	342	684	76
			0.662	0.039	0.592	0.737	0.040	AVR	AR	FS2	9 1	72	684	76
			0.657	0.074	0.566	0.829	0.076	NONE	FFT	FS2	9 1	72	684	76
			0.655	0.030	0.609	0.724	0.031	NONE	PSD	FS3	1 1	342	384	384
			0.651	0.035	0.592	0.711	0.036	NONE	AR	FS2	9 1	72	684	76
R	C	NONE	0.788	0.047	0.684	0.829	0.048	NONE	FFT	FS3	9 1	342	684	76
			0.778	0.042	0.724	0.829	0.043	NONE	PSD	FS3	9 1	342	684	76
			0.775	0.047	0.684	0.816	0.048	AVR	FFT	FS3	9 1	342	684	76
			0.753	0.019	0.711	0.779	0.020	AVR	FFT	FS3	1 1	342	384	384
			0.747	0.010	0.729	0.763	0.010	NONE	FFT	FS3	1 1	342	384	384
			0.745	0.054	0.671	0.816	0.056	SPF	FFT	FS3	9 1	342	684	76
			0.732	0.049	0.645	0.803	0.050	AVR	PSD	FS3	9 1	342	684	76
			0.730	0.037	0.671	0.789	0.038	NONE	PSD	FS2	9 1	72	684	76
			0.730	0.023	0.695	0.760	0.024	NONE	PSD	FS3	1 1	342	384	384
			0.729	0.066	0.632	0.829	0.068	NONE	FFT	FS2	9 1	72	684	76
F	R	C	0.533	0.039	0.474	0.605	0.040	NONE	FFT	FS3	9 1	342	1026	114
			0.504	0.040	0.465	0.579	0.041	NONE	PSD	FS3	9 1	342	1026	114
			0.504	0.018	0.481	0.533	0.018	NONE	FFT	FS3	1 1	342	576	576
			0.502	0.039	0.430	0.561	0.040	NONE	FFT	FS2	9 1	72	1026	114
			0.498	0.053	0.421	0.596	0.055	AVR	AR	FS3	9 1	342	1026	114
			0.488	0.013	0.469	0.509	0.014	NONE	FFT	FS2	1 1	72	576	576
			0.484	0.037	0.430	0.544	0.038	AVR	FFT	FS3	9 1	342	1026	114
			0.481	0.022	0.444	0.507	0.023	AVR	FFT	FS3	1 1	342	576	576
			0.479	0.048	0.430	0.596	0.049	LPF	FFT	FS3	9 1	342	1026	114
			0.478	0.047	0.412	0.535	0.048	LPF	PSD	FS3	9 1	342	1026	114
From total of 256 Strategies: CND=F-R, F-C, FL-FR, R-C, F-R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS2, FS3. SR=9:1,1:1. NP=5. CF=MLP. SN=3.														

Table 5.11: Optimal strategies using the Fisher discriminant classifier.

Optimum strategies for each condition combination and subject using the FISHER classifier															
SN	CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
	Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
1	FL	FR	NONE	0.741	0.036	0.703	0.781	0.074	SPF	FFT	FS3	1 1	486	192	192
	F	R	NONE	0.636	0.022	0.609	0.664	0.045	SPF	FFT	FS3	1 1	486	384	384
	F	C	NONE	0.816	0.021	0.789	0.842	0.043	SPF	AR	FS3	9 1	486	684	76
	R	C	NONE	0.834	0.048	0.776	0.908	0.099	SPF	FFT	FS3	9 1	486	684	76
	F	R	C	0.579	0.051	0.526	0.658	0.105	SPF	FFT	FS3	9 1	486	1026	114
2	FL	FR	NONE	0.641	0.027	0.599	0.667	0.056	AVR	AR	FS3	1 1	456	192	192
	F	R	NONE	0.603	0.055	0.566	0.697	0.112	AVR	AR	FS3	9 1	456	684	76
	F	C	NONE	0.671	0.076	0.592	0.789	0.157	AVR	PSD	FS2	9 1	72	684	76
	R	C	NONE	0.684	0.053	0.618	0.750	0.108	NONE	FFT	FS3	9 1	456	684	76
	F	R	C	0.458	0.052	0.412	0.526	0.106	NONE	FFT	FS3	9 1	456	1026	114
3	FL	FR	NONE	0.665	0.048	0.615	0.740	0.098	AVR	AR	FS3	1 1	342	192	192
	F	R	NONE	0.592	0.057	0.526	0.658	0.117	AVR	FFT	FS3	9 1	342	684	76
	F	C	NONE	0.666	0.083	0.526	0.737	0.171	NONE	FFT	FS3	9 1	342	684	76
	R	C	NONE	0.679	0.026	0.645	0.711	0.053	NONE	PSD	FS3	9 1	342	684	76
	F	R	C	0.430	0.036	0.377	0.465	0.073	NONE	PSD	FS3	9 1	342	1026	114
From total of 1440 Strategies: CND=F-R, F-C, F-C-R, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=FISHER. SN=1,2,3.															

Table 5.12: Optimal strategies using the generalised linear model classifier.

Optimum strategies for each condition classifier and subject using the GLM classifier														
SN	CND		Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
	Class 1	Class 2	Mean	Std.	Min.	Max.								
1	FL	FR	0.688	0.078	0.553	0.868	0.044	SPF	PSD	FS2	9 1	78	342	38
	F	R	0.672	0.050	0.553	0.803	0.028	AVR	AR	FS2	9 1	78	684	76
	F	C	0.819	0.047	0.711	0.882	0.026	AVR	FFT	FS2	9 1	78	684	76
	R	C	0.816	0.029	0.763	0.882	0.016	LPF	AR	FS2	9 1	78	684	76
2	FL	FR	0.608	0.076	0.447	0.763	0.043	AVR	AR	FS2	9 1	72	342	38
	F	R	0.623	0.047	0.526	0.684	0.026	AVR	AR	FS3	9 1	456	684	76
	F	C	0.752	0.041	0.671	0.829	0.023	AVR	AR	FS2	9 1	72	684	76
	R	C	0.735	0.044	0.658	0.829	0.025	AVR	PSD	FS2	9 1	72	684	76
3	FL	FR	0.707	0.070	0.553	0.842	0.039	AVR	ARMO	FS3	9 1	57	342	38
	F	R	0.621	0.068	0.487	0.789	0.038	NONE	AR	FS3	9 1	342	684	76
	F	C	0.684	0.051	0.592	0.789	0.029	NONE	AR	FS2	9 1	72	684	76
	R	C	0.717	0.045	0.658	0.829	0.025	AVR	PSD	FS3	9 1	342	684	76
From total of 1152 Strategies: CND=F-R, F-C, FL-FR, R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS1, FS2, FS3. SR=9:1,1:1. NP=5. CF=GLM. SN=1,2,3.														

Table 5.13: Optimal strategies using the multilayer perceptron classifier.

Optimum strategies for each condition combination and subject using the MLP classifier															
SN	CND			Classification fitness				Confidence Limits (+/-)	PP	FX	FS	SR	NI	NES	NTS
	Class 1	Class 2	Class 3	Mean	Std.	Min.	Max.								
1	FL	FR	NONE	0.805	0.056	0.684	0.868	0.057	SPF	FFT	FS3	9 1	486	342	38
	F	R	NONE	0.709	0.039	0.658	0.776	0.041	SPF	FFT	FS3	9 1	486	684	76
	F	C	NONE	0.805	0.056	0.684	0.868	0.057	SPF	FFT	FS3	9 1	486	342	38
	R	C	NONE	0.883	0.035	0.842	0.934	0.036	SPF	FFT	FS3	9 1	486	684	76
	F	R	C	0.669	0.024	0.625	0.714	0.025	SPF	FFT	FS3	1 1	486	576	576
2	FL	FR	NONE	0.703	0.068	0.579	0.816	0.070	AVR	AR	FS3	9 1	456	342	38
	F	R	NONE	0.639	0.056	0.513	0.697	0.057	AVR	FFT	FS3	9 1	456	684	76
	F	C	NONE	0.768	0.031	0.724	0.829	0.032	AVR	FFT	FS3	9 1	456	684	76
	R	C	NONE	0.751	0.060	0.658	0.842	0.062	AVR	PSD	FS3	9 1	456	684	76
	F	R	C	0.561	0.050	0.456	0.623	0.052	AVR	FFT	FS3	9 1	456	1026	114
3	FL	FR	NONE	0.774	0.053	0.684	0.842	0.054	AVR	AR	FS3	9 1	342	342	38
	F	R	NONE	0.662	0.059	0.579	0.763	0.060	AVR	AR	FS3	9 1	342	684	76
	F	C	NONE	0.730	0.044	0.645	0.789	0.046	AVR	FFT	FS3	9 1	342	684	76
	R	C	NONE	0.788	0.047	0.684	0.829	0.048	NONE	FFT	FS3	9 1	342	684	76
	F	R	C	0.533	0.039	0.474	0.605	0.040	NONE	FFT	FS3	9 1	342	1026	114
From total of 768 Strategies: CND=F-R, F-C, FL-FR, R-C, F-R-C. PP=NONE, AVR, SPF, LPF. FX=AR, ARMO, FFT, PSD. FS=FS2, FS3. SR=9:1,1:1. NP=5. CF=MLP. SN=1,2,3.															

