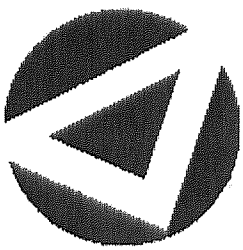
Analysis of Magnetoencephalographic Data as a Nonlinear Dynamical System

Wei Lee Woon

Doctor Of Philosophy



The University of Aston in Birmingham

March 2002

Abstract

This thesis presents the results from an investigation into the merits of analysing Magnetoencephalographic (MEG) data in the context of dynamical systems theory.

MEG is the study of both the methods for the measurement of minute magnetic flux variations at the scalp, resulting from neuro-electric activity in the neocortex, as well as the techniques required to process and extract useful information from these measurements. As a result of its unique mode of action - by directly measuring neuronal activity via the resulting magnetic field fluctuations - MEG possesses a number of useful qualities which could potentially make it a powerful addition to any brain researcher's arsenal. Unfortunately, MEG research has so far failed to fulfil its early promise, being hindered in its progress by a variety of factors.

Conventionally, the analysis of MEG has been dominated by the search for activity in certain spectral bands - the so-called *alpha, delta, beta,* etc that are commonly referred to in both academic and lay publications. Other efforts have centred upon generating optimal fits of "equivalent current dipoles" that best explain the observed field distribution. Many of these approaches carry the implicit assumption that the dynamics which result in the observed time series are linear. This is despite a variety of reasons which suggest that nonlinearity might be present in MEG recordings. By using methods that allow for nonlinear dynamics, the research described in this thesis avoids these restrictive linearity assumptions.

A crucial concept underpinning this project is the belief that MEG recordings are mere observations of the evolution of the true underlying state, which is unobservable and is assumed to reflect some abstract brain cognitive state. Further, we maintain that it is unreasonable to expect these processes to be adequately described in the traditional way: as a linear sum of a large number of frequency generators. One of the main objectives of this thesis will be to prove that much more effective and powerful analysis of MEG can be achieved if one were to assume the presence of both linear and nonlinear characteristics from the outset. Our position is that the combined action of a relatively small number of these generators, coupled with external and dynamic noise sources, is more than sufficient to account for the complexity observed in the MEG recordings.

Another problem that has plagued MEG researchers is the extremely low signal to noise ratios that are obtained. As the magnetic flux variations resulting from actual cortical processes can be extremely minute, the measuring devices used in MEG are, necessarily, extremely sensitive. The unfortunate side-effect of this is that even commonplace phenomena such as the earth's geomagnetic field can easily swamp signals of interest. This problem is commonly addressed by averaging over a large number of recordings. However, this has a number of notable drawbacks. In particular, it is difficult to synchronise high frequency activity which might be of interest, and often these signals will be cancelled out by the averaging process. Other problems that have been encountered are high costs and low portability of state-of-the-art multichannel machines. The result of this is that the use of MEG has, hitherto, been restricted to large institutions which are able to afford the high costs associated with the procurement and maintenance of these machines. In this project, we seek to address these

issues by working almost exclusively with single channel, unaveraged MEG data.

We demonstrate the applicability of a variety of methods originating from the fields of signal processing, dynamical systems, information theory and neural networks, to the analysis of MEG data. It is noteworthy that while modern signal processing tools such as independent component analysis, topographic maps and latent variable modelling have enjoyed extensive success in a variety of research areas from financial time series modelling to the analysis of sun spot activity, their use in MEG analysis has thus far been extremely limited. It is hoped that this work will help to remedy this oversight.

To my family

Acknowledgements

This thesis and the work described herein could not have been completed without the assistance and support of many friends, colleagues and my family.

First and foremost, I would like to thank my supervisor, Professor David Lowe, for his excellent supervision, stimulating ideas, and his uncanny ability to always find time whenever I made an unscheduled appearance at his office. I would also like to record my appreciation to the members of the Neural Computing Research Group, in particular my fellow PhD students Mehdi Azzouzi, Lehel Csató, David Evans, Lars Hjorth, Ragnar Lesch, Renato Vicente and Sun Yi, for providing a wonderful and friendly environment in which to work. Special thanks go to Dr. Christopher James for the many stimulating discussions and ideas throughout the course of this project. I am also grateful to Dr. Ian Nabney and Dr. David Evans for their advice and suggestions on various aspects of this thesis. Also, my sincere thanks to Vicky Bond, our group assistant, for all her help on countless occasions.

I also thank the staff of the Clinical Neurophysiology Unit for their help with the MEG recording sessions. In particular I would like to thank Dr. Ian Holliday for his advice and comments on parts of my thesis, as well as Dr. Arjan Hillebrand and Dr. Gareth Barnes for their help on numerous occasions, as well as those games of football!

A big "thank you" to the Wednesday football group for those weekly workouts which were so invaluable, in particular to alleviate the stress of the thesis writing period, as well as Professor David Saad for all those games of squash. I definitely needed the exercise!

I am also indebted to the many friends which I have made during my three years at Aston, in particular Jifan Li, my flatmate for two years and a special friend. Thank you so much for the daily support, advice and friendship. You are the best "B.M." ever!

Finally, I am eternally grateful to my family for everything. None of this would have been possible without you.

Contents

List of Figures

List of Tables

Declaration

This thesis describes work carried out between January 1999 and October 2001 in the Neural Computing Research Group at Aston University under the supervision of Professor David Lowe.

The work reported in this thesis has been entirely executed by myself. This thesis has been composed by myself and has not, either in its present form or in a derived form, been submitted in any previous application for a degree. The following publications have been produced during the course of the work described herein:

- "A dynamical systems approach to feature extraction and pattern processing in MEG". WL Woon and D Lowe. *IEE Conference on "Biomedical applications of signal processing"*, London 1999.
- "A dynamical systems approach to MEG signal processing". WL Woon and D Lowe. *4th International Conference 'Neural Networks and Expert Systems in Medicine and HealthCare'(NNESMED)*, 20-22 June, 2001.
- "Nonlinear signal processing for noise reduction of unaveraged single channel MEG". WL Woon and D Lowe. *International Conference on Artificial Neural Networks (ICANN)*. 21-25 August, 2001.

Chapter 1

Introduction

"Don't Panic.."

Douglas Adams
The Hitchhikers' Guide to the Galaxy (1979).

1.1 Background

The human brain is the most complex structure known to man and is also the most important in terms of our continued well-being. Apart from the obvious distinction of being the centre of conscious thought, the brain also acts as the central coordinator and regulator of a myriad of critical bodily functions. However, despite the existence of a voluminous amount of brain related research, our knowledge regarding the workings of the brain remains sadly inadequate. While it would be unfair to deny that many interesting and important breakthroughs have been made in recent times regarding the anatomical, physiological and (to a lesser extent) functional attributes of the brain, the fact remains that the fundamental secrets regarding how the brain stores information, processes patterns and features, generates emotions and ultimately gives rise to the sensation that we know as consciousness, still defeat the best efforts of researchers seeking to unlock them.

The synergy between modern advances in engineering and increasing scientific knowledge of the brain have led to the emergence of a variety of brain scanning techniques that have further aided the efforts to unveil the secrets of the mind. The more prominent amongst these are PET (Positron Emission Tomography), SPECT (Single Photon Emission Computed Tomography), MRI (Magnetic Resonance Imaging), EEG (Electroencephalography) and MEG (Magnetoencephalography). The key strength of these methods is their ability to provide

us with detailed information on the internal structure and functionings of the brain without the need for explorative surgery. This capability is obviously an attractive alternative to the more traditional methods of elucidating brain function, such as inference via psychological phenomena or by tests on animal subjects.

This thesis focuses on one of these techniques in particular: MEG or magnetoencephalography. In the course of this work, we will describe how the unique properties of MEG make it a powerful addition to the existing array of methods that can be used to study brain function. However, we also observe that the existing methods of processing MEG are hindered by a number of crucial shortcomings and do not exploit the full capabilities of MEG. This thesis will report on our investigations into the efficacy of a range of new analysis techniques on extracting information from MEG time series.

1.2 Main components

1.2.1 Neocortex and magnetoencephalography

It is generally accepted that conscious thought and actions originate from the outermost layer of neurons that envelope the top of the brain, known as the *neocortex* [56; 28]. This is a layer of about 10^{10} neurons which are massively interconnected to form a vast signal handling network. Magnetoencephalography is a non-invasive technique for investigating magnetic activity emanating from this region of the brain. The processing of information by the brain causes minute electric currents to flow through the neural circuits embedded in the neocortex, which in turn result in fluctuations in the magnetic field incident upon the surface of the scalp. MEG works by detecting and recording these fluctuations.

Besides MEG, there exists a variety of other methods which also permit non-invasive imaging or monitoring of the brain. Methods like CAT and MRI allow the structural details of the brain to be studied in great detail while other methods like PET and fMRI allow functional information to be gathered. MEG, however, possesses a unique advantage: on one hand, its temporal resolution, which is better than 1ms, is similar to the related technique of electroencephalography (EEG) and significantly better than functional Magnetic Resonance Imaging (fMRI), the latter being physically limited by the temporal smoothing of the hemodynamic response. On the other hand, the spatial resolution of MEG, in the region of 2-3mm for cortical sources, is superior to EEG and is similar to that of fMRI. Hence, in this sense MEG combines the best properties of EEG and fMRI.

Despite these advantages, the clinical use of MEG has been slow to catch on due to a variety

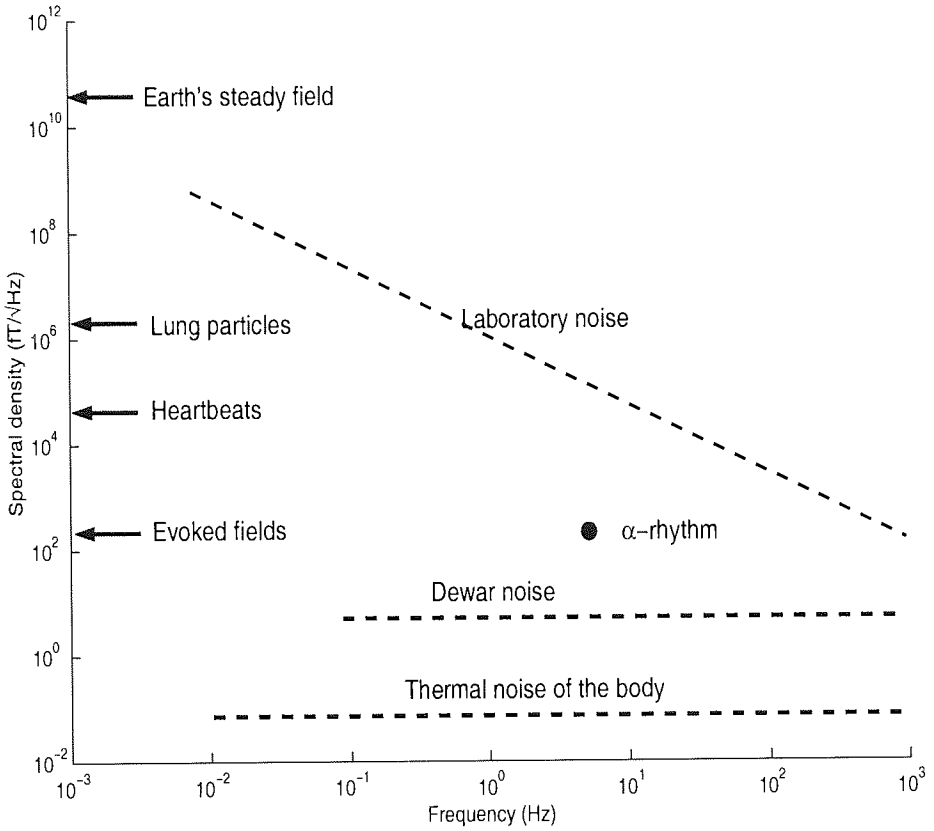


Figure 1.1: Peak amplitudes (arrows) and spectral densities of fields due to typical biomagnetic noise sources. The power of MEG evoked fields as well as magnetic fields due to α-waves has been indicated on the graph (Source:[28])

of problems. In particular, the high sensitivity required of MEG measuring equipment also makes it very susceptible to environmental noise from a variety of sources. Occurrences such as inhaled metallic particles have been known to result in artifactual signals that are orders of magnitude higher than the actual signals of interest, as depicted in figure 1.1. In addition, MEG measuring machines are both bulky and prohibitively expensive, making them unsuitable for a variety of practical applications.

1.2.2 Dynamical systems

The difficulties mentioned in the previous section have motivated a search for a new suite of tools by which MEG may be more reliably analysed. For reasons that will become clear as

we progress through this thesis, we have chosen to adopt an approach grounded in the field of dynamical systems theory.

The dynamical systems framework has been gaining popularity in a variety of fields both industrial and academic. The great strength of this paradigm lies in the flexibility in its way of describing both natural and artificial systems. A researcher working within the dynamical systems framework regards time series as a sequence of observations of the *state* of some generating system. By seeking to understand the rules which govern the evolution of the state from one time step to the other, the researcher attempts to gain deeper insight into the long term behaviour of the system [62]. Historically, the study of dynamical systems has assumed deterministic evolution laws but in practice, the evolution rule may be modified to include stochastic components or external stimuli in order to increase the similarity to the system being modelled.

In this project we view the MEG data as observations of an unobservable and possibly nonlinear dynamical system. This dynamical system, which represents abstract quantifiers of the underlying "brain state", is in turn assumed to be separable into a small number of interacting but functionally distinct subsystems, the combined action of which results in cognition and consciousness. Clearly, these state variables are not directly observable and a method must be found to extract information which will allow us to understand its characteristics.

There are many areas of research in which dynamical systems methods, if not its theories, have been useful. Some examples are:

1. Signal processing of ECG and other biological rhythms [69]
2. Analysis of computer network traffic patterns [1]
3. Ecological models such as population dynamics or the spread of epidemics [51]

1.2.3 Feature extraction

A final aspect of this project which should be mentioned at this point is the process of *feature extraction*. As the name suggests, feature extraction involves the identification and study of features or patterns of interest that may be present in data. Naturally, this broad definition encompasses a wide range of activity and may refer to a variety of automatic data processing techniques. Examples are the detection of edges or objects in image processing, localisation of flaws in structures via neural networks and various others. In the context of MEG, feature extraction is a particularly crucial process since high dimensionality and noise

levels invariably confound most attempts to manually inspect MEG data. In this project we focus on transforming the raw MEG recordings into a feature space which will allow more effective analysis of the data. In particular, emphases will be given to the ICA algorithm.

Independent component analysis, or ICA, has recently been the subject of intense interest from the academic community. The objective of ICA is to find a suitable representation of high dimensional data via the application of a rotational transform (in the case of linear ICA). ICA generalises the more commonly used method of principal component analysis (PCA) by utilising higher order moments of the data (PCA only uses second order information). In this way, ICA finds representations that minimise the *statistical dependence* between the extracted variables or *sources*.

The main strength of ICA is that no prior information is needed regarding the properties of the sources except the assumption that they are statistically independent. As we will see later, finding a suitable rotation based solely on this requirement is sufficient to provide us with a linear transform of the data that is satisfying in a number of ways, depending on the context in which it is viewed. From the signal processing viewpoint, ICA generates a more efficient representation of the data that minimises the mutual information between separated sources, hence allowing most of the information in the data to be represented by a relatively small number of components. In the context of probabilistic modelling, it can be shown that, provided the model possesses certain characteristics, performing ICA is equivalent to finding the maximum likelihood solution to the observation model.

1.3 Motivations and novelty

The fundamental tenet in our work is the assumption that the MEG recordings are only observations of underlying system dynamics. Crucially, we claim that the complexity observed in MEG recordings is the result of the combination of a small number of unobservable system generators. This significant shift from the philosophy adopted in the majority of existing MEG studies is the main motivation behind this project. We strongly believe that the most effective way of analysing MEG data is by moving the focus of the research away from the recorded time series, onto the underlying phase space. Examining and determining the properties of this unobservable system attractor or manifold is likely to produce a range of interesting results not previously observed.

As discussed earlier, MEG possesses a number of advantages over existing methods of brain imaging. In particular, by having spatial resolution comparable to MRI, and temporal resolution comparable to EEG, MEG occupies a niche whereby it is suitable for studying both

functional as well as anatomical aspects of brain function. Furthermore, the orthogonality of the electric and magnetic fields means that MEG is able to monitor signals from regions in the neocortex that are normally hidden from EEG.

Nevertheless, despite these advantages, progress in MEG research has been relatively slow. One of our main contentions is that the main reason for this lack of progress is the widespread adherence to unrealistic assumptions concerning the linearity of the MEG time series, as evidenced by the extensive use of power spectrum analysis and PCA. This situation is all the more untenable in view of the existence of a wide range of tools that allow for the effective analysis of time series with nonlinear characteristics. Furthermore, a number of research areas such as the analysis of financial time series and sunspot activity have already benefitted significantly from the utilisation of nonlinear tools. In the context of MEG research, our motivations for investigating the use of nonlinear methods are twofold: firstly, it is clear that as far as exploitation of the MEG data is concerned, linear signal analysis methods have fallen short. The tried and tested approach of searching for various frequency band activity such as the ubiquitous Alpha and Beta waves has hitherto only produced limited and superficial knowledge regarding the functionings of the brain, while leaving a large variety of signals and structures in the data unaccounted for. Secondly, prior knowledge regarding the functionings of the brain, both on the cellular level as well as on the scale of human behaviour, indicates the presence of nonlinearity.

The combination of these two main factors: the potential advantages of MEG as a clinical tool, and the relatively limited use of nonlinear analytical tools in the analysis of MEG presents a very strong case for a course of research to bridge the gap between the current research activities in MEG, and the availability of modern, nonlinear signal processing methods. We hope that the work described in this thesis will help to stimulate further interest in this direction.

A further problem that has stood in the way of widespread adoption of MEG technology is the complexity of the MEG measuring equipment. The flux variations generated by neuronal currents are extremely minute and can only be measured using very specialised superconducting devices known as "SQUIDS" (Superconducting Quantum Interference Devices). The high costs associated with the procurement and maintenance of these devices means only a relatively small number of research institutions can afford to operate the more powerful and useful multichannel MEG systems. Additionally, multichannel systems tend to be extremely bulky and hence ill-suited to most industrial and practical applications. To address this issue, our analysis will incorporate another novel aspect: we will seek to restrict ourselves to working exclusively with single channel data, as this will help to reduce the dependence on huge and costly multichannel devices.

1.4 Contributions of thesis

As described in the previous section, the overall aim of this project is to study and subsequently present evidence of the applicability of a range of advanced signal processing methods in the analysis of MEG data. These techniques will represent a combination of existing methods drawn from the fields of information theory, dynamical systems and neural networks, and have been selected in response to a variety of problems that are known to plague existing approaches to analysing MEG data.

In support of this aim, this thesis will focus on a number of issues. These include:

1. **Model the MEG generating processes as an unobserved dynamical system** - This is related to the latent variable modelling approach which explicitly treats experimental data as observations of unseen, or *latent* variables. The latent variable modelling approach is extremely general and covers a wide range of techniques and by carefully selecting the parameters which describe the latent variables, it is possible to independently derive a range of signal processing tools such as Kalman filtering and PCA. Models from this class have been applied to a range of problems in physics and engineering but are conspicuously under-utilised in MEG signal analysis. In addition, our adherence to this standpoint means that there is a need to explore methods of reconstructing the underlying system manifold given only the available data.

2. **Use of methods that allow for nonlinear dynamics** - As described in the previous section, existing approaches to MEG analysis focus on linear properties of the data. This implies that MEG recordings are essentially the sum of a number of frequency generators, a position which can potentially exclude possible explanations for observed complexity. Hence, this project will seek, where possible, to use methods that allow for nonlinear dynamics. In addition, to justify the use of these more complex methods, a systematic series of tests will be performed to ascertain the presence of detectable levels of nonlinearity in the MEG time series.

3. **Utilisation of single channel unaveraged MEG data** - Also touched on in the previous section, the use of multichannel MEG systems brings a host of problems. Cost and size of the machines is just one aspect. The computational overhead and sheer difficulty of dealing with huge multivariate data sets (often with more than a hundred dimensions) means that the development of methods that can be used on single channel data could potentially bring a number of very significant benefits and advantages.

4. **Application of topographic projection methods with the aim of visualisation and clustering** - As is always the case with high dimensional data sets, one of the

greatest problems is the need for some means of visualising relationships within the data, and for searching for structures such as clusters in high dimensional spaces. In the course of this project, we will test the usefulness of topographic projection tools in the visualisation and clustering of the MEG data.

5. **Testing of methods for characterising and extracting qualitative properties from the data** - A difficult problem when dealing with neurophysiological data is one of interpretation. It is widely assumed that the MEG observables reflect underlying mental processes, though at a very coarse resolution. If this is indeed the case, how can we understand the functional significance of any signals or features that we are able to extract from the MEG recordings? To address this concern, this thesis will also seek to present and test methods for extracting qualitative properties from MEG signals.

1.5 Plan of the thesis

This thesis will be organised in the following manner:

Chapter 1 is this introduction.

Chapter 2 gives an introduction to the field of Magnetoencephalography and provides a survey of MEG-related areas of study that are relevant to our project. Where applicable, the aims and objectives laid out in this chapter will be linked with related aspects of MEG research as these are encountered. This chapter also presents a qualitative description of a model of neocortical dynamics upon which subsequent analysis will be based. Finally, this chapter examines the data that will be used in this project, with particular emphasis towards the procedures and experimental conditions observed during the data collection process.

Chapter 3 is a detailed study of the dynamical systems context in which the project is grounded. A short introduction is given to the main concepts of dynamical systems theory and aspects that are relevant to this project are highlighted and explained in greater detail. In particular the subject of delay embedding is introduced as a way of reconstructing and hence recovering the system attractor on which the neocortical dynamics are assumed to reside. The final part of this chapter is devoted to a detailed study on the issue of nonlinear dynamics and in particular the methods used to test for nonlinear dynamics in time series data. The results of the application of said methods on MEG are also presented.

Chapter 4 is the first of two chapters devoted to the actual application of the methods proposed to the analysis of MEG. A general overview of the algorithms used, in particular ICA is given. The results of the application of ICA on the data are also presented and discussed

in some detail. A method for the clustering and visualisation of components extracted using ICA is proposed. Again, examples of results obtained from the application of these tools are described.

Chapter 5 investigates two approaches which were used for characterising the results of the previous chapter, and in particular to link them to cognitive function. As there is little prior knowledge about the kinds of components which might be extracted, we first propose a technique which would allow the spatial distribution of the components over the scalp surface to be visualised. Finally, a method for characterising the complexity of MEG time series is applied and evaluated.

Chapter 6 concludes this thesis by presenting a unifying summary of the results obtained throughout the course of this project. The aims and objective laid out in the beginning of the thesis are also briefly reviewed and these are linked to the results which are presented. Finally, directions for further work are suggested and reviewed.

Chapter 2

Neocortex and Magnetoencephalography

"The whole is greater than the sum of its parts"

Ida P. Rolf

Chapter 1 has already provided some insight to MEG and some of the issues and problems in its analysis. As MEG is a relative newcomer in the field of non-invasive brain monitoring, it is characterised on one hand by high expectations and potential for further development, and on the other by the many difficult technical problems that need to be resolved before it can be fully exploited. This chapter provides a more thorough introduction to this field of research as well as a review of traditional signal processing techniques that have been used on MEG. However, a comprehensive review of the theory and many issues surrounding MEG research is beyond the scope of this thesis and the interested reader is referred to the many review articles in the literature, examples being [28; 68] and the references contained therein.

Section 2.1 introduces the physiological properties of the brain. The main brain signals that are measured in EEG and MEG are also listed and explained. Section 2.2 reviews a variety of existing brain scanning techniques and highlights their relative strengths and weakness compared to MEG. Because of their importance and greater relevance to this project, MEG and EEG are described in greater detail in section 2.3. This section also reviews some of the existing MEG analysis techniques. Section 2.4 discusses the MEG generation process, and presents a qualitative model of cortical dynamics. Finally, the practical issues involved in the collection of MEG are discussed in section 2.5. This section also discusses the experiments that were performed and the recording procedures that were observed during the recording sessions.

2.1 Physiology of the brain

The brain is an enormously complex structure composed of a number of functionally distinct organs. In the most commonly adopted scheme, the brain is divided into three main sections: the hindbrain, midbrain and the forebrain; these are situated progressively higher starting from just above the spinal cord. Generally, the more automatic and reflexive an action is, the likelier it is to involve the hindbrain or midbrain. Conversely, the more complex a behaviour, the more likely it is to have originated from the forebrain [78].

The main structures in the hindbrain and midbrain are the brain stem, the reticular activating system, the cerebellum and thalamus. Between them, these components regulate a variety of subconscious bodily functions such as heartbeat, breathing, autonomic responses like shivering and the dilation of blood vessels in response to certain stimuli as well as certain sensations such as hunger and thirst. While the importance of all of these functions to our wellbeing are obvious, in this thesis we will not be concerned with details about either of these sections of the brain as neither generate signals that are significantly measurable with MEG [68].

The forebrain itself consists of two main sections. Immediately above the midbrain is a set of loosely interconnected structures that form a border on the underside of the brain's "cauliflower". Together these structures make up the *limbic system* of the brain. This region is thought to perform two main functions; firstly, it is heavily involved in emotions such as rage and fear, which we share with other animals. An important part of the limbic system, the *hippocampus*, is thought to be involved in a kind of preprocessing of information that is sent from the lower regions of the brain - its main function is to locate stimuli or information that is "unusual" and to relay this to the higher information processing centres.

Situated just above the limbic system and by far the most important structure in the context of MEG is a layer consisting of densely packed cells known collectively as the cerebral cortex. Cell bodies in the cortex are crowded together and produce a grayish tissue; hence the name *gray matter* (in other parts of the brain and the nervous system, long myelinated axons prevail, producing the *white matter*). It has been estimated that the cortex contains almost three fourths of all the cells in the brain, a fact made possible by two main adaptations:

1. The extremely high density of cells in the cortex (each square inch contains an estimated 10000 synaptically connected nerve cells).

2. The surface of the cortex is covered with many deep crevasses and wrinkles, greatly increasing the effective surface area - the total surface area of the cortex is estimated at $2500cm^2$.

The brain itself is divided into two hemispheres by the longitudinal fissure, and each of these halves is further divided by a series of deep crevasses into four distinct pairs of lobes (see figure 2.1[78]). Each of these have been loosely associated with certain cognitive functions:

1. **Occipital lobes** - contains the *visual cortex.*
2. **Parietal lobes** - The main functional component here is the *somatosensory cortex*, which receives information about pressure, pain, touch and temperature from all over the body.
3. **Temporal lobes** - Primarily associated with the *auditory cortex.* Also involved in memory, perception, emotion and language comprehension.
4. **Frontal lobes** - Contain the *motor cortex*, which controls voluntary movement. The frontal lobe also seems to contain faculties that seem to be responsible for the ability to make plans, think creatively and take initiative.

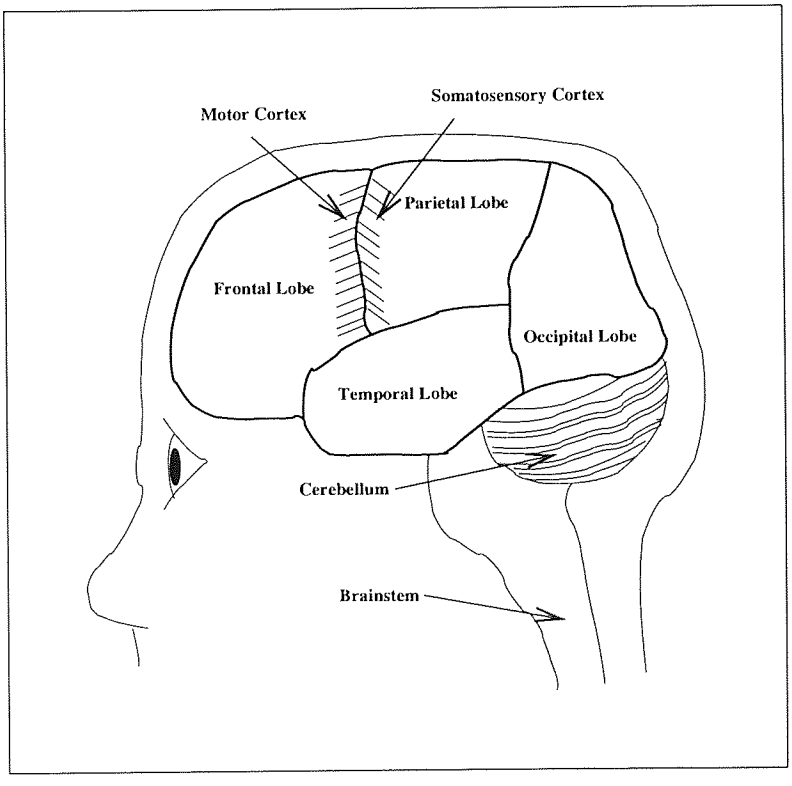


Figure 2.1: Human brain seen from the left side. Main divisions of the cerebral cortex have been indicated

2.2 Noninvasive brain imaging

MEG is a member of a family of *brain imaging* techniques. These are technologies with the primary aim of gathering information about structures and functions of the brain without the need for invasive procedures. We now review some of the more popular imaging techniques and discuss their comparative advantages and disadvantages relative to MEG. The motive is to highlight areas of research where MEG can be of particular use. It is important to note that the purpose of this section is not to prove that one technique is necessarily superior to others. Rather, it must be emphasised that each technique has its own region of applicability, as determined by a range of factors which will be discussed.

Computed tomography (CAT) scan

A CAT scan works via the projection of x-rays through the head. The unique feature of CAT is the rotation of the x-ray projector through an angle of 180^o around the head, thus providing an all-round view of the brain's structure. The multiple x-ray images generated in this fashion are fused using computer algorithms to generate a three dimensional anatomical image of the brain. Though CAT scans provide good anatomical information, they reveal nothing about cognitive function. Also, the use of x-rays makes CAT scanning inappropriate for use in pregnant women and children. In terms of cost of equipment as well as procedural costs, CAT is more economic than MEG.

Positron emission tomography (PET)

During a PET scan, radioactive isotopes of biologically active molecules are injected into the bloodstream, which are then carried into the brain. The subsequent radioactive decay of these molecules result in the emission of gamma radiation which is detected by the PET scanner. Depending on the type of tracer used, various physiological and biochemical parameters can be measured. The versatility of PET is a direct result of the fact that almost all biological molecules can be labelled with positron emitters such as carbon-11, nitrogen-13 and oxygen-15. The spatial resolution of PET reaches a theoretical limit of about 5mm, which is comparable to MEG and superior to EEG. However, the temporal resolution of PET is in the range of minutes to hours, whereas EEG and MEG are sensitive to changes that occur in milliseconds. In addition, PET is even more expensive than MEG as an on-site cyclotron (or other accelerator) is required to generate the required isotopes. Furthermore, the cost for each procedure is also high as a new batch of radiopharmaceuticals need to be made for each experiment.

Single photon emission computed tomography (SPECT)

In common with PET, SPECT also uses radioactive tracers and a scanner to record data that a computer uses to construct two- or three-dimensional images of active brain regions. However, the radiopharmaceuticals used in SPECT are more long-lived and may be obtained from an external source, greatly reducing the operating cost of SPECT, though at the expense of inferior spatial resolution. The cost of SPECT is low compared to MEG.

Magnetic resonance imaging(MRI)

MRI is another method capable of providing detailed anatomical information on the brain. MRI works by placing the subject in a static magnetic field, then further applying an RF alternating magnetic field. The energy from the alternating field is absorbed by nuclei and is re-emitted when the RF field is turned off, resulting in the emission of radio frequency waves. In this way it is possible to gain detailed anatomical information. A more recent development in MRI allows the recovery of functional information, and is known accordingly as *functional* MRI or fMRI. As with MRI, fMRI is capable of excellent spatial resolution but suffers from poor temporal resolution as it depends on analysing changes in blood circulation.

Table 2.1 provides a summary of the characteristics as well as relative strengths of the various brain imaging techniques.

2.3 EEG and MEG

2.3.1 Introduction

The functioning human brain is a hive of electromagnetic activity. Ionic currents associated with a functioning brain produce measurable electric and magnetic fields. The electric potential disturbances at the scalp originating from these currents were first measured using vacuum tube amplifiers in 1929 [77]. On the other hand, the more challenging goal of measuring the corresponding magnetic flux densities was only achieved much later in 1968 [77]. In view of this disparity, one can easily understand why research in MEG lags significantly behind EEG.

Electroencephalography or EEG is the measurement of electric potential at the scalp arising from neurons in the cerebral cortex. In this sense, EEG is most similar to MEG in terms of functionality and mode of operation, as both these techniques share the common underlying cause of synchronised mesoscopic neuronal activity. The way that the data is collected means that intrinsically, both EEG and MEG work in the temporal domain. The measurement of

Method	Res.	Advantages	Disadvantages
CT	5-10mm	Low Cost and good spatial resolution. Good for elucidating brain structures	Invasive (exposure to x-rays). No functional information
SPECT	10 mm	Low cost and easily available. Provides similar capabilities to PET at fraction of cost	Invasive. Inferior resolution (compared to PET)
PET	5 mm	Good spatial resolution. Suitable for metabolic studies and receptor mapping studies	Invasive. Very expensive
EEG	<1cm	Very low cost. Suitable for sleep and operation monitoring. Non-invasive. Good temporal resolution	Not an imaging technique. Poor spatial resolution
MEG	5 mm	Good temporal and spatial resolution. Selectively sensitive to different parts of neocortex than EEG. Non-invasive	Very Expensive, Limited resolution for deep structures
fMRI	3 mm	Excellent spatial resolution. Non-invasive	Expensive. Intense magnetic fields can be a safety issue

Table 2.1: Summary of brain imaging techniques

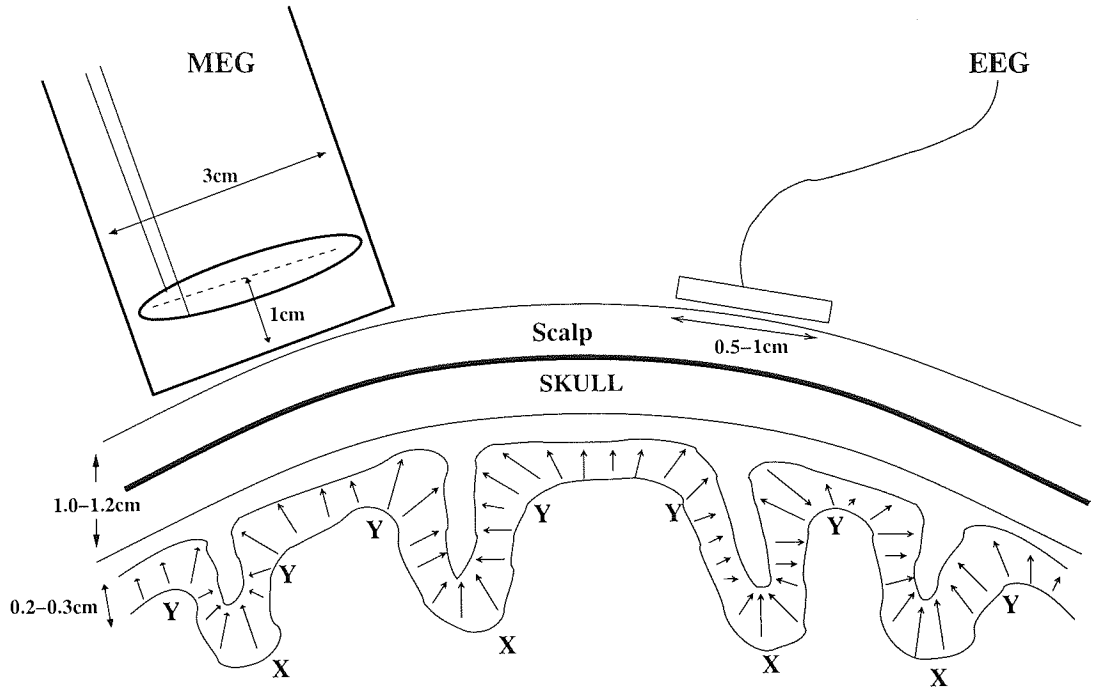


Figure 2.2: Extracranial recording of magnetoencephalographic (MEG) and electroencephalographic (EEG) data. Due to the orientation of the generated fields, MEG is most sensitive to correlated dipole sources in the sulcus (**X-X**) whereas EEG is more sensitive to sources located in the gyri (**Y-Y**).

magnetic flux (as opposed to electric potential) however, affects MEG in two significant ways. The first issue is related to the way in which the neurophysiologically driven EM fields emerge at the surface of the scalp. Figure 2.2 depicts the measurement process in MEG and EEG, giving the typical dimensions of the relevant structures. The arrows in the figure represent the current dipoles that may be active within the cortex. It can be seen that the dipoles are embedded vertically within the cortical matrix. In fact, it is this very fortunate alignment of the neurons that allow MEG and EEG to pick up coherent signals from the brain. If, as is the case with deeper brain structures, the neural columns were oriented in a variety of directions, the EM fields resulting from cortical currents would cancel out and become undetectable by extracranial methods. Furthermore, because existing scanners require a certain amount of neurons to fire automatically (observed magnetic fields indicate that at least about 10^5 neurons are required), it is likely that without the highly ordered arrangement of neurons in the cortex, MEG or EEG would simply not be possible at all. Nevertheless, the cortex is, of necessity, extremely wrinkled, meaning that there is still variation in the orientations of the neurons. Resulting electromagnetic fields may be oriented radially or tangentially with respect to the scalp. As the magnetic and electric fields are at right angles to each

other, MEG can in theory monitor signals from different regions of the neocortex that EEG monitors. MEG is most sensitive to the correlated dipole layer in the sulcus (labelled as **X** in figure 2.2) as the external magnetic field caused by a radial dipole in a medium of spherically symmetric conductive layers is zero. Although this symmetry condition is not fully satisfied in actual heads, it is close enough so that measured magnetic fields caused by radial dipoles are typically much smaller than fields caused by tangential dipoles of the same strength and depth. Conversely, EEG is most sensitive to the correlated dipole layers in the gyri (labelled as **Y** in figure 2.2) as tangential dipoles tend to cancel out.

The second issue that differentiates MEG from EEG is spatial accuracy. Currents that flow in the head can either be primary currents, which are the actual neuronal dipoles that flow along the nerve fibres, or volume currents, which carry the return current and which complete the circuits. These volume currents flow via the medium in which the cells are embedded. Electrical potentials on the scalp, and hence EEG, are known to depend on both of these current components and hence, are more adversely affected by volume conductor properties of the skull and scalp, making accurate determination of the activated area difficult. The magnetic field, in contrast, mainly depends on the primary currents, which only flow inside, or at least in close vicinity to the neurons, which are relatively homogeneous (the volume currents do generate contributions to the magnetic field as well but these are insignificant compared to the primary contributions due to the poor conductivity of the volume conductors). Because of these factors, MEG is capable of spatial accuracy of as high as 3 millimetres under conducive recording conditions. This compares favourably to EEG, which typically only achieves accuracies of around 1cm.

It should however be noted that both of these issues have come under some debate in recent times, with doubts being cast on both the ability of MEG to provide information beyond that already extracted by EEG, as well as the superior spatial accuracy of MEG. A detailed investigation is beyond the scope of this thesis but the interested reader is referred to the variety of comparative MEG and EEG studies in the literature, examples being [55; 50; 18].

2.3.2 Overview of MEG research

Oscillatory brain dynamics

Neuronal activity recorded by MEG and EEG has traditionally been referred to in terms of the frequency bands in which they occur, a practice which can often be justified by the fact that there are known correlations between cognitive events and increased power in certain frequency ranges. The simple fact that these oscillations can be recorded using extracranial sensors documents the important fact that large populations of neurons within the cortex have

to be oscillating at the same frequency and phase. Hence, the signal power of oscillations should be taken as more of a measure of the synchrony of the associated neurons rather than an actual indicator of the very presence of these oscillations.

In recent times, a variety of theories have emerged regarding the origin and purpose of the various oscillatory neuro-electric components. What seems clear is that they play a critical role in the proper functioning of the brain, both in sensory and cognitive aspects. One likely role for these oscillatory components is as resonant communication networks helping to synchronise large populations of neurons. The most commonly observed modes are briefly described below in order of increasing frequency:

- **Delta waves** have frequencies in the range of 0.5 to 3 Hz. Phase locked, steady state delta rhythms tend to have large amplitudes and are prominent in young children, but are rarer in wake, healthy adults. Induced delta activity has been implicated in a range of cognitive activities such as "oddball" experiments and in detection of near-threshold hearing stimuli experiments.

- **Theta waves** occur in the 4 to 8 Hz frequency range. As in the case with delta waves, theta activity has been implicated in a range of cognitive functions such as memory and attention. In "idle" brain states, theta activity is believed to be desynchronised and therefore difficult to detect. Conversely, heightened mental activity results in an increase in theta activity, indicative of greater theta synchronisation. In particular, theta activity is related to both the retrieval and encoding of information from episodic memory. Theta waves are mainly associated with the frontal and central regions of the cortex, though no one region of the cortex is firmly associated with the generation of theta activity (in fact, some theories suggest that memory related theta activity originates from the hippocampus).

- **Alpha waves** refer to brain activity in the 7 to 13 Hz range. It is normally best seen in the posterior of the brain but is generally prominent in most parts. It is the dominant brain wave in adults and is manifested by a peak in the frequency domain. Alpha activity is associated with a relaxed, alert state and is enhanced when the eyes are shut and the subject is sitting comfortably and relaxed. However, mental effort or attention results in alpha desynchronisation, causing the alpha rhythms to be attenuated or "blocked". In this sense, alpha and theta waves behave in opposite ways. Where increasing task demands causes increased synchronisation of theta, the opposite occurs in alpha. Even the spatial distribution of alpha and theta waves are opposite: alpha waves are primarily observed over the occipital cortex, whereas theta waves are mainly associated with the frontal and central regions.