Table 5.14: Grand strategy averages for subject 1.

Strategy Averages - Subject 1						
Processing stage	Variable	Classification fitness				Confidence Limits (+/-)
		Mean	Std.	Min.	Max.	
PP	NONE	0.593	0.041	0.528	0.655	0.055
	AVR	0.61	0.04	0.548	0.674	0.054
	SPF	0.612	0.04	0.549	0.674	0.052
	LPF	0.581	0.038	0.522	0.639	0.051
FX	AR	0.601	0.035	0.544	0.657	0.046
	ARMO	0.512	0.04	0.451	0.574	0.053
	FFT	0.664	0.041	0.579	0.707	0.056
	PSD	0.64	0.042	0.573	0.703	0.057
FS	FS1	0.581	0.044	0.514	0.648	0.062
	FS2	0.62	0.041	0.553	0.684	0.055
	FS3	0.606	0.043	0.537	0.675	0.057
SR	9 1	0.607	0.053	0.523	0.69	0.07
	1 1	0.591	0.026	0.55	0.631	0.036
CF	FISHER	0.608	0.043	0.554	0.661	0.089
	GLM	0.618	0.04	0.536	0.697	0.023
	MLP	0.672	0.036	0.614	0.728	0.037

Table 5.15: Grand strategy averages for subject 2.

Strategy Averages - Subject 2						
Processing stage	Variable	Classification fitness				Confidence Limits (+/-)
		Mean	Std.	Min.	Max.	
PP	NONE	0.545	0.042	0.479	0.61	0.055
	AVR	0.561	0.042	0.493	0.625	0.056
	SPF	0.522	0.039	0.461	0.583	0.049
	LPF	0.49	0.02	0.459	0.523	0.025
FX	AR	0.535	0.03	0.488	0.581	0.039
	ARMO	0.492	0.036	0.435	0.549	0.046
	FFT	0.55	0.038	0.489	0.608	0.05
	PSD	0.541	0.039	0.481	0.601	0.051
FS	FS1	0.523	0.035	0.464	0.58	0.049
	FS2	0.557	0.035	0.495	0.613	0.046
	FS3	0.541	0.036	0.479	0.602	0.045
SR	9 1	0.534	0.049	0.456	0.61	0.062
	1 1	0.526	0.023	0.49	0.56	0.03
CF	FISHER	0.536	0.038	0.488	0.583	0.079
	GLM	0.553	0.037	0.481	0.624	0.02
	MLP	0.587	0.034	0.532	0.644	0.036

Table 5.16: Grand strategy averages for subject 3.

Strategy Averages - Subject 3						
Processing stage	Variable	Classification fitness				Confidence Limits (+/-)
		Mean	Std.	Min.	Max.	
PP	NONE	0.564	0.042	0.498	0.628	0.056
	AVR	0.563	0.042	0.497	0.627	0.055
	SPF	0.54	0.041	0.476	0.604	0.054
	LPF	0.515	0.031	0.467	0.564	0.042
FX	AR	0.543	0.031	0.493	0.591	0.041
	ARMO	0.51	0.04	0.447	0.574	0.053
	FFT	0.568	0.042	0.502	0.633	0.057
	PSD	0.562	0.042	0.497	0.626	0.056
FS	FS1	0.547	0.041	0.482	0.611	0.059
	FS2	0.563	0.039	0.499	0.626	0.056
	FS3	0.563	0.039	0.497	0.631	0.055
SR	9 1	0.551	0.053	0.468	0.634	0.069
	1 1	0.54	0.025	0.501	0.578	0.034
CF	FISHER	0.548	0.042	0.495	0.6	0.088
	GLM	0.577	0.039	0.501	0.653	0.022
	MLP	0.609	0.037	0.549	0.666	0.038

Table 5.17: Grand strategy averages incorporating all subjects.

Grand Strategy Averages						
Processing stage	Variable	Classification fitness				Confidence Limits (+/-)
		Mean	Std.	Min.	Max.	
PP	NONE	0.567	0.042	0.502	0.631	0.055
	AVR	0.578	0.041	0.513	0.642	0.055
	SPF	0.558	0.040	0.495	0.620	0.052
	LPF	0.529	0.030	0.483	0.575	0.039
FX	AR	0.560	0.032	0.508	0.610	0.042
	ARMO	0.505	0.039	0.444	0.566	0.051
	FFT	0.594	0.040	0.523	0.649	0.054
	PSD	0.581	0.041	0.517	0.643	0.055
FS	FS1	0.529	0.040	0.468	0.598	0.057
	FS2	0.558	0.039	0.496	0.616	0.055
	FS3	0.551	0.040	0.489	0.614	0.055
SR	9 1	0.564	0.052	0.482	0.645	0.067
	1 1	0.552	0.025	0.514	0.590	0.033
CF	FISHER	0.564	0.041	0.512	0.615	0.085
	GLM	0.586	0.039	0.508	0.662	0.022
	MLP	0.623	0.036	0.565	0.679	0.037

Optimal strategies

For subject 1, the optimal strategy for the two-way classification FL-FR, $\{SPF, FFT, FS3, 9:1, MLP\}$, yields a classification fitness of 80.5% compared with 74.7% for the same strategy except using AVR (see Table 5.8). LPF and NONE don't even reach the top-ten strategies list for most of the strategies which utilise the MLP classifier for subject 1. This trend is maintained for strategies that utilise the FISHER classifier (see Table 5.2) but less so for GLM based strategies (see Table 5.5).

The Optimal strategies for subjects 2 and 3 for the same two-way classification, FL-FR, also utilise the MLP classifier, the larger channel set (FS3) and larger split ratio (9:1). However, unlike strategies for subject 1, these strategies tend to utilise the AVR method in preference to SPF. Taking the optimal strategy for subjects 2 and 3, $\{AVR, AR, FS3, 9:1, MLP\}$, subject 2 yields 70.3% compared with 66.8% for NONE and 66.3% for SPF. As with subject 1, LPF doesn't make it into the top-ten strategies for the FL-FR classification. A similar trend is evident for subject 3, giving 77.4% for AVR, 70% for SPF and 68.9% for NONE. Again, LPF doesn't make it into the top-ten. The variable performance of NONE and SPF can be somewhat enlightened upon by considering the occurrence of strategies which utilise these methods in the top-ten strategies lists for each condition combination. It can be seen that the prevalence of strategies utilising the NONE method far outnumber those which use SPF. This is also confirmed by the strategy averages tables, which are discussed below.

Strategy averages

The strategy averages for each subject mostly confirm the findings of the optimal strategies. The general trend puts AVR slightly above NONE, which is slightly better than SPF, which is in turn slightly better than LPF. The main point of interest here is that, according to the strategy averages, AVR and SPF perform neck-to-neck for subject 1. It is not until the optimal strategies are considered that the differences between these pre-processing variations becomes more clearly defined. The comparatively poor performance of LPF is reflected in all four strategy averages tables.

5.4.2 Feature extraction

As is the case for the pre-processing methods, there is no one feature extraction method which is significantly better than the other for all three subjects. Instead, it is found that FFT is most successful for subject 1, compared to AR for subjects 2 and 3.

Optimal strategies

When considering the top-ten strategies for the important two-way classifications FL-RL and F-R, it can be seen that FFT is the best method for subject 1, as opposed to AR for subjects 2 and 3. For condition comparisons F-C, R-C and F-R-C, the optimal strategies for subjects 2 and 3 are less clear cut (in terms of an obvious preferred feature extraction method), in these cases, FFT, AR and PSD all yield similar results. In all cases however, ARMO performs comparatively poorly.

Taking the optimal strategy of subject 1, $\{SPF, FFT, FS3, 9 : 1, MLP\}$, and varying the feature extraction method gives the following results (for FL-FR condition pair): 80.5% (FFT), 77.1% (PSD), 72.4% (AR). A similar trend is apparent For the F-R condition pair, although to a lesser extent: 70.9% (FFT), 67.6% (PSD), 68.3% (AR). Note that ARMO doesn't appear in the top-ten strategies list for any of the other condition combinations. When looking at the three-way classification, F-R-C, the difference in performance between FFT and AR diminishes towards being negligible: 65.8% for FFT, compared to 65.1% for AR, and 63.2% for PSD. Similar results are found for strategies utilising the FISHER classifier, but to a lesser extent with the GLM classifier. These trends are also apparent when considering the strategy averages data.

Strategy averages

Looking at the strategy averages, it can seen that for subject 1, FFT (which is closely followed by PSD) significantly outperforms AR, whereas for subjects 2 and 3, FFT's improvement over AR is only slight. Strategies using ARMO perform comparatively worse (at least 10% worse in most cases) than any of the other three methods. This is not that surprising since the ARMO representation has only one feature per channel, compared to five features for FFT and PSD, and 6 for AR.

5.4.3 Feature selection

When looking at the frequency and position of individual strategies which utilise one of the three feature selection methods in the top-ten tables, the following trends are apparent:

- FS3 is significantly better than FS2 in strategies which use the FISHER classifier. FS1[11] performs poorly in comparison to both FS3 and FS2.
- FS3 is slightly better than FS2 in strategies which use the MLP classifier. FS1 performs poorly in comparison to both FS3 and FS2.
- FS2 is slightly better than FS3 in strategies which use the GLM classifier. FS1 performs poorly in comparison to both FS3 and FS2, but not quite as poorly as in the above two cases.

Optimal strategies

In the case of strategies using MLP (the most successful classifier), FS3 generally scores some 5% higher than FS2. For example, consider subject 1 with the FL-FR condition combination. The strategy $\{SPF, FFT, FS3, 9:1, MLP\}$ yields 80.5%, compared with 70.5% when using FS2. This is also the case for strategies using the

[11]This was realised before the analysis computations had been completed, hence the FS1 strategy variation was not obtained for the MLP classifier as the analysis for which was computed last.

FISHER classifier. However, almost all the optimal strategies using the GLM classifier favour FS2 over FS3. In most cases FS1 dose not appear in the top-ten tables.

Strategy averages

It is clear to see from Table 5.17 that FS2 performs slightly better than FS3: 55.8% as opposed to 55.1%. FS1 performs significantly less successfully at 52.7%. This trend is reflected in all the single subject averages tables, with the exception of subject 3 (Table 5.16) where it can be seen that FS3 and FS2 perform equally well. Note that this finding dose not agree with the trends discovered by looking at the optimal strategies[12].

5.4.4 Training set size

It is generally the case that strategies trained with a larger training set size, resulting from the 9:1 split ratio, perform only marginally better than those trained on the smaller data sets resulting from a 1:1 split ratio. This is apparent in both the top-ten strategy lists and the strategy averages tables.

This result suggests that the additional information provided by the larger training set sizes is either not necessary (in terms of modelling the data), or that the classifiers are not sensitive enough to this extra information. If the former is true, which further analysis would show, then this eludes the question: "What is the lower limit at which the number of training set patterns will no longer achieve a suitable classification solution?" The answer, as discussed before, could be a critical factor in deciding whether the existing classification techniques might be applicable to a working BCMI system.

5.4.5 Classifier

The best classification fitnesses are obtained from strategies which utilise the MLP classifier. These are followed by strategies employing the GLM, and lastly the FISHER classifier. This trend is apparent in both the top-ten strategy lists and the strategy averages tables. Looking at the optimal strategy for each classifier and subject, it can be seen that although strategies employing the MLP result in the best performance,

[12]This is not taken to seriously because of the large number of sub-optimal strategies which combine to arrive at the strategy averages data presented here. The point of computing the strategy averages is to gain some insight into the overall performance of individual method variations - as opposed to simply identifying the 'winning formula' as it were.

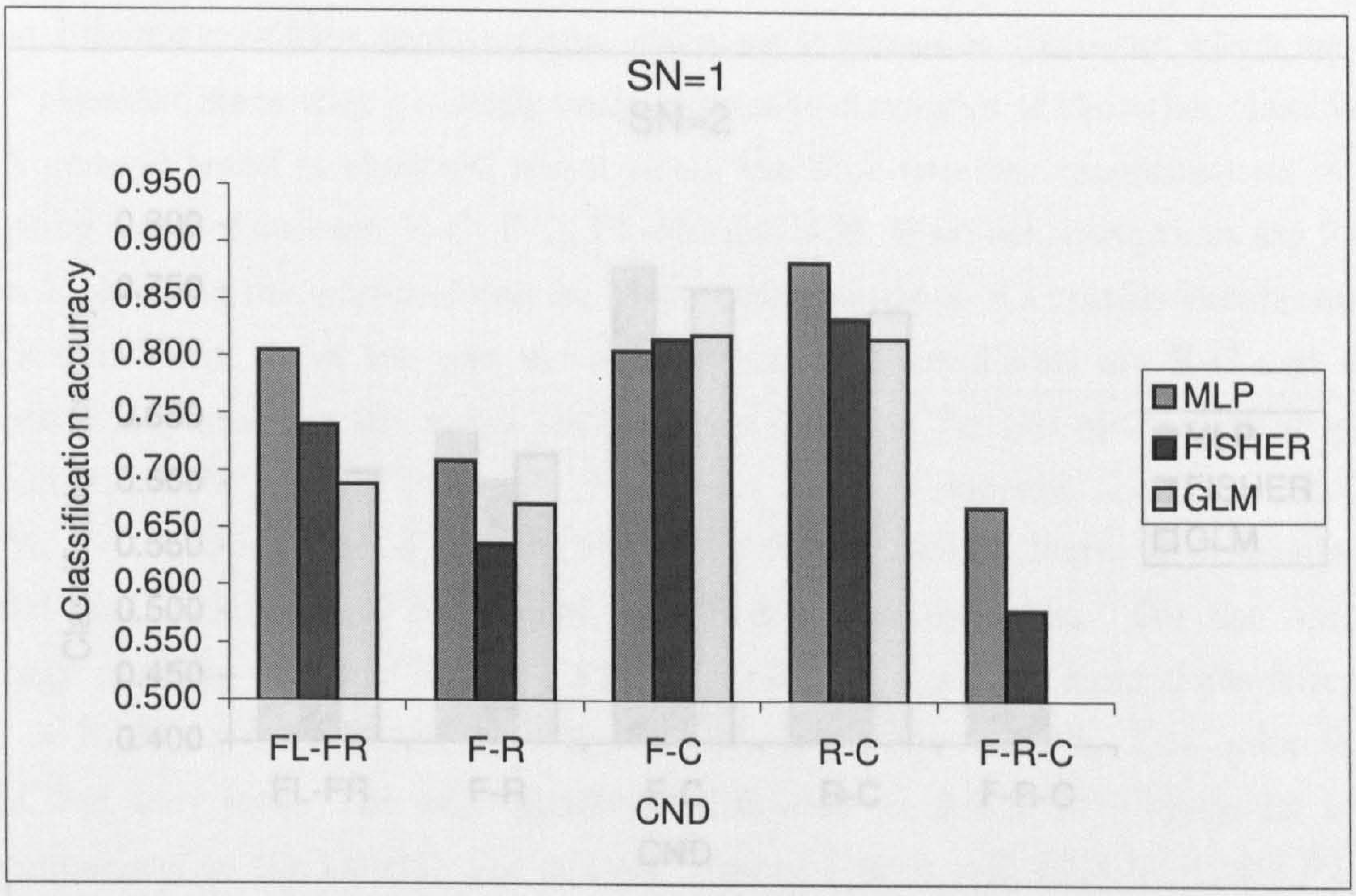

Figure 5.3: Bar plot of classification fitness for all four condition comparisons for optimal strategies, as a function of classifier. SN=1.

the GLM and FISHER classifier based strategies are still capable of achieving significantly better than chance classifications. For example, consider the FL-FR condition pair. The optimal strategies using the MLP yield 80.5%, 70.3% and 77.4% for subjects 1, 2 and 3 respectively, GLM yields 68.8%, 60.8% and 70.7%, and FISHER, 74.1%, 64.1% and 66.5%. To further gauge the performance of the 3 classifiers, the grand average classification fitness of the first four optimal strategies in the top-ten tables have been computed: MLP = 75.2% , GLM = 70.4% and FISHER = 68.6%. This trend is also reflected in the averages and grand averages tables. Figures ?? – ?? also display these results.

5.4.6 Conditions

The condition combinations of most interest from the BCMI perspective are F-R and FL-FR, as these both incorporate musical tasks, as opposed to the counting task, which is included in this experiment primarily as a measure to control for modality (see earlier discussion). However, since the analysis has been performed for all the other combinations, they will be discussed here as well. Tables 5.11, 5.12 and 5.13 summarise the optimal strategies for each subject and condition combination for the FISHER, GLM and MLP classifiers respectively. When comparing the relative success