- **Beta waves** are higher in frequency than Alpha waves (they have a frequency of 13 Hz and greater) and are often associated with anxious or alert adults with open eyes. They also often occur when the subject is concentrating hard, in problem solving, for instance. It is usually seen on both sides in a symmetrical distribution but is most evident in the frontal regions. In areas where there is cortical damage, it may be reduced or absent altogether.

- **Gamma waves** have received a lot of attention in recent times. The term generally refers to high frequency activity in the range of 20-100 Hz, though the most prominent examples occur in the region of 40-60Hz. While the exact nature of their function is still a subject of some debate, a growing body of research now suggests that they are somehow related to the process of linking disparate features from distinct neural assemblies to form a coherent percept[6], a process dubbed "perceptual binding" or simply the "binding problem" [74]. However, gamma activity also has other cognitive correlates and is known to directly follow sensory detections of stimuli.

Equivalent current dipoles

There are a number of ways in which the magnetic flux distribution in multichannel MEG recordings can be interpreted. One of the most common approaches is to assume that the cortical current is well represented by a single or a small number of *current dipoles*. A given sensory stimulus is assumed to initially activate a relatively small area of the cortex, resulting in a focused, local current moment. This process is associated with a primary current source related to the movement of ions due to their chemical concentration gradients, which in turn are a direct result of the depolarisation of the neuron cell membranes.

One popular technique is to fit a single equivalent current dipole (ECD) to the observed magnetic field map. This approach requires a forward model for magnetic field generation, which commonly assumes a spherical head with homogeneous volume conductivities. This model will take as an input the location, amplitude and orientation of a current dipole occurring in the neocortex and generate a predicted field mapping over the surface of the scalp. Of course, in practice, the problem has to be solved in reverse, *viz* the scalp flux distribution is known, but not the parameters of the dipole. This inverse problem can be solved by optimising the dipole parameters with respect to some measure of similarity between the predicted and observed magnetic flux patterns. Both EEG and MEG suffer from non-uniqueness as there are many different current distributions within the head which could produce the measured field. However by adding extra information (such as restricting the maximum magnitude of current, or restricting the spatial extent of the current sources) it is possible to reduce the problem to a soluble one. This rather approximate approach has still

proven successful in many applications for several reasons. Firstly, recall that the majority of MEG experiments are performed using averaging of many responses to a repetitive stimulus. This has the effect of cancelling out most of the fine or high frequency structure of the signal. Secondly, the millisecond time resolution means that if the activation of different areas are separated by more than a millisecond, they can be resolved as singly active areas. However, the very aspects that make the ECD approach practical, also lead to certain disadvantages:

1. Firstly, for a number of reasons it is only possible to work with a very limited number of dipoles (the method is only practical up to about three dipoles). This limits the effectiveness of the method because it is uncertain how well this approximates the actual underlying sources.

2. Secondly, an ECD only takes into account the MEG distribution at a single time step. It is likely that any current sources within the cortex might have varying spatial distributions which the ECD fit will be unable to properly account for.

3. It is becoming increasingly likely that most cognitive operations are not the result of focussed sources, but rather of distributed cell assemblies with complex and nonstationary spatial patterns. It is uncertain how well the ECD model is able to adequately accommodate such distributed processing models.

Event related brain signals and signal averaging

An important subject in the analysis of neuro-electric signals has been the study of event related signals. The registration of a stimulus by the brain can often be measured by EEG/MEG. Here we adopt the scheme described in [58] by distinguishing between two types of observed changes: event related potentials (or ERPs - the MEG versions are called ERFs), and event related synchronisation/desynchronisation (ERS/ERD). The main discriminant between these two forms of activity is that ERPs are tightly phase-locked to the stimuli, whereas ERDs and ERSs tend to have varying latencies with respect to the signal of interest.

Of these, ERPs have received the most attention. ERP research generally involves the study of a number post-stimulus peaks or troughs in the recorded data. Interest in this area stems partly from the various neuro-pathological correlates of these waveforms. Another application is to fit equivalent current dipoles to ERP spatial patterns, which enables the researcher to localise cognitive or sensory functions. Examples of ERP waveforms include the N100 component (so called because it manifests itself as a negative signal peaking approximately 100ms after the onset of the stimulus). Another waveform that has attracted a lot of interest is the P300 wave. Unlike shorter-latency ERPs like the N100, this signal is only manifested with task-related stimuli and hence is widely believed to be the earliest indication in EEG that a stimulus has been cognitively "registered" [56]. There are also examples which are

known to precede events, for example the so-called "contingent negative variation" or CNV, a cognitive process related to anticipation and motor preparation.

The study of ERPS and ERFs has been the subject of a lot of useful research and has certainly played a prominent role in the study of brain function. However, research in this field is plagued by a few problems:

1. These signals are extremely weak compared to the background MEG and are difficult to analyse in raw data. As such, signal averaging is widely used in the study of ERFs. This involves repeating an experiment a large number of times and averaging over all the trials. Typically, to aid signal strength, the epochs are first inspected manually and epochs with artifacts inside are discarded.

2. In light of recent findings [43; 5], it is becoming increasingly clear that ERFs are in fact composite signals - the superimposition of multiple components. This fact casts doubts on the practice of estimating equivalent current dipoles of ERPs since a composite signal is unlikely to originate from a single locality in the cortex.

3. Finally, the dependence on signal averaging carries certain disadvantages. In the first place, it is uncertain whether averaged data are equally representative of responses on individual trials. Indeed, it seems reasonable to hypothesize that analysis of individual trials will provide information that is otherwise obscured by the grand-average, and that the very variability of responses might carry valuable information regarding the issue that is being studied.

4. As mentioned, certain signals, in particular ERD and ERS associated waveforms are not phase-locked to the stimulus. As an example, *induced gamma* waves [9], are known to follow stimuli at varying latencies and would be cancelled out by signal averaging. Also, experiments which lack a fixed stimulus preclude the use of signal averaging.

2.4 Neuronal generation of MEG

2.4.1 The linear observation paradigm

MEG works by detecting fluctuations in the magnetic field strength at the scalp, which result from neuro-electric signals flowing in the cortex. Of course, it is not possible to monitor the activity of single neurons, or even small groups of neurons, as the currents involved would be vanishingly small. Rather, what is actually measured is the space averaged signals of a very large number of neurons firing simultaneously. A direct result of the linearity of Maxwell's

equations is that the magnetic field strength is related linearly to the strength of the current elements which set up the field. Of course, the sensitivity of the sensor to spatially distributed current elements will not be uniform but rather will depend on factors such as orientation of the current dipole as well as distance from the sensor. This relationship can be written as [28]:

$$b_i = \int \mathcal{L}_i(v).\mathbf{J}^P(v)dv \tag{2.1}$$

Where b_i is the net magnetic field at point i, $\mathbf{J}^P$ is the primary current density and $\mathcal{L}_i$ is a weighting coefficient known as the *lead field*, which represents the sensitivity of a sensor at position i to current flowing in volume element v. As the cortex is assumed to comprise a single layer of dipoles, this equation basically states that the resultant magnetic flux density measured at a given point is a weighted summation over all the current elements in the cortex.

2.4.2 Neocortical dynamics: A qualitative model

The most general starting point is that the brain is an information processing device. As a gross summary, we can regard the brain as a system in which the incidence of certain inputs (sensory information) gives rise to a response (motor activity) based on certain propagation rules. However, this "black box" view is a drastic oversimplification in that it implies a feedforward system with the direct progression of signals from input to output. A more realistic overview is that, while fundamentally still a signal processing device, the brain also contains a very large number of internal feedback and lateral connections. The result is that the number of paths which a signal can follow, and consequently the duration of time which it might have to take before affecting some form of response, can vary hugely. Indeed, it is often true that certain inputs may elicit no discernible outputs at all.

In this sense, it is useful to think of the brain not as a simple filter as such, but more as a kind of interactive dynamical system. Indeed, this viewpoint is more consistent with experimental results which indicate the presence of nonstationary and highly complex dynamical processes driving the electro-magnetic fluctuations which are measured. On a related issue, one of the strongest differentiating characteristics of the human brain, compared to other mammalian brains is the extremely high ratio of cortico-cortical and intracortical fibres to input connections from the midbrain and other regions. This points to the critical role of neocortical dynamics and feedback connections in the generation of cognitive and higher analytical processes.

In this section, we describe a model for neuronal generation of MEG signals at the scalp. Rather than present a comprehensive treatment of this issue, the objective is to present a short summary of existing theories of neocortical dynamics insofar as these are relevant to

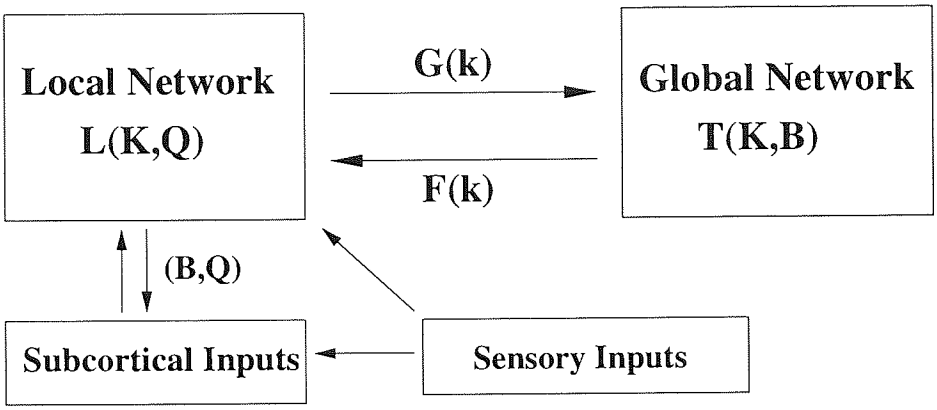


Figure 2.3: Figure depicting inter-relationships between the components of the model for MEG generation. $L(k, Q)$ and $T(k, B)$ are the local and global transfer functions respectively while $G(k)$ and $F(k)$ represent the activation levels that are passed between the components. The addition of external inputs, both from subcortical sources, as well as directly from sensory modules completes the model

our approach to MEG signal processing. Also, a full quantitative analysis of the processes of interest is beyond the scope of this discussion. Rather, emphasis is on gaining a functional, qualitative understanding of the neocortical dynamics which contribute to scalp MEG recordings.

Figure 2.3 outlines the basic model which we will use. Note the division of neocortical dynamics into local and global contributions. Local contributions refer to properties or behaviour of neurons or groups of neurons that may be inferred from their individual characteristics. These include subcortical structures like the thalamus and a variety of proposed "pacemakers". The outputs of these subcortical modules are fed into and directly affect the next subcategory of cortical dynamics - the local cortical components. These are based on intracortical interactions between neighbouring neurons or groups of neurons, and generally span spatial ranges of a few mm up to the cm range. More formally, local cortical components refer to circuits in which the dominating source of time delay is the rise and fall times of post synaptic potentials (PSP).

Global contributions are cortical interactions that are based on the longer ranged cortico-cortical fibres. These differ from intracortical connections in that they generally leave the cortical matrix en route to their destinations. The influence of boundary conditions on the overall dynamics of the cortex is also classified under global effects.

Of course, the reality is that there is no clear-cut distinction between local and global com-

ponents. Indeed many circuits may well cross several spatial scales. It is widely held that the combination of global boundary conditions resulting from the finite size of the neocortex, and the dense interconnecting network of cortico-cortical fibres results in the presence of standing or travelling waves in certain frequency bands. These are controlled by a variety of neuromodulating factors such as local conductivities, fibre propagation speeds and neurotransmitter densities, which may be controlled via external controls from the subcortical bodies. In general, it is these globally driven effects that are measured by MEG and EEG. In particular, the traditional use of global coherence indicators is a strong measure of the activity of global neocortical circuits.

The emerging picture is of a variety of distributed circuits, possibly based on neural networks driven by local or regional connections, embedded in the background of macroscopic oscillatory activity. Exactly how these local circuits operate is still unclear though these are likely to consist of a variety of cooperative assemblies each handling specific aspects of various cognitive tasks. It is proposed that the role of globally coherent oscillations, apart from maintaining general stability, is to synchronise and bind the activities of distributed processing elements into a coherent processing unit.

This leads us to the issue of MEG generation. To summarise, it follows from the preceding discussion that the dynamics that we are seeking to extract consist of a mixture of semi-independent, spatially distributed neural networks. These are likely to evolve according to local dynamical rules, yet operate cooperatively to affect and indeed give rise to overall cognitive function. It is also likely that any interactions, both within individual circuits as well as between spatially distinct circuits are going to be nonlinear and nonstationary as the control parameters of the cortical medium shift in response to external requirements.

In view of the complexity of the problem, it is unlikely that existing methods, which consist mainly of the study of coherence in different frequency bands, or of the isolation of ECDs, will be able to yield information regarding the true dynamics underlying cognitive processing.

2.5 MEG data recording

2.5.1 Recording equipment

The MEG data used in this project was recorded using the CTF Omega-151 system installed at the Clinical Neurophysiological Unit at Aston University. A total of 150 MEG channels were recorded at a sampling frequency of 1250 Hz. The MEG channels were recorded from 150 separate measurement sensors uniformly distributed over the head in a closed packed

hexagonal grid. All of the measurements were taken in a magnetically shielded room to minimise contamination from environmental magnetic sources while any required visual stimuli were conveyed into the room via a system of mirrors. While the CTF system uses an adjustable dewar and gantry which permits the subject to be seated in a range of positions from fully upright to supine, all measurements used in this thesis were taken with the subject in a slightly reclined upright position. Sampling was performed at 1.25kHz without the use of any hardware filtering. However, the channels were subsequently subsampled by a factor of five to an effective sampling rate of 250Hz. To avoid aliasing artifacts, a low pass filter with a cutoff frequency of 100Hz was applied prior to the downsampling operation. Because it is unlikely that signals of biological origin will have a frequency of higher than 100Hz, this was deemed to be an acceptable preprocessing step (in fact, many existing studies into MEG and EEG use low pass filters with much lower cutoff frequencies. e.g.: 40Hz in [54] and 50Hz in [61]).

A sample of the raw MEG recordings is shown in figure 2.4. The four channels shown in this figure correspond to four channels specifically selected from regions in the frontal, occipital and two auditory cortices and were recorded over a period of sixty seconds. These four channels are labeled 'RT','LT','O' and 'F' for right and left temporal, occipital and frontal respectively. During this period, the subject was required to sit idly and to blink his eyes every five seconds.

2.5.2 Experimental setup

Given limited time and computing resources, it is important that the experiments to be performed should be selected carefully in order to derive maximum benefit. One consideration was that this project does not aim to study a particular cognitive or physiological process, but rather the general effectiveness of an approach to signal processing in extracting features from the recordings. As such, the experiments were designed to be as unspecific as possible, and to cover a range of possible cognitive functions. In addition, because of the high noise and nonstationary behaviour of MEG recordings, it is important that the selected tasks fulfilled two further criteria:

1. They should not be too complex so that the variation in the responses would not be too large as to hinder the discovery of any patterns or regularities.

2. The tasks should be relatively short, which would enable multiple trials to be recorded to help counter the problems posed by noise and artifact corruption, as well as the natural variability of MEG.

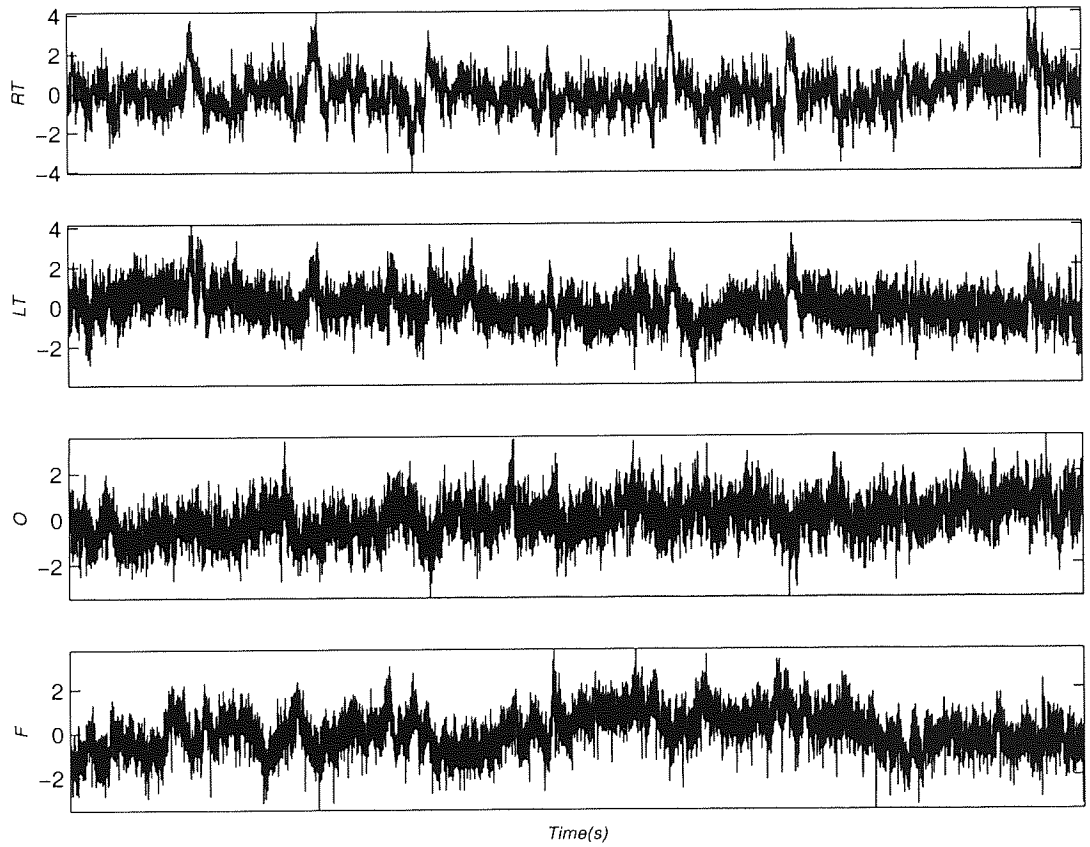


Figure 2.4: Example of raw MEG data, recorded from channels in the (from top) right temporal, left temporal, occipital and frontal regions

Towards this end, the subjects were required to perform a set of simple visual pattern recognition tasks, coupled with some baseline control experiments.

Volunteer subjects who were non-experts in MEG experiments were selected from amongst the staff of the research group for the experiments. To allow for inter-subject variation, as well as variations carried out on the same subjects but over a period of time, the experiments were spread over a period of around a year. Two volunteers were recruited to be subjects. Henceforth, they shall be referred to as subjects A and B. Subject A was female and subject B was male.

In each measurement session, they were required to perform a set of three tasks of gradually increasing difficulty. Two recording sessions were held with all of the tasks repeated at least twice during each sessions. Recording sessions were carried out with approximately a month separating them.

Passive paradigm

In this first experiment, the subject was required to sit still and to fixate at a point on a screen. He/she was then required to close and open his eyes in alternate periods of fifteen seconds each, for a total duration of one minute. The main purpose of this first task is to provide a baseline against which the recordings from the more difficult tasks may be compared. A further objective is to to allow the subject to settle into the session and for any nervousness to subside.

Visual recognition paradigm

1. **Facial recognition**. In this experiment, the subject is shown a series of pictures of unfamiliar faces[1]. The pictures are then removed and after a short pause, a pair of faces are shown to him, one of which corresponds with the initial face. The task is to identify the initial face by clicking one of two buttons. The test images, along with the corresponding pairs of alternative solutions are given in appendix A. For the first two recording sessions, a total of ten such sets were shown over a period of one minute. In the final recording session, only five sets were shown in the same period.

2. **Natural scene recognition**. In the final task, the subject is required to identify images of natural scenery. The way in which this is conducted is the same as in the facial recognition task above, but this time, instead of faces, images of natural scenes

[1]The test images were obtained from the Yale Face Database, accessible via the World Wide Web at: http://cvc.yale.edu/projects/yalefaces/yalefaces.html

were used[2]. This task was considerably more difficult than the face recognition task for two main reasons: Firstly, the images used here were significantly less clear than the face images (all subjects reported difficulty in discerning features in these images). Secondly, there is evidence that the visual system is highly specialised to the recognition of human faces [29; 11; 42]. The images which were used in this test, as well as the corresponding solutions can be found in appendix A.

2.6 Summary

In this chapter our aim has been to introduce relevant theoretical aspects of neurophysiology in general and MEG in particular. As we have pointed out, the single most challenging aspect of MEG signal processing is that the magnetic signals produced by the brain are extremely weak (femtoTesla range). This necessitates the use of extremely sensitive sensors for their detection. The consequence is that environmental magnetic noise sources such as the earth's magnetic field, electrical equipment and power lines generate much stronger signals than those produced by the brain. Hardware measures such as shielded environments and gradiometric sensors help greatly in reducing these contaminants but the resulting signal is still extremely noisy. The traditional approach to addressing these issues has largely hinged on averaging the MEG recordings over a large number of signals. Even when further processing operations are to be applied, the analysis is often preceded by the averaging step. The trouble is that not all event-related signals are known to be time locked to the stimuli, and under certain experimental paradigms, the exact moment of recognition can not be determined with accuracy. Hence, this project is aimed at using only unaveraged MEG data. Further, in response to issues of cost and sheer size of MEG multichannel equipment, we will make the further restriction of performing signal analysis using only single channel MEG.

This approach of using single-channel, single-trial MEG is hitherto completely unique and has not been attempted in any other research effort to date.

[2]the test images for this task were obtained from the Groningen natural images database, which may be accessed via the web at http://hlab.phys.rug.nl/archive.html

Chapter 3

Dynamical Systems

The previous chapter introduced MEG and explained some of the problems encountered in MEG research. This project is an attempt to develop a novel framework in which these problems may be more effectively addressed. Towards this end, we adopt a methodology motivated and developed within the dynamical systems framework, which we will now focus on. Section 3.0.1 introduces the basics of dynamical systems theory. Related terminology that will be used throughout the rest of the thesis is also covered here. Section 3.1 looks at the method of delays, also known delay embedding. This approach is used extensively in this project and is a crucial processing step when analysing single channel data. We also devote some attention to the determination of the embedding dimension, an important parameter which helps determine the success of the embedding. Section 3.2 outlines the differences between linear and nonlinear dynamical systems and discusses the shortcomings of frequency based methods, which implicitly assume linear dynamics. We also perform a series of experiments to explicitly test for the presence of nonlinear dynamics in MEG data. The methodology and results for this are presented in section 3.3

3.0.1 What are dynamical systems?

A system is said to be dynamic if it evolves in time. If we wish to characterise such a system satisfactorily, or even predict its future, it is necessary to study not just overall statistical quantities of the available data, but also the path on which it evolves, or *trajectory*. It follows that dynamical processes are never instantaneous but rather occupy real time. Elements in the data set are no longer simply realisations of a random variable, but are related to each other with respect to their temporal orderings. Hence, data vectors are often referred to as *states* in a dynamical context. The general field of study which concerns itself with systems of

this sort is what is known as *dynamical systems* theory. Briefly stated, the aim of dynamical systems research is to achieve one of two ends:

1. To learn about the dynamic rules governing the evolution of the state from one time instant to another, as well as the relationship between observed data and the actual system dynamics

2. To understand the long-term behaviour of the system via indicators which characterise the shape of the system attractor

More formally, dynamical systems theory is concerned with the study of the iteration of functions from a space onto itself. These iterations can either be in discrete repetitions or in continuous time, but in this thesis we will concern ourselves with only the discrete case, as discretisation is a necessary processing step when performing analysis by digital computers. In its simplest form, this can be stated mathematically as:

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}), \quad t = 0, 1, \ldots, n \tag{3.1}$$

Where $\mathbf{x}$ is the system state vector, indexed by the time variable t, and f is the *transition function.* For a deterministic dynamical system, the behaviour of the system is totally determined by f and the initial conditions, $\mathbf{x}_0$. A more general expression must allow for the presence of dynamical noise:

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \nu_t, \quad t = 0, 1, \ldots, n \tag{3.2}$$

where ν is the dynamical noise term, indexed by t. This quantity represents the inherent uncertainty in a system and can be used to account for spurious signals, background noise or simply complex dynamical effects which we are unable to model adequately. In fact, equation 3.2 is still not entirely generic as its present form assumes an additive noise term.

3.1 Dynamical embedding

To recap our previous assertions, two of the main objectives of this project are to work exclusively with data collected from a single recording site chosen from the MEG array, and to conduct the analysis in the context of dynamical systems theory, wherein the recordings taken at the surface of the scalp are regarded as indirect observations of the unobservable state evolution. Further, it is our intention to use a variety of methods such as PCA and ICA to process the data. However, techniques such as these are applicable only to data sets containing a large number of channels.

Remarkably, there exists a method which allows all of these concerns to be addressed simultaneously. Takens embedding theorem [70] was formulated precisely with the aim of analysing

nonlinear dynamical systems in cases where the actual state variables $\mathbf{x}$ are inaccessible, but rather can be studied via some indirect measurement function $g(.)$. The evolution of the observation with time in this case is then entirely dependent on the evolution of the underlying state dynamics and is given by $g(f(\mathbf{x}(t)))$. Since we are making the observations at discrete time intervals, the observations form a finite series of recordings $g(f(\mathbf{x}_i))$ where $i = \tau, 2\tau, 3\tau, .., n\tau$ and n is the number of observations. In this case, Takens theorem states that, provided certain conditions are met, a data matrix composed of delayed values of a scalar recording will be an embedding of the unobservable dynamical system which generated the measured time series. The conditions that need to be fulfilled are:

1. If the fractal dimension of the system attractor is D, then an embedding dimension of at least $m = 2D + 1$ is required to guarantee the embedding (Whitney's embedding theorem)

2. Both $f(.)$, the system dynamics, and $g(.)$, the observation function, have to be smooth.

This approach, known as the method of delays, uses a series of delay vectors containing consecutive segments from the time series to be analysed. These are combined to form a multidimensional data set. For example, to construct an m-dimensional matrix for an n-valued scalar data set $\{x_i\}_{i=1,..,n}$, one simply extracts segments of length m from the time series, and use these as vectors in the reconstructed matrix, as follows:

$$v_n = \{x_k, x_{k+1}, \ldots, x_{k+m-1}\}$$

The embedding matrix V is formed by simply obtaining v_K for successive values of n and combining these to form:

$$\mathbf{V} = \begin{bmatrix} x_t & x_{t+\tau} & \cdots & x_{t+m\tau} \\ x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(m+1)\tau} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \tag{3.3}$$

The method of delays provides a surprisingly simple procedure whereby it is possible to obtain a topologically correct reconstruction of the system manifold, the validity of which is guaranteed by Takens embedding theorem. Figure 3.1 illustrates the relationship between the true system manifold, the observations and the reconstruction.

3.1.1 Channel selection

While the embedding theorem dictates that the choice of any recording channel would yield a topologically equivalent reconstruction of the system manifold, it is intuitively clear that certain channels would be more appropriate than others for analysis. In its original form, Takens embedding theorem was proved for the case of a noiseless system with smooth observation

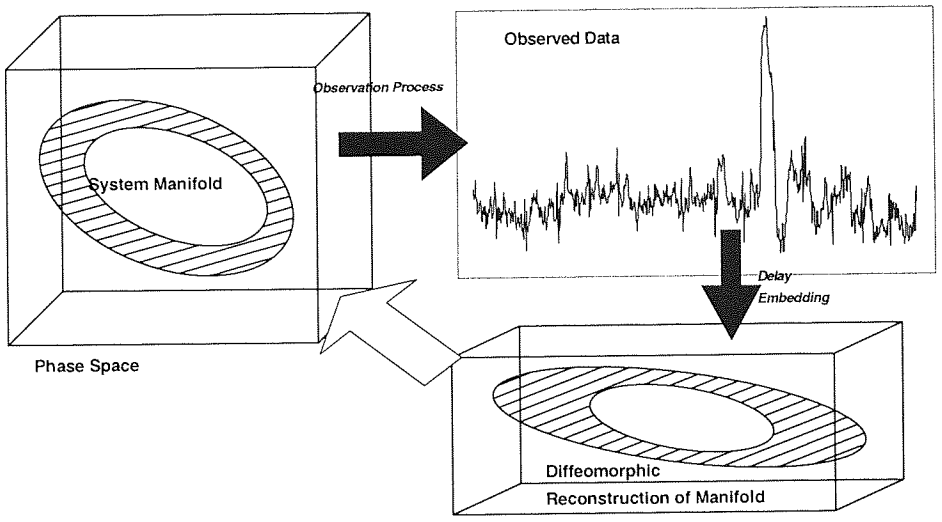


Figure 3.1: Reconstruction of system manifold via dynamical embedding.

and transition functions. The degree to which these assumptions are true, however, will vary between the recording channels. Clearly, practical considerations such as noise levels and the relative power of signals of interest mean that the choice of channel used when constructing the embedding is still of considerable importance. Also, it is worth bearing in mind that two *topologically* equivalent reconstructions of the same manifold may still differ greatly from each other. An inappropriate choice of embedding parameters or recording channel could easily result in an embedding where the reconstructed manifold is restricted to the diagonal of the embedding space, or is swamped by recording noise and artifacts.

For the bulk of results presented in this thesis, we will be working with the time series in channel MRT22 in the right temporal region. This channel was selected due to two main reasons:

1. Relatively low noise levels, and strong signal power, based on subjective assessment.
2. The experiments that we performed required both visual and cognitive processing, which is likely to involve activity in the visual cortex (rear of the head) and the frontal cortex. The choice of MRT22 both presents a compromise recording site, as well as demonstrates the capability of the embedding approach to extract signals which are spatially removed from the recording site.

3.1.2 Determination of embedding parameters

The effectiveness of the delay embedding in the study of the system dynamics depends critically on a suitable choice of embedding parameters. In particular, the choice of the embedding dimension is important as the embedding dimension has to be sufficiently large to embed the system manifold. Recall that Takens theorem specifies that the the quantity $2D + 1$ forms a lower limit to the embedding dimension used. The implication is that any choice of embedding dimension above this value will give a valid reconstruction. On the other hand, when dealing with noisy data sets, it should be noted that because noise is effectively infinite dimensional, increasing the embedding dimension after a valid embedding has been achieved will add further noise into the reconstruction while the power of the signal of interest remains constant. Since noise is almost universally present in real systems, an inappropriate choice of m - both too small and too large - is likely to have a negative effect on the quality of the reconstruction, and all subsequent analysis. It is important, therefore, that an embedding dimension that is at least close to the optimal value is chosen in practice.

Selection of delay time, τ

The choice of the delay time τ can also be important. When deciding on a suitable value for this quantity, the general approach has been to aim for a optimal point where

1. τ is sufficiently large such that the reconstructed attractor is expanded away from the diagonal of the embedding space
2. τ is not so large until the reconstructed attractor is folded in upon itself

Often these two requirements can be met by setting τ such that the components of the delay vector, x_k, are uncorrelated. However, the determination of the value of τ is, in general, uncertain and in combination with the need to obtain a good value for m sets up a difficult problem of estimating two unknown but related parameters, both of which are troublesome to estimate. For our analysis, we choose rather to use the singular spectrum approach (SSA). SSA works similarly to the basic delay embedding algorithm in using a delay vector and also in basing the reconstruction on a delay embedding matrix. However, instead of searching for an optimum τ, SSA starts by setting $\tau = 1$, and forming a large embedding matrix, $\mathbf{W}$, which in general is larger than the equivalent matrix constructed when using basic delay embedding. However, this matrix is then reduced to a smaller m-dimensional matrix via a projection onto the first m principal components of $\mathbf{W}$, to form the SSA reconstruction, $\mathbf{V}$. While these two methods can be shown to be theoretically equivalent, it has been demonstrated in [49] that

the SSA can provide better reconstruction of the system manifold in the presence of noise by virtue of the noise-rejecting properties of the singular vector projection.

Selection of the embedding dimension, m

The standard way to find m is via the estimation of some statistic or quantity which characterises the geometry of the system attractor. This quantity is evaluated at increasingly higher values of m and its evolution studied. Depending on the nature of this statistic, the embedding dimension is assumed to be optimal at the point where the statistic either reaches a minimum, maximum or stabilises.

The most commonly used approach is probably the *false neighbours* algorithm. This works by evaluating the number of so-called "false neighbours" in a data set for a range of m. The idea is that if the embedding dimension used is too low, points that are distant from each other on the true manifold will be embedded close to each other. By comparing the distances of neighbouring points at successive values of m, it is possible to obtain an estimate of the proportion of points that are false neighbours. The embedding dimension m^* whereat this proportion first drops to zero or almost zero, is the minimum m that will generate a valid embedding of the system manifold. Examples of other indicators that might be used are the correlation dimension as well as the singular spectrum of the delay embedding matrix. Yet another approach to estimating a suitable m is by setting the length of the embedding window to be equal to the period of the lowest frequency feature observable in the time series.

However, we note that the first two methods are known to be sensitive to noise and rely on the time series being relatively smooth. Because the MEG data is known to be corrupted with high levels of noise, we will base our selection of m on a combination of the last two methods, which will be discussed next.

- It is noted in [46] that what is important in determining the embedding parameters is mainly the size of the embedding window τ_w. It is suggested that the value of τ_w should be set to be at least as big as τ_p, the *mean orbital period*. This is defined as the mean time between two consecutive visits to the same neighbourhood on the system attractor.

  τ_p can be determined in a number of ways. One approach, used in [37] is to set m based on:

  $$m \geq \frac{f_s}{f_L}, \tau = 1 \tag{3.4}$$

  where f_s is the sampling frequency and f_L is the lowest frequency of interest. It has also been proposed to use the frequency spectrum to find dominant frequency components

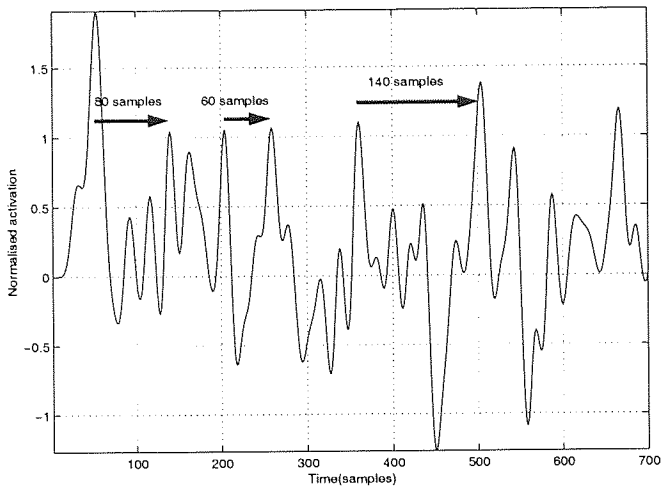


Figure 3.2: Segment of MEG data showing three of the values used to estimate the orbital period of the MEG.

to give an idea of f_L. However, as nonlinear time series may not show well defined frequency peaks, a more suitable approach [46] might be to work in the time domain. The simplest way would be to simply inspect the time series and look for the duration of the longest observable feature. Alternatively, we can calculate the mean *time between peaks* (*tbp*) for the time series. In general, due to noise and other uncertainties, an estimate of τ_p obtained in this manner is normally smaller than the actual value and should only be considered as an approximation.

Figure 3.2 shows a segment of MEG data which has been low-pass filtered to remove high-frequency components which are irrelevant to this procedure (since we are only interested in the *lowest* frequency components in this case). The three values of *tbp* shown have an average of 70 samples. A similar analysis of other segments of MEG recording reveal similar values for *tbp*. As can be appreciated from figure 3.2, the estimation of *tbp* is a subjective process and so should not be assumed to be exact. Also, as stated before, this value will often be smaller than the actual value and so this will be taken as a lower limit on what the optimal value of m might be.

- The other method for estimating m which we will use has been presented in [12]. It is based on the use of the singular spectrum of a dataset to characterise its overall structure. The procedure may be summarised as follows:

  1. The singular values of the data are calculated for a range of embedding dimensions
  2. As the embedding dimension is increased from a low starting point, the structure in the data that is captured by the embedding progressively increases and

consequently, the shape of the singular spectrum evolves.

3. When an appropriate embedding dimension is reached, all of the structure in the data is captured and increasing the dimension further will only result in noise: hence, the width of the "noise floor" in the spectrum increases, whilst the shape of the spectrum itself remains constant.
4. When this occurs, the singular spectrum is said to have "converged".

Unlike in the previous method, this approach utilises a far coarser characterisation of the time series data in order to approximate m. It is felt that the combination of this method with the previous method of finding the mean orbital period, which works on extremely short-term, local properties of the time series is able to provide us with a more reliable approximation of m than is possible with either method individually.

For the same segment of data from which the section in figure 3.2 was extracted, the singular spectra were calculated for embedding matrices with a range of m from 50 to 120 (this range was chosen based on the estimate obtained using the mean orbital period). In increments of ten dimensions, the singular spectra obtained in this way are shown in figure 3.3. As can be seen, the initial shape of the singular spectra contain a pronounced "shoulder" within the first few singular values. As the embedding dimension is increased, this shoulder grows less pronounced but is still discernible up until about 80 dimensions. It can be seen that somewhere between here and the 100-dimensions point, the shape of the spectrum has converged to almost a smooth line, terminating in the noise floor. This is similar to the singular spectra obtained for pure white noise (on close inspection, there is still some structure in the spectrum, though this too seems to have converged and doesn't change much after $m = 90$). However, this observation is also consistent with the assumption that the MEG time series is a combination of strong stochastic and possibly nonlinear components (in the extreme of a chaotic attractor, the singular spectrum should resemble that of white noise). The real noise floor, however, can be clearly seen beyond the sloped section of the singular spectrum. After the embedding dimension passes about 95, this can be seen to gradually elongate while the rest of the spectrum remains constant. Based on these observations, it is concluded that a reasonable value of m for the MEG time series is in the region of 90 dimensions.

The analysis is then also applied to other recordings. Figures 3.4 and 3.5 show singular spectrum plots obtained for recordings of the face recognition and natural scene recognition tasks respectively. In both cases, 90 seemed to be a reasonable embedding dimension beyond which the spectra ceased to evolve significantly. Of course, the optimal choice of embedding dimension is both difficult to determine exactly, and is likely to show some variation between datasets. However, estimating m individually for each data set is both impractical and renders comparisons difficult. In practice, it was ob-

served that the choice of $m = 90$ proved to be a sensible choice for most data sets and unless mentioned otherwise, this is adopted in subsequent experiments.

To quantify the changes in the singular spectra, we would like to calculate a convergence function which reflects the change between the singular spectra calculated at subsequent embedding dimensions. A simple way to achieve this is by calculating the Euclidean distance in the following manner:

1. Calculate the singular spectra for the target range of embedding dimensions (in our case 50 to 120).
2. To make the spectra comparable, we first *truncate* the singular spectra to 40 dimensions. We choose 40 because almost all of the structure in the singular spectra appear in this region.
3. Next, we *normalise* the values of the singular spectra by dividing with the largest singular value. This has the effect of setting the first singular value as a fixed reference point.
4. Finally, the convergence function can be calculated simply as the element-wise sum-squared error between subsequent singular spectra.

The normalised sum squared error curves generated in this fashion are given in figure 3.6.

3.2 Linear vs nonlinear dynamical systems

The study of dynamical systems can be divided into two main categories depending on the nature of the transition function f. The first class, where the transition function is linear, is accordingly termed linear dynamical systems. Similarly, the term nonlinear dynamical systems is used to refer to cases where the transition function is nonlinear.

Why the division into these two classes? At least part of the reason is historical: it is only recently that the ability to effectively analyse nonlinear systems has become widely attainable. The nature of linear dynamics is such that it permits only one of three classes of behaviour:

1. Convergence to a stationary point
2. Sustained oscillation at one or a number of fixed frequencies
3. Exponential growth (instability)

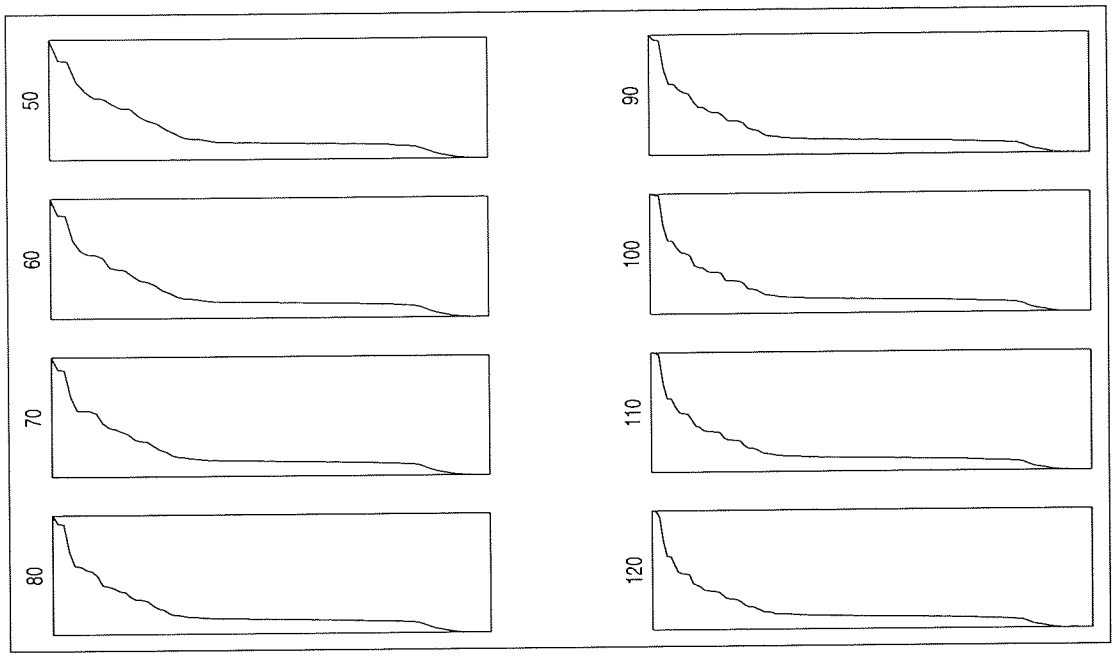


Figure 3.3: Series of graphs showing the singular spectra evaluated for a range of embedding dimensions from 50 to 120, in increments of 10 dimensions. The numbers on the left of each graph indicate the embedding dimension

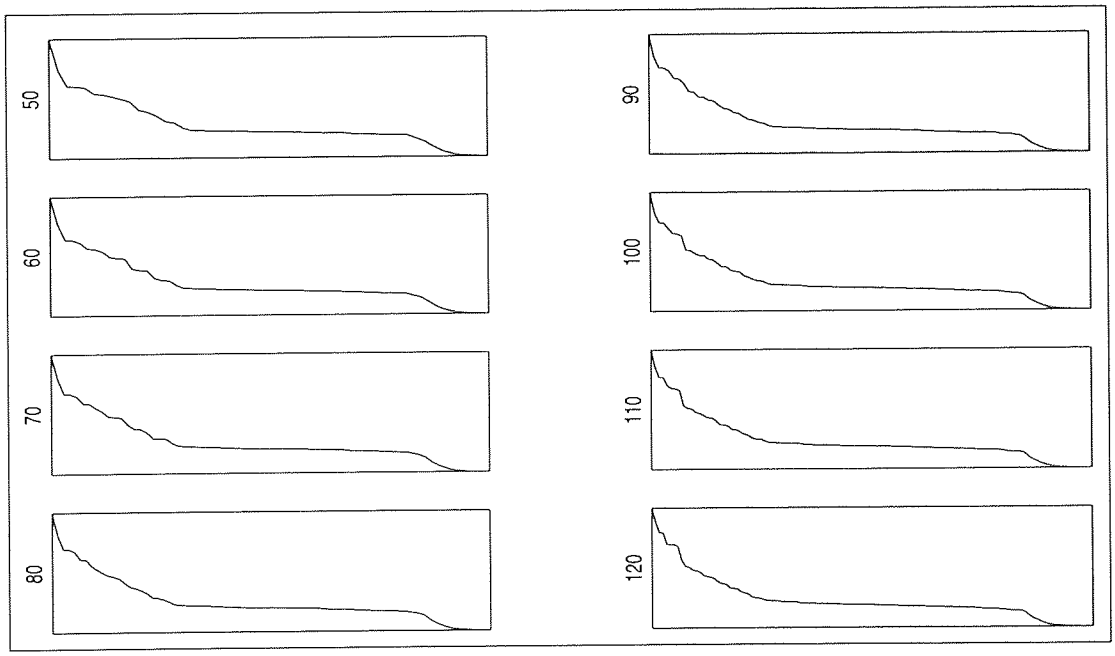


Figure 3.4: Multi-plot showing the singular spectra evaluated for a range of embedding dimensions from 50 to 120, in increments of 10 dimensions. Results shown are for the face recognition task

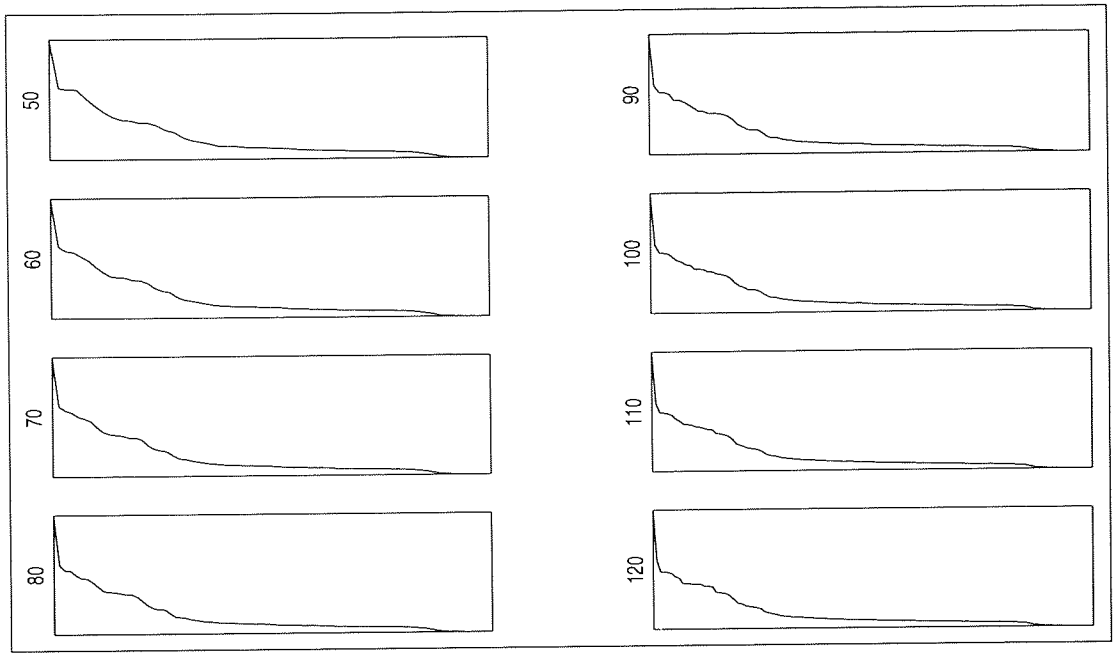
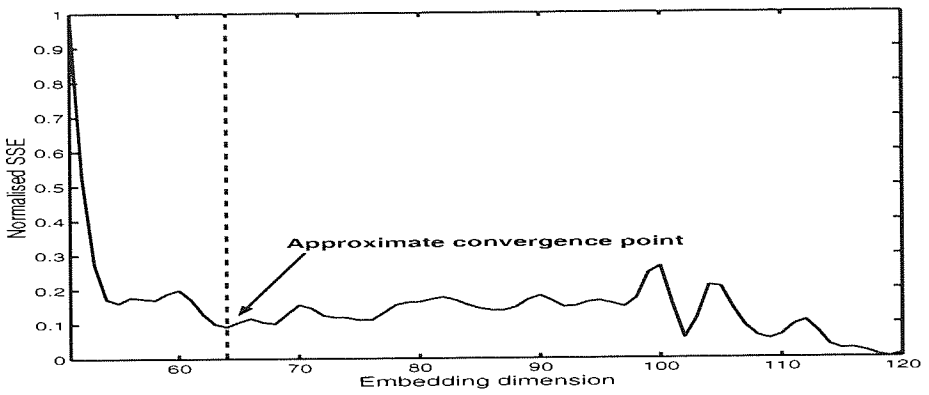
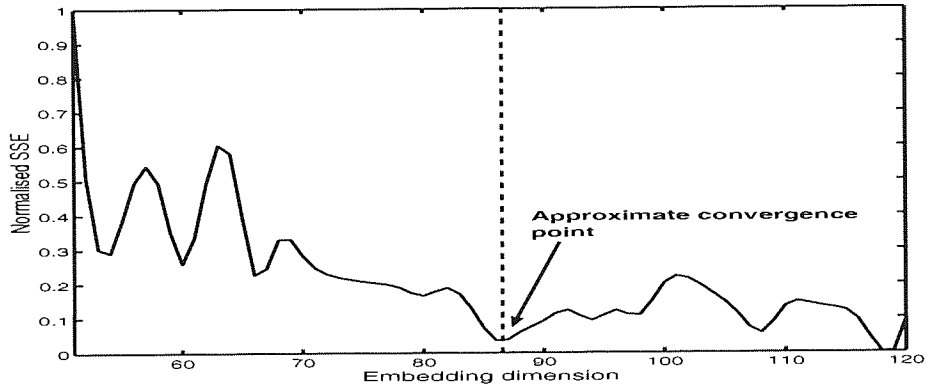


Figure 3.5: Multi-plot showing the singular spectra evaluated for a range of embedding dimensions from 50 to 120, in increments of 10 dimensions. Results shown are for the natural scene recognition task

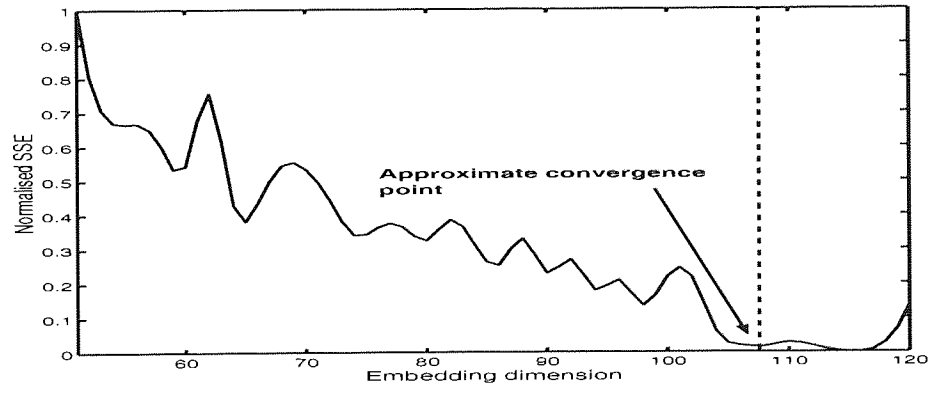
As systems belonging to the first or last classes are of little interest, the focus of linear dynamical systems analysis has been on the study of *frequency components*. Under this paradigm, the signal of interest is described as being the combination of a large number of separate frequency generators. For long periods of time, the analysis of dynamical systems using linear tools was probably the only really effective method of characterising noisy data. Consequently, a large body of research has grown up around a number of methods which either implicitly or explicitly assume linear dynamics. These include all methods related to Fourier transforms, frequency and power spectral methods, as well as explicitly linear modelling techniques such as auto-regressive and ARMA (Auto Regressive Moving Average) models and Kalman filters. Certainly, the characterisation of time series via power spectra has had a long and distinguished history of successful applications in an exceedingly wide variety of research areas and remains one of the most reliable methods for time series analysis. At the same time, it is not true to say that nonlinear dynamical systems have received no attention at all. However, the bulk of the interest in this area has been focused on the issue of chaotic dynamics. This phenomenon refers to a particular class of nonlinear dynamical systems where the behaviour of the system comes to almost resemble that of white noise. While totally deterministic, systems exhibiting chaotic behaviour are characterised by an extremely high sensitivity to initial conditions. In fact, the standard method of detecting nonlinear behaviour, the "Lyapunov exponents", is actually a measure of the rate of divergence of



(a) Idle task



(b) Face recognition



(c) Natural scene recognition

Figure 3.6: Convergence functions of singular spectra. The y-axis gives is the Euclidean distance between singular spectra measured at successive embedding dimensions, after first normalising the spectra to the first singular value

points that are nearby in a state space. An exponential rate of divergence is taken as an indication of chaos or nonlinear dynamics. Nevertheless, for the most part, the study of chaotic dynamics has been more fruitful as a mathematical, rather than a practical interest.

Nevertheless, the last few years have seen a massive expansion in the scope of research into dynamical systems that are not adequately handled by these traditional methods. The existing body of linear tools has been augmented by a host of new approaches that are capable of modelling systems that show nonlinear behaviour but not chaos. The behaviour of such systems vary from limit cycles with varying frequency content to fluctuations between modes of dynamical behaviour, which may include chaotic dynamics. Because these dynamics can contain a mix of linear, nonlinear (both chaotic and non-chaotic) as well as stochastic behaviour, modelling them using methods that assume strictly linear, or strictly chaotic characteristics, is likely to produce unsatisfactory results. To complicate matters, systems can be chaotic at one level of organisation, and much more simply representable at another, reflecting the existence of micro and meso-scopic levels of dynamical behaviour in complex systems such as the brain.

3.3 Testing for nonlinearity of MEG dynamics

It has been a major assertion of this thesis that allowing for nonlinearity in MEG recordings is likely to help yield more information regarding the underlying brain function. To summarise again, our motivations for pursuing this avenue of research are twofold:

1. As far as exploitation of MEG data is concerned, linear signal analysis methods have fallen short. Traditional methods based on frequency information have hitherto only produced limited and superficial knowledge regarding the functioning of the brain, while leaving a large variety of signals and structures in the data unaccounted for.

2. Prior knowledge regarding the functioning of the brain, both at the cellular level and on the scale of human behaviour, indicates the presence of nonlinearity.

Nevertheless, the latter reasoning can prove insufficient to justify the use of nonlinear methods. The fact that a system contains nonlinear components does not prove that this nonlinearity is also reflected in a specific signal we measure from that system. We could, of course, forgo theoretical rigour and simply forge ahead with the application of nonlinear tools, citing successes in other fields as ample justification. In this case, the observation of interesting results would provide *a posteriori* justification for our approach. However, bearing in mind the increased computational and analytical demands of nonlinear analysis, formal prior validation of nonlinearity is still desirable. It is also notable that although a number of nonlinear

analysis studies have been carried out on EEG (examples being [71; 57; 16; 17]), these have tended to be either inconclusive, or focused more on epileptic electroencephalograms and the search for chaos, rather than nonlinearity (though the two are certainly related). To the best of our knowledge, the literature contains no formal studies of the linearity properties of MEG. In this section, we seek to address this shortcoming by seeking formal proof for the presence of measurable nonlinearity in MEG data.

How can we test for the presence of nonlinear generating components in the data? The naïve way is to calculate some general measure of nonlinearity or chaos, examples being various high order statistics and Lyapunov exponents. One popular method has been to calculate both the correlation dimension and first Lyapunov exponent - a combination of convergence of the correlation dimension and a positive exponent was then taken as an indication of nonlinearity. However, it is now generally accepted that measures such as these are not sufficient, by themselves, to establish chaotic or nonlinear behaviour in the data [26; 72]. In particular, the measurement of such statistics is often prone to noise contamination and requires large input data sets, which increases the computational overhead. In addition, using longer data sets increases the likelihood of encountering nonstationarities. Errors associated with the acquisition of data such as inappropriate sampling frequency, noise filtering and digitisation error can all lead to erroneous values of these statistics being returned. Finally, even if we were able to determine these values with sufficient accuracy, the actual distribution for the nonlinear statistic in question is generally not known except for the simplest of models.

The method of surrogate data analysis [72] solves some of these problems by providing a suitable statistical framework in which nonlinearity tests may be performed more reliably. It is based on the principle of Monte Carlo methods: the idea is to sample in the space of possible time series matching some carefully chosen *null hypothesis*, then perform a standard statistical t-test to reject this hypothesis.

The first step is to specify a null hypothesis against which to test. In our case, a suitable null hypothesis might be that the MEG data was generated by a linear stochastic Gaussian process. Next, a nonlinear parameter is extracted from the data. In theory, this can be anything, provided it is able to reject data not belonging to the class of models defined by the null hypothesis. However, it is often desirable that this be a physically meaningful quantity. In the past, Lyapunov exponents and correlation dimensions have been popular choices. The most important requirement when choosing this parameter, however, is that it should have a relatively localised distribution when applied to data that conforms to the null hypothesis. In other words, the value of this quantity should be sensitive to changes near regions of interest since this will increase our ability to reject data with different characteristics from the null hypothesis.

For our purposes, we will not use a single null hypothesis or discriminating statistic but rather a combination. The idea is to use a hierarchy of tests, starting with relatively simple assumptions then gradually narrowing down the scope of our testing. It is hoped that using this approach will allow us to gain a clearer understanding of the true nature of the MEG data.

3.3.1 Null hypotheses

To test the MEG data, two sets of null hypotheses have been formulated:

1. *Independent identically distributed (I.I.D.)* - The purpose of this null hypothesis is simply to test if significant dynamics can be discerned in the data. Surrogate data generated for this hypothesis should still have the same static distribution as the original data set.

2. *Linear Stochastic process* - This a much more realistic null hypothesis than the previous one. In this case we are directly testing against the null hypothesis that the data contains only linear components (and Gaussian noise). We will not be making any prior assumptions about the distribution of the data.

3.3.2 Generating surrogate data

In the case of the I.I.D. null hypothesis, the simplest way of generating surrogate data is to shuffle the data points so as to destroy all temporal structure. At the same time, this will ensure that the resulting surrogate data set retains the static distribution of the data.

For the second case, we want to ensure that the surrogate data retains the linear properties of the original data, i.e.: it should have the same power spectrum. There are two main approaches: The first method is to directly model the data with an appropriately specified ARMA or autoregressive model, then to use this to generate new data sets. However, this "*typical realisations*" approach, as it is commonly known, is troubled by the need to successfully extract the model describing the underlying system. Parameters such as model class and order need to be estimated and this can be a delicate process.

The other approach, commonly known as *constrained realisations*, is to directly modify the power spectrum of the data. This is commonly done by taking the Fourier transform of the data, randomising the phases before taking the inverse Fourier transform again. Generally, an extra processing step is performed whereby the amplitude of the data is transformed to

a Gaussian before phase randomisation. Finally, the distribution is transformed back to the original distribution of the data. However, this method has recently been criticised for being responsible for occasional false rejections of the null hypothesis [48; 47]. Because the rescaling process also affects the power spectrum of the data, surrogate data sets generated using this method have power spectra that are only approximately equal to the original time series. Furthermore, rescaling to a Gaussian distribution invariably imposes a flatness bias in the resulting power spectrum, which can affect the end result of the surrogate test. Hence, for the experiments described here, another method, proposed in [66], was used to generate the surrogate data.

In brief, the idea is to directly optimise the spectrum of a randomly generated time series such that it matches that of the original data. First, an initial data set is chosen either by generating some random numbers, or by using the phase randomisation method mentioned above. Next, the power spectra of the new set is optimised via a two step iterative procedure:

1. The Fourier transform is performed on the surrogate data, and the magnitude of the Fourier coefficients are set to that of the original data.

2. The inverse Fourier transform is calculated. However, changing the Fourier amplitudes will have affected the distribution of the data. To remedy this, a two step procedure is applied: firstly, both the candidate surrogate set and the real dataset are arranged so that the largest data points come first, second largest second, and so forth. However the original time indices of the surrogate set are retained, and are used to rearrange the ordered version of the original data set. This ensures that the static distribution of the new time series now matches that of the original data set, while the dynamical ordering of the data points are changed to match the surrogate dataset. However, this reordering will again change the power spectra. This so step one is carried out again and so on. With each iteration, the changes required become increasingly smaller until convergence is reached.

Further, as recommended in [48], we terminated the training after the adjustment of the spectra, rather than the amplitude distribution, as this method results in surrogate data with the same autocorrelation structure as the data. The slight deviation in the amplitude spectra that this results in was deemed to be acceptable. For a more detailed description of this algorithm, or of surrogate data testing in general, please refer to [66; 72; 67].

3.3.3 Discriminating statistics

As with the null hypotheses, we will be using a number of discriminating statistics with the intention of testing different aspects of the data. These are:

Time asymmetry

This is one of the most powerful signatures of nonlinearity. It uses the fact that the statistics of linear stochastic processes are always symmetric under time reversal, since the power spectrum itself does not contain any information about the direction of time. There are a number of ways for measuring time asymmetry but one of the more useful methods, and also the one which we will be using, is given by the following equation:

$$\phi^{rev} = \frac{E[(s_n - s_{n-\tau})]^3}{E[(s_n - s_{n-\tau})^2]^{3/2}} \tag{3.5}$$

Nonlinear prediction error

While time asymmetry is one of the strongest indicators of nonlinearity, in itself it does not necessarily imply the presence of nonlinear dynamics. For example, it is known that a non-invertible measurement function can result in time-asymmetry even though the dynamics of the system may be linear. To directly test for nonlinear dynamics, one may use nonlinear prediction errors. The idea is that a nonlinear predictor will be able to exploit nonlinear characteristics of the data to produce a better prediction than data with identical linear properties but with the nonlinear characteristics removed.

For our experiments, we have chosen to use the locally linear prediction method proposed in [21]. Suppose, for a point in the time series, $x(t)$, we wish to predict $x(t+1)$, the algorithm would proceed as follows:

i) Select N closest points to the point $x(t)$, $x(t'_n)|_{n=1\ldots N}$.

ii) Collect the set of all immediately following points, $x(t'_n+1)|_{n=1\ldots N}$, and find the linear map $x(t'_n)|_{n=1\ldots N} \rightarrow x(t'_n+1)|_{n=1\ldots N}$ using simple least squares fitting.

iii) Use the linear map calculated in (iii) to predict the immediate future point for $x(t)$.

As a test for nonlinear dynamics, the nonlinear prediction error is particularly suitable since

it directly tests the temporal evolution of the data for signs of determinism.

Correlation dimension

This is one of the oldest measures of nonlinearity, though it is still commonly used and popular. A closely related statistic is the *correlation integral*, which is normally approximated using:

$$\hat{C}_2(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||x_i - x_j||)$$

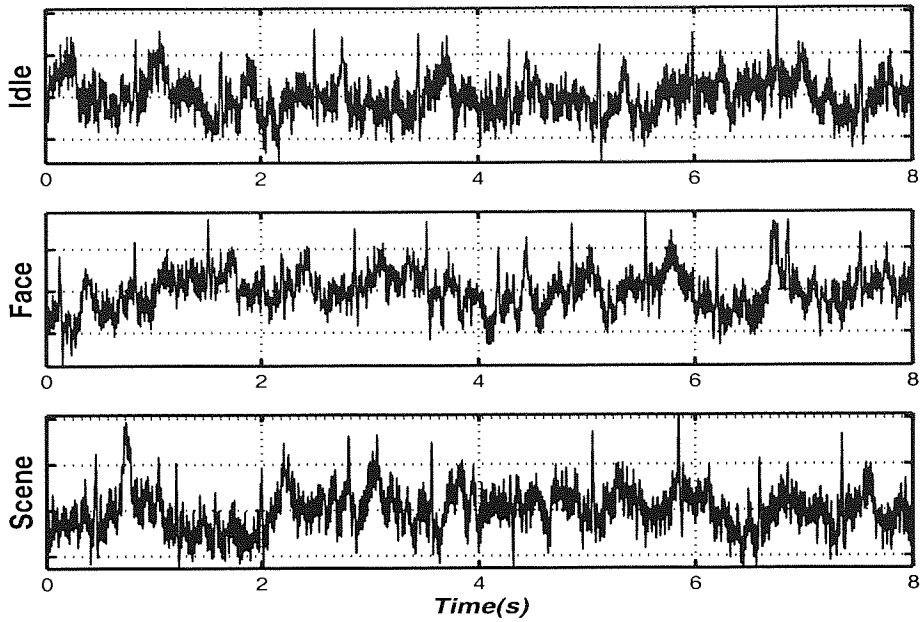
where $\Theta(.)$ is the Heaviside function, which returns a one when its argument is positive, and zero when it is negative. Hence, $\hat{C}_2(\epsilon)$ approximates the probability that the distance between two points in the data set is smaller than the parameter ϵ. The correlation dimension is found by calculating the gradient of the $\log(\hat{C}_2)$ versus $\log(\epsilon)$, and is defined for the limit when the embedding dimension approaches infinity. Of course, in practice this is not possible and hence any estimate of correlation dimension will be parameterised by the embedding dimension m and τ, the delay lag.

In the context of surrogate data analysis, the idea is that a nonlinear deterministic system, while appearing like noise, will converge to a stable correlation dimension whereas white noise will effectively have a correlation dimension of infinity.

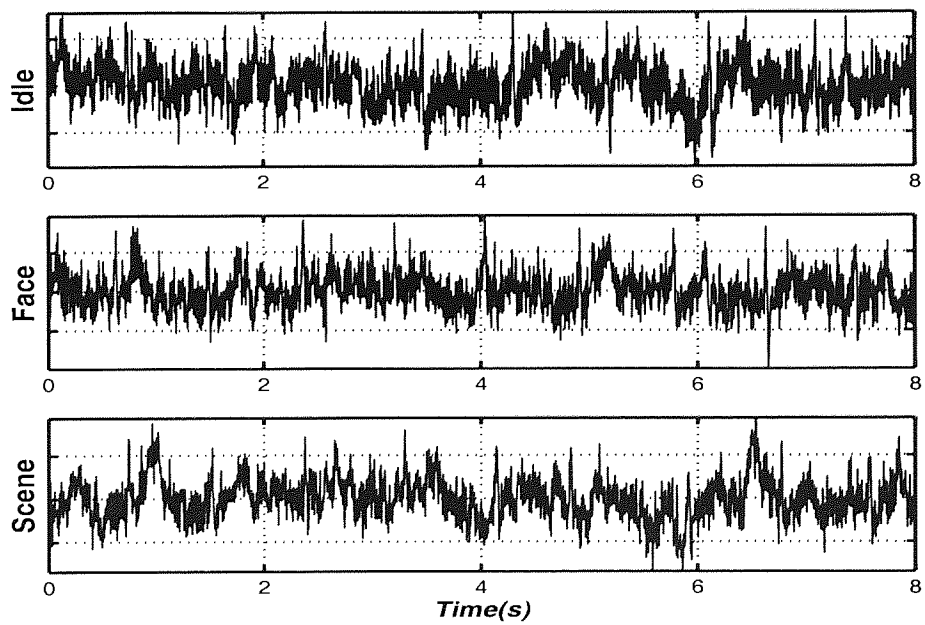
The problem with using the correlation dimension as a discriminating statistic is that real world data often has high stochastic content, which can confound the test. Nevertheless, it remains a widely used indicator for the presence of low dimensional chaos and should be included if only for completeness.

3.3.4 Experimental setup

For our tests, we selected samples from each of the three tests: idle, facial and natural scene recognition. The time series to be tested were all from channel MRT22 in the right temporal region. Furthermore, as the nonlinearity tests generally depend on signal stationarity, we extracted an 8-second segment from each channel as the signals can be seen to be roughly stationary within this time period. Segmentation into periods of roughly the same time-length has been used previously by others [16; 57; 73]. The segments to be analysed are shown in figure 3.7. One set each has been chosen from subjects A and B.



(a) Subject A



(b) Subject B

Figure 3.7: Segments of MEG recorded in the right temporal lobe

For the correlation dimension and prediction error tests, the data had to be embedded. The embedding dimension used, as determined in section 3.1.2, was 90, and the delay time parameter τ was set to one. As the determination of the embedding dimension carries an element of subjectivity, all experiments were performed for a range of embedding dimensions from 85 to 95 dimensions. For the time asymmetry statistic, the delay between the two evaluation points was varied from 1 to 10.

How can we evaluate the significance of the results? The most common way is calculate the deviation of the data from the mean in terms of numbers of standard deviation. The so-called p-value is then given by the well known error function erfc(). However, this depends on the assumption that the distribution of the statistic is according to Student's t distribution, which is either difficult to prove or simply untrue. For our work, we use the method proposed in [72]. This is based on the idea of rank statistics, the basic idea being as follows: first, we select a residual probability α of a false rejection. Then, for one sided tests (such as correlation dimension and prediction error, where we are only looking for small values), a total of $1/\alpha$ surrogate data sets need to be generated (well, to be exact, we only need $1/(1-\alpha)$ sets, but by using $1/\alpha$ we are being more stringent). In this case, the probability that the data has the smallest value in the set by pure coincidence, is then exactly α, as desired. For a double-sided test such as the time asymmetry measure of equation (3.5), $2/\alpha$ surrogate data sets are needed. Hence, in order to attain the commonly used 95% significance mark, we generated 20 surrogate data sets for the prediction error and correlation dimension tests, and 40 surrogate sets for the time asymmetry test.

3.3.5 Test results and analysis

The surrogate tests were run and the results are given in figures 3.8, 3.10 and 3.12, for the idle, facial and scene recognition tests respectively. For the recordings from subject B, the results are shown in figures 3.9, 3.11 and 3.13 respectively. To demonstrate the repeatability of these results, an additional set of test results were generated and are included in appendix B.

As expected, the null hypothesis of I.I.D data was clearly rejected using all of the test statistics. This confirms the existence of measureable dynamics in the raw data. This result is clearly not surprising since the I.I.D. null hypothesis implies that the MEG data is a series of values randomly drawn from a static distribution.

However, in the case of the Fourier surrogates, the results were more mixed:

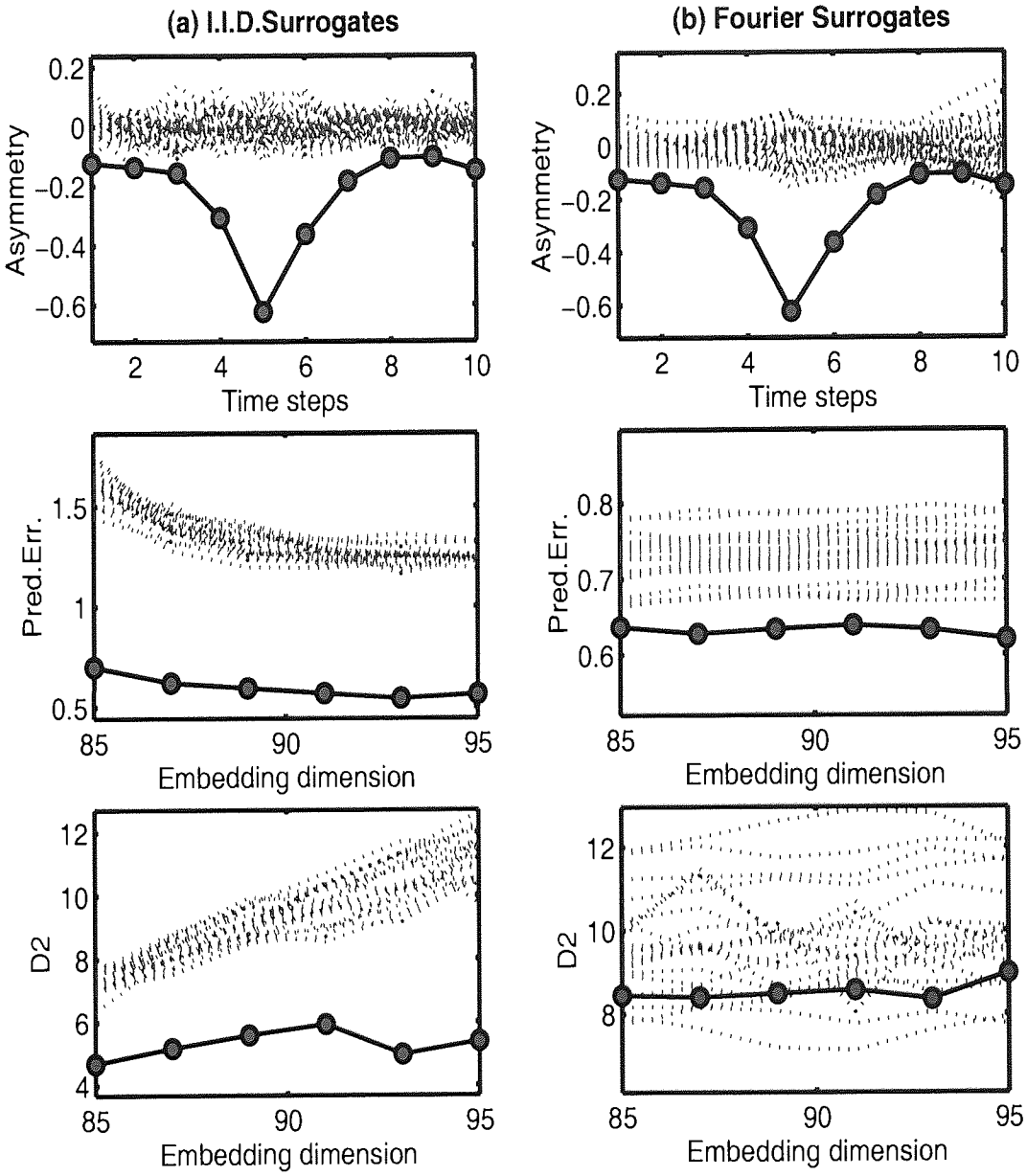


Figure 3.8: Result of surrogate tests on the MEG recordings from subject A for the idle task. The blue dotted lines indicate the surrogate data, and the red circles are the statistics obtained for the actual dataset. The three rows contain results for the time asymmetry, RMS prediction error and correlation dimension surrogate tests respectively.

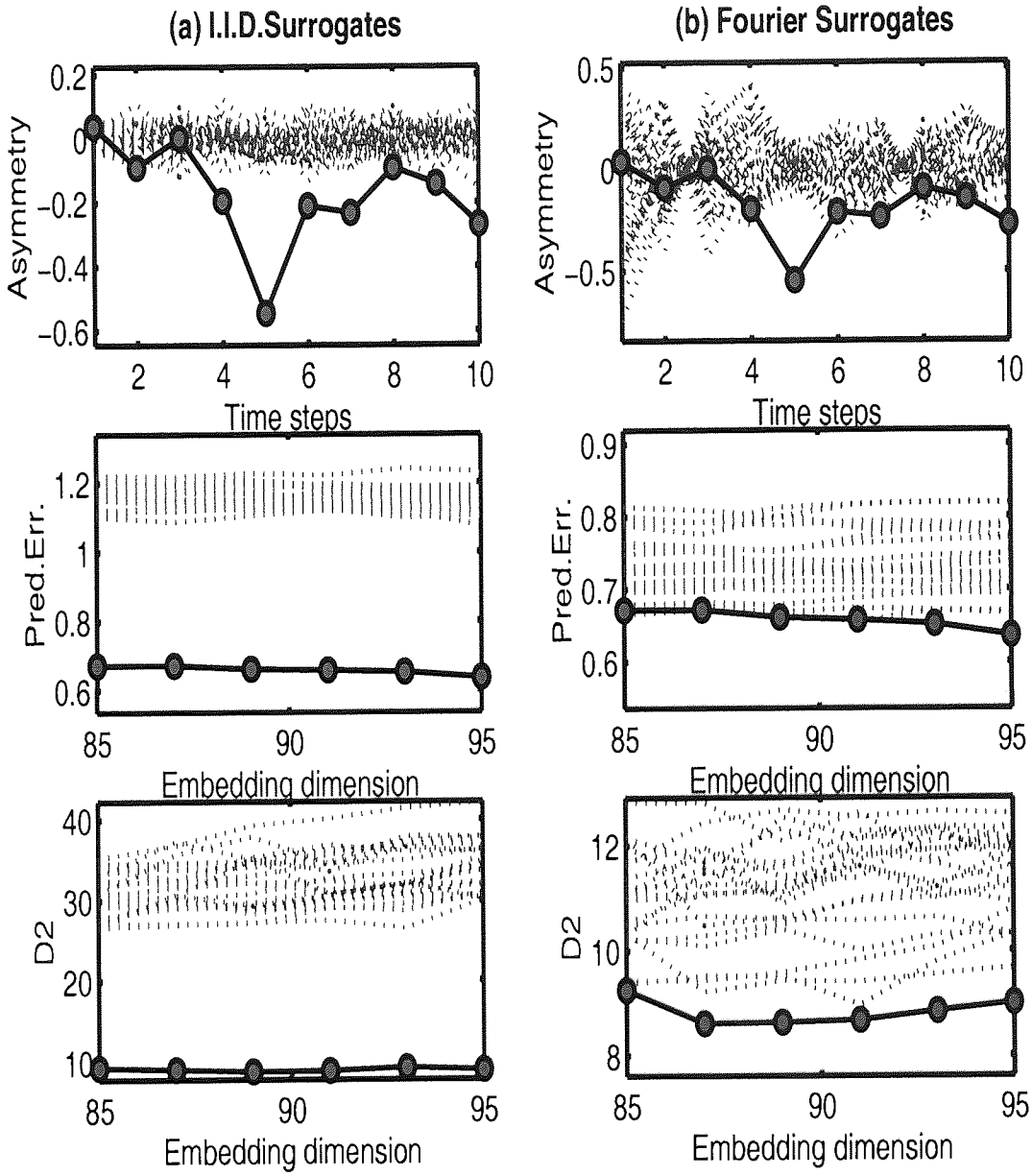


Figure 3.9: Result of surrogate tests on the MEG recordings from subject B for the idle task. The blue dotted lines indicate the surrogate data, and the red circles are the statistics obtained for the actual dataset. The three rows contain results for the time asymmetry, RMS prediction error and correlation dimension surrogate tests respectively.