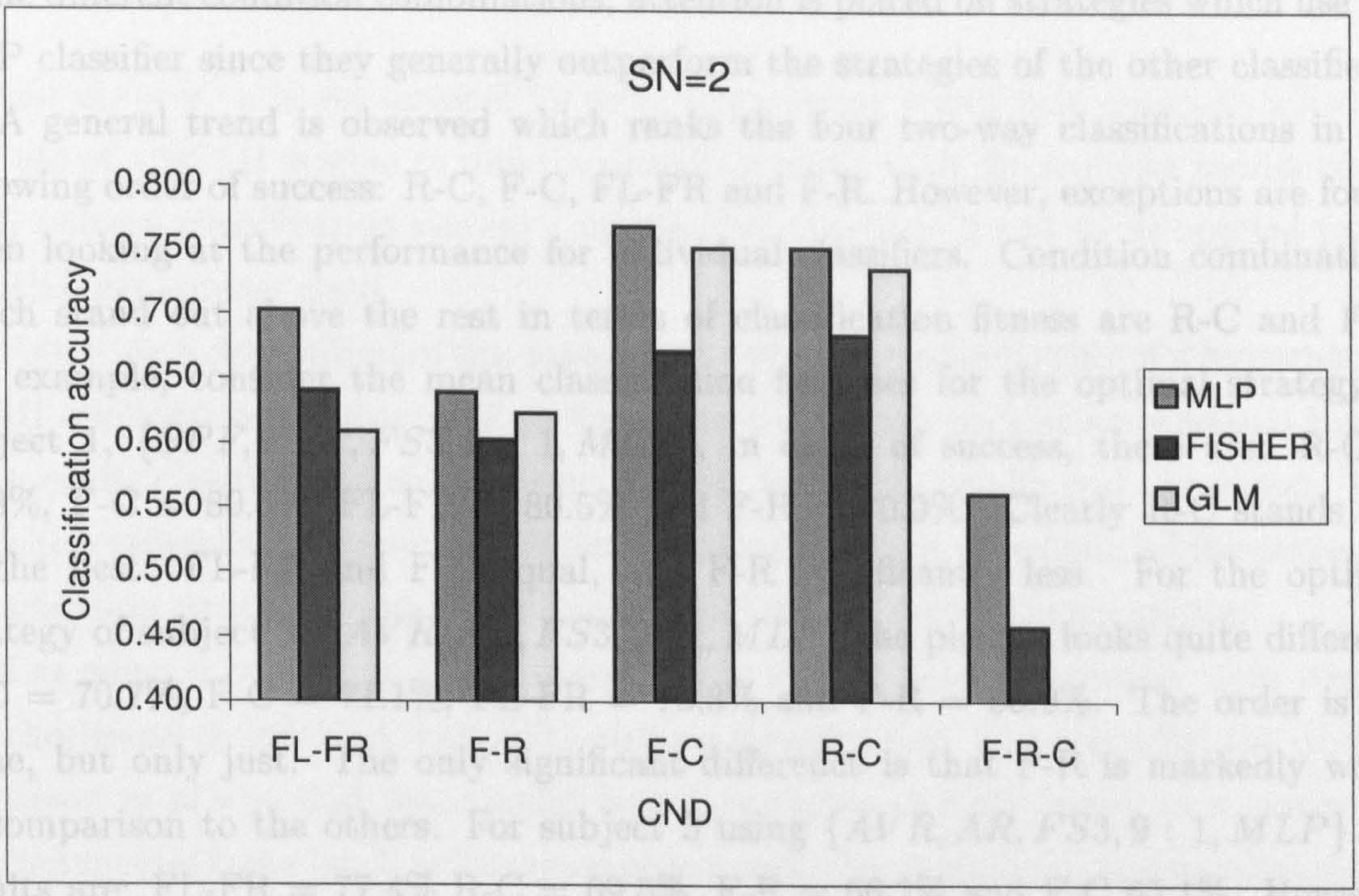

Figure 5.4: Bar plot of classification fitness for all four condition comparisons for optimal strategies, as a function of classifier. SN=2.

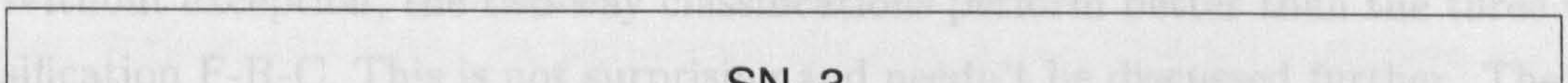

Figure 5.5: Bar plot of classification fitness for all four condition comparisons for optimal strategies, as a function of classifier. SN=3.

of the different condition combinations, attention is placed on strategies which use the MLP classifier since they generally outperform the strategies of the other classifiers.

A general trend is observed which ranks the four two-way classifications in the following order of success: R-C, F-C, FL-FR and F-R. However, exceptions are found when looking at the performance for individual classifiers. Condition combinations which stand out above the rest in terms of classification fitness are R-C and F-C. For example, consider the mean classification fitnesses for the optimal strategy of subject 1, $\{SPF, FFT, FS3, 9 : 1, MLP\}$, in order of success, these are: R-C = 88.3%, F-C = 80.5%, FL-FR = 80.5% and F-R = 70.9%. Clearly R-C stands out as the best: FL-FR and F-C equal, and F-R significantly less. For the optimal strategy of subject 2, $\{AVR, AR, FS3, 9 : 1, MLP\}$ the picture looks quite different: R-C = 70.7%, F-C = 71.1%, FL-FR = 70.3% and F-R = 60.9%. The order is the same, but only just. The only significant difference is that F-R is markedly worse in comparison to the others. For subject 3 using $\{AVR, AR, FS3, 9 : 1, MLP\}$ the results are: FL-FR = 77.4% R-C = 69.5%, F-R = 66.2% and F-C 65.1%. However, when using $\{NONE, FFT, FS3, 9 : 1, MLP\}$ which performs well for all 4 condition comparisons, one sees: R-C = 78.8%, FL-FR = 68.9% F-C = 67.6%, F-R = 65.4% respectively.

Without exception, the two-way classifications perform better than the three way classification F-R-C. This is not surprising and needn't be discussed further. The following strategy: $\{NONE, FFT, FS3, 9 : 1, MLP\}$ yields 53.3% which is 20% better than the average random guess for a three-class problem which is of course 33.3% (assuming equal a priori probabilities for each class).

5.4.7 Subjects

Strategies tested on subject 1 yield the best results, followed closely by those for subject 3 and less closely by those of subject 2.

Consider for example the optimal strategies (see Table 5.13) for the two-way classification FL-FR. For subjects 1, 2 and 3 these are 80.5%, 70.3% and 77.4% respectively. Similarly, for the F-R condition pair: 70.9%, 62.9% and 66.2%.

For the three-way comparison, F-R-C, one has: 66.9%, 56.1% and 53.3%, which implies that subject 2 outperforms subject 3 for this condition combination[13].

[13] Due to the order of standard deviation and confidence limits which are both approximately 3%, it cannot be said that there is a significant difference between the performance of these strategies.

5.5 Conclusions

In this chapter, a novel musical focusing experiment which was designed with the end use of a BCMI system in mind is described along with the analysis results of the EEG pattern classification system - described in Chapters 2 and 4 - as applied to the focusing experiment data.

Following a description of the experiment (which includes details of a hypothetical BCMI application utilising musical focusing) a thorough performance-related breakdown of the classification sub-systems is given, by way of results and discussion. This leads to a refined set of recommendations relating to which variations (of the numerous classification strategies employed) might be worthy of further investigation.

The general behaviour of the classification system acts accordingly with respect to other research involved with EEG pattern classification. In particular, the work of Anderson *et al* who find that similar length segments of EEG based on a number of distinct mental tasks (such as imagined object rotation, relaxation, mental arithmetic) can be successfully classified using similar classification strategies as those employed here [AS96].

The classification fitnesses achieved by the optimal strategies reported in this chapter, being in the order of 60 - 80% for two way classifications, should warrant further work in this field. In particular, the use of pre-classifier processes such as spatial filtering and autoregressive modelling, as a way of reducing the dimensionality of multi-channel EEG data before classification proper is performed by a nonlinear classifier (such as static-multilayer perceptrons or generalised linear models).

The classification methodology described in Chapters 2 and 4 is evaluated for many sub-system variations. The key findings relating to these sub-systems (stages of the classification methodology) are detailed below:

Pre-processing

An improved classification performance due to pre-processing filters, especially Hjorth's Laplacian spatial filter, and to a lesser extent, the average reference filter (also known as common reference filtering) is observed. Of the four pre-processing variations investigated (NONE, AVR, SPF and LPF), AVR and SPF are found to considerably improve the classification performance when compared to NONE. LPF doesn't even match the performance of NONE.

Feature extraction

The success of linear autoregressive model coefficients (for two out of three subjects) suggests a possible superior feature extraction method over traditional FFT based measures. Of the four methods investigated (AR, ARMO, FFT and PSD), AR and FFT are found to considerably improve the classification performance when compared to NONE. Furthermore, ARMO falls short of the performance of NONE.

Feature selection

Classifiers trained on data from a large number of channels, such as the 128-channel dense array system used here can achieve significantly better results than those trained on data from a handful of electrodes. However, in some cases, as few as 12 electrodes (in combination with the GLM classifier) achieved comparable results. Classifiers presented with features made up of the largest channel set (FS3) perform the best most of the time. However, when using the GLM classifier, the optimal strategies are slightly more in favour of the medium set, FS2 (based on the international 10-20 set of electrodes). Strategies employing the smallest set, FS1, which consist of the 4 temporal electrodes perform comparatively worse in all but a few cases[14]. In this work, the classification system attempts to classify 2-second segments on an individual basis. However, this time window could be increased several fold without compromising the needs of a BCMI system, the result of which might lead to a considerably improved classification fitness, as other studies have shown [AS96, PPF97, PP99, PRCS00]. This in turn might allow, among other things, a reduced channel set to be employed. These and other ideas will discussed in more detail in Chapter 6.

Training set size

Of the two data split ratios, 9:1 and 1:1), the former (which results in 90% of the available data being used for training) only results in a marginal improvement over strategies which employ the 50%/50% data split, 1:1. The fact that the smaller training set size performs almost as well as the larger one is encouraging.

[14]Since one of the BCMI engineering ideals is to reduce the number of EEG electrodes to a minimum, more work is required in exploring solutions which employ smaller channel sets. This and other issues concerning future developments of this work are given in Chapter 6.

Classifiers

Non-linear classifiers, namely the static-multilayer perceptron and generalised linear model neural networks can outperform linear classifiers, such as the Fisher discriminant. The three classifiers listed in order of success are: MLP, GLM and FISHER.

Subjects

All three subjects achieve good classification results. Subject 1 yields the best results when utilising the following strategy: $\{SPF, FFT, FS3, 9:1, MLP\}$, whereas subjects 2 and 3 utilise $\{AVR, AR, FS3, 9:1, MLP\}$. These strategies take into account the importance of the FL-FR and F-R condition comparisons over the remaining comparisons which include the counting task.

5.6 Summary

The results are very encouraging and help validate the hypothesis stated in the introduction, that is: a person's EEG contains information allowing one to ascertain - to a reasonable degree of success[15] - which one of 3 mental tasks (musical focusing, passive listening, and counting) they are performing.

Although there exist so far no studies exactly like this one, the results agree with the findings of studies which deal with DSP realised EEG pattern classification of mental tasks. Without a doubt, the results from this experiment bode well for the case in favour of a BCMI which utilises specific music related mental tasks, as opposed to more abstract mental activities such as learned control over the alpha wave (8-12 Hz) EEG component, via relaxation.

[15] In this case, 'reasonable' means statistically better than chance.

Chapter 6

Conclusions and Future Work

The key point of this work has not been to try and solve the mystery of the brain, rather, to evaluate whether the EEG can be harnessed *in new ways* during certain musical situations, in a way that would allow thought-related control of an interactive musical environment. The work described in this thesis started with a thorough search for studies in the field of *thought-controlled musical devices* [DMS98a]. With the exception of one study [Ros90], no existing work was found that specifically sought to address the author's concept of a BCMI, which is briefly defined as:

> "A musical synthesis device that uses the knowledge of the *presence* or *absence* of certain musical thoughts or experiences, by means of a *brain-computer interface* and *EEG analysis engine*, so as to allow thought-control of the music that is subsequently created."

Although this concept has been suggested by Rosenboom, [Ros90], work has yet to be carried out which attempts to address the engineering demands of such as system in the context of modern day technologies, especially the rapidly expanding field of brain-computer-interfacing. Hence, to the best of the author's knowledge, this thesis opens up a new topic, fusing the domains of *experimental computerised musical instruments* and *brain-computer interfacing.*

6.1 Major contributions

6.1.1 A new area of research

Besides opening a way forward in the field of BCMI systems, the major contribution of this research has been the building and testing of a *BCMI evaluation protocol* incorporating the design and implementation of novel EEG experiments, and the development and evaluation of various EEG pattern classification strategies, some of which are novel. This was achieved by the following iterative procedure, the fruits of which are reported in Chapters 3, 4, and 5 respectively:

1. *Hypothetical BCMI applications outlined:* The purpose of this stage was to establish classification problems that formed a basis upon which EEG experiments were designed. To this end, three problems were defined, each based on the basic BCMI framework outlined in Chapter 1.

2. *Novel EEG experiments designed and implemented:* Given the hypothetical BCMI applications and associated classification problems, suitable experiments were designed and implemented that provided data for the evaluation of various state-of-the-art classification methods taken from the BCI field.

3. *Data analysis:* Involving a systematic evaluation of EEG pattern classification methods, with a view to both validating the BCMI concept, and locating successful classification strategies and experimental paradigms for future iterations/developments.

The results of this work have led to a number of insights relating to the plausibility of the novel BCMI concepts described in this thesis. These insights are detailed below.

6.1.2 Insights gained

BCMI concept is plausible

The primary finding of this research is that the concept of the BCMI, as defined above (and elsewhere in the thesis) is without doubt plausible with current technology, such as those methods found to work in this thesis.

ERP based experiments worthy of further investigation

The ERP-based auditory stimulus experiment described in Chapter 3, demonstrates that successful classification of single 1-second segments of pre and post-stimulus onset EEG is possible by means of a novel correlation-based feature extraction technique. The experiment also shows that the EEG contains information concerning the experience of music (in this case, the perception of simple tones heard over silence), and that this information is accessible in a reasonable time frame. The results warrant further investigation into ERP-based BCMI systems.

Music related mental tasks can be classified

Both the musical imagery and musical focusing experiments (Chapters 4 and 5) boast positive results in favour of BCMI systems that utilise discrete musically relevant mental tasks. In particular, the successful pair-wise classifications (imagery versus relaxing, focusing versus relaxing, and focusing-left versus focusing-right) show that both imagery and focusing could be considered as viable candidate tasks for any future BCMI research.

Optimal classification methods correlate with BCI findings

Optimal strategies correlate with finding from BCI research, in particular, the success of Laplace spatial filtering for pre-processing raw EEG data, linear autoregressive modelling for feature extraction, and *static feedforward multilayer perceptron neural networks* for classification.

Need to consider both large and small-channel-set systems

The fact that the successful classification results reported in this thesis were based on features obtained from large channel sets (between 20 and 128 electrodes) may not detract from the case in favour of the BCMI concept. The reasons for this are as follows. First, the experiments confirm the principle behind the BCMI concept is viable. Second, although the ideal real-world BCMI would be a portable easy-to-use device, harbouring a small EEG sensor array (perhaps 2 to 4 electrodes as part of a head band), there is no reason why large-channel-set systems could not be realised. For example, the 128-channel device used in this thesis can be set up in less than 30 minutes. Thirdly, it is likely that the methods described in this thesis can be adjusted to account for small-channel-set situations. For details, see the section on future work.

6.2 Future work

6.2.1 Refine classification methods

Advanced search techniques for refining optimal classification strategies

As a result of the great number of possible pre-processing options available, and since the consequence of the choice of these options is crucial to the success of the classifier, it is necessary to adopt a systematic approach that enables a large number of these options to be compared. One particular technique known to be useful in engineering problems where an exhaustive search becomes practically infeasible, is evolutionary methods such as genetic algorithms. These could be employed as a method of quickly honing in on optimal strategies, allowing a greater number of classification strategies to be tested.

Multiple segment averaging

One possible solution that might improve the classification performance of the optimal strategies would be multiple segment averaging which involves making classifications over a number of short segments, then combining the results by choosing the most prevalent class. This methodology has been shown to increase accuracy significantly in other studies. For example, Anderson et al. [AS96] find that, for a five-class problem, averaging the results of 20 half-second segments improves the classification accuracy by up to 16%.

Committees of networks

Committees of networks have been successfully employed by Pfurtscheller's group [PPF97]. Rather than training a single network on features from all the channels, they used a separate network for each channel, then formed a committee that consisted of the strongest networks (i.e. those that performed the best). The class attributed to new EEG segments is thus formed by taking a 'vote' which simply consists of labelling the segment as belonging to the class which received the highest number of votes.

6.2.2 Refine experimental paradigms

Mental tasks

More experiments are required to validate the usefulness of musical imagery and focusing. Particular attention needs to be placed on integrating ideas about how a real-world BCMI would utilise the ability to detect the presence or absence of these mental tasks in a subject's EEG. It would also be worthwhile to devise experiments that incorporated both musical imagery, focusing, and passive listening. A successful result in this scenario would certainly be exciting!

Experimental artefacts

Further experimentation should be undertaken where emphasis is placed on the need to rule out the chance of contamination due to experimental artefacts, such as those experienced in the musical imagery experiment. For example, the musical imagery task could be tested in a paradigm similar to that of the focusing experiment.

Towards adaptive systems

More work needs to be undertaken to establish experimental paradigms which reflect the sort of challenges that would be expected in a real-world, on-line system. For example, the fact that during the course of time, the underlying statistics of the EEG are changing. For example, as the subject becomes tired, alpha (frequencies in the 8-12Hz range) tends to increase [NLDS98]. Therefore, experiments should be set-up which allow for adaptive classification methods (such as Hidden Markov Models [PR99]) to be evaluated to compensate for the changeable nature of the EEG during longer periods of time.

6.2.3 On-line prototypes

The findings of this thesis indicate that a prototype BCMI which incorporates imagery and focusing mental tasks would be worthy of investigation. However, in order to achieve suitable classification accuracies in a real-time environment, further off-line experimentation would be advisable.

Appendix A

Artefact detection algorithms

A brief description of the artefact detection algorithms (including pseudo-code) is given below. The algorithms are designed to detect eye-blink and eye-movement artefacts, and bad channels, from a segment of multi-channel EEG (acquired using the 128-channel geodesic net of EGI (http://www.egi.com).

A.1 Eye-blink artefact detection algorithm†

This algorithm is applied twice (once for each pair of eye-blink channels {8 126} and {127 128}) to detect eye-blink artefacts. It compares the deviation between fast and slow running averages of a pair of eye-blink channels with a threshold. An eye-blink is detected when the deviation exceeds the threshold level.

```
Fast = 0
Slow = average of difference of 1st 10 samples
for each sample:
  Diff = difference in voltage of eye channels
  Fast = 0.8 * Fast + 0.2 * (Diff - Slow)
  Slow = 0.975 * Slow + 0.025 * Diff
  if |Fast| > Eye Blink Threshold (70 μV), reject segment
```

(See figure A.1 for example plot.)