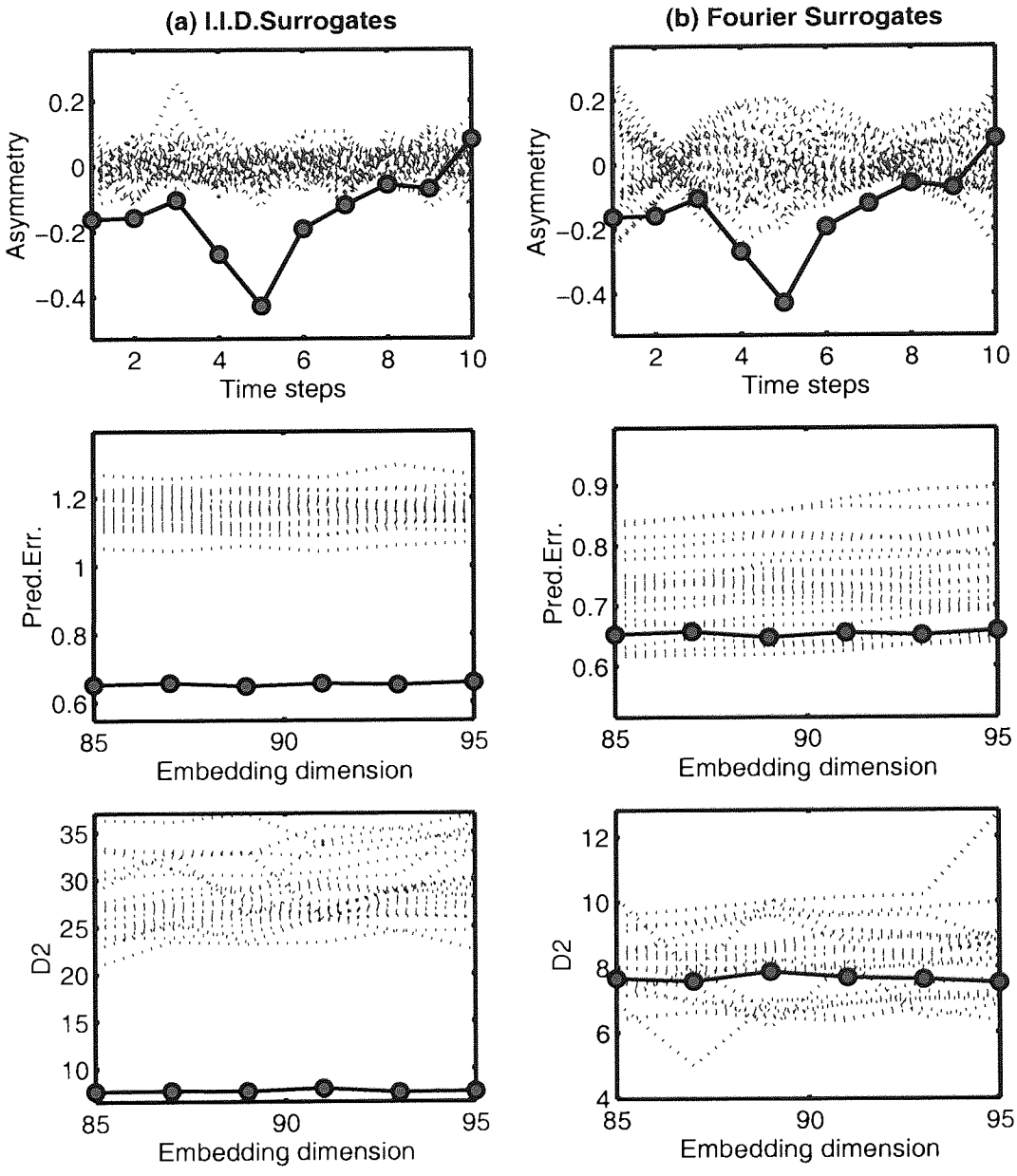


Figure 3.10: Results of surrogate testing on facial recognition MEG recordings. Subject A

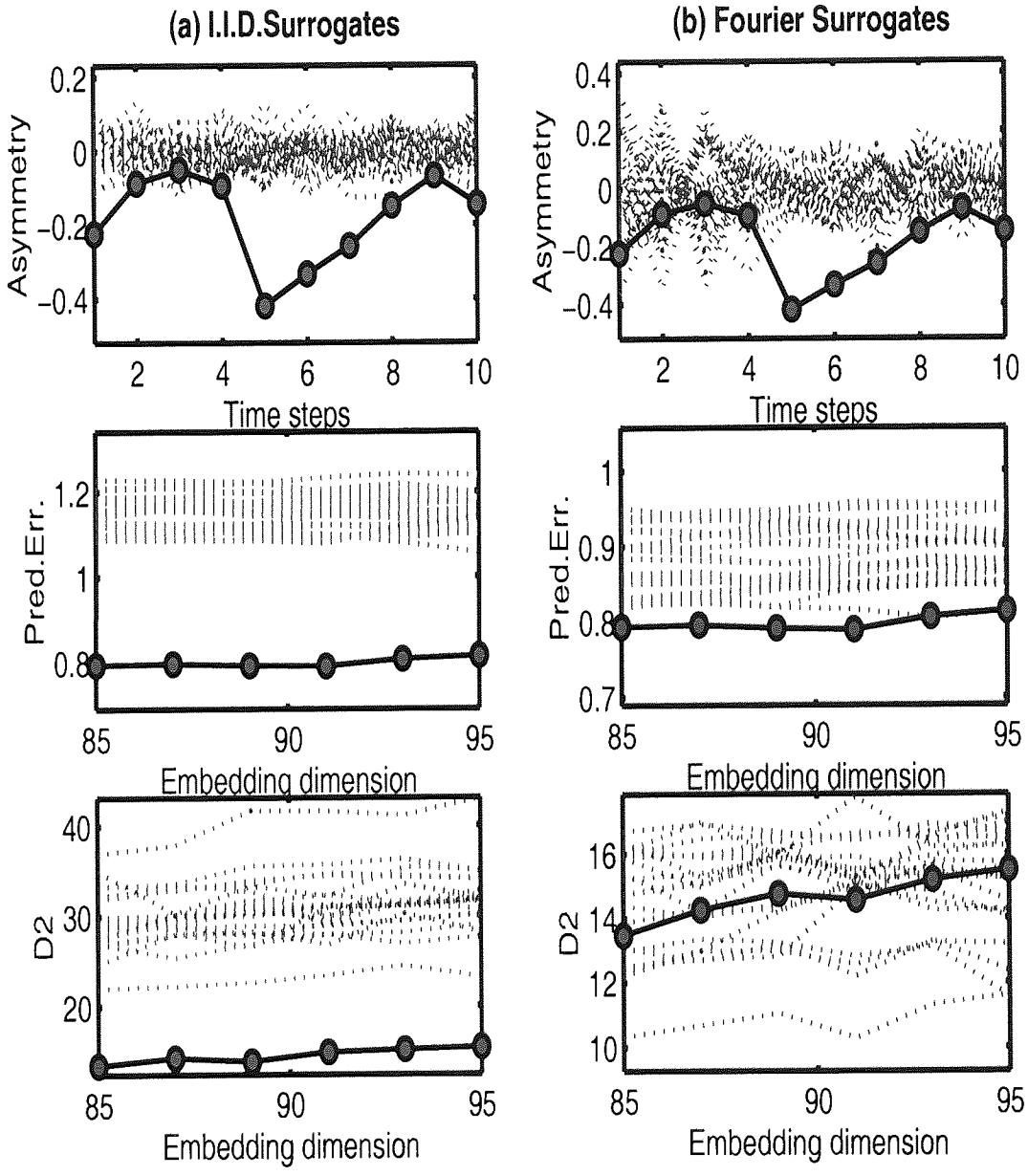


Figure 3.11: Results of surrogate testing on facial recognition MEG recordings. Subject B

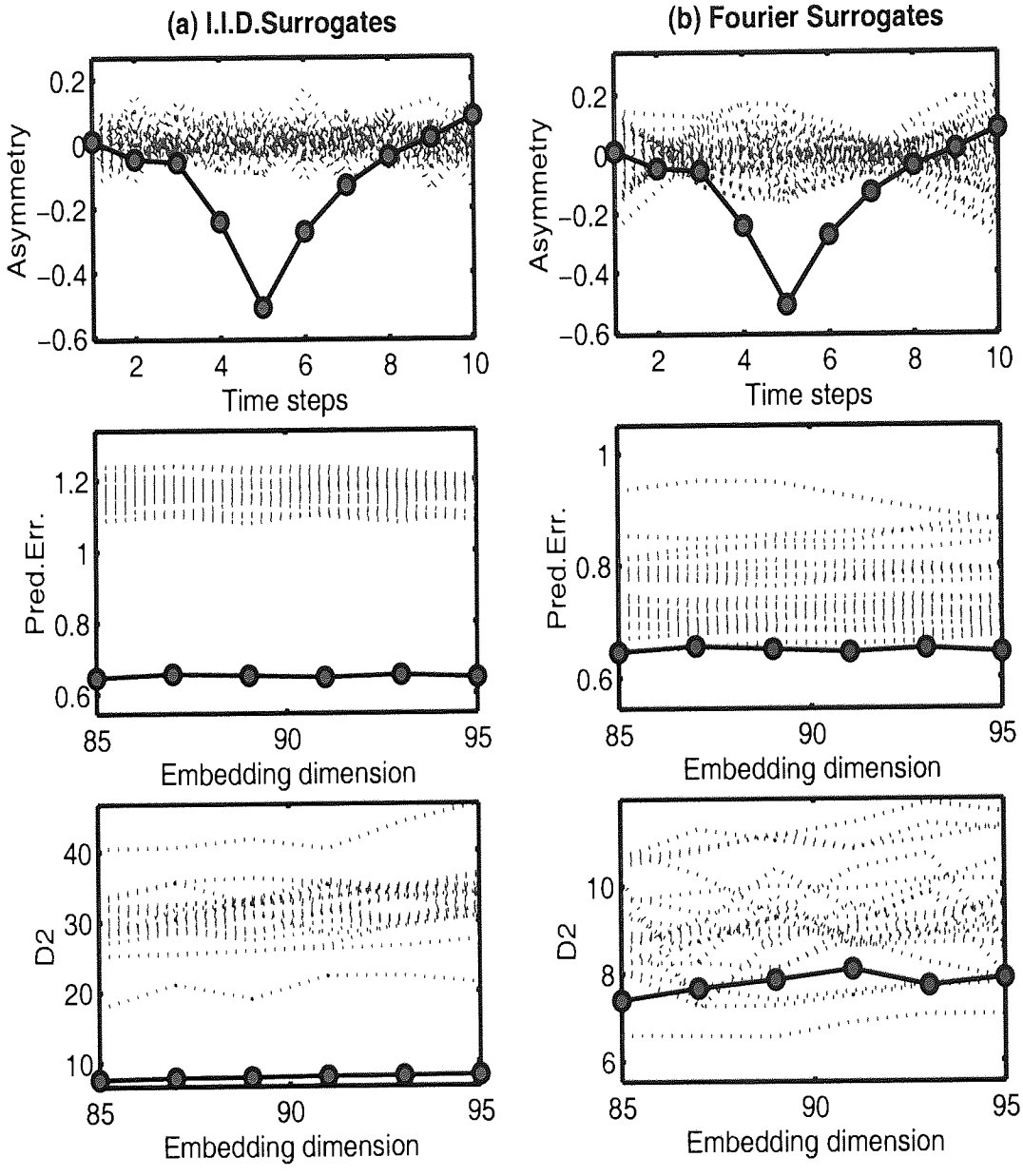


Figure 3.12: Results of surrogate data tests performed on natural scene recognition MEG. Subject A

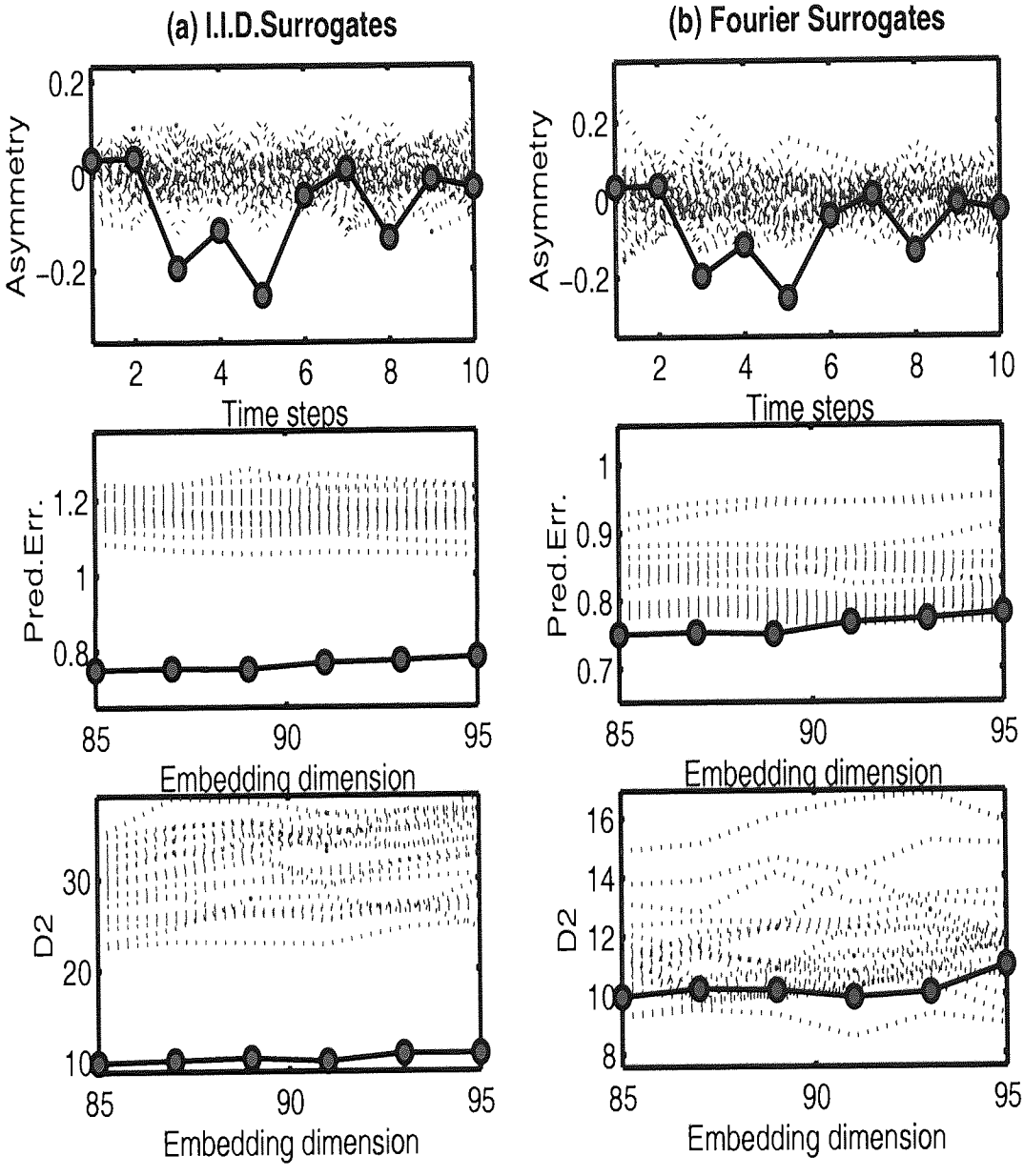


Figure 3.13: Results of surrogate data tests performed on natural scene recognition MEG. Subject B

- The **time asymmetry** statistic seemed to perform well, clearly and consistently indicating the presence of nonlinearity for all six of the time series that we tested (though this was only for a range of delay values).
- The **nonlinear prediction error** test was also successful in rejecting the linear null hypothesis, though only marginally so - we can see areas where the surrogate data sets occasionally have lower prediction error than the data. However, this is to be expected as the linear null hypothesis is much harder to reject than the i.i.d. null hypothesis, and noise and artifacts will inevitably affect the test results. For the most part though, the prediction error statistic was able to reject the linear null hypothesis, indicating the presence of nonlinear dynamics.
- The results for the **correlation dimension** were far less satisfactory. It was only able to cleanly reject the null hypothesis in one test set, the idle task recording for subject B. It exhibited marginal rejections for some of the other sets and completely failed to reject the null hypothesis for the facial recognition tests for both subjects A and B.

From these results, we might draw the following conclusions:

1. Overall, the surrogate test results pointed to the presence of nonlinear dynamics in the MEG data.
2. The reduced rejection margins when comparing the results using Fourier surrogates with those of the I.I.D. surrogates, as well as the failure of the correlation dimension test to reject the linear dynamics null hypothesis, however, indicates that the nonlinearity is likely to be embedded in a background of noise and linear oscillations. So, while there is justification for use of nonlinear signal processing methods, there is also a need for the nonlinear components to be isolated and extracted in order for them to be properly analysed.
3. The results are also consistent with the well known fact that the correlation dimension is susceptible to noise and artifacts and is hence unsuitable for application on unprocessed MEG data.

3.4 Summary

The objective of this chapter has been to present a brief introduction to relevant elements of dynamical systems theory. In particular we have tried to focus on the idea of the MEG generator as an unobservable dynamical system. In this framework, the MEG recordings are

simply observations of this dynamical system. Early in the chapter, we present the dynamical embedding approach as a means of shifting the analysis of MEG from the study of the recorded time series, to the phase space in which the actual system dynamics are assumed to reside. However, while delay embedding is a powerful tool capable of generically reconstructing an unobservable manifold, it is most effective only when properly applied. Hence, in this chapter we also demonstrate the use of two techniques to help guide our choice of the most important embedding parameter - the embedding dimension (and by extension the embedding window length). Our tests indicate that a reasonable choice of embedding dimension is around 90, corresponding to a window length of 360 milliseconds. Finally, we address another important issue: the presence of nonlinearity in MEG data. We apply a surrogate testing methodology to ascertain the presence of nonlinear components in the MEG data. The results of our tests on samples of MEG data are presented and appear to indicate that the MEG time series are a mixture of linear and stochastic components but with definite nonlinear components mixed in.

This chapter concludes the theoretical introduction to the project. In the next two chapters, we review the analysis of single channel MEG recordings with a novel suite of signal processing tools.

Chapter 4

Blind source separation of single channel MEG

"Is grandmother an oscillation?"

M.P.Stryker

In this chapter we investigate the application of a variety of signal processing tools on the MEG data. In particular, we wish to make use of independent component analysis, or ICA. This is a popular signal processing tool which has been used previously in the analysis of multichannel MEG data. In the context of this project, ICA is used as a method of separating the reconstructed system phase space into independent subspaces. The idea is to isolate the underlying generators of the MEG data so that they may be studied individually. Besides this, methods for clustering and visualising signals are also employed and are discussed in some detail.

Section 4.1 is devoted to independent component analysis, and in particular the FastICA algorithm, which is used extensively throughout this work. ICA is a member of a general class of linear feature extraction tools, which includes other popular algorithms such as PCA and factor analysis, and these are also described briefly. Section 4.2 presents the results of the application of ICA onto MEG data. In addition, we propose a novel use of topographic projection tools as a means of visualising the relationships between extracted ICs. We also show how these topographic maps can be used to assist the process of clustering the extracted independent components into categories. In this work, we impose a subjective classification scheme which mainly corresponds to well known classes of neuro-electric activity, such as Alpha and theta waves. The motivation for this is mainly to allow for easy comparisons to be made with existing EEG and MEG theory. Finally, section 4.3 discusses a number of

practical considerations. Firstly, the results of the ICA extraction are compared with the more conventional multichannel ICA approach. We also compare our results with similar ICA analysis applied to other channels. Finally, we demonstrate a simple noise reduction application on a sample of MEG data, and compare the results with low-pass filtering.

4.1 Independent component analysis

4.1.1 Linear transformations

A common goal in a variety of fields is feature extraction: the task of transforming points and vectors in the data space into a feature space that presents this data in a more efficient or comprehensible manner. As the delay embedding of MEG results in a 90 dimensional dataset, this is clearly an issue of particular relevance to this thesis.

Of the variety of feature selection algorithms that have been developed, the large majority fall into the linear category. This means that they work by projecting data onto *basis vectors* in the data space. The directions assumed by these basis vectors are determined via optimisation algorithms, the nature of which dictate the characteristics of the transform in question. Note that while a given transformation may be linear, the process by which the optimum projection direction is found may utilise highly nonlinear properties of the data, as we shall see in a while. The number of basis vectors used determines the dimensionality of the feature space, and while this can be the same as the data space, it is frequently smaller. Nonlinear methods generalise this approach to projections onto curves or surfaces in the data space. However, the computational and analytical difficulty, as well as the intrinsically ill-posed nature of this problem, means that only a relatively small amount of research effort has been dedicated to the development of nonlinear methods. As these problems will only be intensified by the high dimensionality and poor signal strength of MEG data, we have chosen to restrict the scope of this project primarily to linear transformations.

One popular criterion for finding the optimal projection direction is to search for directions of maximum variance. Two of the most popular linear feature extraction methods, PCA and factor analysis use this principle. The benefit of working with variance is that it allows the data to be characterised in a least-squares optimal sense. i.e.: for a given number of projection directions, second order methods generate representations that capture the maximum signal power. In addition, variance based methods like PCA are generally computationally cheap, and are well understood. However, using only second order properties can cause signals with larger amplitudes to be emphasised over weaker signals. In many cases, "interesting"

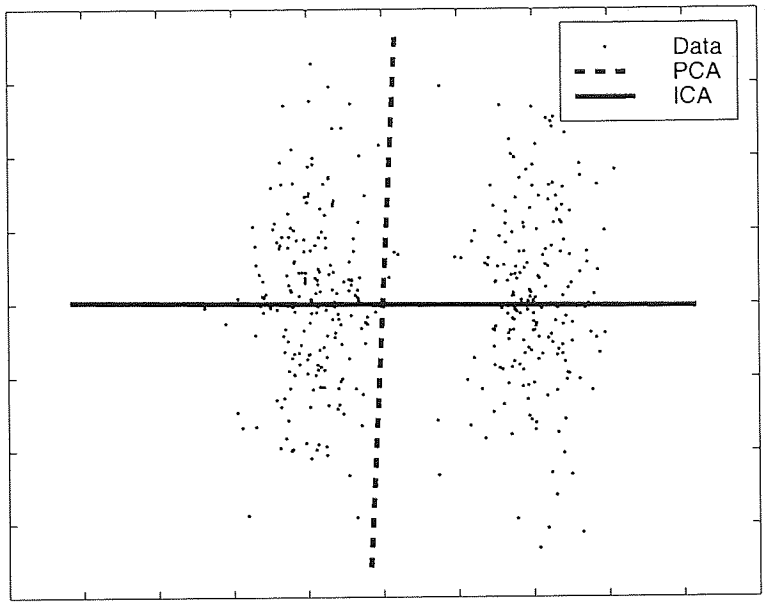


Figure 4.1: Comparison between directions in the data space using PCA and ICA. Only the directions for the first principal axis and first independent component are shown.

structures in the data do not necessarily have large variances. Also, searching for directions of maximum variance does not result in a unique optima, as higher order dependencies in the data remain undetected. PCA solves this problem by imposing an orthogonality constraint on the principal components. However, this restriction is both arbitrary and often inappropriate. Independent component analysis, or ICA[19], provides one solution to this problem.

The distinction between the two approaches is best illustrated by the graphs in figure 4.1. The figure shows data points drawn from a mixture of two gaussian distributions which are elongated along the y-axis, parallel to each other. The two lines in the figure depict the directions of the first principal axis and independent component (labelled accordingly in the figure). As can be seen, the direction of the first principal component is almost parallel with the y-axis, as this has the greater variance. Unfortunately, this is also the less "interesting" direction as it fails to reveal the existence of the two clusters. ICA, on the other hand, correctly detected the direction through the centres of the two gaussians as the first independent component.

The original motivation behind ICA was to solve the so-called "cocktail party" problem. Imagine two people speaking simultaneously in a room, with their combined voices recorded by two spatially separated microphones. The problem then, is to separately extract the two mixed voices using just the microphone recordings. As it turns out, it is possible to solve

this problem using just the statistical properties of the mixed signals (henceforth denoted as $x_1$and x_2) provided that two conditions are met: first, the recordings should be *linear sums* of the original "generators" s_1 and s_2 (i.e. the voices of the individual speakers), and secondly, s_1 and s_2 must be *statistically independent*, defined as:

$$p(s_1).p(s_2) = p(s_1, s_2) \tag{4.1}$$

Of course, this is a difficult condition to meet completely but even in cases when strict independence is not met, ICA is still known to work [32].

Note that, while it is still a linear approach, the ability of ICA algorithms to exploit high-order information in the data grants them an extra level of effectiveness. Since its introduction, interest in ICA has emerged in a wide variety of research areas. Indeed, ICA has already been used previously in EEG [34; 35; 64; 75] and MEG [76], though to a lesser extent in the latter. In these studies, ICA has been applied to multichannel data, where the use of ICA can be theoretically justified via the linearity of equation (2.1), itself a direct consequence of Maxwell's field equations. However, the applicability of ICA in this manner still requires that the spatial distribution of the sources over the surface of the cortex remains stationary, a condition that has not been proven true.

4.1.2 Motivations

Why ICA for MEG analysis? The answer can be expressed from both signal processing and neurophysiological perspectives. If we were to view the MEG problem purely as a signal processing application, the use of ICA helps to achieve the commonly pursued objective of finding a suitable representation of the data. By splitting the MEG data space into a set of independent subspaces, ICA provides a convenient basis set which identifies directions of interest in the data, as determined by the independence criterion. Within this basis set, the MEG signal can be effectively characterised or subjected to further analysis. Additionally, this approach possesses the advantage of having minimal dependency on arbitrary or unproven assumptions about the data.

From a neurophysiological standpoint, there is also a strong case in support of the use of ICA. Recall from the discussion in section 2.4.2 that MEG recordings can be considered as mere observations of the underlying dynamical system. We further postulate that this dynamical system is a combination of a number of spatially distributed and functionally distinct neural assemblies operating semi-independently. Current knowledge supports the hypothesis that cognitive function is the result of the interaction between these generators via a set of global wave dynamics occurring in several well-researched bands. While detailed knowledge on the

dynamics of these local assemblies, and the nature of their interactions, is still preliminary, it seems sensible to assume that their activities would be reasonably independent as the density of local cortical interactions are significantly greater than long range connections [56], suggesting that the local circuits are likely to possess at least moderate degrees of autonomy. From this perspective at least, the independence criterion used in ICA is a suitable tool if attempting to separate signals originating from local neural assemblies.

Furthermore, recall that conclusions of the surrogate tests in chapter 3 indicate the presence of nonlinear activity which is deeply embedded in a background of both linear and stochastic activity. This would certainly support the theory that the nonlinear activity that was detected is the result of these localised circuits. This would explain why these dynamics were so difficult to detect, since signals from such circuits would suffer degradation for two main reasons:

1. A localised circuit is more likely to be physically removed from the measuring site, and hence suffer attenuation proportional to the inverse squared distance.

2. Due to the smaller scale of such circuits, the spatial blurring due to limited resolution is likely to be more severe. This would obscure some of the details of the dynamics.

What is clear is that as long as these nonlinear signals are mixed with the rest of the background time series, performing any kind of analysis on them will be incredibly difficult since they will be extremely weak in comparison to the other signals. What is equally likely is that these dynamics are almost certain to exhibit both linear and nonlinear, as well as stochastic characteristics. Hence, attempts to analyse them using traditional approaches that tend to segregate systems into either purely linear or purely nonlinear processes are likely to be unsatisfactory. ICA might be the tool which allows us to effectively separate signals emanating from these localised circuits from the background activity, enabling us to analyse them in more detail and unveiling their secrets.

4.1.3 Calculation of the independent components (ICs)

For this we need a more rigorous definition of ICA, which is best presented in the context of the statistical latent variable model. Again, we start with the assumption that the data is an observation of a set of underlying variables. This general framework is the basis for a number of signal processing methods, depending on what assumptions one chooses to make regarding the probability distributions of the latent variables. In the case of ICA, the latent variables are assumed to be statistically independent, nongaussian I.I.D. random variables. The nongaussian assumption is necessary because as we will see later, it is not possible to separate a mix of gaussian random variables based solely on their joint static distribution.

In matrix form, this relationship is expressed as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{4.2}$$

Further, as most real measurements are obtained in the presence of noise, it is common to add a noise term, ν:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \nu \tag{4.3}$$

Where $\mathbf{x}$ is the observed data matrix, $\mathbf{A}$ is the mixing matrix and s is the matrix containing the latent variables. Also, note that the time index in equation (4.3) has been omitted. This reflects the fact that ICA uses only the static distribution and the resulting insensitivity towards temporal ordering. The ICA problem is basically to find a rotation matrix $\mathbf{W}$ that reverses this transformation, given only the measured signals $\mathbf{x}$.

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{4.4}$$

Where the optimum point is reached when $\mathbf{y} = \mathbf{s}$. To derive suitable ICA expressions, we can start by expressing the likelihood of the observed data conditioned by the estimated de-mixing matrix. Starting with the generative model in equation (4.3), we proceed by marginalising over the hidden sources:

$$\begin{aligned} P(\mathbf{x}|\mathbf{W}) &= \int P(\mathbf{x}, \mathbf{s}|\mathbf{W}) d\mathbf{s} \\ &= \int P(\mathbf{x}|\mathbf{s}, \mathbf{W}) P(\mathbf{s}) d\mathbf{s} \\ &= \int P(\mathbf{x}|\mathbf{s}, \mathbf{W}) \prod_i p_i(\mathbf{s}_i) d\mathbf{s} \end{aligned} \tag{4.5}$$

The exact form for this equation is of course determined by both the noise model used, as well as the assumed distribution of the sources $p_i(s_i)$. In the special case where the observations are assumed to be noiseless, this reduces to:

$$\begin{aligned} P(\mathbf{x}|\mathbf{W}) &= \int P(\mathbf{x}|\mathbf{s}, \mathbf{W}) \prod_i p_i(\mathbf{s}_i) d\mathbf{s} \\ &= \int \prod_j \delta(\mathbf{x} - \sum_k \tilde{\mathbf{A}}_{jk} \mathbf{s}_k) \prod_i p_i(\mathbf{s}_i) d\mathbf{s} \\ &= \prod_i p_i(y_i) \end{aligned} \tag{4.6}$$

where, $\tilde{\mathbf{A}}$ is the inverse of the estimated de-mixing matrix $\mathbf{W}$, and $y_i = \sum_j \mathbf{W}_{ji} x_i$. Up to this point, this expression holds for almost all ICA algorithms. However, the way in which the likelihood function is then parameterised, and the way in which the optimum $\mathbf{W}$ is obtained, can vary greatly and is the determining factor in the manner in which ICA is performed.

Finally, there exist two fundamental limitations to ICA:

1. It is not possible to determine the variance or power of the extracted sources. This can be seen from equation (4.3) where both $\mathbf{A}$ and $\mathbf{s}$ are unknown. Hence, any scaling of sources in $\mathbf{s}$ can be just as easily cancelled out by a corresponding scaling of $\mathbf{A}$ and is hence indeterminate. A direct result of this is that it is also not possible to determine the sign of extracted components.

2. The ordering of the ICs also cannot be determined. Again, this may be inferred from equation (4.3) where any ordering of the rows of $\mathbf{s}$ can be similarly cancelled out by a reordering of the rows of $\mathbf{A}$. More intuitively, we can reason that the process of mixing the sources to form the observations is devoid of a concept of ordering since it is assumed to be an addition operation, which obeys the commutative law.

Equation (4.5) defines a general expression that should be optimised in order to find a solution to the ICA problem. In practice, the design of ICA algorithms can be approached from a number of directions. Depending on the assumptions that we make regarding the model, a variety of methods can be derived, each based on different contrast functions which can be shown to equate to the independence requirement. A few of the more notable prominent approaches are:

1. **Maximum likelihood** - ICA can be performed by directly optimising the likelihood $P(\mathbf{x}|\mathbf{W})$ in equation 4.5. For this to be tractable, integrable non-gaussian approximations to the source density have to be assumed. A number of different algorithms have emerged which achieve this in different ways.

2. **Infomax** - One of the earliest approaches to ICA. The infomax principle tries to maximise the information transfer from the input to output of an MLP network. An optima is achieved when the outputs of the network are distributed uniformly, at which point the weights of the linear section of the network give the ICA separating matrix $\mathbf{W}$.

3. **Mutual information** - The mutual information between a set of scalar variables is an information theoretic quantity which quantifies the amount of redundancy within these variables. Because the mutual information between two independent variables is zero, ICA can also be performed by minimising the mutual information between the separated components.

4. **Higher order cumulants** - Another way by which ICA may be performed is via the maximisation of the higher order cumulants of the outputs of the ICA network.

A more detailed discussion of each of these approaches is included in appendix C.1. In the following section we discuss the FastICA algorithm in greater detail as we will be using this extensively for our subsequent MEG analysis.

4.1.4 FastICA

First proposed in [31], FastICA is a method for performing ICA which searches for extrema in the fourth order cumulant or kurtosis of the extracted outputs. A brief description will now be given:

For a zero mean random variable, kurtosis is defined as:

$$kurt(v) = E\{v^4\} - 3(E\{v^2\}^2) \tag{4.7}$$

The kurtosis provides a measure of how "peaked" a distribution is relative to a gaussian; i.e. the kurtosis of a gaussian distribution is zero; for densities peaked at zero, kurtosis is positive, and finally, the kurtosis is negative for densities that are flatter than gaussians.

The general goal is to find the optimal rotation of the input variables which maximises the output kurtoses, subject to the requirement that the norm of the weights matrix be constrained to one (because otherwise it is always possible to get arbitrarily high kurtoses). While the above aims are common to many ICA algorithms, the key to the superior performance of the FastICA algorithm is the use of a relatively simple yet highly efficient fixed point iteration scheme for finding the local extrema of the kurtosis function. The algorithm can also be used to extract the independent components in parallel, or sequentially. Because we use the sequential procedure for the MEG data, only this will now be described:

1. Start by sphering the input variable $\mathbf{x}$, such that the covariance becomes the identity matrix. Initialise the separating weights $\mathbf{w}(\mathbf{0})$ by sampling from a zero mean gaussian distribution then normalising to unit norm. Set $k = 1$.

2. From (4.7), the kurtosis of the output is

$$kurt(\mathbf{w}^T\mathbf{x}) = E\{(\mathbf{w}^T\mathbf{x})^4\} - 3||\mathbf{w}||^4 \tag{4.8}$$
(since $\mathbf{x}$ has been pre-whitened)

   We obtain a cost function by writing the kurtosis expression as a function of $\mathbf{w}$, $\mathbf{J}(\mathbf{w})$, and adding a penalty function to enforce the unit norm constraint on the weights:

$$\mathbf{J}(\mathbf{w}) = E\{(\mathbf{w}^T\mathbf{x})^4\} - 3||\mathbf{w}||^4 + F(||\mathbf{w}||^2) \tag{4.9}$$

3. The gradient of this function with respect to $\mathbf{w}$ is given by:

$$\frac{\partial \mathbf{J}(\mathbf{w})}{\partial \mathbf{w}} = 4 \times \left[E\{\mathbf{x}(\mathbf{w}^T\mathbf{x})^3\} - 3||\mathbf{w}||^2\mathbf{w}\right] + 2f(||\mathbf{w}||^2)\mathbf{w}] \tag{4.10}$$

   where f is the derivative of F.

4. The fixed points of the cost function are obtained by equating the gradient term to zero and moving the third expression on the right over, giving:

$$\mathbf{w}(k+1) = scalar \times \left[E\{\mathbf{x}(\mathbf{w}(k)^T\mathbf{x})^3\} - 3\mathbf{w}(k)\right] \tag{4.11}$$

   Where the expectation on the right hand side is estimated by taking the sample average (and using a large sample size).

5. The scalar term can be removed since what is important is the direction of the weights vector. The constraint that the weights matrix must have unit norm is then enforced by dividing $\mathbf{w(k+1)}$ by it's norm.

6. If $|\mathbf{w(k)^T w(k-1)}|$ is not close enough to 1, let $k = k + 1$ and repeat steps 4 and 5. Otherwise, output the vector $\mathbf{w(k)}$.

The final vector $\mathbf{w(k+1)}$ given by the algorithm equals one of the columns of the separating matrix. To separate the next component, the procedure is started again after re-initialising the weights vector $\mathbf{w}$. We can ensure that the new algorithm does not find the same direction by projecting the current solution onto the space orthogonal to the space defined by the previously found weight vectors, $\mathbf{W} = [\mathbf{w}_1\mathbf{w}_2 \ldots \mathbf{w}_i]$. This is achieved by applying the following projection before step 5 is performed:

$$\mathbf{w}(k) = \mathbf{w}(k) - \mathbf{W}\mathbf{W}^T\mathbf{w}(k)$$

The fixed-point algorithm used by FastICA possess a number of advantages [31]:

- Convergence of the algorithm can be shown to be cubic, which is superior to most other ICA algorithms.
- Because it does not use gradient descent based learning, there is no need to tune a learning rate parameter.
- It is able to detect and extract independent components with both positive and negative kurtoses.
- It allows the extraction of components sequentially, instead of in parallel like most other algorithms. This makes working on larger datasets easier.

In our test, we observed that FastICA generally performed faster and was more stable than a variety of other ICA algorithms. Due to this superior performance, ease of implementation, as well as the other reasons outlined above, the analyses undertaken in the following sections were performed using this algorithm. Other ICA algorithms that were considered are:

1. JADE
2. Infomax
3. Natural gradient optimisation
4. RICA

In general, only the Infomax algorithm produced similar performance (JADE also produced similar results but was much slower than the other two methods, whereas the natural gradient optimisation method was unable to separate recognisable sources). The ICs extracted using Infomax proved to be almost identical to what we obtained using FastICA, and some examples are provided in appendix D as a comparison.

4.2 Results and discussions

4.2.1 Methods

Embedding was performed on single channel MEG in the manner described in chapter 3. The embedding dimension, as determined via methods discussed in section 3.1.2 was set to 90 dimensions with the value of τ set to 1 (as before, to ensure stationarity, we extract an eight second segment from the channel prior to embedding). Finally, the embedding matrix is projected onto the principal components to decorrelate the elements. We can exploit this additional step to help remove some of the noise from the data. From the singular spectrum of the data (shown in figure 4.2) it is clear that the spectrum has reached the noise floor after 40 components, so we truncate the principal component reconstruction after the 40^{th} component. FastICA is then performed in serial mode to sequentially extract ICs from the embedding matrices. The ICA extraction was performed for a maximum of 1000 iterations, or when the norm of the weights update term is less than 0.0001. The data was pre-whitened prior to component extraction and the ICA weights were initialised by sampling from a zero mean Gaussian.

After convergence of the algorithm, the final step is to project the extracted ICs back into the measurement space, using the component-wise mixing expression:

$$\mathbf{Y}^i = \mathbf{a}_i \tilde{\mathbf{s}}_i \tag{4.12}$$

Where $\mathbf{a}_i$ is the i^{th} column of the estimated mixing matrix $\mathbf{A}$ and $\tilde{\mathbf{s}}_i$ is the i_th IC. Hence, $\mathbf{Y}_i$ gives the contributions of the i^{th} IC to the embedding matrix to which the ICA was applied. As the embedding matrix is simply composed of time-shifted versions of the original time

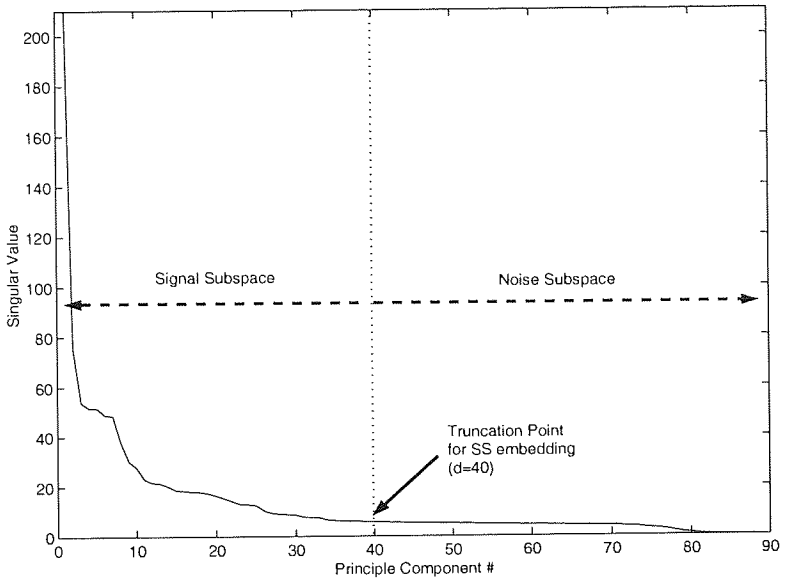


Figure 4.2: Singular Spectrum of MEG embedding matrix with $m = 90$, showing the truncation point

series, we can obtain an estimate of the component of the original time series which is due to IC i by averaging over the rows of this reconstructed embedding matrix, after removal of the time delays. This step is akin to the "Hankelization" procedure which is routinely employed in singular spectrum analysis (SSA) [25] and has previously been applied to EEG analysis in the context of ICA [37].

4.2.2 Component clustering

Once the ICs have been extracted the question that arises is, how do we select components of interest? The application of ICA on the MEG embedding matrix generates a large amount of data, and analysing this presents a huge problem. In addition, while in theory ICA only extracts components which are independent of one another, the reality is that the actual number of components (which is generally close to the dimensionality of the embedding matrix) will often greatly exceed the number of possible underlying generators. While some components will undoubtedly be due to noise or artifacts, others will have originated from the same source but be phase shifted with regard to each other. Yet other components will be mixtures of two or more sources which have not been cleanly separated.

What is required is a way of sorting through the independent components and clustering them into categories. Ideally, this will allow us to only work with individual examples from

each class. If possible, we would like to be able to extract some quantity from each of the ICs which would characterise the dynamics of each component in some way. The variance of each independent component, for example, can be used as a gross descriptor with which ICs may be compared. However, this is obviously a poor choice for a number of reasons. It is very probable that two completely different signals could share a similar variance. Conversely, two signals sharing a common generator might possess very different variances as one of the signals might have been attenuated prior to measurement. Higher order measures such as the kurtosis might be more selective, however the estimation of high-order moments are notoriously prone to noise and outliers and are unlikely to give satisfactory results.

Our approach was to use the frequency spectra of the independent components. While strictly speaking, the frequency spectrum of a time series characterises only linear dynamics, this does not imply that we are restricting our analysis to linear phenomena. Since our aim is to use the entire power spectrum and not just dominant peaks, the only assumption that we make is that components with a common generator share a similar characteristic spectra. It is known that even simple nonlinear processes may possess complex, broadband or multimodal power spectra. Note, because the use of linear ICA on the embedding matrix means that we are only able extract signals which can be differentiated via their spectra in any case, this is not an unreasonable assumption.

How can we cluster the power spectra of the ICs? The most simple approach is simply by visually inspecting the spectra and noting similarities. These can then be corroborated with the morphologies of the signals in the time domain to form a more reliable classification of the IC. However, this process will be slow, and is really effective only with a relatively small number of ICs. Hence, what is needed is a way of generating a low dimensional representation which will allow the visualisation of similarities or "distances" between the FFTs of the independent components. A number of topographic projection tools have been proposed which might be able to address this need. We have considered two such methods in particular, the Sammon map and the Kohonen self-organising map.

Sammon Mapping

Proposed in [65], the Sammon mapping is a visualisation technique used to generate low dimensional representations of high dimensional data. This is achieved by mapping data-vectors from a high dimensional space into a low dimensional visualisation space based on the requirement that topographical relationships within the data are preserved.

This requirement can be enforced by minimising the following cost function:

$$E_{ss} = \frac{\sum_i \sum_{j<i} \left[d^*_{ij} - d_{ij}\right]^2}{\sum_i \sum_{j<i} d_{ij}} \tag{4.13}$$

Where d^*_{ij} and d_{ij} are the interpoint distances in the data and feature spaces respectively. E_{ss} is known as the *Sammon Stress*. Expression 4.13 may then be optimized using any standard nonlinear optimization tool (for our work we used the scaled conjugate gradients algorithm (SCG), which is described in appendix C). For further details regarding Sammon Maps and topographic mapping techniques, refer to [65; 53] and to [10] for details of the SCG algorithm.

All that is needed to generate a Sammon map is some measure of the dissimilarity (distance) between data points, or in our case, independent components. For this, we can simply use the Euclidean distance between the FFTs. The FFT of each component was calculated and then subsampled by a factor of ten to reduce noise, before being normalised to unit variance. The Euclidean distance between each component is calculated by simply subtracting the FFT values at corresponding frequencies. The Sammon mapping algorithm was then used to generate two dimensional representations of the ICs, which can be used to visualise the relationships between the extracted ICs.

In practice, the Sammon map representation of the ICs performs two useful functions:

1. Used as a guide, in conjunction with a subjective analysis of the morphologies of the extracted ICs, the Sammon map allows the reliable classification of ICs into classes or categories representing common generators.

2. Provides a qualitative summary of the extracted ICs. A brief glance of the Sammon map of an IC extraction gives a rough idea of the number of both extracted ICS as well an intuitive feel for the complexity of the extracted IC.

In terms of previous work, the only attempt to categorise ICs in a similar manner was proposed in [41]. In this work, multichannel ICA was used to separate evoked potential from unaveraged EEG data. To cluster the ICs, an inter-component distance matrix was generated in a similar fashion but used both the frequency spectra as well as spatial distribution maps (obtained from columns of the ICA mixing matrix) to obtain a distance measure. Further, the clustering was achieved via an automated system which involved recursively searching for nearest neighbours for each of the ICs. However, parameters such as number of members in each class, as well as number of classes, all needed to be set via trial and error. Given the huge variability which we have observed in these parameters, it is felt that the use of the Sammon Map provides a more reliable approach to component clustering. The cost is that manual inspection of the components is still required.

Kohonen Self-Organising Map

An alternative topographic projection technique which was considered is the self-organising map (SOM). Proposed in [45] by Teuvo Kohonen, the principal goal of the SOM is to transform an incoming signal pattern of arbitrary dimension into a one- or two-dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion. The SOM is composed of a two dimensional lattice of neurons, each of which are fully connected to the inputs of the network.

The "connections" from the inputs to the neurons are implemented as model vectors, defining a mapping where each neuron is assigned to an image in the data space. The objective when training the SOM is to organise these connections such that the final distribution of these neurons effectively approximates the distribution of the data itself. Two aspects of the SOM training algorithm help ensure that this happens:

- **Competition.** When an input pattern is presented to the network, the neurons "compete" for activation on the basis of a discriminant function. This is generally some measure of similarity between the neuron and the input pattern presented. The neuron that is most similar is deemed to have "won".
- **Cooperative adaptation.** The winning neuron is allowed to change its synaptic weights, thereby increasing its resemblance to the input pattern. However, neighbouring neurons are also allowed to adjust their weights, though normally to a lesser degree. This process helps to form topological relationships between the neurons in the lattice.

In brief, the SOM training algorithm is as follows [30]:

1. *Initialisation.* The initial weight vectors $\mathbf{w}_j(0)$ are set to random initial values. Alternatively, the weights vectors could be set to a random subset of the data set, $(\mathbf{x}_i)_{i=1..j}$. It is also desirable that the weights vectors not be initialised to overly large values.