A.2 Eye-movement artefact detection algorithm†

The algorithm used for detecting eye-movement artefacts is exactly the same as for eye-blinks, except that it uses the two horizontal eye channels {128 125}. (See figure

A.2 for example plot.)

A.3 Bad channel artefact detection algorithms

Two algorithms were used at different times to detect bad channels.

Bad channel detection algorithm A

This algorithm rejects channels which sample's exceed a 200 μV threshold more than 10% of the time.

```
for each channel:
  count = 0
  for each sample:
    if |Sample| > 200 (μV), count = count + 1
  end
  if count > 0.1 * number of samples in segment, label channel as bad
end
```

Bad channel detection algorithm B†

This algorithm measures the difference between fast and slow running averages of channel amplitudes, and compares this with a Voltage Threshold and Transit Threshold. It has the effect of detecting both high frequency noise (such as mains hum) and low frequency drift.

```
for each channel:
  Fast = 0
  Slow = average voltage of 1st 10 samples
  for each sample:
    Fast = 0.8 * Fast + 0.2 * sample voltage
    Slow = 0.975 * Slow + 0.025 * sample voltage
    Diff = Fast - Slow
    if |Diff| > Transit Threshold (100 μV), label channel as bad (signal transit rejection)
    else if |Fast| > Voltage Threshold (200 μV), label channel as bad (signal voltage rejection)
  end
end
```

† These algorithms are the same as those described in the EGI Averager software documentation (**reference**).

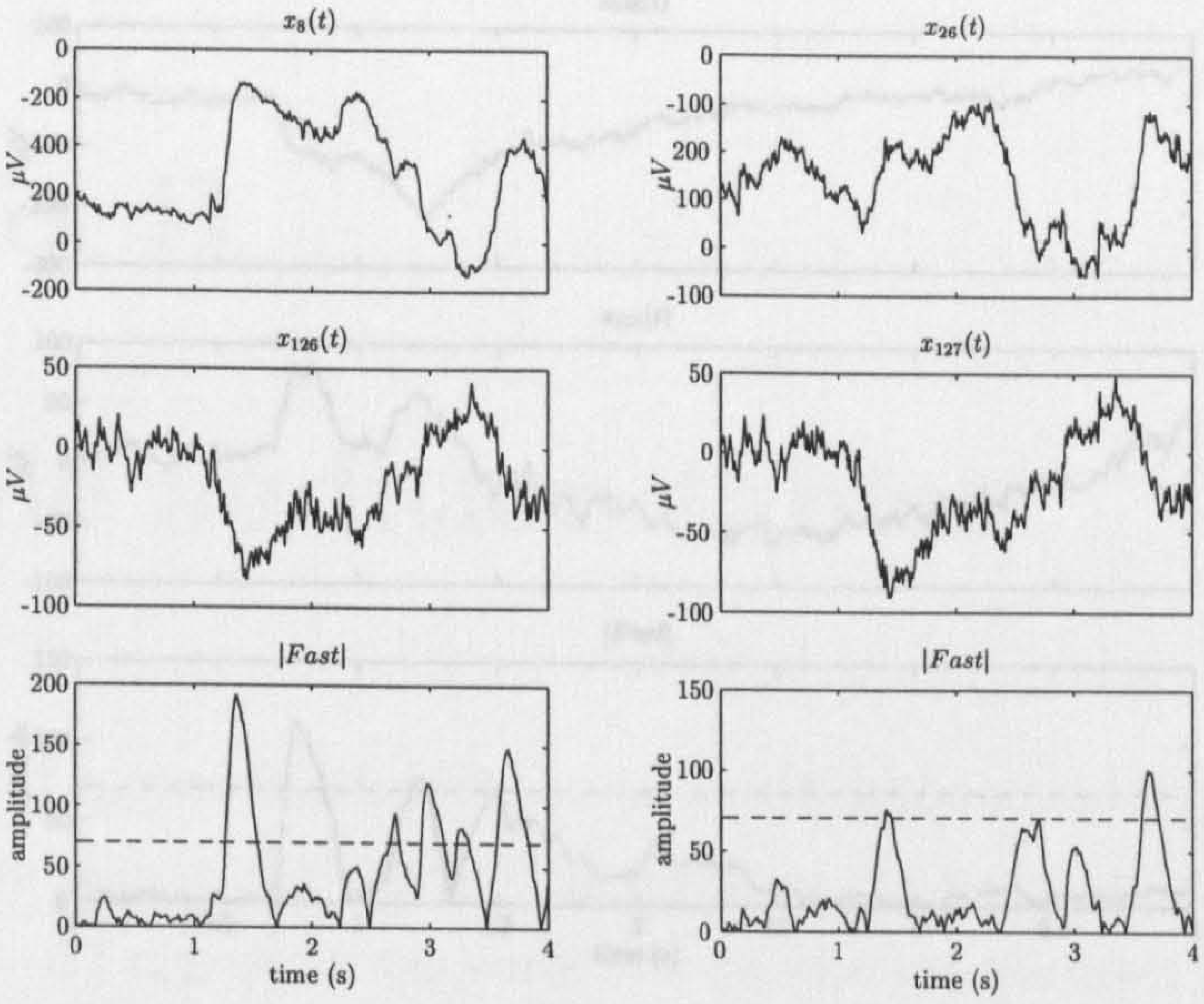

Figure A.1: Eye-blink detection algorithm finds possible artefact in both sets of eye channels. Dotted line on lower plots indicates the 70 μV threshold.

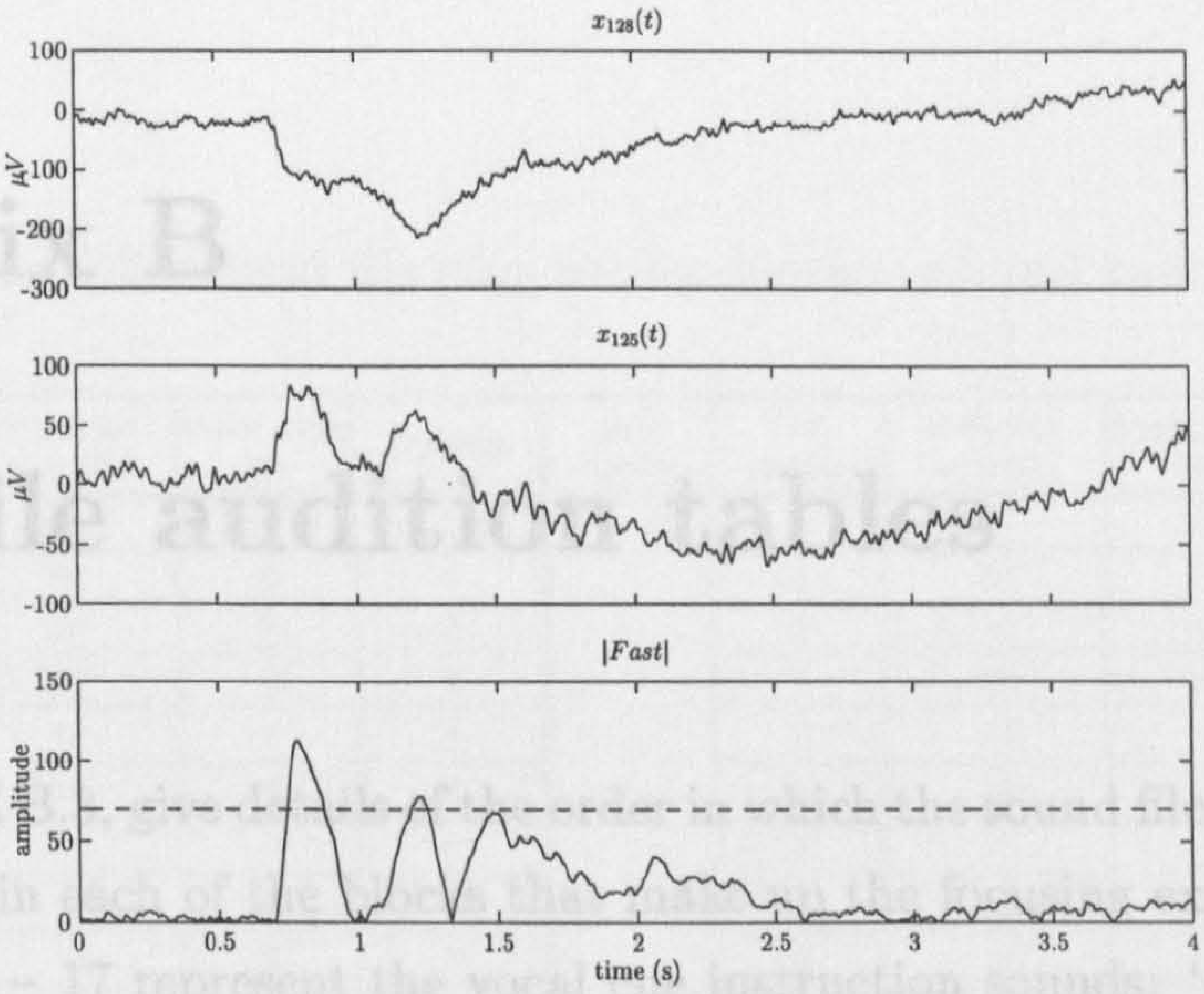

Figure A.2: Eye-movement detection algorithm finds possible artefact. Dotted line on lower plots indicates the 70 μV threshold.

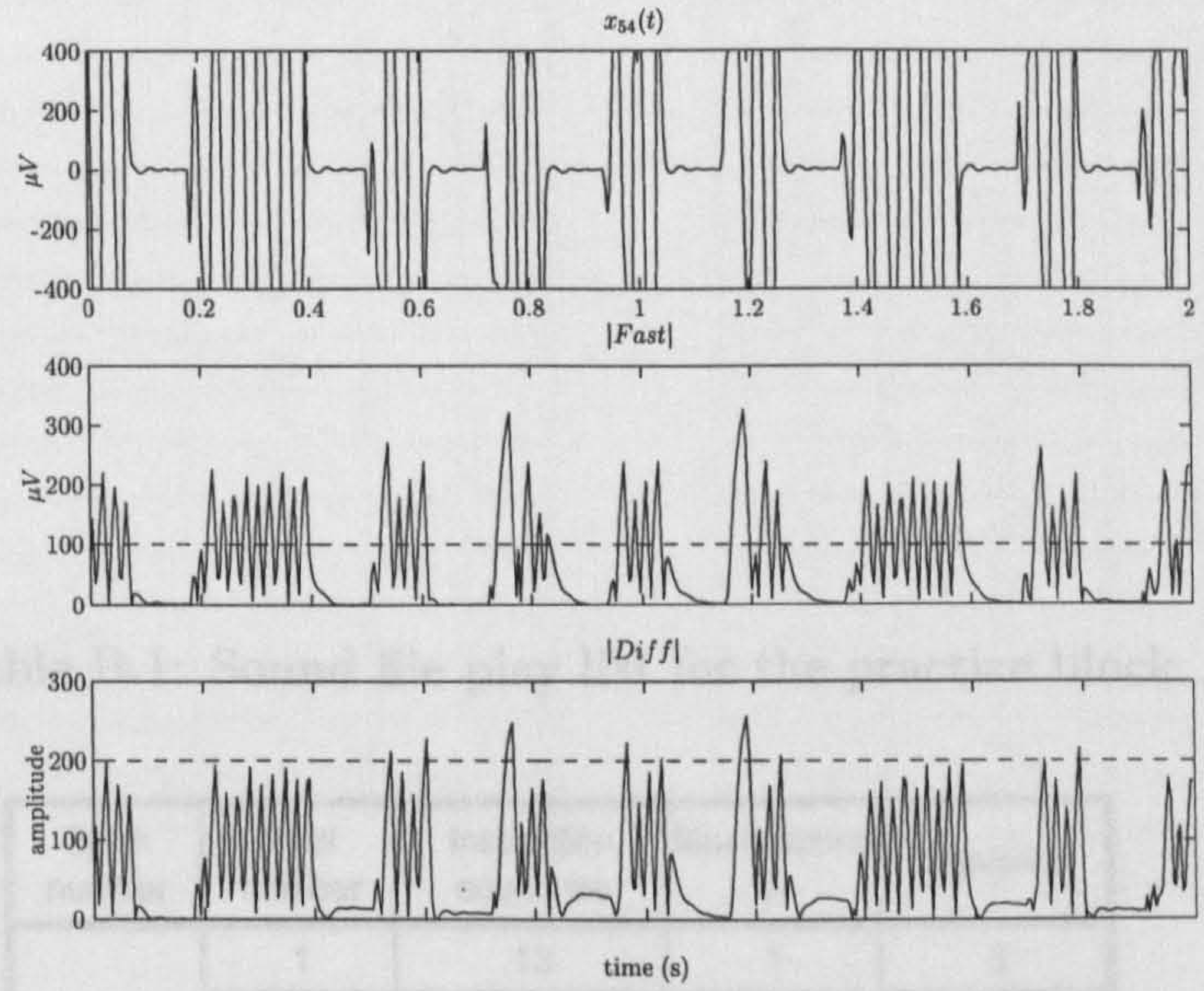

Figure A.3: Bad channel detection algorithm finds a bad channel.

Appendix B

Sound file audition tables

Tables B.1, B.2 and B.3, give details of the order in which the sound files listed in Table 5.1 are auditioned in each of the blocks that make up the focusing experiment. Note that sound files 13 – 17 represent the vocal cue instruction sounds: 'guitar', 'synth', 'piano', 'relax' and 'count' respectively, and conditions 1 - 5 refer to conditions: FL, FR, R, C and Practice respectively.

Table B.1: Sound file play list for the practice block.

Block number	Trial number	Instruction sound file	Music sound file	Condition
0	1	13	1	5
	2	14	2	5
	3	15	3	5
	4	16	4	5
	5	17	5	5

Table B.2: Sound file play list for blocks one and two.

Block number	Trial number	Instruction sound file	Music sound file	Condition	Block number	Trial number	Instruction sound file	Music sound file	Condition
1	1	14	4	1	2	1	13	12	2
	2	15	6	1		2	13	7	1
	3	16	10	3		3	16	1	3
	4	17	10	4		4	16	8	3
	5	13	2	1		5	14	8	2
	6	15	12	1		6	17	10	4
	7	17	6	4		7	16	12	3
	8	16	12	3		8	13	4	2
	9	15	9	2		9	16	5	3
	10	16	8	3		10	16	7	3
	11	17	9	4		11	15	11	1
	12	17	11	4		12	13	1	1
	13	15	3	2		13	17	4	4
	14	17	7	4		14	14	3	1
	15	16	9	3		15	17	5	4
	16	16	4	3		16	16	2	3
	17	14	5	2		17	17	7	4
	18	14	11	2		18	16	4	3
	19	15	7	2		19	16	6	3
	20	17	4	4		20	17	9	4
	21	15	1	2		21	17	3	4
	22	16	5	3		22	17	8	4
	23	16	7	3		23	14	9	1
	24	13	8	1		24	17	12	4
	25	17	3	4		25	17	2	4
	26	17	8	4		26	16	9	3
	27	17	12	4		27	16	10	3
	28	17	5	4		28	17	11	4
	29	16	11	3		29	17	1	4
	30	17	2	4		30	15	5	1
	31	16	6	3		31	13	10	2
	32	16	3	3		32	14	2	2
	33	16	2	3		33	13	6	2
	34	16	1	3		34	16	3	3
	35	17	1	4		35	16	11	3
	36	14	10	1		36	17	6	4

Table B.3: Sound file play list for blocks three and four.

Block number	Trial number	Instruction sound file	Music sound file	Condition	Block number	Trial number	Instruction sound file	Music sound file	Condition
3	1	17	3	4	4	1	16	4	3
	2	17	2	4		2	16	10	3
	3	14	11	2		3	17	11	4
	4	15	7	2		4	17	6	4
	5	14	4	1		5	14	8	2
	6	17	10	4		6	14	9	1
	7	14	5	2		7	16	6	3
	8	17	9	4		8	17	10	4
	9	17	12	4		9	16	9	3
	10	17	4	4		10	16	8	3
	11	15	12	1		11	14	2	2
	12	16	3	3		12	17	9	4
	13	16	7	3		13	13	10	2
	14	17	6	4		14	14	3	1
	15	14	10	1		15	16	3	3
	16	16	6	3		16	16	5	3
	17	16	9	3		17	17	3	4
	18	17	8	4		18	13	4	2
	19	17	1	4		19	16	12	3
	20	16	1	3		20	15	11	1
	21	17	5	4		21	17	7	4
	22	13	8	1		22	17	1	4
	23	13	2	1		23	15	5	1
	24	15	6	1		24	13	7	1
	25	16	12	3		25	16	1	3
	26	17	7	4		26	13	1	1
	27	15	9	2		27	16	2	3
	28	16	5	3		28	17	4	4
	29	16	2	3		29	16	11	3
	30	17	11	4		30	13	12	2
	31	16	10	3		31	17	8	4
	32	15	3	2		32	13	6	2
	33	16	4	3		33	17	5	4
	34	16	8	3		34	16	7	3
	35	16	11	3		35	17	12	4
	36	15	1	2		36	17	2	4

References

[ABST93] H. D. I. Abarbanel, R Brown, J. J. Sidorowich, and L. S. Tsimring. The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65(4):1331–1391, October 1993.

[ADS95] C. Anderson, S. Devulapalli, and E. Stolz. EEG signal classification with different signal representations. *Neural Networks for Signal Processing V*, pages 475–483, 1995.

[AS96] C. Anderson and Z. Sijercic. Classification of EEG signals from four subjects during five mental tasks. *Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks (EANN'96)*, pages 407–414, 1996.

[BGH+99] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398:297–298, 1999.

[Bis95] C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[Car98] R. Carter. *Mapping the Mind.* Weidenfield and Nicolson London, 1 edition, 1998.