2. *Sampling.* A sample vector $\mathbf{x}$ is drawn randomly from the input data and is presented to the neuron lattice. The neuron that best matches this vector is identified:

$$i(\mathbf{x}) = arg\ \min_j ||\mathbf{x}(n) - \mathbf{w}_j||, \quad j = 1, 2, \ldots, l$$

3. *Updating.* The synaptic weight vectors of all the neurons are updated using the update formula:

$$\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \nu(n) h_{j,i(\mathbf{x})}(n)(\mathbf{x}(n) - \mathbf{w}_j(n))$$

where $\nu(n)$ is the learning rate parameter, and $h_{j,i(\mathbf{x})}$ is called the *neighbourhood function.* This is normally a radial function such as a gaussian, and is centred around the winning neuron $i(\mathbf{x})$.

4. *Parameter adjustments.* In this step, both $\nu(n)$ and $\sigma(n)$, the radius of the neighbourhood function, are modified such that they gradually decay over the training period.

5. *Continuation.* Repeat from step 2 until no further change is noticeable.

In this project we have only used maps with either 25 or 36 neurons. The neurons have been arranged in a hexagonal lattice rather than a square lattice as this promotes a better symmetry during training (a neuron in a square lattice will discriminate between adjacent and diagonal neighbours, which might result in a distorted map. As was recommended in [30], we perform the training in two stages. The first stage, called the *ordering* stage, is meant to "untangle" the neurons from their initial random values so that they assume topologically correct orderings. As such, we set the learning parameter ν to 0.1, and the neighbourhood function to include almost all the neurons in the network (centred, however, on the winning neuron). In the second, or *convergence* stage of the training, we start with a learning rate ν of 0.01 and restrict the neighbourhood to only the winning neuron and its immediate neighbours. Finally, for the neighbourhood function we use a Gaussian simply because it is smooth and easily calculable.

Brief, preliminary experimentation with the two techniques established that a practical procedure would be to first cluster the ICs using Sammon Mapping, as this is a continuous mapping which will give us a better representation of distances and relationships between components and also to identify the presence of clusters. During this step we classify the ICs into categories and label them accordingly. The ICs are then trained using the SOM, also using the frequency spectra as the input space. We can then use the trained SOMs to verify the class labels that were previously assigned.

4.2.3 Component extraction and clustering

ICA decomposition of synthesised data

To validate the effectiveness of the ICA algorithm in extracting sources from the embedding matrices, it was first tested on a set of artificial MEG data. To generate the artificial data, the following three-stage procedure was performed:

1. Generate the underlying "sources" of the MEG. Three sources were generated:

   - **Heartbeat**: Real ECG measurements were obtained from the *Physionet* database [24]. These were sampled at 100Hz.
   - **Theta waves**: Fake Theta waves were obtained by generating sinusoidal waves at 5Hz.

- **Alpha waves**: Alpha waves were generated using 10Hz sinusoids.

For each set, 2000 samples (20 seconds) of data were generated and the first three seconds from each of the three sources are given in figure 4.3.

2. Obtain a MEG noise signal by performing the signal-noise subspace decomposition as illustrated in figure 4.2. However, this time we discard the *signal* subspace, leaving only the noise components.

3. The three sources are added together to form the signal, which is then normalised to unit variance. The noise signal is also normalised to unit variance. The signal and noise are then added together in varying ratios to obtain noisy "MEG recordings" at different SNR levels (the levels of noise added to the signal are comparable to noise levels found in real MEG signals [77]). The final (noisy) time series is given in figure 4.4.

The choice of the three sources was motivated by two main factors:

- All three of these signals are commonly found in MEG recordings. In particular, ECG contamination can be a serious problem in MEG signal processing.

- The three signals are fairly close together in frequency. Furthermore, the ECG signal is extremely broadband (refer to figure 4.26 which appears later in this chapter) and has a frequency spectra which envelopes both of the other two sources. Hence, the separation of these signals is a challenging task, and would be difficult to achieve using normal lowpass or bandpass filters.

FastICA was applied to the embedding matrices generated from the test data corrupted with a variety of noise levels. In terms of noise to signal power ratios, the levels of noise were varied from 1:10 to 2:1. Note that this refers to the noise power relative to the *combined* power of the MEG sources, hence the ratio of noise power to the power of any of the individual sources is greater than this. The ICA decomposition was successful in extracting the three artificial MEG sources up to about 3:2 noise/signal ratio (as the noise was increased beyond this, it was still possible to extract the signals but the quality of the decomposition steadily deteriorated). Samples of the ICs that we obtained are shown in figure 4.5. As can be seen, the quality of the decomposition is surprisingly consistent, and the rising noise levels do not appear to affect the general appearance and structure of the extracted signal. Most importantly, this demonstrates that the combination of ICA and dynamical embedding is capable of decomposing the scalar time series into its underlying sources.

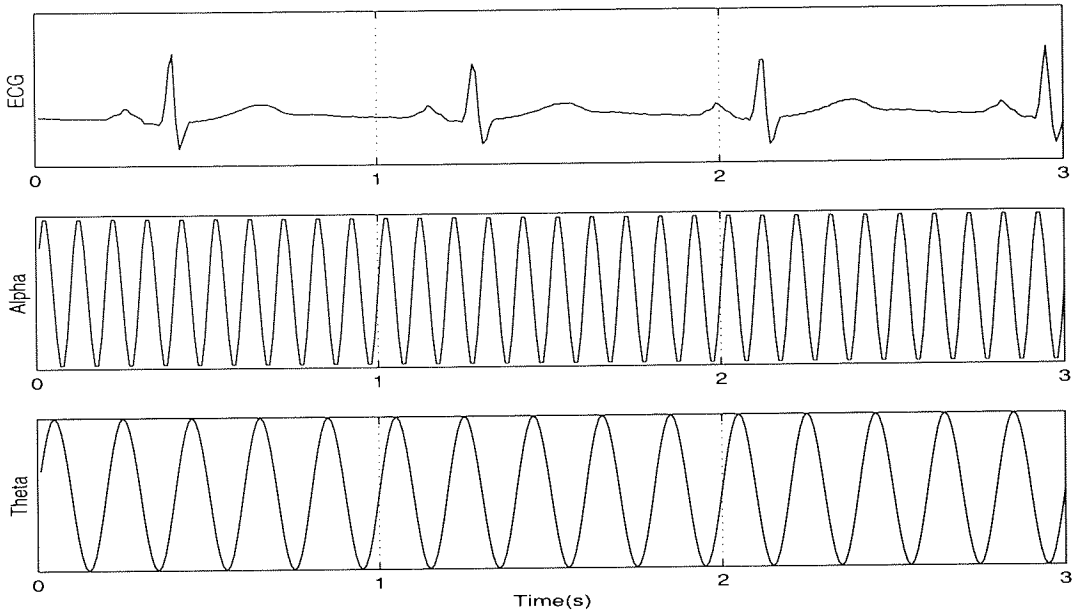


Figure 4.3: Artificial MEG sources.

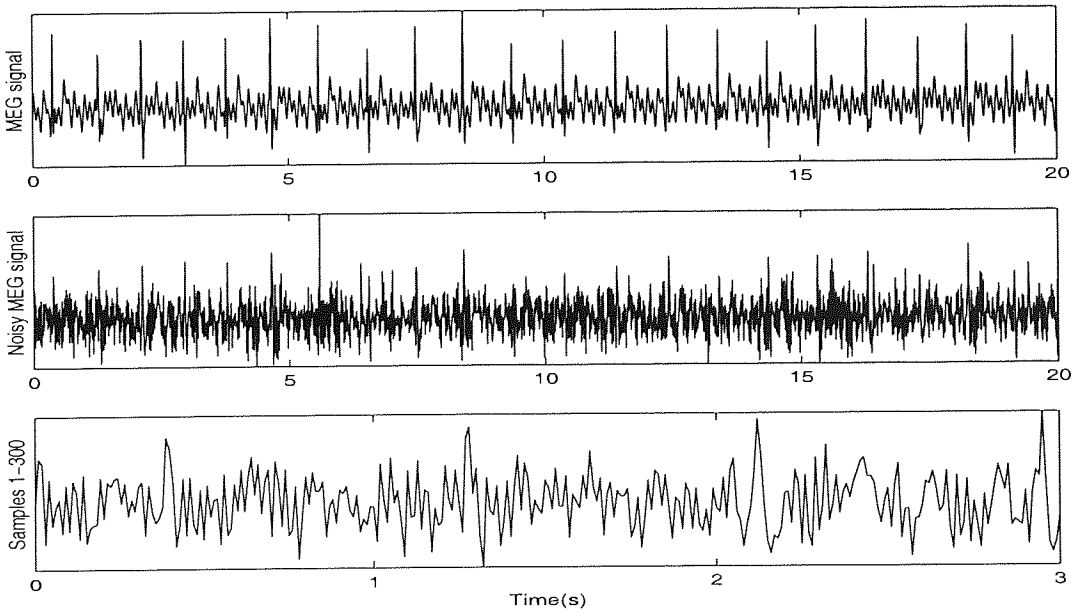
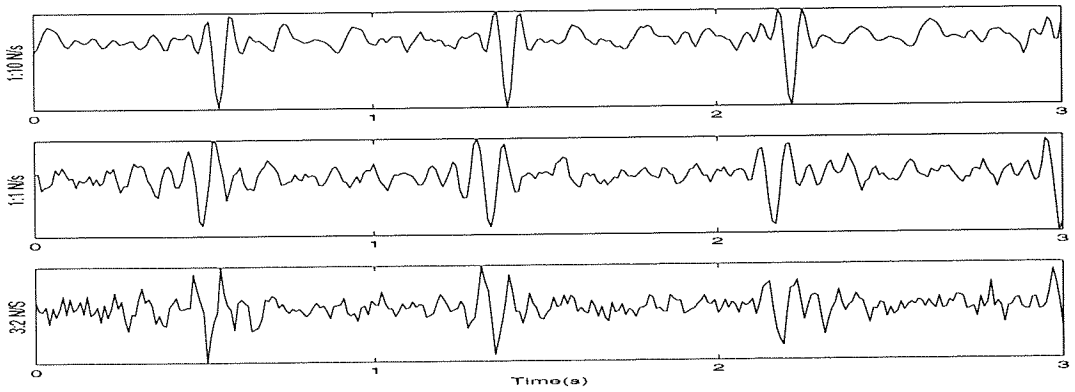
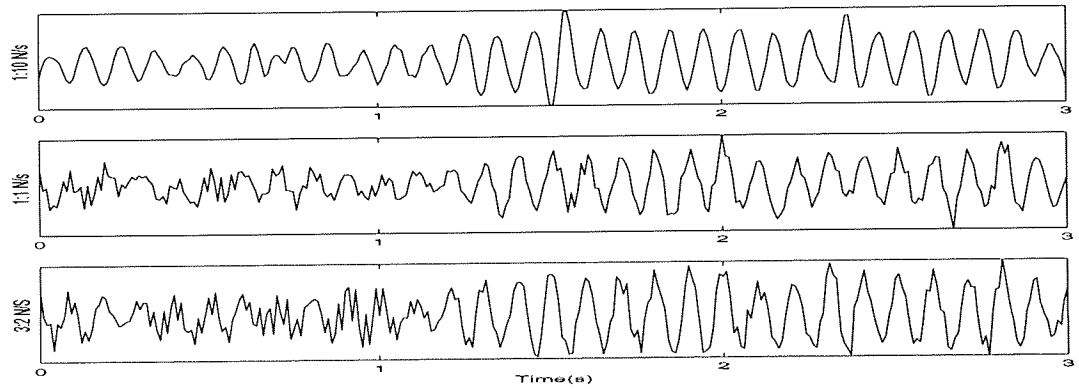


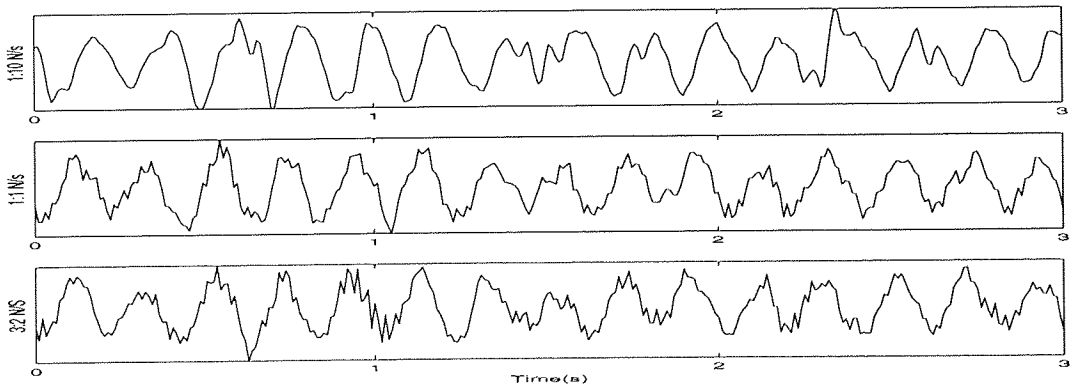
Figure 4.4: Artificial MEG time series generated by mixing the MEG "sources" (shown in figure 4.3). From the top: Noiseless time series, time series after corruption with 150% MEG noise (-1.8DB SNR) and time steps 1 to 300 of the noisy signal.



(a) ECG



(b) Alpha (10Hz)



(c) Theta (5Hz)

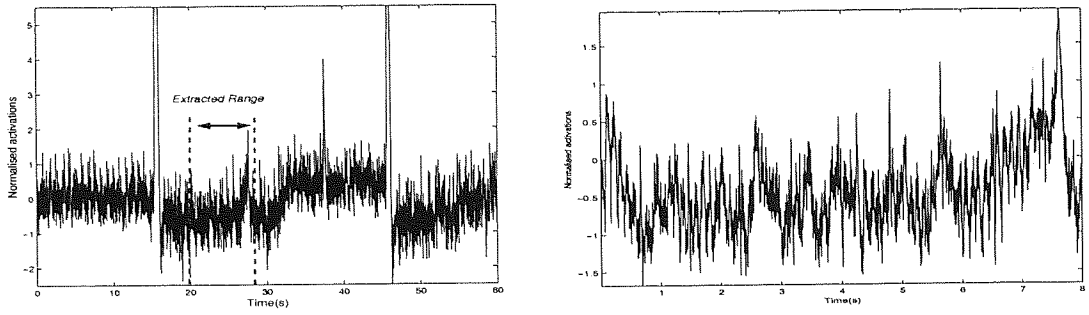
Figure 4.5: Examples of ICs extracted from the synthesised data. Three examples have been included for each source. These are for 1:10, 1:1 and 3:2 ratios of noise to signal power respectively. The figures only show the first 300 samples from each IC to allow easier inspection.

ICA decomposition of real data

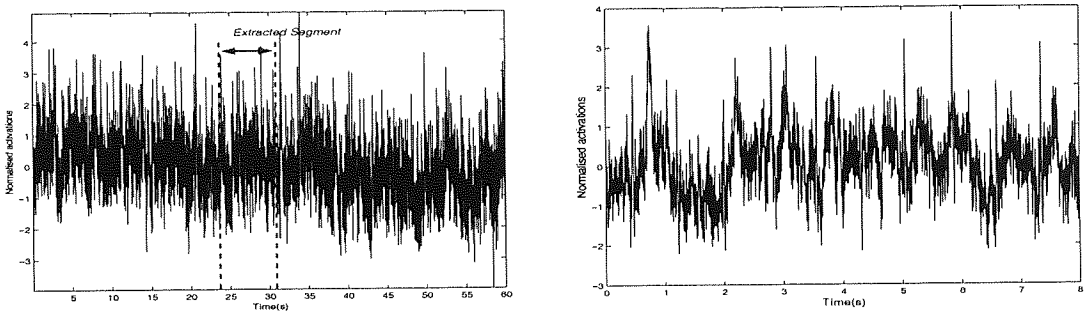
For the ICA analysis, we choose to analyse selected MEG recordings from subject A. For consistency, we will again use recordings from the right temporal site MRT22. Raw MEG recordings for each of the experiments are shown in figure 4.6. A segment from each of the selected recordings is then extracted from the data sets and a 90 dimensional embedding matrix is created as described in section 4.2.1. ICA is applied and a total of 35 ICs were extracted from the data before convergence of the ICA algorithm. As can be seen, even a single segment of MEG can generate a significant quantity of information.

To help visualise the ICs that have been generated, we use the clustering approach discussed in section 4.2.2. PCA is used to initialise the Sammon Maps and these are then trained using the scaled conjugate gradients algorithm for 100 epochs each, after which the Sammon stress had clearly converged. Initialisation with PCA helps the training to converge faster but as always there is a danger of being trapped in a local minimum. To guard against this, each Sammon map was trained with 10 different random initialisations. Figure 4.7 shows the convergence of the error function, and compares this with the convergence when the map is trained with random initialisations. It can be seen that the convergence of the algorithm when initialised using PCA is extremely rapid and constantly achieves a low final error. Figure 4.8(b) shows the resulting Sammon map while figure 4.8(a) shows samples from each of the classes which are labelled in the Sammon map.

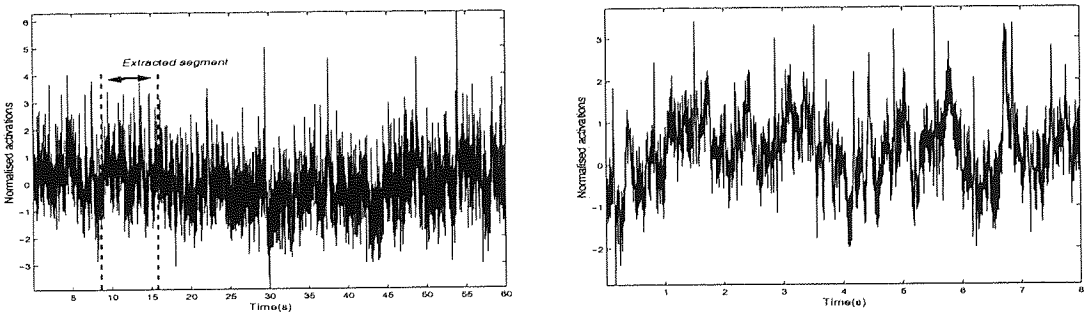
Finally, the frequency spectra of the independent components are used to generate kohonen maps. To visualise these results, we associate each of the lattice points of the Kohonen map with one of the IC classes. This is done by first finding a best matching IC for each neuron, i.e. the IC with the frequency spectra that best matches the weights of the neuron, in a Euclidean sense. This neuron is then labelled with the class label which was determined earlier with the help of the Sammon map. This is displayed in a grayscale shaded map and is given in figure 4.9. Also, the time-domain signal of the respective ICs have also been plotted in their respective positions on the Kohonen lattice and these are also given in the figure.



(a) Idle task

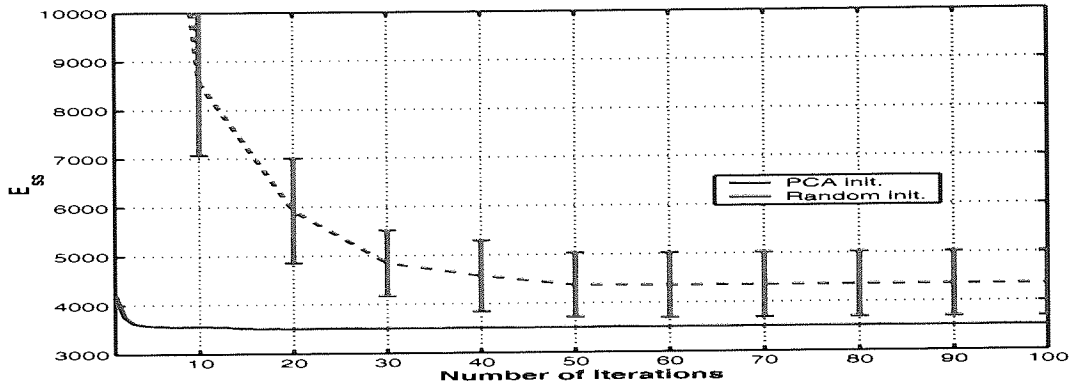


(b) Face recognition

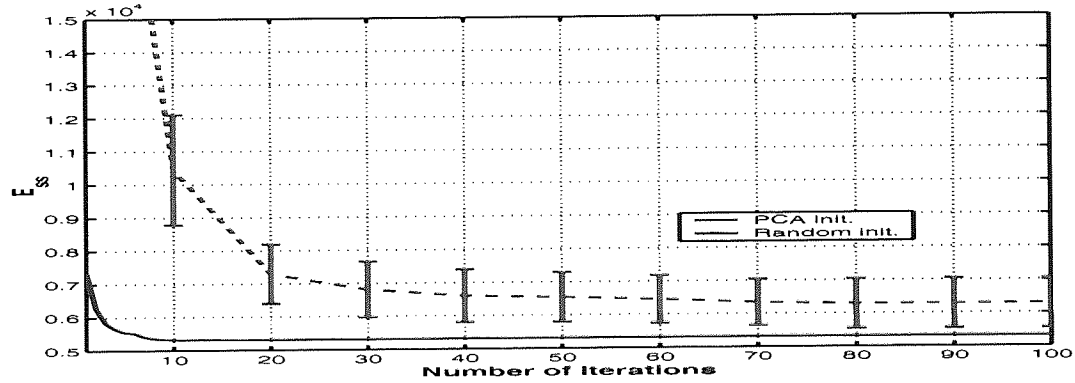


(c) Natural scene recognition

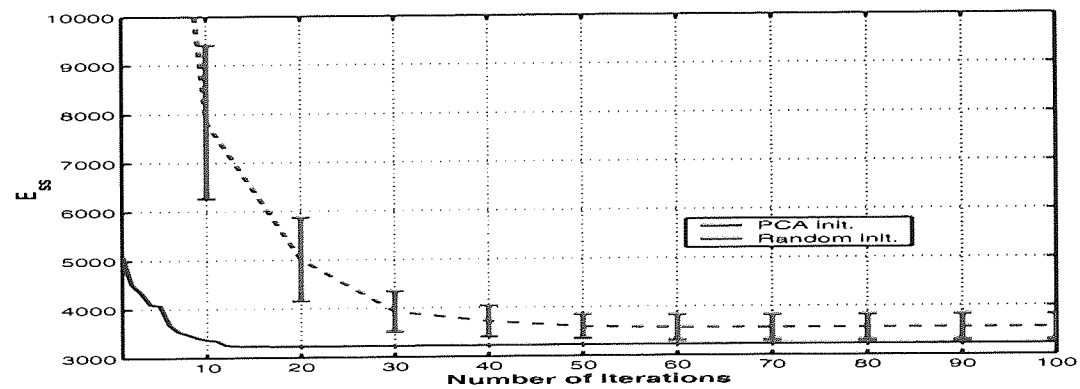
Figure 4.6: Figures showing raw MEG data from selected recordings in each of the three experimental paradigms. The figures on the left show recordings of the whole 1-minute session, whereas the figures on the right show the segments that have been selected for ICA decomposition. The corresponding segments from the original time series have been marked accordingly.



(a) Idle



(b) Face recognition



(c) Natural scene recognition

Figure 4.7: Convergence of Sammon stress function. The solid blue line is with initialisation using PCA, whilst the red dotted line shows the convergence when training is initialised using random initial values.

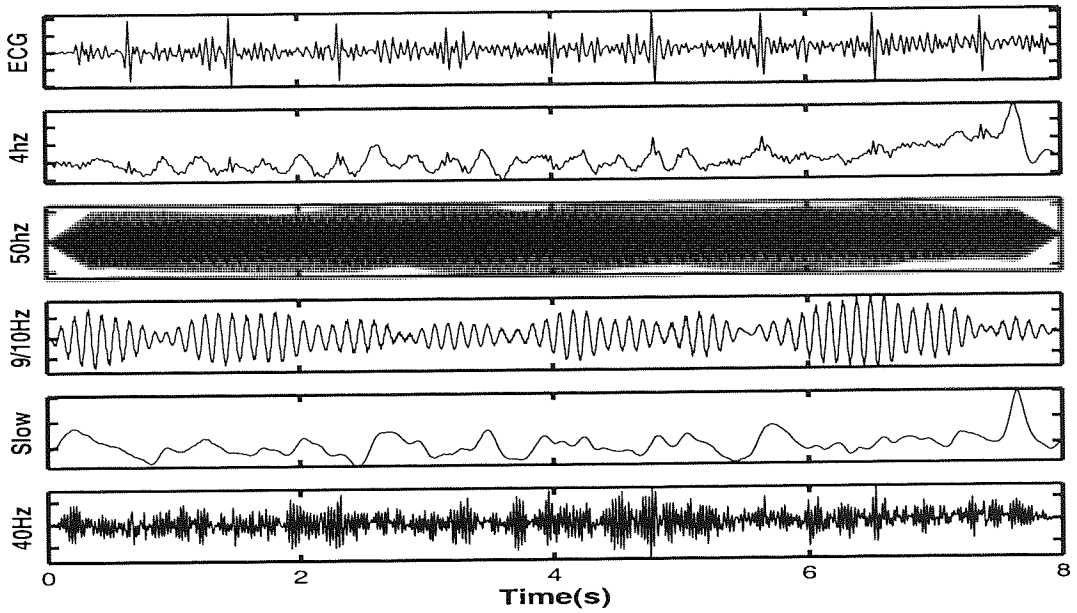
We repeat the procedure with the face and image recognition datasets. The segments are again selected from channels in the temporal region and are from recordings taken of subject A. Results for the facial recognition task are shown in figures 4.10 (samples of extracted ICs) and 4.10 (Sammon visualisation of the components). For the natural scene recognition task, the results are shown in figures 4.12 and 4.12 respectively. A full printout of the all the ICs that were generated in all three of these experiments is given in appendix D. Sample ICs as well as the corresponding Sammon maps extracted from recordings taken of subject B are also provided in appendix D, where it can be seen that broadly similar components were extracted.

Discussion

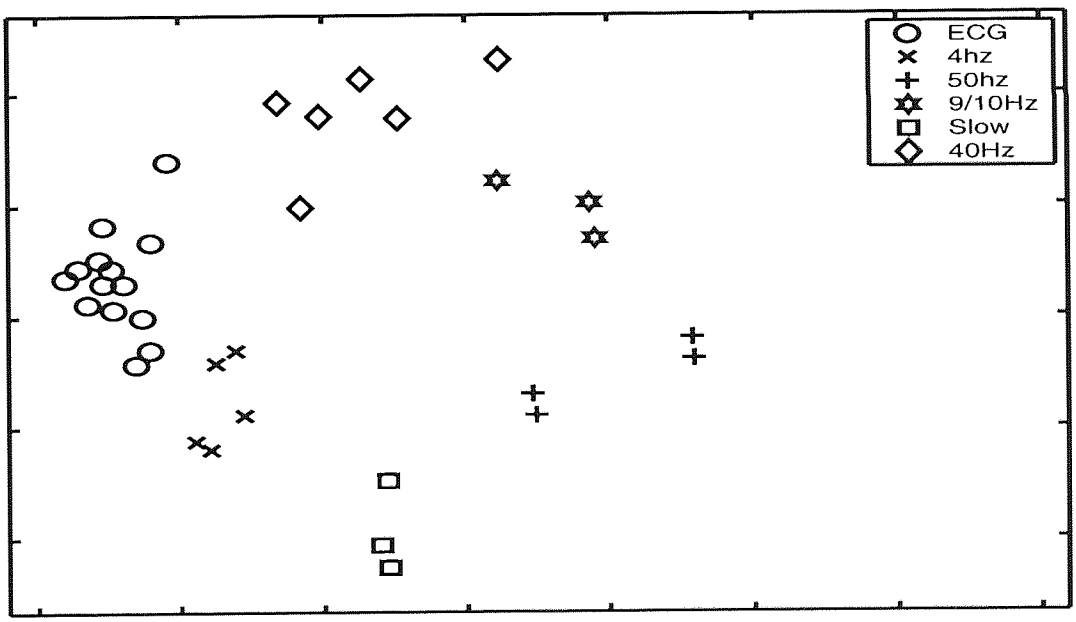
In general, ICA was successful in extracting a number of interesting components from each of the three time series analysed. As the morphologies of even similar components sometimes contained significant variation, or noise contamination, the projection of the components onto the Sammon visualisation space provided a valuable tool for visualising the relationships and dissimilarities between components, and was a great help in the classification of components into categories. Once the components had been labelled, the class labels of the ICs were used to demarcate lattice maps generated using Kohonen mapping, as a means of confirming the validity of the Sammon mapping. Except for a few exceptions, Kohonen maps confirmed the topological correctness of the Sammon space visualisations.

The question now is, what is a suitable scheme for classifying the ICs? As has been stated previously, we would like to avoid the conventional frequency based description of MEG components. Certainly, a brief inspection of the frequency spectra of some of the ICs that were extracted indicates that while certain frequency ranges may dominate, the spectra were almost invariably multimodal and broadband (despite this fact, traditional methods work only with the dominant peaks and classify signals based on the presence or absence of certainly frequently observed frequencies). Nevertheless, a classification scheme based on established classes of neuro-electric signals is still useful as it would allow meaningful comparisons to be made with existing theory. Hence, bearing in mind that the classification of these components is a subjective process, and that it was necessary to "pigeon hole" ICs which sometimes appeared to possess different characteristics, the observed IC classes were:

1. ECG (heartbeat)
2. Alpha waves (9Hz)
3. 4Hz
4. "Dynamic" 7Hz activity
5. Gamma activity (25Hz and above)

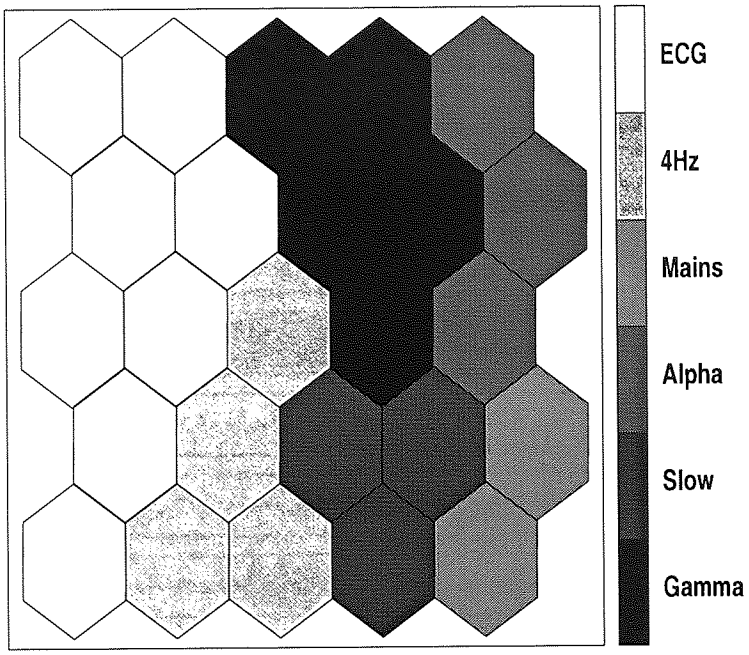


(a) Selected ICs: Idle MEG recordings

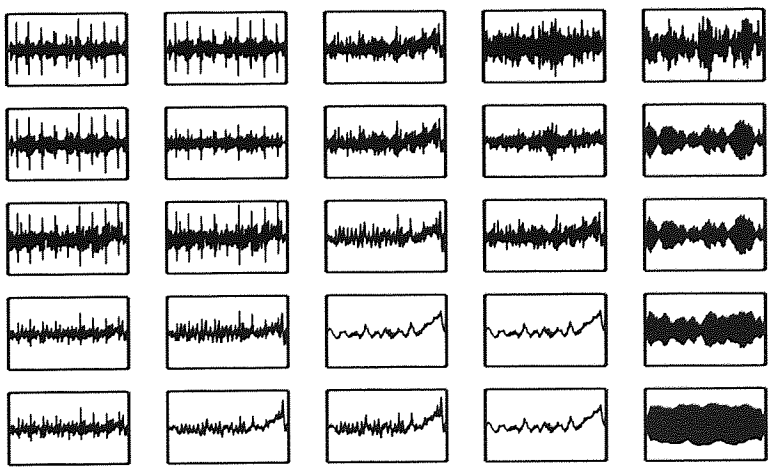


(b) Sammon Map representation

Figure 4.8: Result of blind source separation on MEG recording of subject A performing the idle task. Note the structure of the Sammon mapping, where there is a central cluster of lower-frequency ICs such as alpha and theta activity. This structure is consistently observed in the Sammon maps of the other channels.

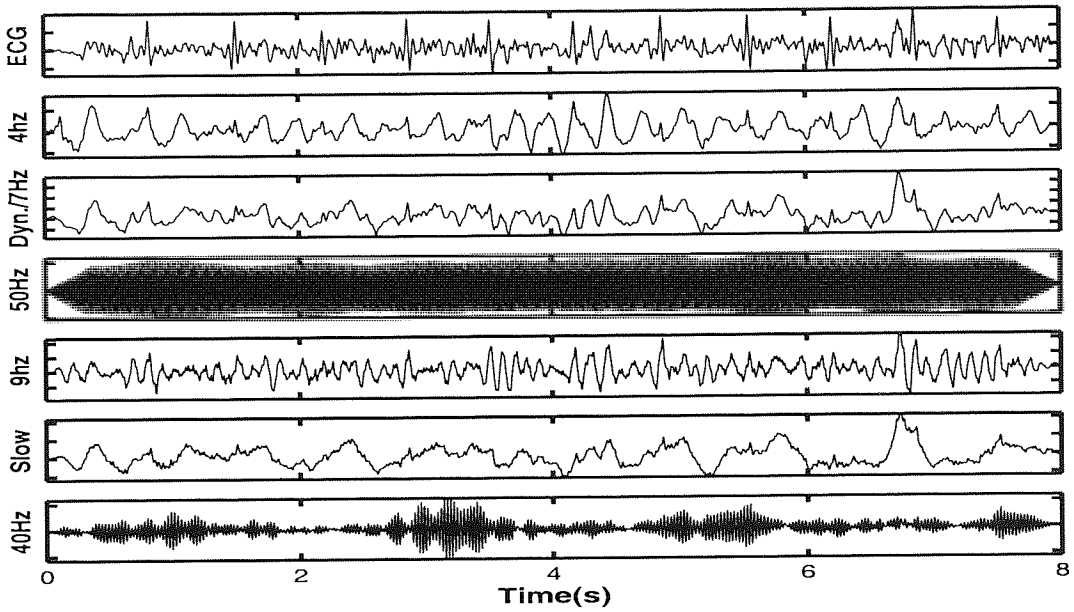


(a) Grayscale-shaded map showing the class assignments of each of the neurons. Notice that the IC clusterings have been preserved, but that the relative size of the clusters have changed. This time cluster size is evidently proportional to *number of ICs* and not to the relative volume of the feature space occupied

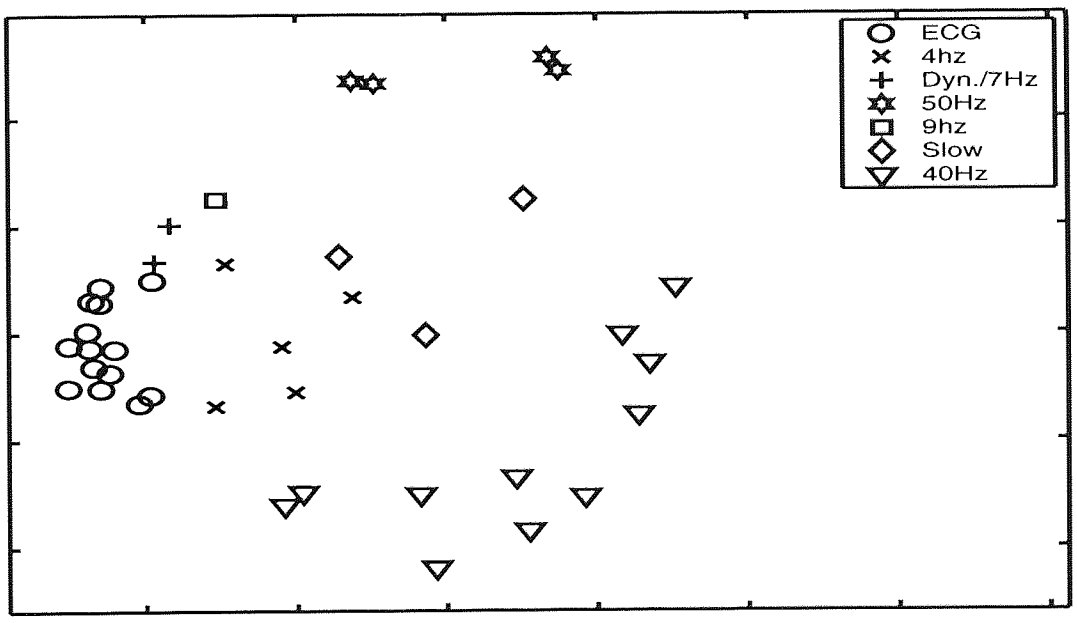


(b) ICs corresponding to the SOM lattice neurons. This figure clearly shows the fact that the ICs are not always cleanly separated into clear-cut classes. Rather, due to noise and signal attenuation, certain ICs often contain combinations from different components

Figure 4.9: Visualisation of the Kohonen SOMs trained using the idle task IC spectra. Note from the SOM representation in (b), that components classes are not always clearly defined, but rather can be seen to vary gradually across a range of morphologies.

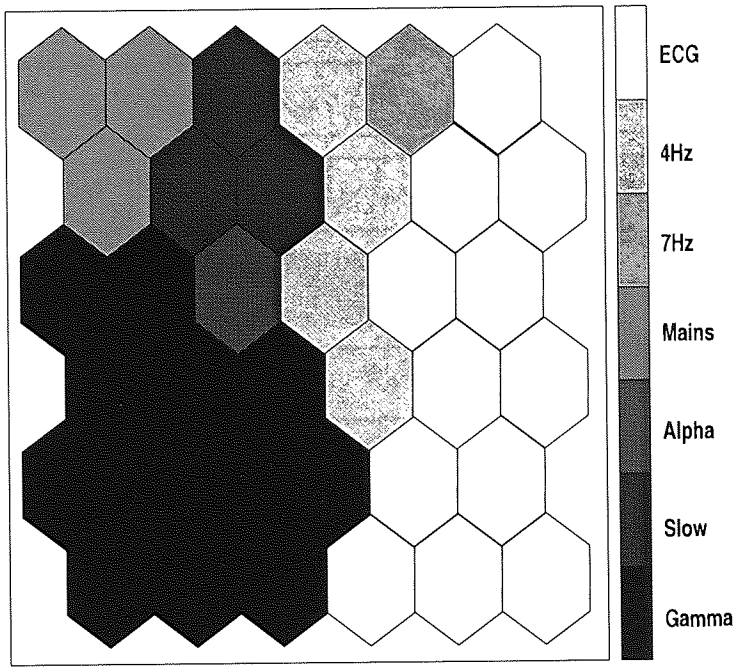


(a) Sample ICs extracted from each class

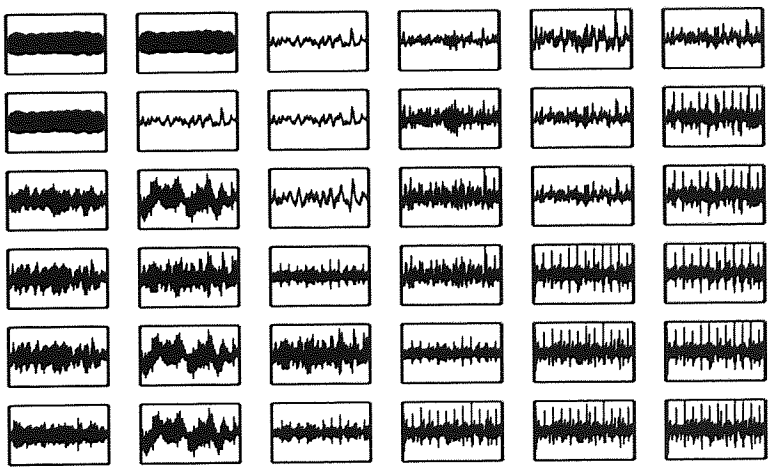


(b) Sammon map depicting relationships between ICs extracted from MEG recording of subject A performing the face recognition task

Figure 4.10: Results for face recognition data

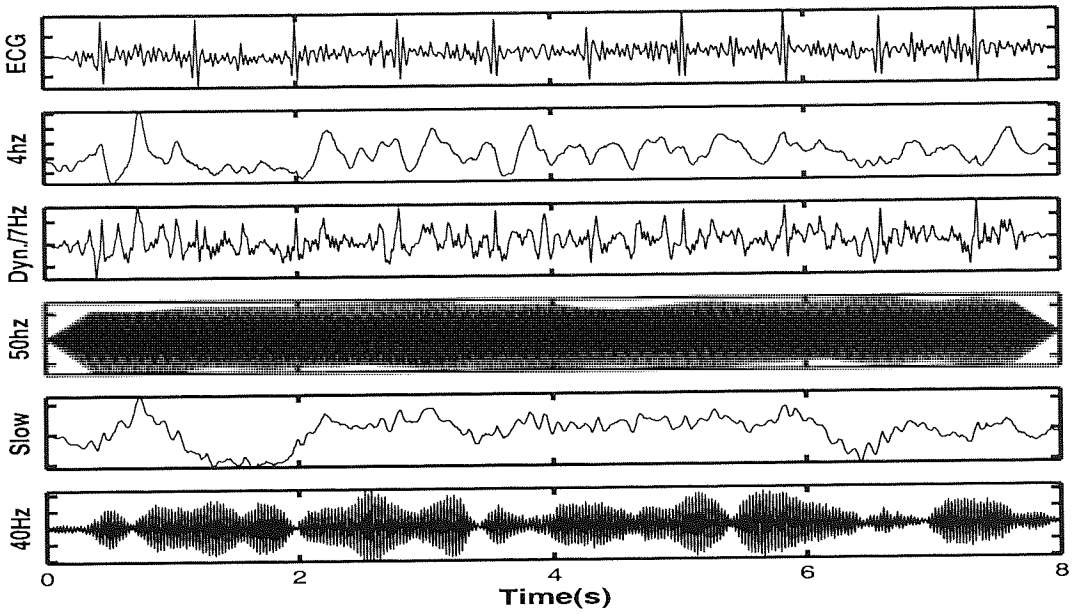


(a) Class assignments

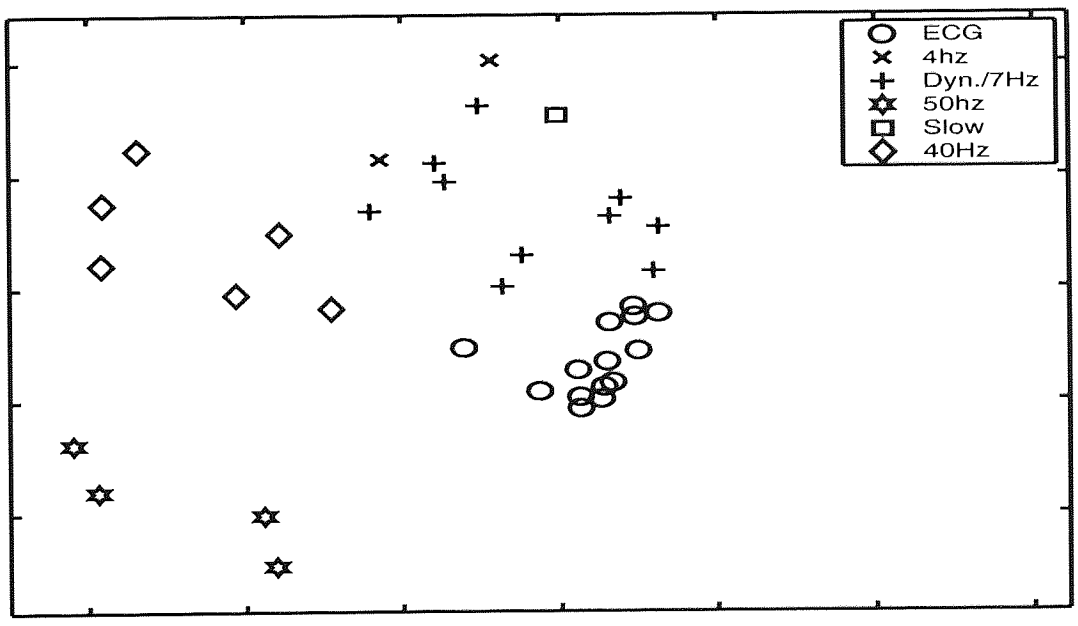


(b) ICs corresponding to the SOM lattice neurons

Figure 4.11: Visualisation of the Kohonen SOMs trained using the face recognition IC spectra.

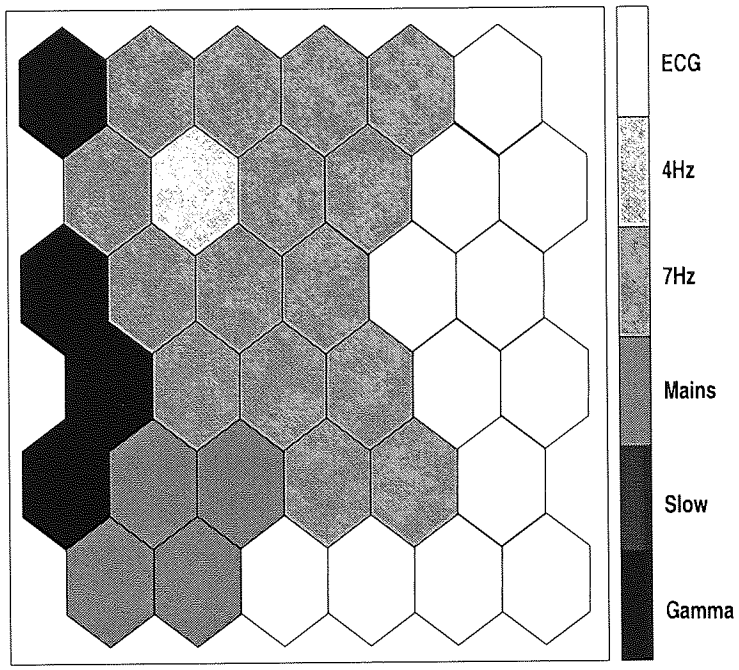


(a) Samples ICs extracted from each class

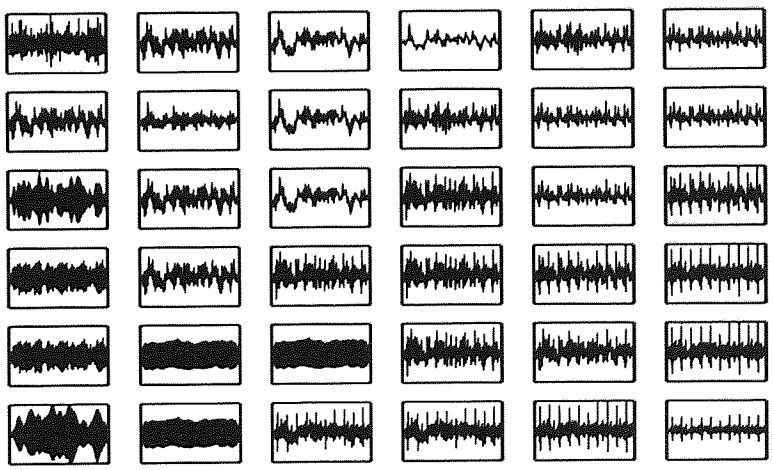


(b) Sammon map of the ICs

Figure 4.12: ICs extracted from MEG recording of subject A performing the natural scene recognition task.



(a) Class assignments



(b) ICs corresponding to the SOM lattice neurons

Figure 4.13: Visualisation of the Kohonen SOMs trained using the natural scene recognition IC spectra.

6. Low frequency drift (probably eyeblink and other muscular artifacts)
7. 50Hz noise

The number of each type of ICs extracted, as well as the relative powers of each of the components, obtained from the variances of the back-projected ICs, is given in figure 4.14.

Other observations:

- The Sammon mappings consistently obtained a characteristic pattern, with a central mass containing the 4Hz, 7Hz, Alpha and ECG activity - these components generally had overlapping frequency spectra and were hence more difficult to differentiate.
- In all three cases, the ICA was successful in clearly extracting the following components: 50Hz mains hum, ECG (heartbeat), 4Hz, a drift ("Slow") component as well as a 40Hz component. Invariably, the ECG signals accounted for the largest number of ICs, with 14 components extracted from each of the three tasks.
- Alpha waves were only recorded in the idle and facial recognition set. As is expected, the number and signal power of these components were significantly less in the case of the facial recognition experiment, while alpha activity is undetected in the image recognition experiment. This trend reflects the alpha desynchronisation process. This is often reflected in a phasic decrease of alpha power which is correlated with increasingly demanding mental activity.
- ICA was also able to detect two classes of signals which fall in the **theta band**. One is a relatively low frequency, and strongly oscillatory 4Hz component, which is detectable in all three of the experiments. The other is a 7Hz component, which is significantly less distinct and which only shows short windows of oscillation. Components from this class were characterised by a noisy, "random walk"-like morphology of varying frequency. It is this latter theta component, however, which appears to correlate more strongly with increasing cognitive demand, appearing only in the recognition paradigm experiments, and exhibiting significantly higher signal power in the more difficult image recognition task. This agrees with previous knowledge regarding the behaviour of theta band signals, in particular the well known association between theta and short term (episodic) memory. The 4Hz component, on the other hand, seems to be relatively independent of the difficulty of the task, and we might speculate that it represents a general sensory or alertness related theta component.
- More interestingly, however, ICA was also able to extract 40Hz oscillations from the data, which seems to indicate activity in the **Gamma band**. The significance of this is that Gamma has been implicated in a variety of cognitive binding roles [8; 27] and also

in the storage of short term memories [38], where a cooperative mechanism involving theta and gamma waves has been proposed for the latter process. Noting that both of these faculties are required in the performance of the recognition experiments, this then seems to correspond with the results as theta band ICs were not even extracted from the idle task recordings. However, Gamma activity seemed to be present in all three of the experiments, with roughly similar numbers of ICs extracted. However, the signal power for both the Gamma and Theta components in both the recognition experiments show significant increases when compared to the powers in the idle task, whereas the powers of these two signals are also greater in the natural scene recognition experiment, compared to the face recognition experiment. As the natural scene recognition experiment is significantly more difficult than the face recognition experiment, the results seem to indicate a correlation between Theta and Gamma power, and cognitive function.

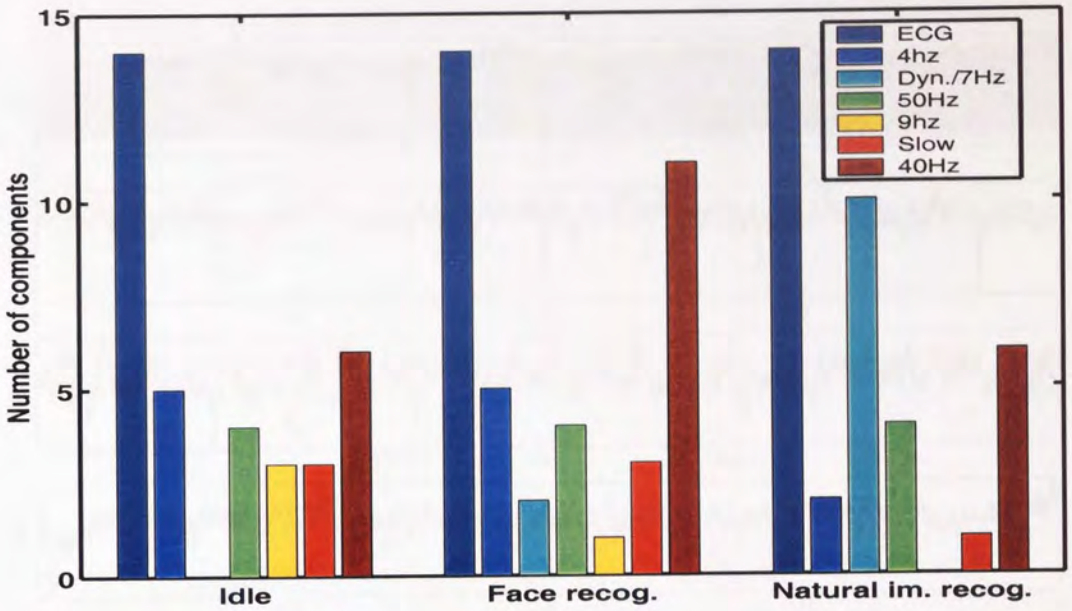
- It has been reported in [13] that induced Gamma activity is either not present, or very weak, in MEG recordings. This is contrary to our results which indicate an increase in event related Gamma activity. While evoked Gamma activity has been detected prior to this in MEG, it is time locked to the stimulus instants and is unlikely to be detectable in unaveraged data. While more detailed future research is required on this issue, it appears that the embedding approach which we are advocating is capable of detecting induced Gamma activity in MEG data.

4.3 Practical Issues

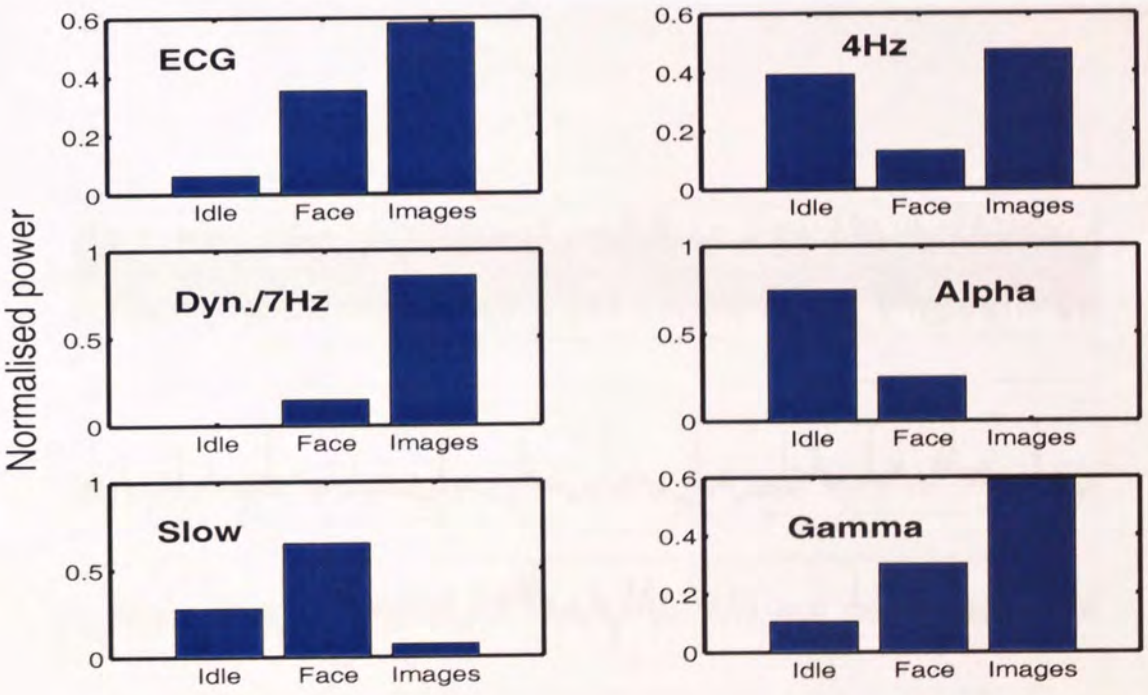
4.3.1 Comparison with ensemble ICA

To evaluate the effectiveness of the embedding approach compared to alternative decomposition schemes, ICA was performed on ensemble (multichannel) ICA. This is the method commonly used for analysis of EEG or MEG data, where, for example, it has been used for ocular artifact rejection in EEG [40; 75] and MEG [33].

Using the full 150-channel MEG recordings, and working with the same segments of data that were used in the previous sections, we again perform blind separation of the ensemble data set using the same procedure that was used with the embedding matrix: first, a signal noise subspace decomposition was performed using SVD, where the data matrix was projected onto the first 40 principal vectors (again based on the "kink" in the singular spectrum). Next, the FastICA algorithm was again used to perform the blind separation of the independent components. The results of the ICA decomposition is shown in figures 4.15,4.16 and 4.17, which give the ICs extracted from the idle, face and natural scene recognition tasks respectively.



(a) Number of IC classes. Note the gaps in the graphs corresponding to the idle and natural scene recognition experiments. These correspond to the component classes (Theta and Alpha respectively) which were not detected in these two classes



(b) Relative power of IC classes

Figure 4.14: Bar graphs showing (a)the number of components from each class which were extracted from each of the recordings; (b) The relative power of each IC class. Signal powers have been normalised to one.

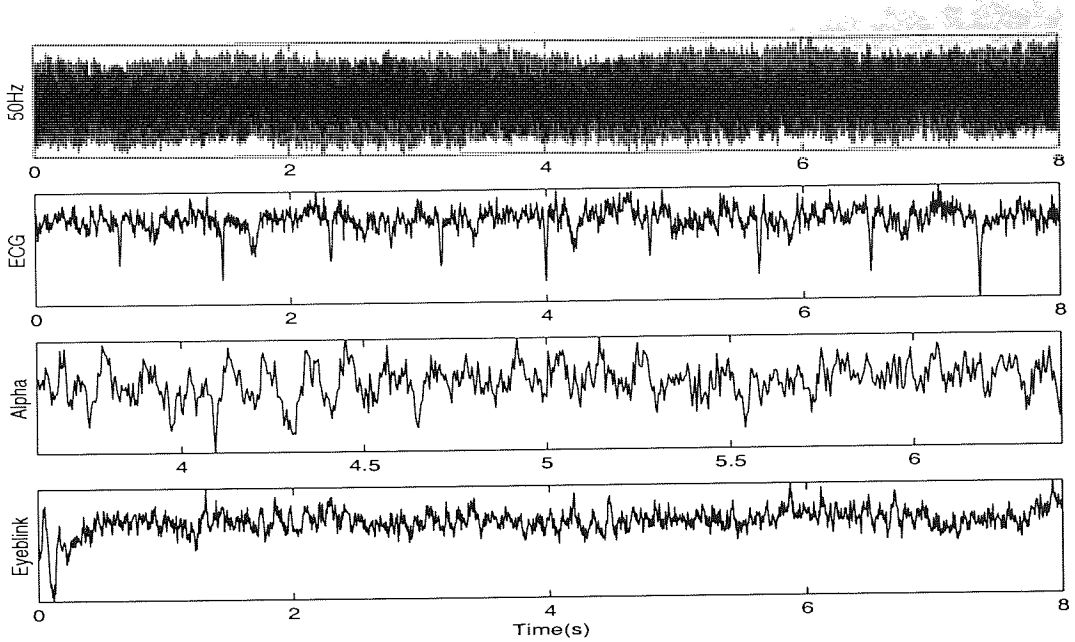


Figure 4.15: ICs extracted from the idle task multichannel dataset. The IC for Alpha activity (third row) has been magnified to focus on the region with clearest oscillatory activity.

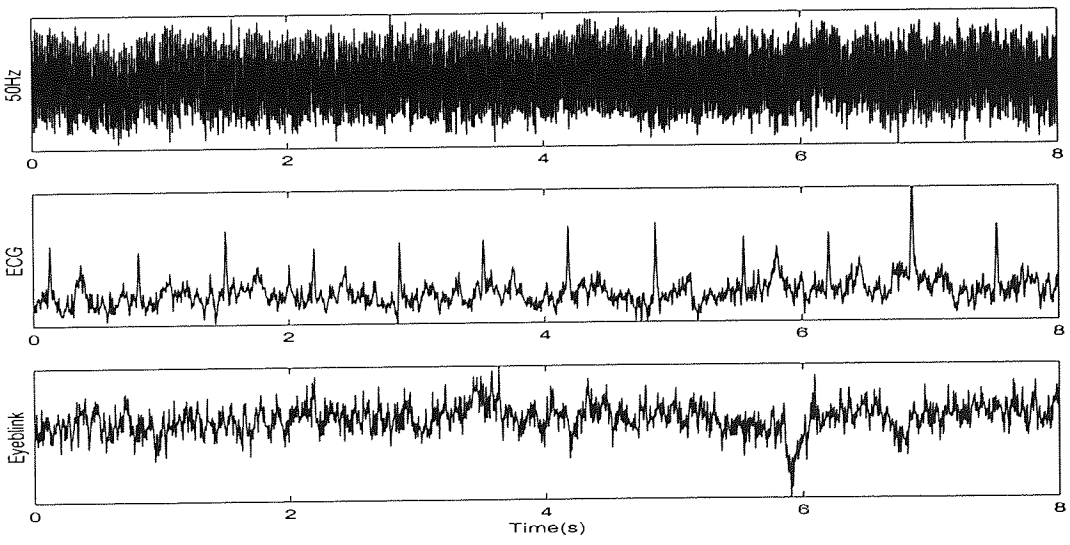


Figure 4.16: ICs extracted from the face recognition task multichannel dataset.

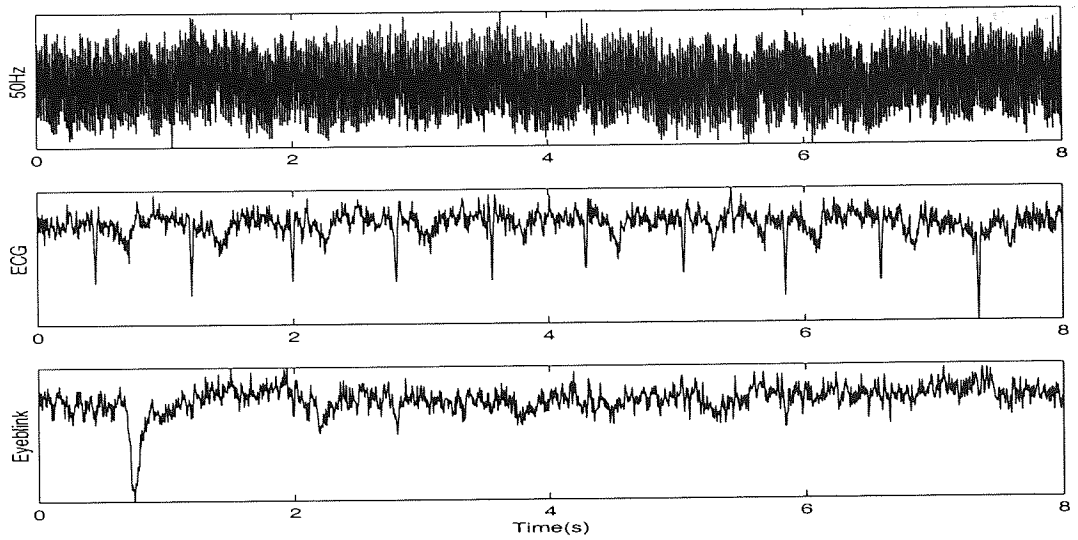


Figure 4.17: ICs extracted from the natural scene recognition multi-channel dataset.

From the decompositions, the following observations were made:

1. Upon visual inspection of the ICs, it is obvious that the spatial filters generated with ICA generally performed poorly compared to the embedding approach. While it was possible to identify some of the same components which were extracted using the embedding approach, the extracted ICs were a lot noisier and many of the components could not be identified at all.

2. Only the following components were positively identified:

   - 50Hz system mains noise
   - ECG artifacts
   - Alpha waves
   - Eyeblink artifacts
   - "Slow" activity

3. No occurrences of Theta, Gamma or Delta waves were found. However, this does not necessarily mean that these components were not extracted, merely that it was difficult to recognise them as the signals were not well separated from the background MEG. This is not surprising, as the Theta and Gamma waves extracted with the embedding approach were already difficult to observe, it follows that they would have been even more difficult to recognise when using ensemble ICA. Unexpectedly though, the Delta waves were also not extracted.

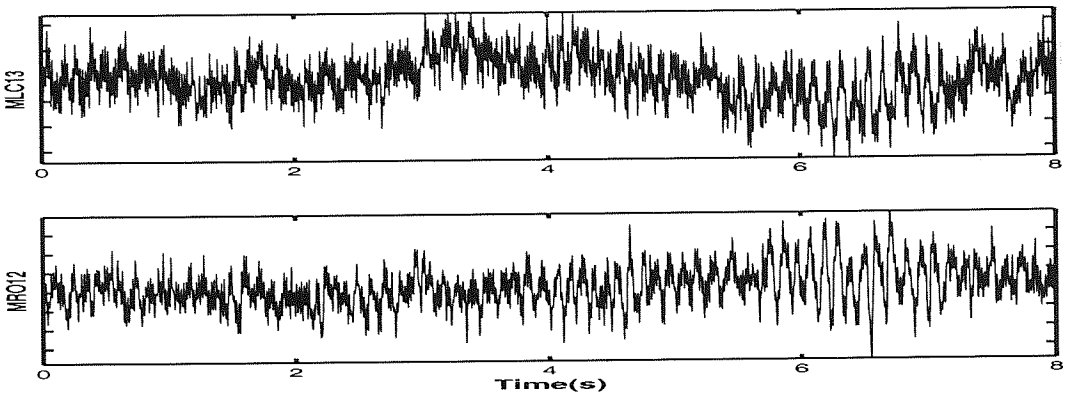
4. Clustering the ICs using the Sammon Map approach was not a practical step as there were large numbers of ICs that could not be categorised as belonging to any class in particular. In general, the components extracted were generally very noisy,

The ICs extracted seem to agree with those obtained from many of the previous studies which use ICA to analyse MEG or EEG. This seems to indicate that the embedding approach is more effective at separating signals from MEG recordings.

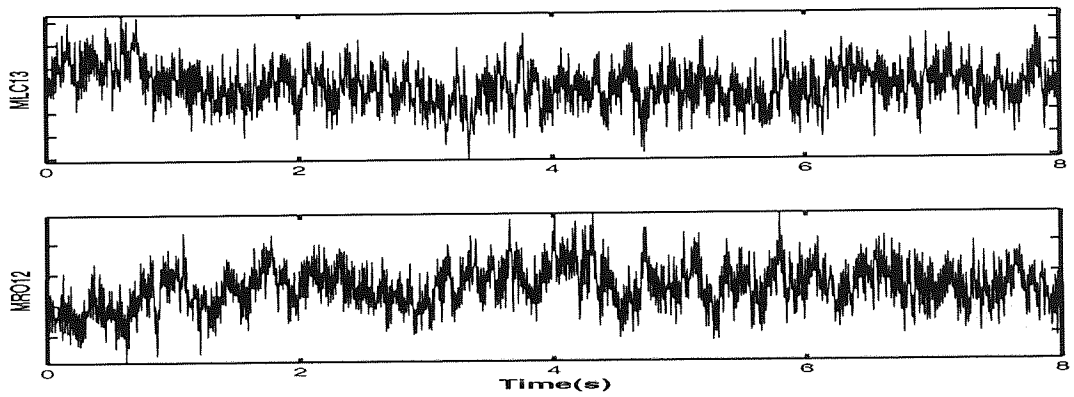
4.3.2 Inter-channel variability of extracted ICs

In this section some of the results from other channels are shown. The main aim is to study the effect on ICA when performing the embedding on different channels. As explained previously, the embedding theorem guarantees a topologically accurate reconstruction of the system manifold provided that smoothness condition is fulfilled for the system dynamics and observation function. Clearly, the dynamical properties of the system cannot change depending on which measuring channel is used. Equally, the observation function relating scalp measurements at any site, and the underlying dynamics, is almost certain to be smooth. However, it is possible that differing noise levels might adversely affect recordings at various sites, rendering them unusable. Also, while the reconstructed manifold for two different sites might be topologically equivalent, their suitability for analysis via ICA (in particular linear ICA) might be compromised as a result of stretching and distortion, perhaps resulting from highly nonlinear observation relationships.

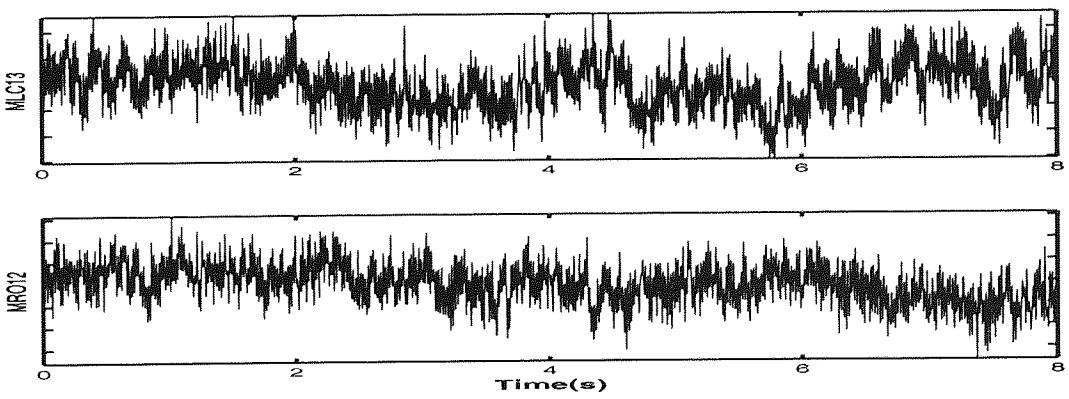
For this purpose, two other channels have been chosen. These were channels MLC-13 and MRO-12, situated in the left-central and occipital regions of the head respectively. The choice of these channels was motivated by their locations in different sections of the head. The raw time series which were to be analysed are given in figure 4.18. For each of the channels shown, the embedding and ICA extraction was performed in the same manner as with channel MRT-22. Samples of the ICs extracted from each of these channels are given in figures 4.19, 4.20 and 4.21 for the idle, face and natural scene recognition tasks respectively. As can be seen, the ICs extracted from the other channels are broadly similar to the case with MRT22 though there were some differences. For example, we were able to extract ICs in the Beta frequency range (around 20Hz). This is because these channels appeared not to be contaminated by ECG activity, which in the case of MRT22, seemed to obscure the beta activity due to the overlapping frequency spectra. Results from the decomposition of these channels were also subjected to clustering analysis using Sammon Mapping. The sammon maps for each of these channels are given in figures 4.22, 4.23 and 4.24.



(a) Idle

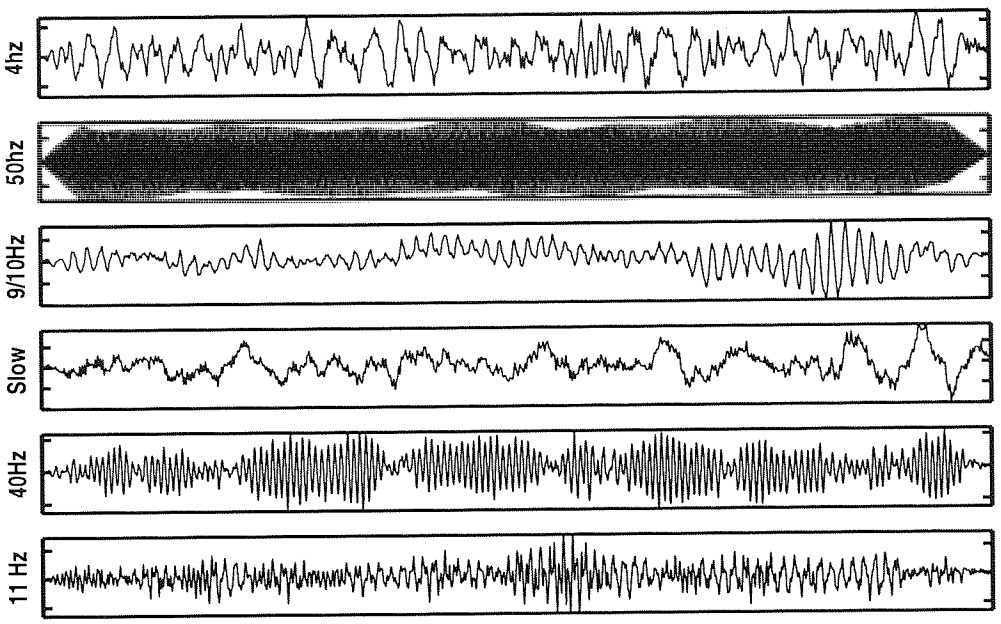


(b) Face recognition

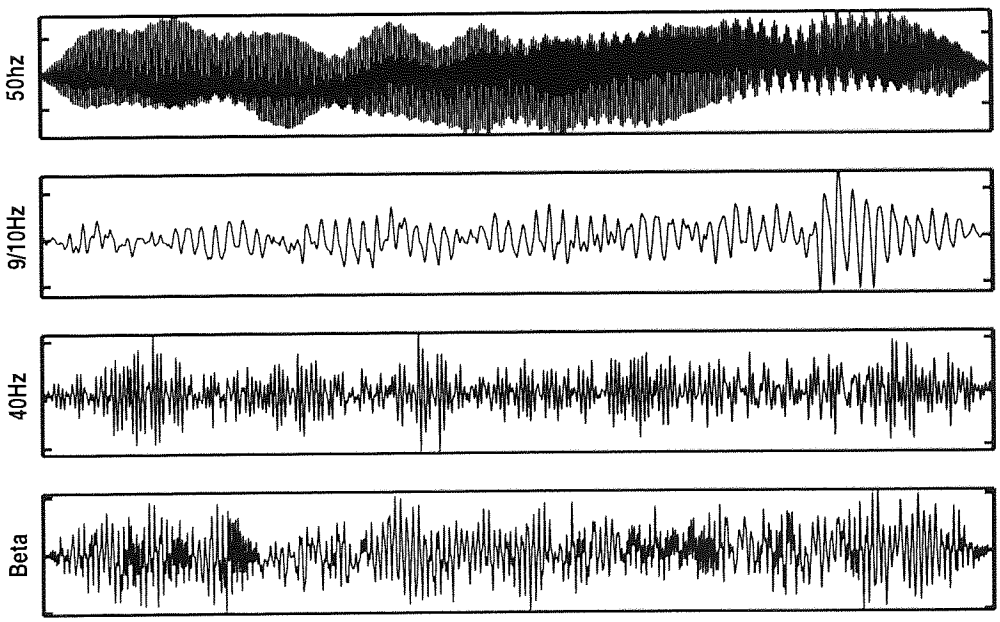


(c) Natural scene recognition

Figure 4.18: MEG recordings from MEG channels MLC-13 and MRO-12.

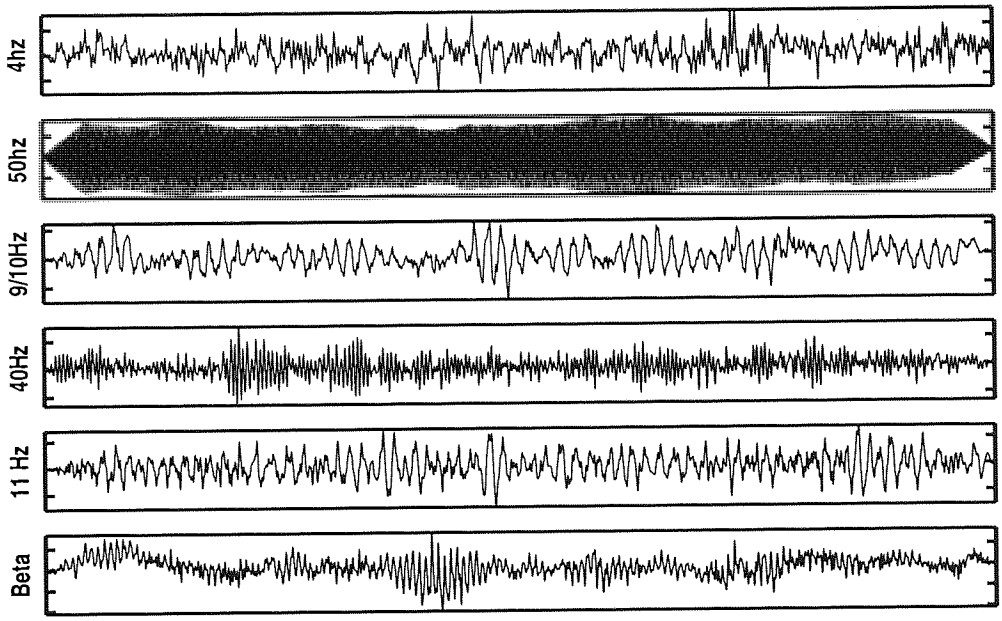


(a) MLC-13

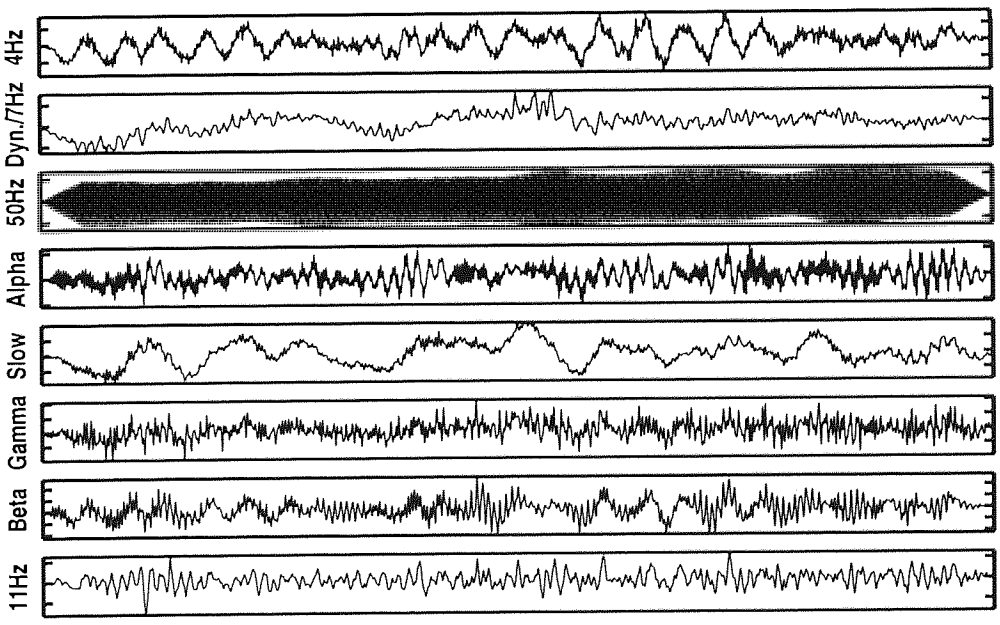


(b) MRO-12

Figure 4.19: ICs extracted from idle task recordings: MLC-13 and MRO-12.

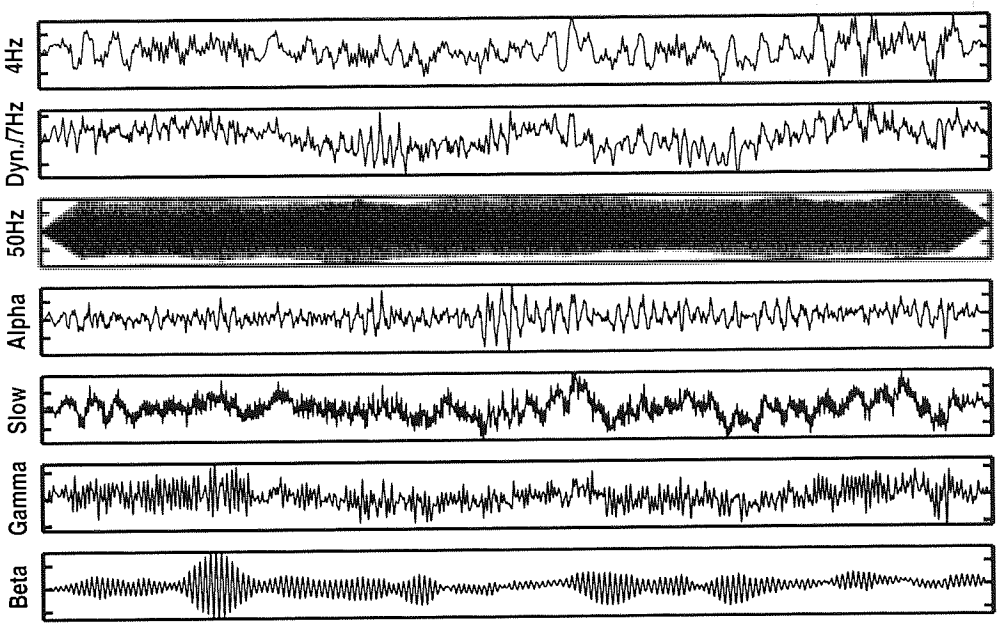


(a) MLC-13

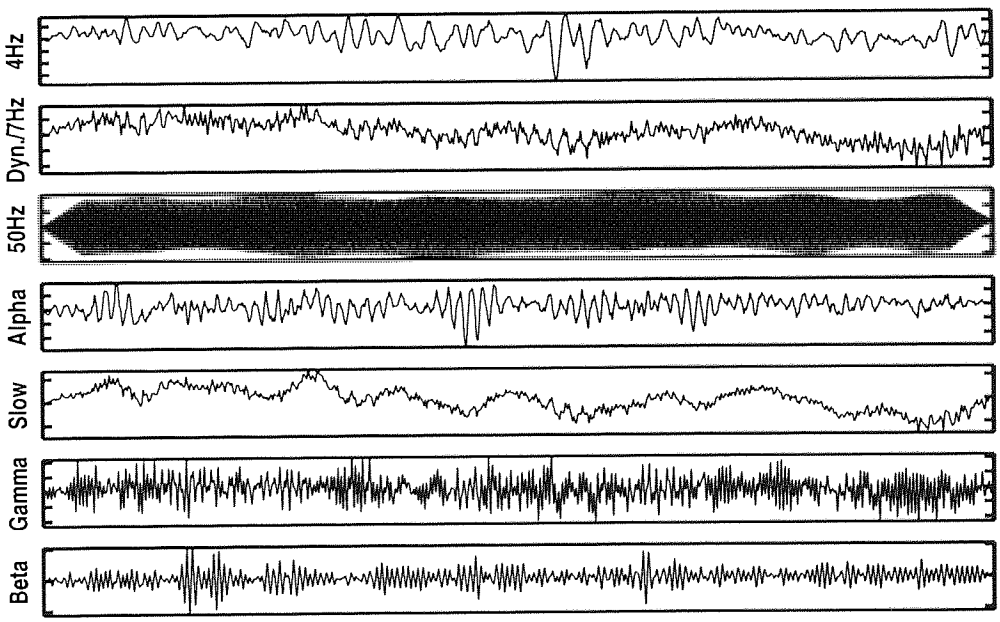


(b) MRO-12

Figure 4.20: ICs extracted from face recognition recordings: MLC-13 and MRO-12.