[CDA93a] E. Ciaccio, S. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems - part 1 of 4: Fundamental concepts. *IEEE Engineering In Medicine and Biology*, 12(3):89–97, September 1993.

[CDA93b] E. Ciaccio, S. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems - part 2 of 4: Methods for feature extraction and selection. *IEEE Engineering In Medicine and Biology*, 12(4):106–113, December 1993.

[CDA94a] E. Ciaccio, S. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems - part 3 of 4: Methods of classification. *IEEE Engineering In Medicine and Biology*, 13(1):129–135, February 1994.

[CDA94b] E. Ciaccio, S. Dunn, and M. Akay. Biosignal pattern recognition and interpretation systems - part 4 of 4: Review of applications. *IEEE Engineering In Medicine and Biology*, 13(2):269–283, April/May 1994.

[CH77] M. Critchley and R. Henson. *Memory and Attention in Music.* Willian Heinemann Medical Books Limited, 1977.

[DDO98] R. Dafters, F. Duffy, and P. O'Donnell. Changes in brain function following long-term MDMA ("ecstasy") use revealed by multi-channel quantitative EEG. Department of Psychology, Glasgow University, 1998.

[Deu77] D. Deutsch. *Memory and Attention in Music.* In [CH77], 1977.

[Die98] J Dien. Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*, 30(1):34–43, 1998.

[DMS98a] A. A. Duncan, E. R. Miranda, and K. C. Sharman. Mind the music: Towards thought-controlled music systems. Proceedings of the V Brazilian Symposium on Computer Music, 1998.

[DMS98b] A. A. Duncan, E. R. Miranda, and K. C. Sharman. Towards a neurocompatible interface for music research. In *Journees d'Iformatique Musicale 98*, volume 148, pages F3–1–F3–6, Marseille, France, 1998. Publications du LMA.

[FH98] S. Fels and G. Hinton. Glove-talkii–a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Netwoks*, 9(1):205–212, January 1998.

[Fot98] E. Fottorino. Une machine clibataire. *Le Monde*, pages 12–13, feb 1998.

[Fre98] W. J. Freeman. Local phase gradients in neocortical EEGs show repeated phase transitions that form domains of spatially coherent chaotic activity. *to appear in: International Journal of Bifurcation and Chaos*, 1998.

[Gur97] K. Gurney. *An Introduction to Neural Networks.* UCL Press, 1997.

[GWD+97] M. Galicki, H. Witte, J. Dorschel, M. Eiselt, and G. Griessbach. Common optimisation of adaptive preprocessing units and a neural network during the learning period. application in EEG pattern recognition. *Neural Networks*, 10(6):1153–1163, 1997.

[HH77] G Harrer and H. Harrer. *Music, Emotion and Autonomic Function.* In [CH77], 1977.

[HJ93] U. Hegerl and G. Juckel. Intensity dependance of auditory evoked potentials as an indicator of central serotonergic neurotransmission: a new hypothesis. *Biological Psychiatry*, 33:173–187, 1993.

[Hjo75] B. Hjorth. An on-line transformation of EEG scalp potentials into orthogonal source derivations. *Electronencephalography and Clinical Neurophysiology*, 39:526–530, 1975.

[HPGM95] F. Hassainia, D. Petit, S. Gauthier, and J. Montplaisir. Topographical study of the heterogeneity of impairments in early Alzheimer's disease patients. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, volume 17 of *2*, pages 1009–1010, 1995.

[Hug95] K Hugdahl. *Psychophysiology - The Mind Body Perspective.* Harvard University Press, 1995.

[JMS97] T. Jung, S. Makeig, and M. Stensmo. Estimating alertness from the EEG power spectrum. *IEEE Transactions on Biomedical Engineering*, 44:60–69, 1997.

[JP93] P. Janata and H. Petsche. Spectral analysis of the EEG as a tool for evaluating expectancy violations of musical contexts. *Music Perception*, 10(3):281–304, Spring 1993.

[LKL94] H. Lusted, R. Knapp, and A. Llyod. Biocontrollers for virtual environments. Discover, May 1994.

[LPI95] D. H Lange, H Pratt, and G. F. Inbar. Segmented matched filtering of single event related evoked potentials. *IEEE Trans. Biomed. Eng.*, 42:317–321, 1995.

[MEJS00] S. Makeig, S. Enghoff, T-P. Jung, and J Sejnowski, T. A natural basis for efficient brain-actuated control. *IEEE Transactions on Rehabilitation Engineering*, 8(2):208–11, 2000.

[Mir97] E. Miranda. La msica, las mquinas, la inteligencia y el cerebro. *Letra Internacional*, 53:40–43, 1997.

[MKB97] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1997.

[MN83] P McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1983.

[Mul73] L. J. Mullins. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2:157–180, 1973.

[NK81] P. L. Nunez and R. Katznelson. *Electrical Fields of the Brain – The Neurophysics of EEG*. New York: Oxford University Press, 1881.

[NLDS98] E. Niedermeyer and F. Lopes-Da-Silva. *Electroencephalography : Basic Principles, Clinical Applications, and Related Fields*. Lippincott, Williams and Wilkins, 4 edition, October 1998.

[NS] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *http://www.aos.Princeton.EDU/WWWPUBLIC/tapio/arfit/*.

[Oki95] K. Oki. Brain wave rider: A human-machine interface. *Leonardo*, 28(4):307–311, 1995.

[Pet96] H. Petsche. Approaches to verbal, visual and musical creativity by EEG coherence analysis. *International Journal of Psychophysiology*, 24:145–159, 1996.

[PGH97] L. Pelosi, J. Geesken, and M. Holly. Working memory impairment in early multiple sclerosis - evidence from an event-related potential study of patients with clinically isolated myelopathy. *Brain*, 120:2039–2058, 1997.

[Por97] B. Porat. *A Course in Digital Signal Processing*. John Wiley and Sons, Inc., 1 edition, 1997.

[PP99] M Pregenzer and G Pfurtscheller. Frequency component selection for an EEG-based brain to computer interface. *IEEE Transactions on Rehabilitation Engineering*, 7(4):413–419, December 1999.

[PPF97] B. O. Peters, G. Pfurtscheller, and H. Flyvberg. Prompt recognition of brain states by their EEG signals. *Theory in Biosciences*, 116:290–301, 1997.

[PR99] W.D Penny and S. J. Roberts. Dynamic models for nonstationary signal segmentation. *Computers and Biomedical Research*, 32(6):483–502, December 1999.

[PRCS00] W.D Penny, S. J Roberts, E Curran, and M. Stokes. EEG-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, 8(2), June 2000.

[RL97] J. Rosenfeld and J. Lubar. From the guest editors. *Biofeedback - Newsmagazine of the Association for Applied Psychophysiology and Biofeedback*, 25(1):2, 1997.

[Roa96] C. Roads. *The Computer Music Turorial.* The MIT Press, 1996.

[Ros90] D. Rosenboom. Extended musical interface with the human nervous system. *Journal of the International Society for the Arts, Sciences and Technology, Leonardo Monograph No.1*, 1990.

[RP00] S.J Roberts and W Penny. Real-time brain computer interfacing: a prelininary study using bayesian learning. *Medical & Biological Engineering and Computing*, 38(1):56–61, 2000.

[RR98] I.A. Rezek and S.J. Roberts. Stochastic complexity measures for physiological signal analysis. *IEEE Transactions on Biomedical Engineering*, 44(9):1186–1191, 1998.

[SH99] M. I. Schlesinger and V. Hlavac. *Ten lectures from the statistical and structural pattern recognition theory.* CVUT, Prague, 1999.

[SKI97] N. Saiwaki, K. Kato, and S. Inokuchi. An approach to analysis of EEGs recorded during music listening. *Journal of New Music Research*, 26:227–243, 1997.

[SN] T. Schneider and A. Neumaier. Algorithm: Arfit - a matlab package for the estimation and spectral decomposition of multivariate autoregressive models. *http://www.aos.Princeton.EDU/WWWPUBLIC/tapio/arfit/.*

[Sou] Soundbeam. http://www.soundbeam.co.uk.

[Str90] F. G. Stremler. *Introduction to Communication Systems.* Addison-Wesley, 1990.

[Swi96] K. Swingler. *Applying Neural Networks: A Practical Guide.* Academic Press, 1996.

[Tuc93] D. Tucker. Spatial sampling of head electric fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87:154 163, 1993.

[WK96] W. Weng and K. Khorasani. An adaptive structure neural networks with application to EEG automatic seizure detection. *Neural Networks*, 9(7):1223-1240, 1996.

[WMNF91] J. R. Wolpaw, D. J. McFarland, D. J. Neat, and C. A. Forneris. An EEG-based braincomputer interface for cursor control. *Electronencephalography and Clinical Neurophysiology*, 78:252–259, 1991.

[ZEM94] R. Zatorre, A. Evans, and E. Meyer. Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14(4):1908 1919, April 1994.

[ZHP96] R. Zatorre, A. Halpern, and D. Perry. Hearing in the mind's ear: A pet investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, 8(1):29–46, 1996.