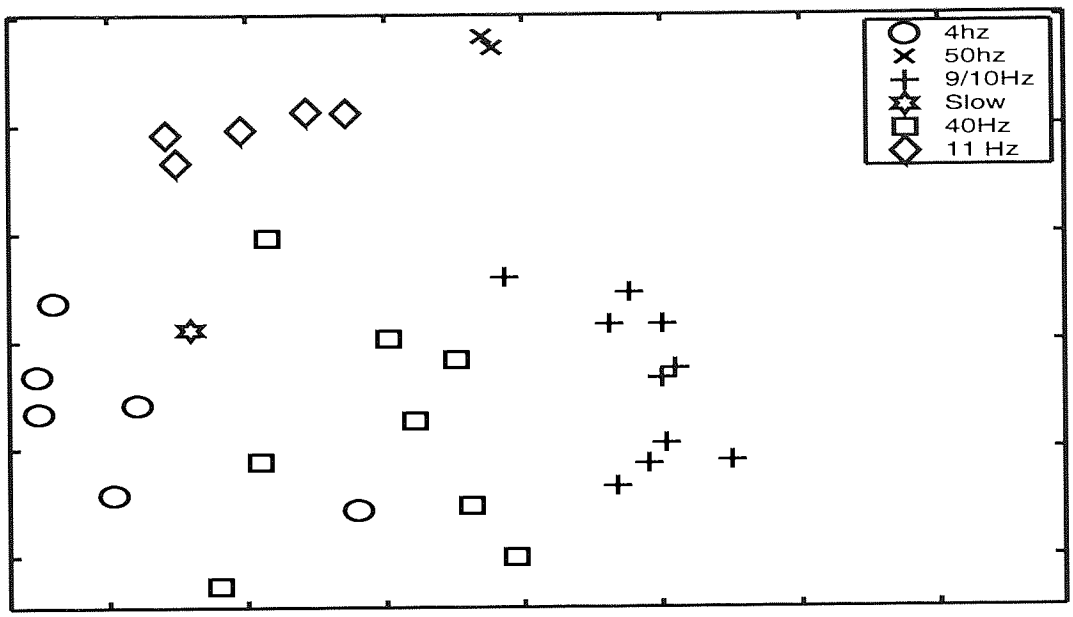


(a) MLC-13

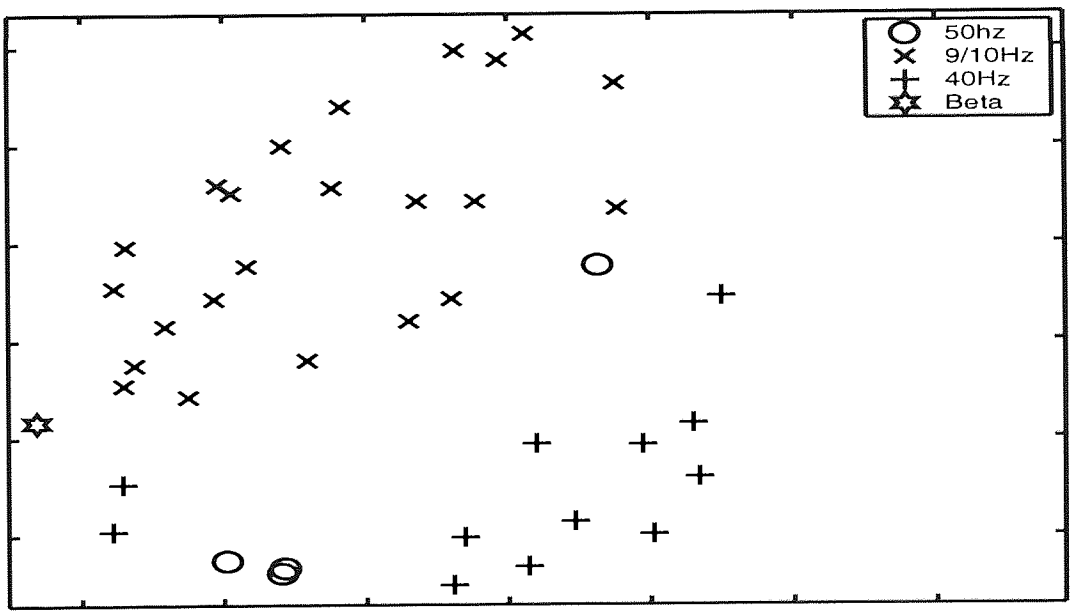


(b) MRO-12

Figure 4.21: ICs extracted from natural scene task recordings: MLC-13 and MRO-12.

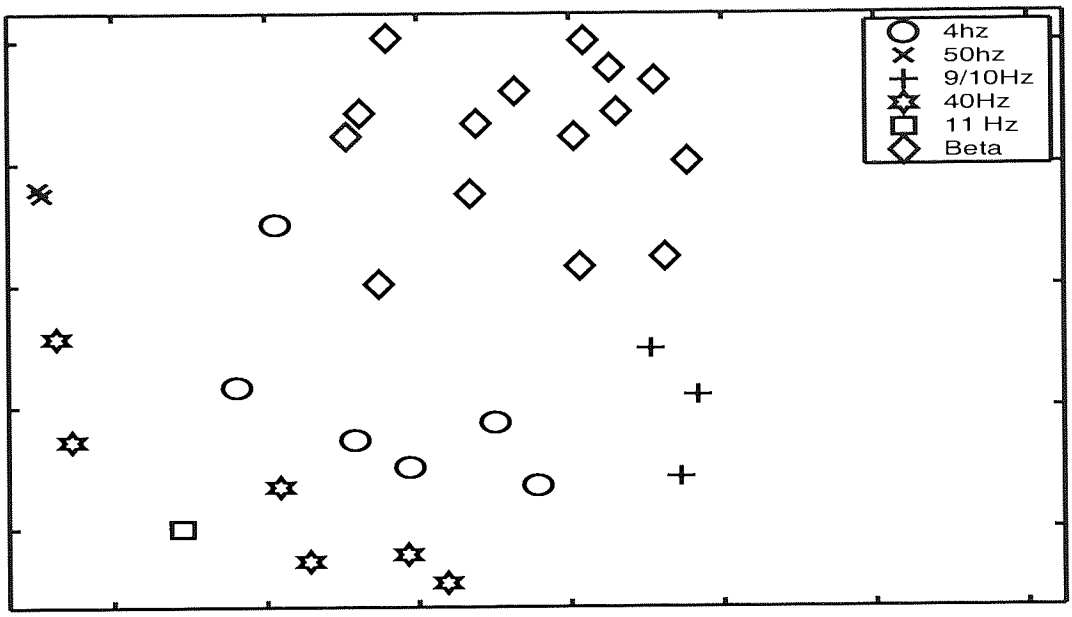


(a) MLC-13

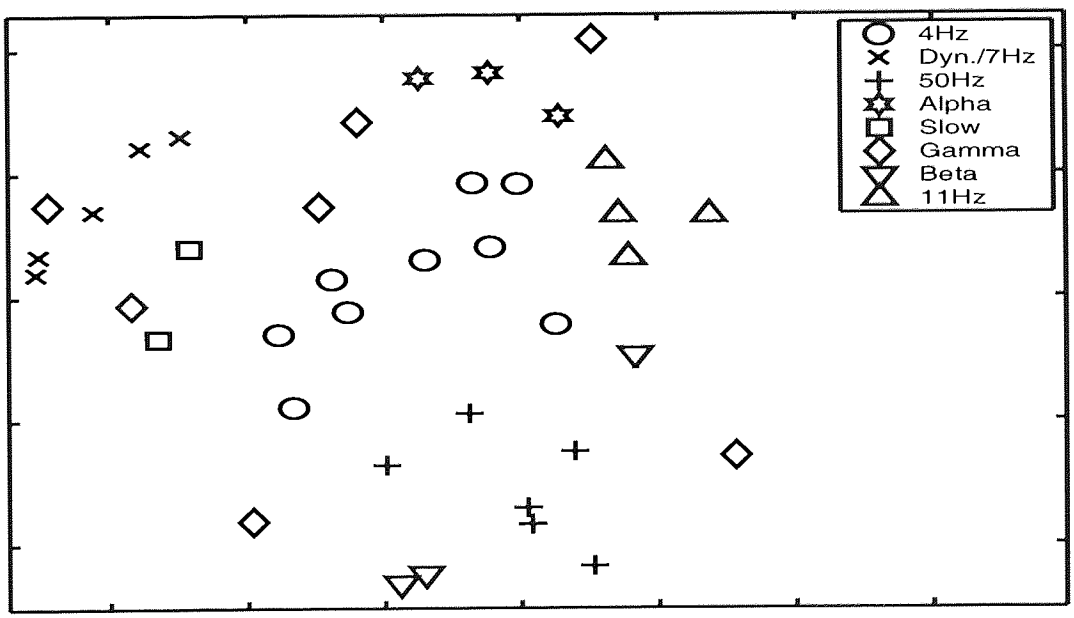


(b) MRO-12

Figure 4.22: Sammon maps for idle task recordings. Note the large number of Alpha components extracted from the occipital recordings (MRO-012).

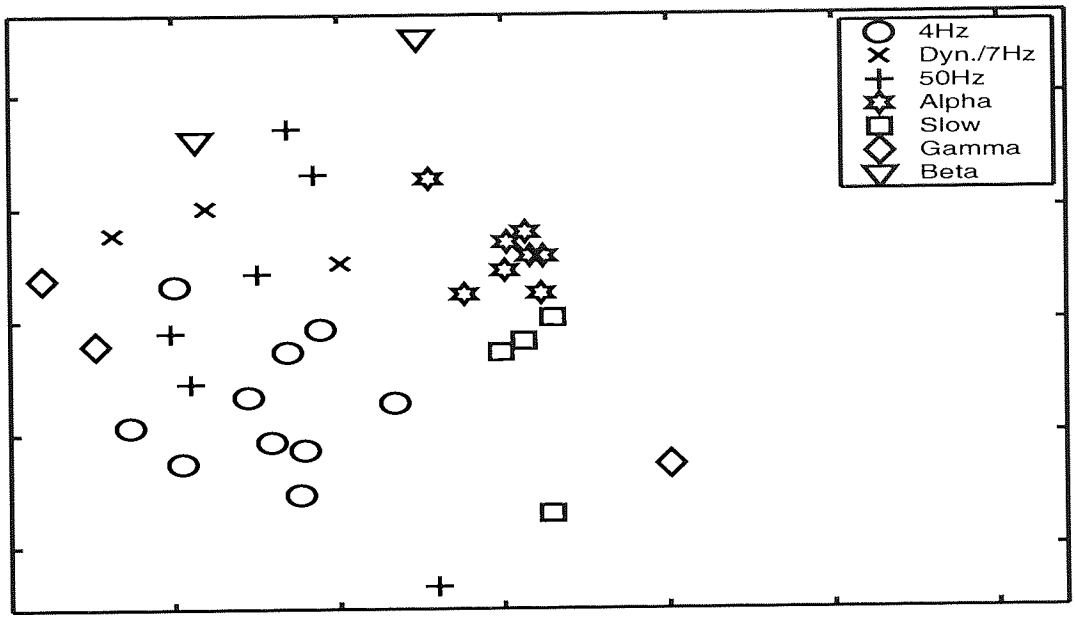


(a) MLC-13

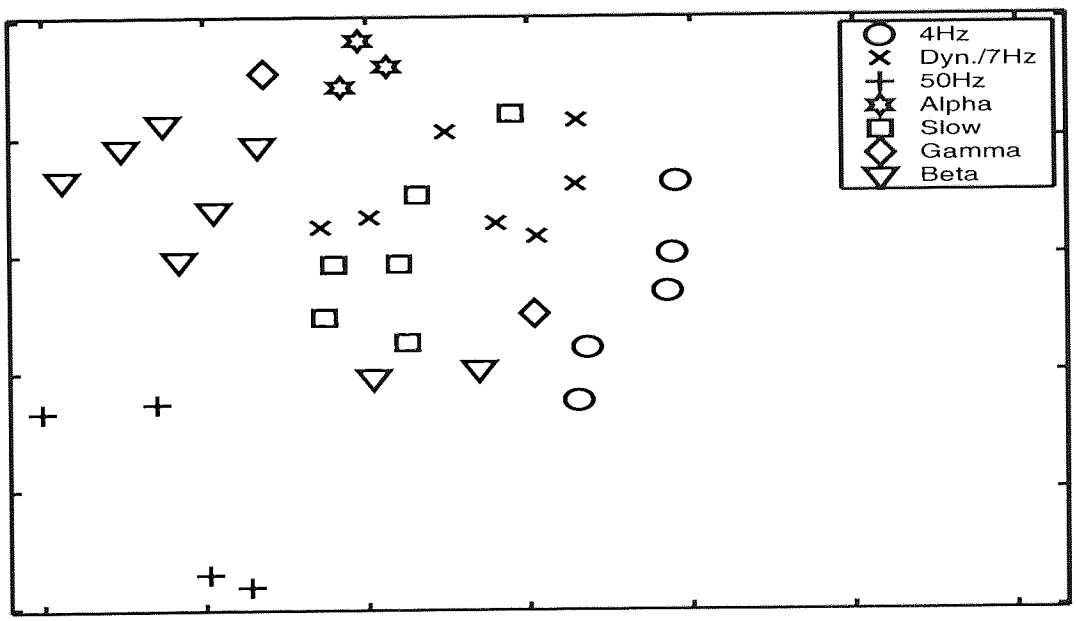


(b) MRO-12

Figure 4.23: Sammon maps for the face recognition task.



(a) MLC-13



(b) MRO-12

Figure 4.24: Sammon maps for the natural scene recognition task.

4.3.3 Example application: Noise and artifact rejection

As a demonstration of the usefulness of the framework described, a simple noise and artifact rejection application will now be discussed (though noise reduction is just one of the techniques we are studying in this framework). This application has been chosen as a proof of concept for three main reasons:

1. Noise and artifact contamination is a very serious problem in MEG signal processing. The levels of noise are so high that visual examination of MEG time series is extremely tricky and time consuming. At the same time, the predominant method of noise suppression in MEG is simply to average a large number of trials of a given experiment, which is clearly not an optimal solution. Other methods require additional input signals such as EOG sensors to provide a reference signal for regression matching.

2. The ICA extraction procedure which we have described appears to perform well in the separation of noise and artifact from the MEG. This is not surprising since we expect that external contaminants are almost certain to be independent of the true cortical magnetic signals which we are interested in. As such, they can be expected to conform well to the independence criterion.

3. One of the major problems with MEG analysis is that very little is actually known about the signals that are being measured. Even if the ICA algorithm was able to cleanly extract the underlying brain dynamics, we would still not have a very clear idea of what to look for. A noise reduction system, apart from being easily implemented, is easier to evaluate as its effects can be clearly seen.

The procedure which we shall use to separate the noise and artifacts is quite straightforward. After the delay embedding, we perform ICA decomposition and clustering to find components that are associated with noise/artifacts. For this demonstration, we will use another segment of MEG which has been chosen specifically because it contains a large number of ocular artifacts (eyeblinks). ICA is applied on the embedding matrix and the resulting ICs corresponding to ECG, 50Hz noise as well as the eyeblinks are removed by setting the corresponding mixing matrix columns to zero. The ICs are then projected back into the observation space. Figure 4.25 compares the original (noisy) data, as well as the de-noised time series. As a comparison, a low-pass filtered version of the data has been generated using a low-pass filter with a cutoff of 25Hz, and is also shown in figure 4.25

From figure 4.25, we can see that the denoised time series is almost entirely free from ECG and eyeblink contamination. The removal of the 50Hz noise also means that the time series

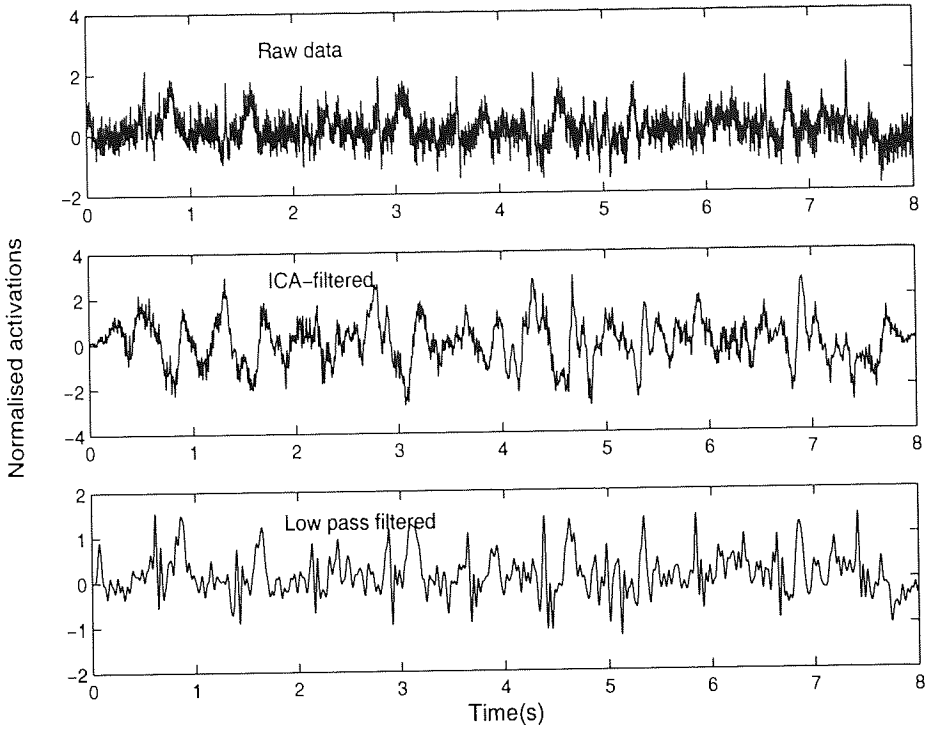


Figure 4.25: Subplots of (top) Original data (for comparison), (middle) the same time series, after application of de-noising procedure, and (bottom) the data after filtering with a low pass filter.

looks a lot cleaner than prior to processing. It is also clear that the low pass filtered signal is still heavily contaminated by ECG as well as eyeblink content. Also, while the 50Hz mains noise has been removed, it is also likely that all high frequency cortical signal content has been removed as well. Figure 4.26 shows the frequency spectra of the ECG and eyeblink components. Note that the ECG component in particular contains frequency content from 0 all the way up until 60 Hz, making it difficult to design a bandpass/lowpass filter that will not also remove other components of interest.

4.4 Summary

This chapter investigated the use of independent component analysis for the blind decomposition of *single channel* MEG. The basic concept behind our method is quite simple and involves two steps: first, we use dynamical embedding as as method for transforming the scalar MEG time series into a multidimensional trajectory matrix. The second step is to find an optimal linear transformation of this matrix such that the underlying MEG sources

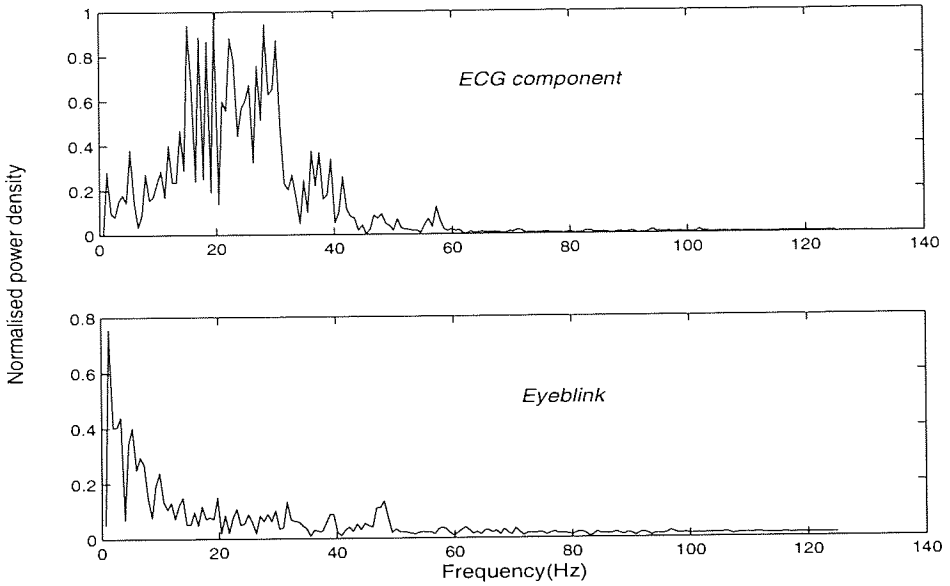


Figure 4.26: Frequency spectra of: (top) ECG artifact component, and (bottom) Eyeblink component.

are extracted. We use the *statistical independence* of the separated sources as the criterion to perform this transformation. Independent component analysis provides a tool by which this can be achieved. To validate the embedding-ICA approach, we successfully performed a trial using realistically synthesised MEG data which demonstrated the applicability of the method.

We also proposed a novel application of topographic projection techniques for clustering the independent components. While ICA has been used previously in MEG and EEG analysis (though only on multichannel MEG), the problem of analysing and classifying the extracted components into classes has not been properly addressed prior to this. As MEG invariably generates a large number of ICs, this is clearly an important issue.

Of course, the extraction of the independent components is only the first step. The interpretation of these independent components presents a massive challenge, and a full analysis of the neurophysiological significance of these signals is clearly beyond the scope of this thesis. Nevertheless, in the next chapter we develop two possible routes by which further information may be derived from the results obtained in this chapter.

Chapter 5

Qualitative analysis of single channel MEG

In this chapter we extend the analysis carried out previously by subjecting the MEG data, and in particular the extracted independent components to more detailed analysis. One major question that confronts us now is the interpretation of these components, which presents a serious challenge. Certainly, from the components that we have studied, it has been possible to identify contributions from established frequency bands, namely theta, alpha, beta and possibly even gamma waves. Other components have been associated with artifactual or spurious signals such as heartbeat, mains noise and eye blinks. However, the classification of the ICs into these categories hides a richness of variability in the morphology and structure of these independent sources. Even within the same "class" of ICs, there are often components that, upon visual inspection, clearly differ in appearance and functional correlates. Given that we do not have a good understanding of the nature of the signals that we are looking for, identifying or interpreting them poses a significant challenge. In this chapter we propose two main approaches by which we are able to gain a greater degree of qualitative knowledge regarding both the ICs, and MEG time series in general.

As we have hitherto been working exclusively with single channel data, we have little knowledge of the physiological properties of the signals that we are studying. While the dynamical systems framework imposes an abstraction layer that tends of obscure the physical processes giving rise to the observed data, information regarding the spatial distribution of the ICs is nevertheless likely to be useful. In particular, this knowledge will allow the components to be associated with specific regions of the scalp, which in turn can help to elicit their functional origins. We address this need in section 5.1, which describes a method for the generation and use of scalp distribution maps corresponding to the ICs. Section 5.2 investigates a different

way of characterising signals: via signal complexity. The calculation of various complexity measures has previously been applied to EEG, with varying degrees of success. However, it is unclear if MEG is amenable to this mode of analysis, given the poor SNR characteristics. We explore the applicability of some of the newest complexity measures on both the raw MEG data, as well as the independent components which we have previously extracted.

5.1 Spatial distribution of extracted features

One of the shortcomings of single channel analysis is that it does not yield any information regarding the spatial distribution of the sources over the scalp surface. Knowing the topology of a given IC is useful in three main ways:

1. It provides an additional set of features by which the ICs may be compared and differentiated.

2. It provides clues regarding the functional correlates of ICs via corroboration with prior information regarding functional attributes of regions of the neocortex. Agreement with existing theories also presents further support for the validity of the single channel ICA procedure.

3. Supplies valuable qualitative information regarding an IC - we are interested not only in the locations where the spectral power of an IC might be concentrated (even if these existed), but also in properties such as the spread of the distribution, whether it is multimodal, diffuse or focussed in specific regions.

5.1.1 Generation of spatial amplitude maps

To generate the spatial maps, we again need some way of measuring the similarity between two time series. This time however, the aim is to measure the similarity between the IC (or ICs) in question, and the MEG time series at each of the recording sites on the scalp. In accordance with our previous analysis, we retain the assumption that signals originating from a common underlying generator will possess a characteristic frequency spectrum. However, performing an element-wise differencing between the frequency spectra of the ICs and the MEG recordings fails to generate a good mapping as the extremely broadband nature of the raw time series, as well as the presence of artifactual signals, mean that any similarities between the spectra are likely to be masked by frequency contributions from other sources.

In an earlier work [79], we proposed the use of segments extracted from the ICs as FIR filter coefficients, which are applied to recordings from the other channels to obtain an approx-

imation of the similarity between the time series. However, subsequent analysis using this method has indicated that it is sensitive to random noise and artifacts. Better results were obtained using a surprisingly simple alternative: by simply filtering the multichannel data with FIR filters formed from the columns of the ICA de-mixing matrix corresponding to ICs of interest. As an example, the similarity between IC i, and MEG recorded from channel j (denoted as x_j), can be found in the following manner:

$$p = w_i \mathbf{x}_j \tag{5.1}$$

where w_i is the estimated de-mixing matrix $\mathbf{w}$ with all but the ith column zeroed out. $\mathbf{x}_j$ is the embedding matrix formed from channel x_j. p is then subjected to the Hankelization operator described in section 4.2.1. The amplitude of the resulting reconstructed time series then gives an indication of the signal power corresponding to channel j. This is repeated to generate an intensity-mapping depicting the contributions from each of the MEG recording channels.

The method described in the previous section was applied to each of the three segments of data and spatial maps corresponding to each of the ICs were generated in the said fashion. The spatial maps for one of each class of ICs will now be analysed.

5.1.2 Results

4Hz waves

Table 5.1 summarises the key points regarding component spatial distributions, as well as some of the known functional correlates of the various classes of neocortical activity.

An example of an IC containing 4Hz activity was selected from each of the experimental sets and the spatial maps for these components, along with the corresponding time domain signal, are shown in figure 5.1. Studying the maps for the three experiments, the most striking feature is clearly the large concentration of signal power in the frontal right region. However, as will become clear later, the signal power in this area is mostly due to the muscular artifact originating in the eyes and forehead, and should be interpreted with caution. We focus instead on the signal distributions in the other regions of the scalp. In this case, it should be observed that the power distribution becomes both more pronounced, and also envelopes a larger spread of scalp surface, when we compare the face recognition data, to the idle data. A similar trend is observed when comparing the image recognition to the face recognition map. We may attribute this to two effects. Firstly, as can be seen from the ICs in question, the smooth sinusoidal oscillations which were extracted in the case of the idle task become increasingly unclear and noisy in the other two data sets, resulting in the spatial

Component:	Probable classification	High signal concentrations	Known functional correlates:	Observed in:
4Hz	Delta/Theta	Frontal, central and left temporal	General cognition	All
Dyn./7Hz	Theta	Frontal and central	General cognition, attention, episodic memory	Face and scene recognition
9Hz	Alpha	Posterior	Attentive inactivity	Idle and face recognition tasks
40Hz+	Gamma	Frontal	Cognitive binding, attention, general sensory, memory	All

Table 5.1: Summary of key properties of the spatial distribution maps.

maps becoming smeared out. The other possible contributing factor is that the spread in delta activity corresponds to an actual increased recruitment of resources across the cortex, resulting in increased signal strength.

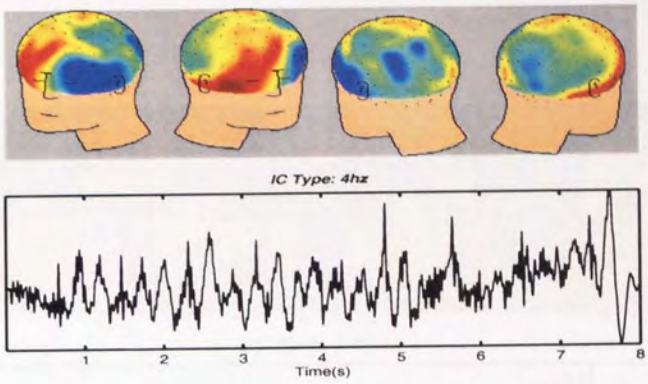
Alpha waves

As can be seen in figure 5.2, the spatial map for the alpha wave IC extracted from the idle task recordings showed a good correspondence with existing theoretical knowledge. Clearly, the alpha signal power is greatly concentrated in the posterior (occipital cortex) region of the scalp. However, in the face recognition set, we observe that alpha signal power is much more diffusely distributed; while a larger concentration of the power still appears over the occipital cortex, there is now significant signal power over the central regions as well.

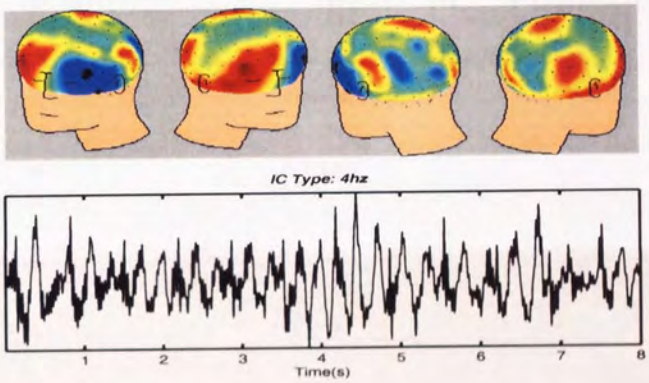
The observations correspond with the much-reported fact that alpha activity is most pronounced in the occipital cortex during idle but attentive brain states. During the more active brain-states, such as the face recognition task, a phenomenon known as "alpha-blocking" occurs, where the power and coherence of cortical alpha activity is significantly reduced. Interestingly, it is reported in [58] that visual stimuli often result in the simultaneous desynchronisation of occipital alpha rhythms, combined with an enhancement or synchronisation of central mu (or upper alpha) rhythms. This correlates extremely well with our observations.

Gamma waves

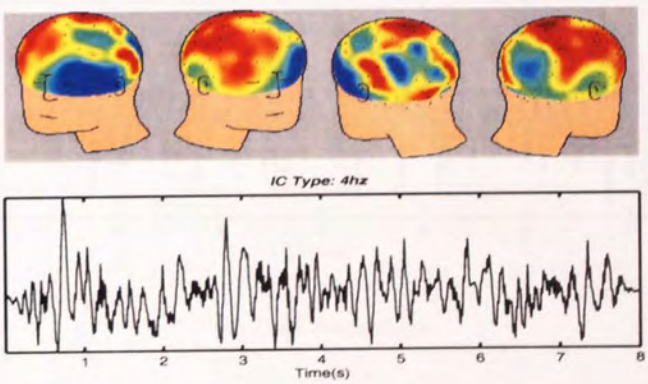
The spatial maps for gamma related ICs are shown in figure 5.3. It is observed that the power of the signals is concentrated mainly in the frontal and central regions of the scalp. This is consistent with the proposed relationship between gamma activity and cognitive function, which is most closely associated with the frontal cortex. Attention related gamma, for example, is known to be focused in the frontal and central areas [4] of the brain. Again we observe a general trend in which the distributions become increasingly diffuse as we move from idle to face recognition, and from face recognition to image recognition recordings. However, the observed differences are not sufficient to form a firm conclusion as they could merely be due to a general increase in complexity of the ongoing background MEG associated with the increasing difficulty of the recognition tasks. A final point of interest is the observation that with increasing task difficulty, the power and spread of posterior gamma also increases, both in the area of the parietal cortex, as well as at a very focused point around the occipital cortex. This might be indicative of increased *induced gamma* activity, as reported in [8].



(a) Idle

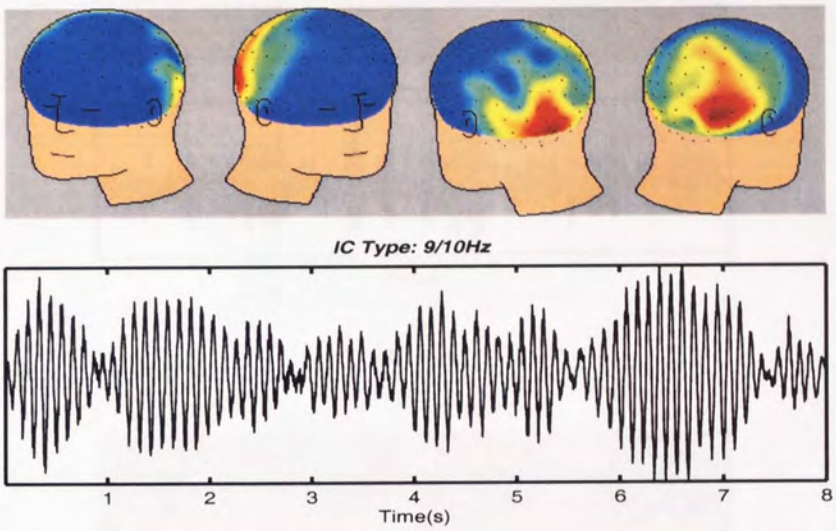


(b) Face recognition

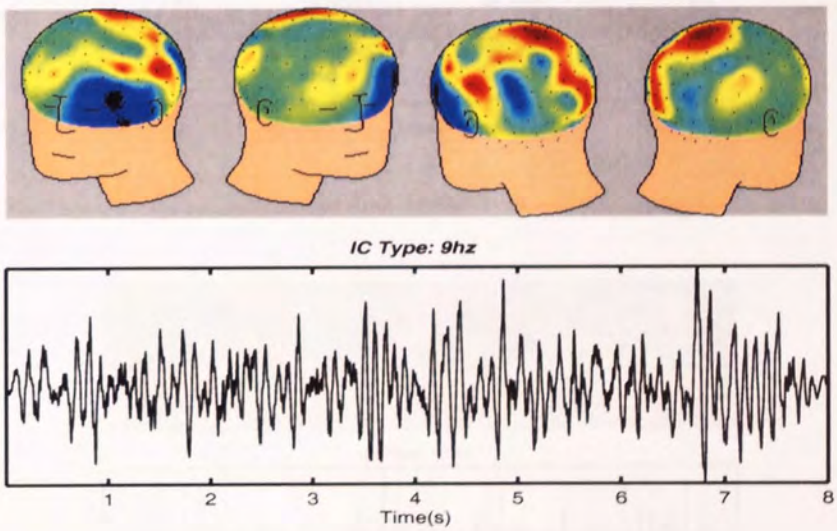


(c) Image recognition

Figure 5.1: Spatial maps: 4Hz activity

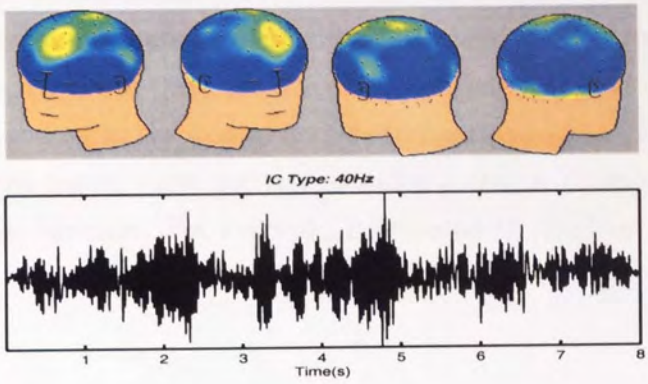


(a) Idle

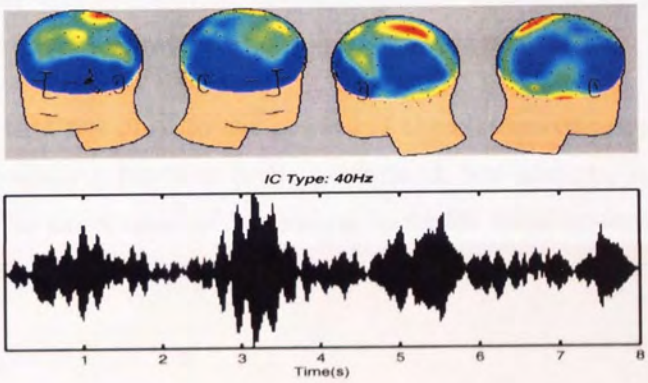


(b) Face recognition

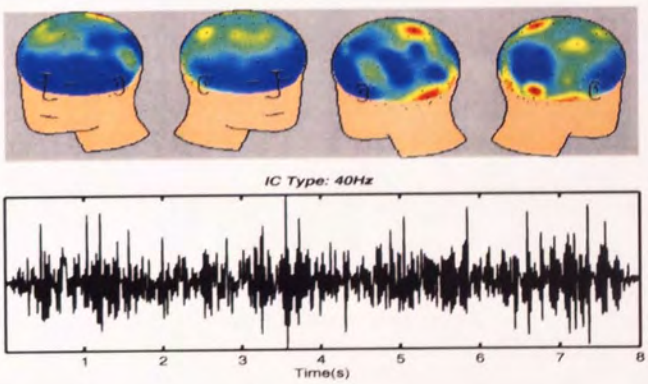
Figure 5.2: Alpha activity spatial distribution maps. Alpha components were only identified in the idle and face recognition tasks.



(a) Idle



(b) Face recognition



(c) Image recognition

Figure 5.3: Spatial maps: gamma activity

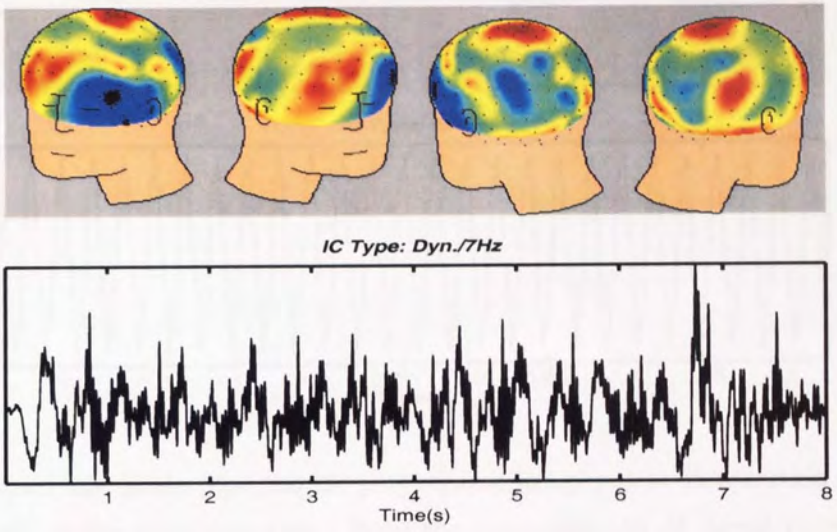
Theta activity

The spatial maps for the theta-related ICs are shown in figure 5.4. We observe that the distribution for these components is certainly not focused in one region but is scattered throughout the scalp. However, there are perceivably higher concentrations of activity in the scalp regions corresponding to the frontal and central cortical areas. This again agrees with existing theories which regard theta activity as being a crucial component in cognitive and memory related brain function. For example, it is noted in [39] that increased short term memory loading can stimulate proportional increases in frontal theta power.

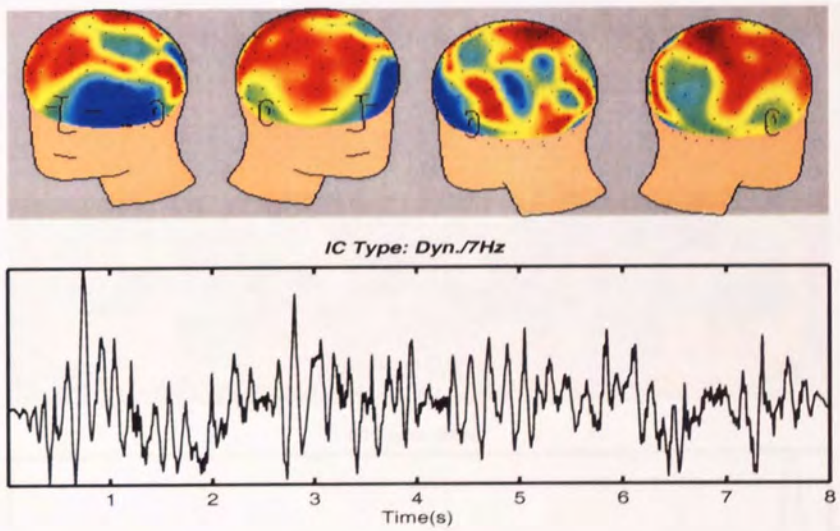
Artifact and noise signals

Three main identifiable artifact and noise sources have been identified. These are mains hum, eyeblink and/or ocular artifacts and ECG artifacts. The spatial maps for each of these have been generated and are shown in figures 5.5, 5.6 and 5.7 respectively.

It may be concluded that not only do the *powers* of the respective components show changes with respect to the cognitive function being performed, but also the spatial distributions of these components. The exact neurophysiological basis for these observed changes is beyond the scope of this thesis.



(a) Face recognition



(b) Natural scene recognition

Figure 5.4: Theta activity spatial distribution maps. No theta components were extracted for the idle task.

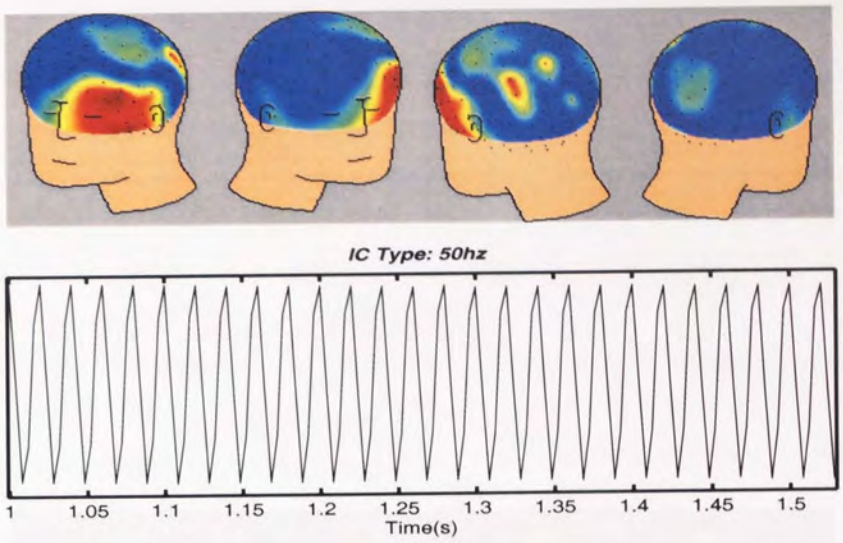


Figure 5.5: Scalp topographic plot representing spatial distribution of 50Hz mains contamination. Note the concentration of signal power to a few channels - examination of the channels confirm that these channels are almost entirely composed of mains noise.

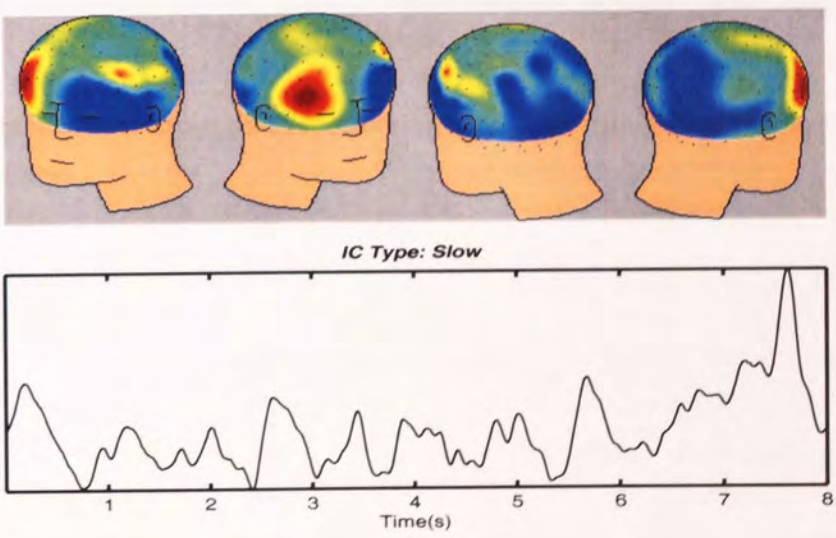


Figure 5.6: Contour plot depicting the contribution of different regions of the head to the slow-moving trend IC.

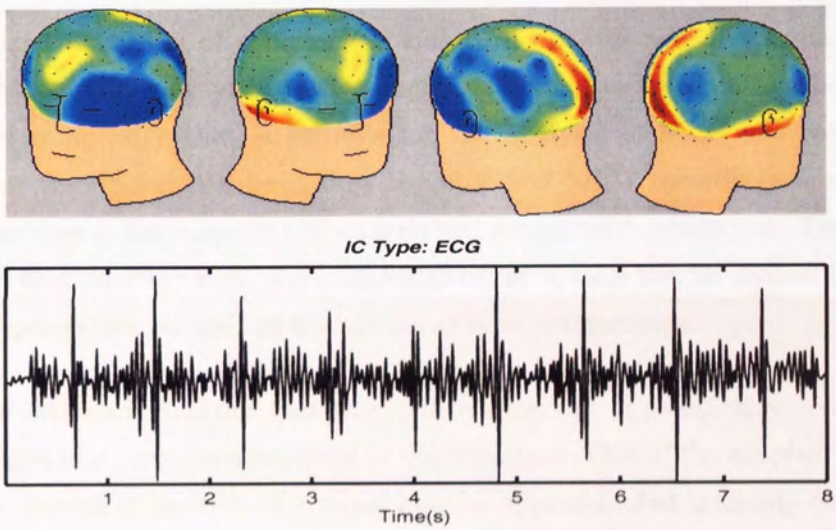


Figure 5.7: Contour plot representing spatial distribution of heartbeat artifacts.

5.2 Complexity

When inspecting a signal, apart from the signal power and perhaps the dominant oscillatory behaviour (if present), probably the first thing to strike a casual observer is the "complexity" of a signal. Intuitively, this is the general appearance of disorder in the signal. On one extreme, a white noise process will lack any apparent structure, while on the other, a pure sinusoidal signal is clearly evolving according to a fixed rule. In general, we are not interested in either of these two cases but rather in the qualitative properties of signals which fall between these two extremes. Signals may vary quickly, or show bursts of oscillatory behaviour amidst apparently random behaviour, and so on. An issue which has received considerable attention in the last few years is the search for appropriate quantifiers of this property of signals.

There is no fixed definition of complexity; indeed, a measure which is suitable for use in one application domain may yield misleading results when applied in another. To obtain a definition of complexity that is suitable for the analysis of MEG, we recall that from the outset, our standpoint has been that the observed MEG recordings are the result of interactions between a finite number of underlying, unobserved generators. Then, intuitively we may reason that on some level, the complexity of the system can be defined by the number of underlying generators, as well as the nature of their interactions.

A survey of related work confirms that this intuitive concept of complexity is consistent with existing measures that have been reported in the literature. One of the simplest ways in which the number of degrees of freedom of a signal may be approximated is simply to perform SVD and to look for the point at which the singular spectrum levels off. This approach yields an estimate of the number of significant components in a signal, and has been explored for use in the EEG domain in [44], where the complexity of a signal is quantified by the so-called "KL-complexity", defined as the number of singular values with variances higher than 1% of the first singular value. A similar example was presented in [60], which reviews a set of complexity measures. One of those measures include a quantity termed the "Fractional Spectral Radius" (FSR), which is defined as the power of the data projected onto the first j eigenvectors, as a percentage of the total signal power (in the paper he defines $j = 1$). This is similar to the "KL-complexity" term from [44] in that it quantifies the degree of "spreading" of power amongst the singular values. In more recent work [52; 36], more sophisticated measures of complexity have been utilised which are better able to exploit the detailed structure of the singular spectra of EEG signals. We will use two complexity measures which were proposed in these publications. These will be described next.

5.2.1 Characterising the singular spectrum: information theoretic measures

Entropy

From our observations in section 3.1.2, we saw that the structure of the singular spectrum is not random, but rather most of the power, and therefore the information content, is concentrated in the first few singular values (figure 5.9(a) shows this is more detail, in particular the evolution of the singular spectrum in time). In addition, because the SVD decomposition automatically searches for directions of high variance in the data, it acts as a data-adaptive basis which can characterise the structure of the data in a highly efficient way. We can hypothesise that subtle changes in the structure of the singular spectrum due to a shift in the number of degrees of freedom intrinsic in the MEG data is indicative of underlying changes in the brain state of the subject. To characterise this change in the singular spectrum, we use a quantity borrowed from the field of information theory: entropy.

Shannon's entropy, or just entropy, is the classical measure of disorder or uncertainty in a system. It operates on the probability density function (PDF) of the system and is defined as:

$$\begin{aligned} H(x) &= -E[\log(p(x))] \\ &= -\int_{-\infty}^{\infty} p(x)\log(p(x))dx \end{aligned} \tag{5.2}$$

While the singular spectrum is not a PDF in the conventional sense, it is still a distribution function of the power of the signal concentrated in each eigenspace. We may thus apply the discrete version of equation (5.2), to characterise the structure of the MEG singular spectra.

$$H(x) = -\sum_{i=1}^{m} p(x_i)\log(p(x_i)) \tag{5.3}$$

Fisher information

One potential shortcoming of the entropy complexity measure is that it has no notion of the arrangement of the singular values. As a global measure of disorder, entropy is inherently incapable of detecting more subtle changes in the spectrum.

In [36] an alternative quantifier of complexity is used which works on the local gradients of distributions. Known as the *Fisher information*, this measure was devised by the statistician R. A. Fisher in the 1920s. Its original aim was to tell us how easy it is to learn about a probability distribution by sampling from it. Given some parameter θ which describes this

distribution, the Fisher information is defined as [20]:

$$I(\theta) = E\left[\left(\frac{\partial(log(p(x|\theta)))}{\partial(\theta)}\right)^2\right] \tag{5.4}$$

Where $p(x|\theta)$ is the likelihood of some variable x, given the parameter value θ. To make this applicable for our uses, we can assume that the singular spectrum is a set of samples of $p(x|\theta)$ obtained uniformly from the space of θ. We can then find the Fisher information as:

$$\begin{aligned} I(\theta) &= \sum_{i=1}^{n}\left[\left(\frac{\partial(log(p(x|\theta)))}{\partial(\theta)}\right)^2\right] \\ &= \sum_{i=1}^{n}\frac{(p_i'(\theta))^2}{p_i(\theta)} \end{aligned} \tag{5.5}$$

That is to say, we can obtain a "Fisher complexity" measure of the MEG time series by calculating the local gradients along the singular spectrum, then summing over these.

5.2.2 Generation of complexity measures

We now apply the complexity tests detailed above to the characterisation of the MEG recordings. Previous work on complexity has concentrated on the characterisation of raw EEG data. We expect that similar changes may be observable in MEG data though this is not to be assumed since MEG data possesses different noise and signal characteristics. In addition to the conventional complexity approximations, we can take advantage of the ICA separation to study the complexity of individual independent subspaces of the MEG data individually. In this way it is hoped to learn more about the exact components which are responsible for any observed complexity variations.

To generate a running estimate of the signal complexity, the time series to be analysed is first divided into segments. For each segment, we evaluate the entropy and Fisher complexity measures, thus forming new time series showing how the complexity of the signal evolves over time. The determination of the window length is problematic as there is no straightforward way in which an appropriate value may be inferred from the data. In fact, because the signals that we are trying to characterise are likely to be the result of interactions between several generators, each with differing properties and time scales, there is probably no "optimal" window length as such. Rather, a pragmatic choice must be made which will almost certainly depend on the desired application, as well as computing resources. In [61], a window length of 4 seconds (500 samples) is used, with an overlap of 50 samples between successive windows. More recent studies [52] have used shorter time-scales, of the order of 1 second. For our study, we will test four different window sizes: 250, 500, 750 and 1000 (1-4 seconds). We can expect the increasing window lengths to affect a smoothing of the extracted complexity values.

From the results of preliminary testing, it was decided that a window size of 750 (3 seconds) seemed to strike an reasonable balance between temporal resolution and noise rejection. All results henceforth will be generated with this window size. For comparison, a selection of results generated using the other window sizes are provided in appendix E.

In calculating the complexity measures, we follow the methodology prescribed in [52; 36]: first, segments are extracted from the data, with consecutive jumps of 50 samples. As stated above, the length of the segments were varied from 250 to 1000 samples. However, we perform an additional pre-processing step by multiplying each segment with a Hamming window. This has two main effects:

1. It helps to reduce fringe effects due to the discontinuities at the edges of the windows.
2. It assigns a higher importance to the centre of the window, thus helping to improve the localisation of the complexity function. This permits larger window lengths to be used without sacrificing resolution.

Next the singular spectra of each of the segments is calculated, and is normalised to the value of the largest eigenvalue. This alternative normalisation scheme helps to emphasise the structure of the singular spectra and is shown in [22] to be superior to the standard normalisation scheme. Finally, the complexity measures are calculated via the application of equations (5.3) and (5.5), where $p(x)$ and $p(\theta)$ in these equations are set to the normalised singular values. Henceforth, we shall refer to the entropy complexity measure as C_e and to Fisher complexity as C_f.

5.2.3 Results

In terms of the two complexity measures, a general inspection of the results indicated that C_f appeared the noisier of the two, and showed a quicker variation in response to minute changes in the singular spectra (Unexpectedly however, we also note that C_f is surprisingly resistant to artifact contamination). Because Fisher information is based on the local gradients of a distribution, we can expect that C_f will be more sensitive to the fine structure of the spectra, in particular the relationship between the first few singular values where the greatest variation in gradients occur. C_e on the other hand, is a measure of global signal distribution, and is likely to take a larger region of the singular spectra into consideration.

In general, it was noted that the complexity of the recognition data sets were slightly higher than for the idle set (the differences are small but similar values are obtained consistently). Table 5.2 summarises the statistics for the complexities. In addition to the mean values, which were strongly affected by the artifacts, the median values were also calculated (this gave a

much more accurate estimate of the general complexity level for the idle tasks). This indicates that the complexity measures were able to detect subtle changes in the signal content of the MEG recordings. The higher complexity of the recognition data sets points to the increased signal content associated with the more challenging recognition task.

Task	Mean C_e	Median C_e	Mean C_f	Median C_f
Idle	15.6619	17.0338	0.1167	0.1160
Face rec.	17.9939	17.8238	0.1050	0.1048
Scene rec.	17.4717	17.8064	0.1099	0.1084

Table 5.2: Comparison of signal complexities for the three experimental tasks

Figures 5.9, 5.11 and 5.12 show the results of complexity analysis on the idle, face recognition and image recognition data sets respectively. A detailed inspection confirms the above intuition: C_f is noticeably better at tracking variations near the bottom of the contour plot of the singular spectra (where the larger values reside), whereas C_e is more sensitive towards changes in the middle of the spectra. This is because of the weighting of the singular spectrum coefficients imposed by the $p(x)logp(x)$ term (equation (5.3)), which has the effect of emphasising singular values with mid-range variance. Figure 5.8 illustrates this occurrence and shows that the entropy term places a high importance on the singular values in the region of the first "shoulder" of the singular spectrum. A few sample singular spectra have been super-imposed onto the figure to illustrate the degree of variation in the singular spectra in the approximate region where the entropy function is most sensitive. It is clear from the figure how C_e provides an estimate of the number of components above the "shoulder". As these components comprise a significant part of the variance of the signal, it is easy to understand why the entropy measure is known to track the power of a signal closely [52]. In this sense C_e is likely to have similar properties to the KL-complexity measure discussed earlier in that it provides a measure of the number of components that exist in the upper region of the singular spectra.

The complexity analysis is also performed for the independent subspaces. To find the complexity of the independent subspaces, we project the entire raw MEG time series onto the columns of the ICA separating matrix corresponding to the components to be analysed, then applied the complexity measures to these separated signals. While many of the complexity profiles showed little correlation to the opening/closing of the eyes, there were certain notable exceptions. The most obvious are the Alpha-wave associated components, which generally

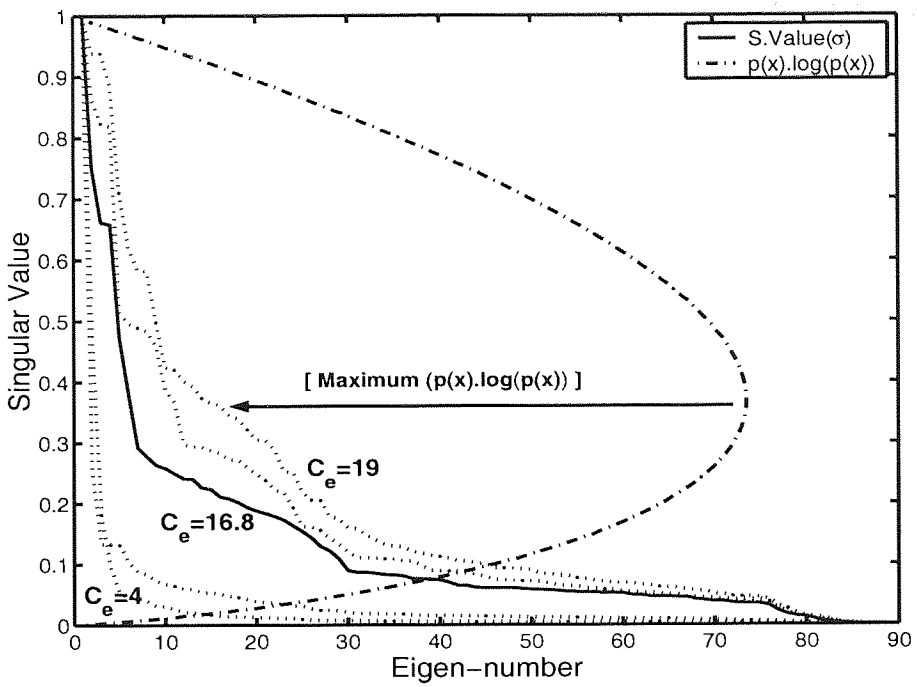


Figure 5.8: Entropy weighting function superimposed over a sample of singular vector plots. The value of C_e for the innermost, outermost and central singular spectra have been calculated and are indicated in the figure.

showed decreased C_e during the eyes closed periods. This reflects the synchronisation, and subsequent increase in signal power of Alpha components when the eyes are shut. Apart from the Alpha components, the only other component of note was IC29, one of the Gamma band components. The complexity profiles for IC29 showed definite correlations to the opening and closing of the eyes. The reconstructed time series, as well as complexity curves for IC29 and IC35 (one of the alpha components), is given in figure 5.10. Figure 5.10(a) shows two time series generated by projecting the entire recording from the recognition task onto the columns of the ICA mixing matrix for ICs 29 and 35, which were gamma and alpha ICs respectively. Apart from the obvious correlation pattern, the complexity profiles of the two independent subspaces provides us with a good illustration of the distinction between C_e and C_f which was discussed above. Firstly, note that the two complexity measures behave in almost opposite ways when applied to these two signals. We observe that C_f appears to be less sensitive to the variation in the alpha components, but better able to detect the complexity variation in IC29. Conversely, C_f seems less sensitive to the change in the alpha component, whereas C_e very clearly reflects the variation in alpha power. One way which we could explain this is by the fact that C_e tends to track the power of a signal a lot more closely than C_f. Looking at the time series in figure 5.10(a), it is clear that the power of the alpha IC increases greatly

during the eye close periods. Also, it is of note that while most of the alpha components had complexities that showed correspondence with the opening/closing of the eyes, only one of the gamma components showed significant correlation. We may hypothesise that this is an indication of a number of different signals or processes being simultaneously active in the gamma band, each with its own functional correlates. This also helps to explain the observed ubiquity of gamma components.

For the two recognition tasks, it was possible to discern some very rough correlation between the complexity measures and the instant of stimuli presentation. However, this was very indistinct and should be interpreted with great caution. Nevertheless, bearing this in mind, we may speculate that the complexity curves do still reflect some underlying shifts in brain function with respect to the recognition tasks. The fact that the variations certainly do not seem to be time locked is not critical as it is unlikely that evoked fields (the type of brain activity which is time-locked to stimuli) are either able to affect a significant influence on the complexities, or that they are even significantly observable in unaveraged data. Our target from the outset, in view of this, should be the search for signs of *induced* brain activity, which includes event related synchronisations/desynchronisations (section 2.3.2). Recall that these are genuine shifts in brain dynamic activity related to external events/stimuli. The critical point is that these signals are known to occur with significantly variable latencies (as well as frequencies) with respect to the stimuli, which might explain the highly tenuous nature of the observed stimuli-complexity correlations.

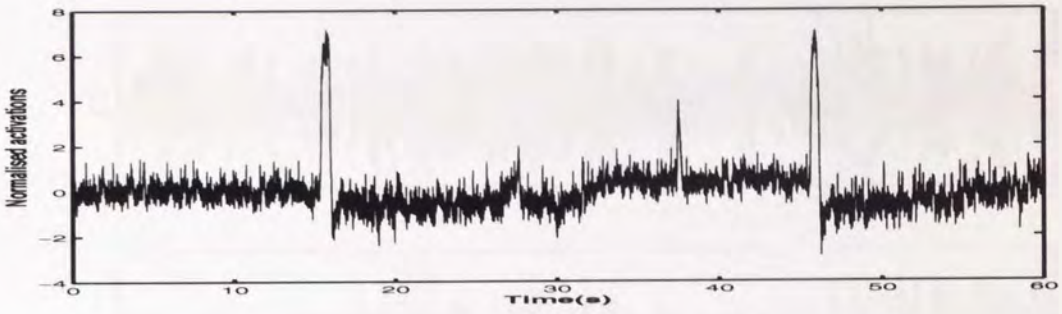
In the final analysis, we apply the complexity measures to the independent components which were extracted from the recognition datasets. The recognition task is assumed to depend on two main faculties: short term memory and feature extraction/binding. These activities have been strongly associated with activity in the gamma and theta bands, so we present results of the complexity analysis on a sample of 4Hz, 7Hz and 40Hz ICs. These are presented in figure 5.13 and 5.14 for the face and natural scene recognition tasks respectively.

From the graphs in the figure, we see that again, as in the case with the raw data, the correlations are extremely tenuous. Unfortunately, the conclusion of this analysis must be that the complexity measures are incapable of satisfactorily tracking complex cognitive events (in this case a visual recognition paradigm). It is possible that the complexity profiles are indeed tracking the induced brain rhythms associated with the recognition task, or that they are monitoring some abstract thought processes but it is unproductive to speculate further about these issues without additional information.

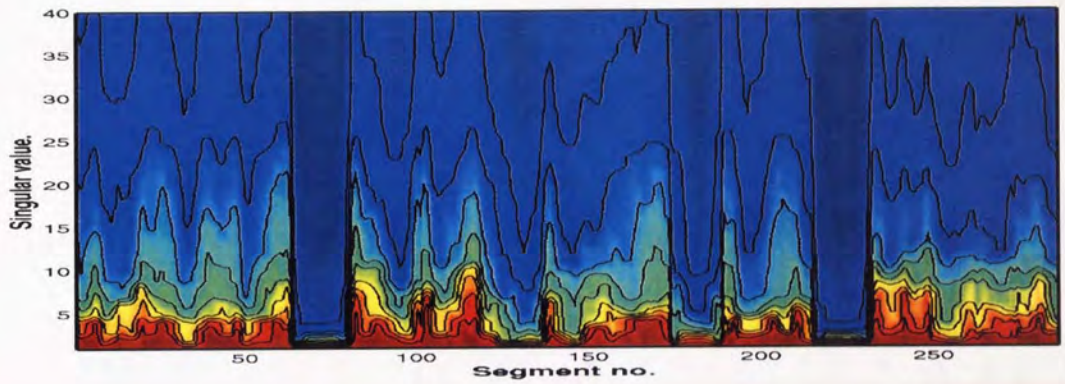
5.3 Summary

The focus of this chapter has been on the qualitative interpretation of both the raw MEG data, and the ICs which were extracted. Two methods are proposed: first, we describe a technique by which an independent component extracted from single channel MEG may be projected back into the multichannel MEG recording space to generate a spatial distribution map. The variance of the projection onto a certain location is taken as an indicator of activity. The study of these maps provides further insight into the functional origins of each IC, particularly when there is correspondence with existing neurophysiological knowledge. The second method covered is the characterisation of MEG via its "complexity". We describe the use of two complexity measures, both based on information theoretic quantities, for this purpose. The first method uses Shannon's entropy term as a means of quantifying the "spread" of power amongst the singular values. This is the most commonly used tool when seeking to describe the amount of disorder present in a probabilistic system. However, it is a global measure and is incapable of detecting local shifts in curvature, for example. The second measure which we will use is the Fisher information measure. This statistical quantity is traditionally used to gain an indication of how much we can learn about a probability distribution by sampling from it. In this project we use it as a pragmatic measure capable of detecting and quantifying minute structural changes in the singular spectra of MEG embeddings. The results obtained from application of each of these methodologies on the MEG are presented and discussed. They indicate that complexity measures extracted from the signal can differentiate between general cognitive states (in our case idle as well as visual recognition brain states). The complexity measures also seem to be a good way of detecting artifacts, a fact which might be used for the automatic labelling of artifactual signals, for example. However, our attempts to correlate shifts in MEG complexity with the cognitive tasks were unsuccessful given the noise and signal characteristics we encountered.

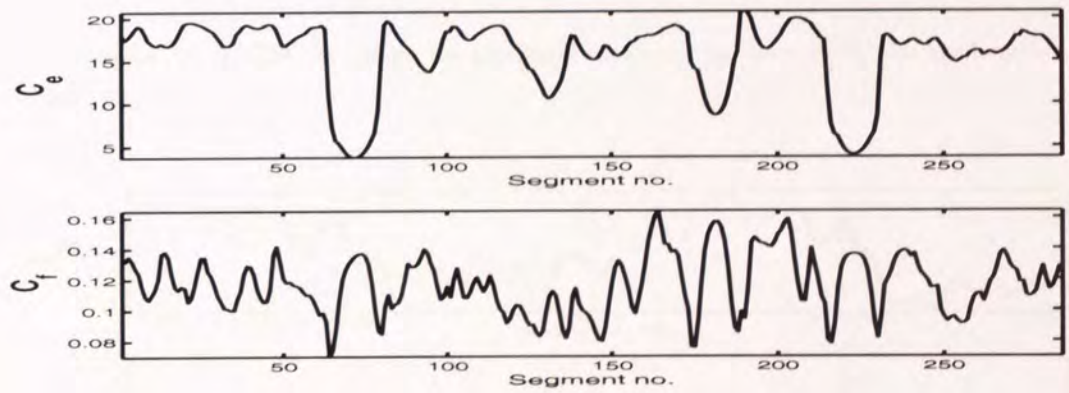
This chapter concludes the applications and results section of this thesis. The following and final chapter concludes the thesis by providing a summary of the key findings uncovered during the course of this project.



(a) Idle recording of subject A

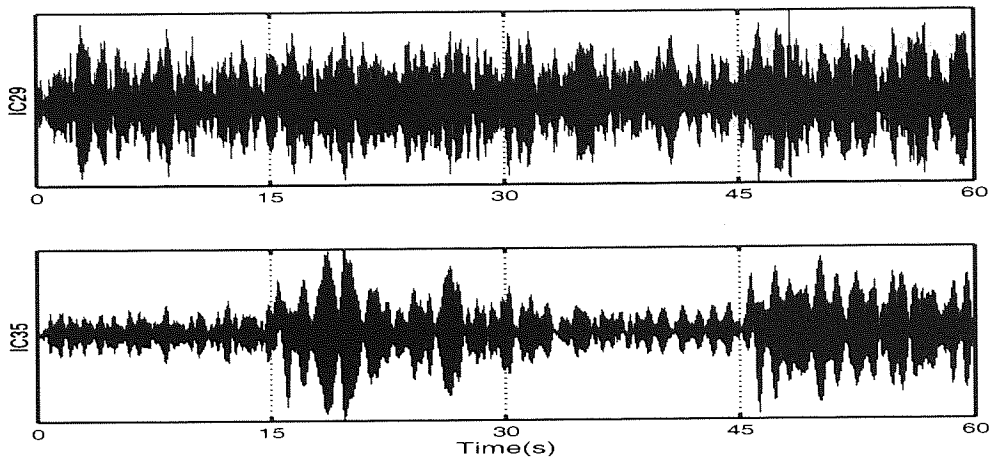


(b) Contour plot of singular spectra evaluated at different time intervals

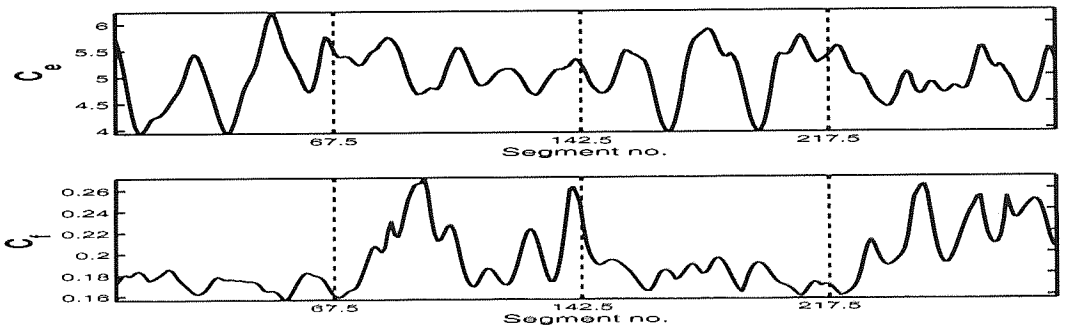


(c) Complexity values corresponding to the singular spectra. The top graph is for C_e and the bottom graph is for C_f

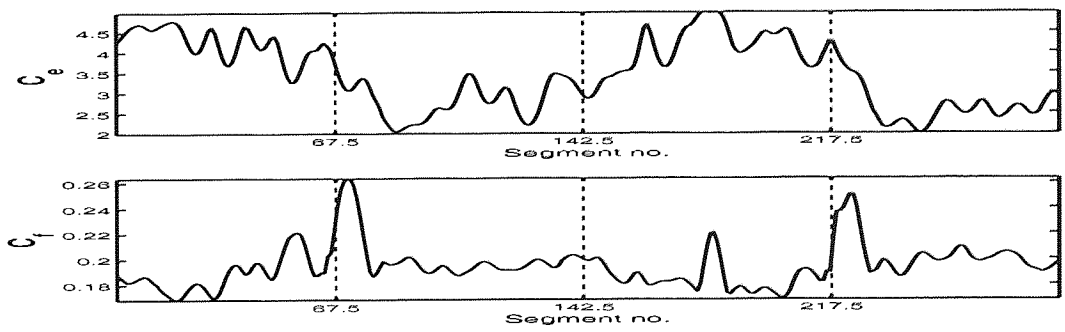
Figure 5.9: Idle recordings: correlation between observable features in the time series, and the evolution of the singular spectrum. The bottom graph shows how this evolution is reflected in the two complexity measures. Both C_e and C_f are able to detect the muscular artifacts (clearly visible in the time series and singular spectrum as well). However, only C_f reflects the more subtle variations in the singular spectrum.



(a) Gamma and alpha components extracted using ICs 29 and 35

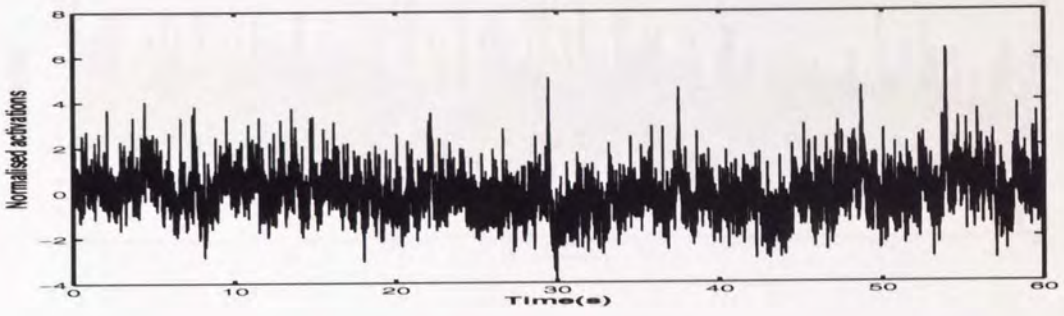


(b) Complexity of IC29. As before, the top figure shows the evolution of C_e and the bottom figure C_f

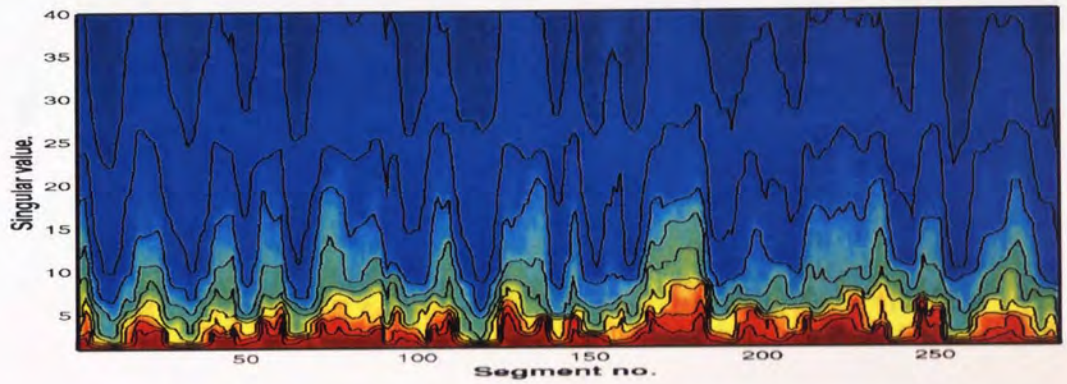


(c) Complexity profiles for IC35

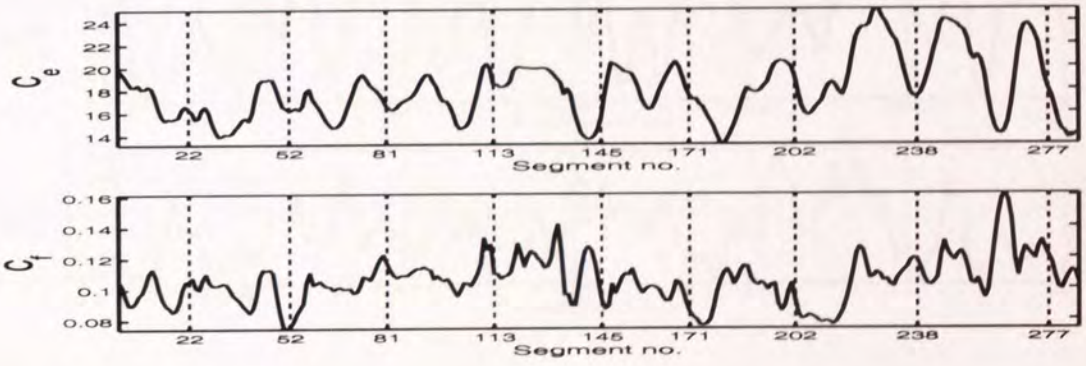
Figure 5.10: Complexity profiles of subspaces corresponding to ICs 29 and 35 of the idle dataset. Dotted lines indicate the time-points where the subject was requested to either closed (first and third line), or opened her eyes.



(a) Idle recording of subject A

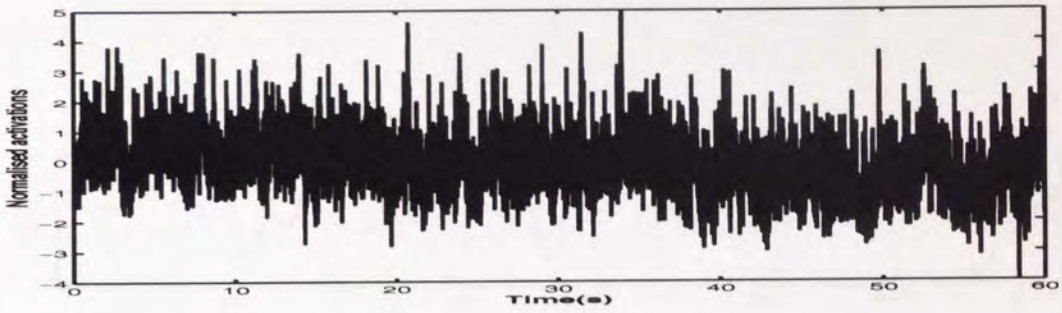


(b) Contour plot of singular spectra evaluated at different time intervals

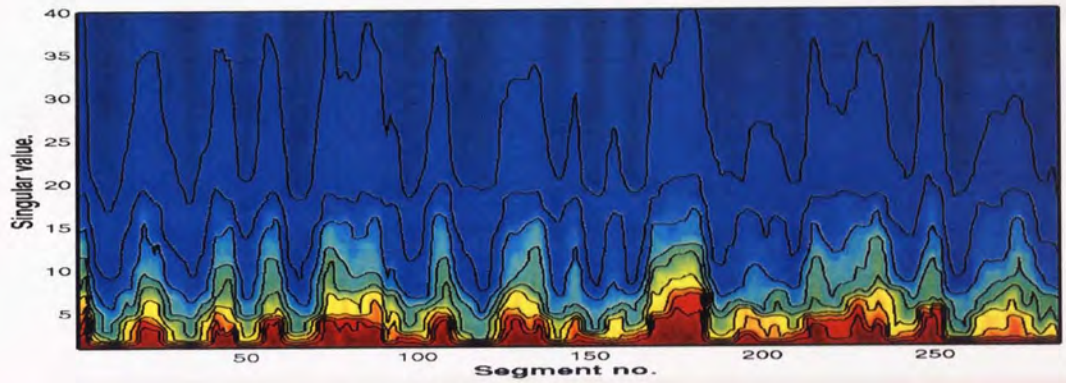


(c) Complexity values corresponding to the singular spectra. The dotted lines indicate time points when the choice of images are presented to the subject

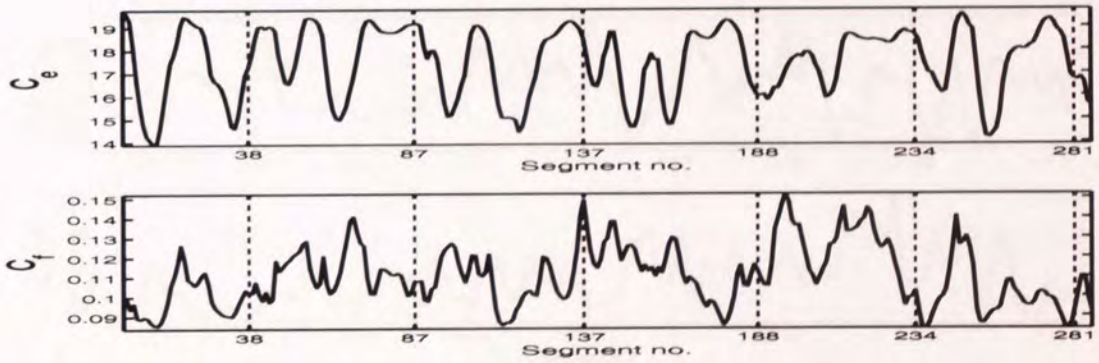
Figure 5.11: Complexity analysis of the facial recognition data. Note that only C_f reflects the more subtle variations in the singular spectrum.



(a) Idle recording of subject A

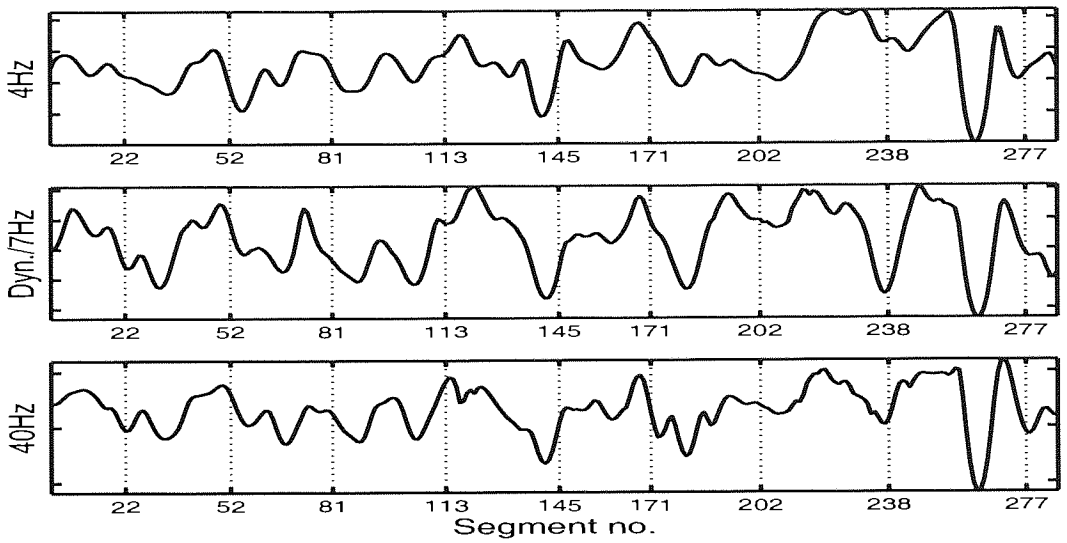


(b) Contour plot of singular spectra evaluated at different time intervals

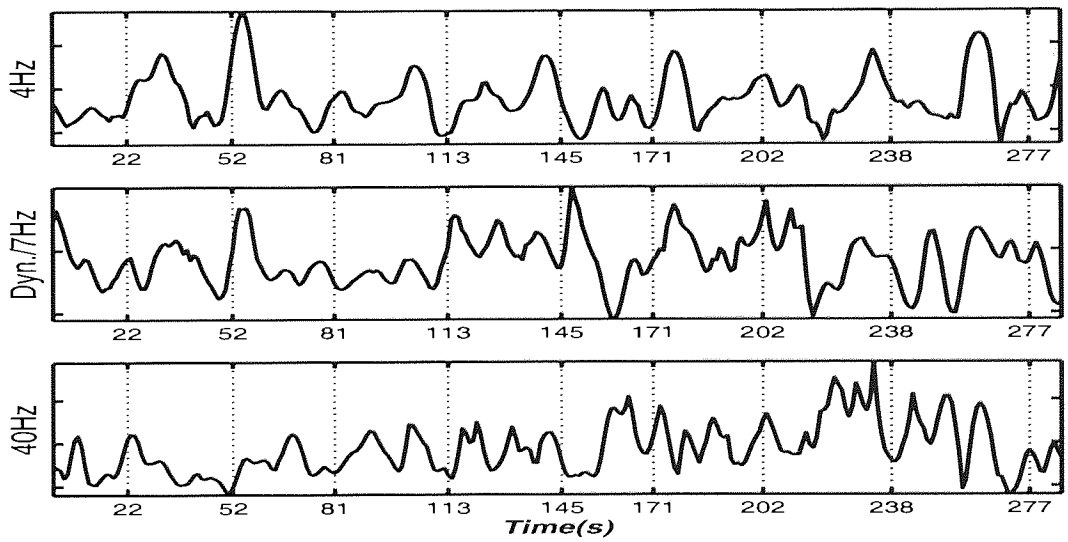


(c) Complexity values corresponding to the singular spectra. The dotted lines indicate time points when the choice of images are presented to the subject

Figure 5.12: Complexity analysis of the natural scene recognition data. Again, the evolution of the singular spectrum, and consequently the complexity measures only approximately tracks the recognition task. C_e seems to show a closer correlation to the stimuli but this is not clear and should be interpreted with caution.

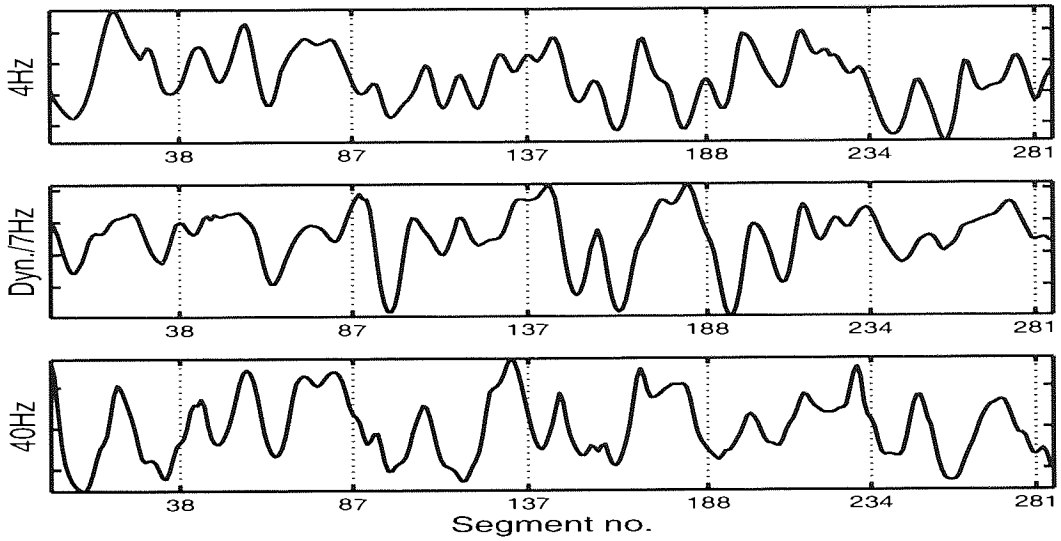


(a) Entropy

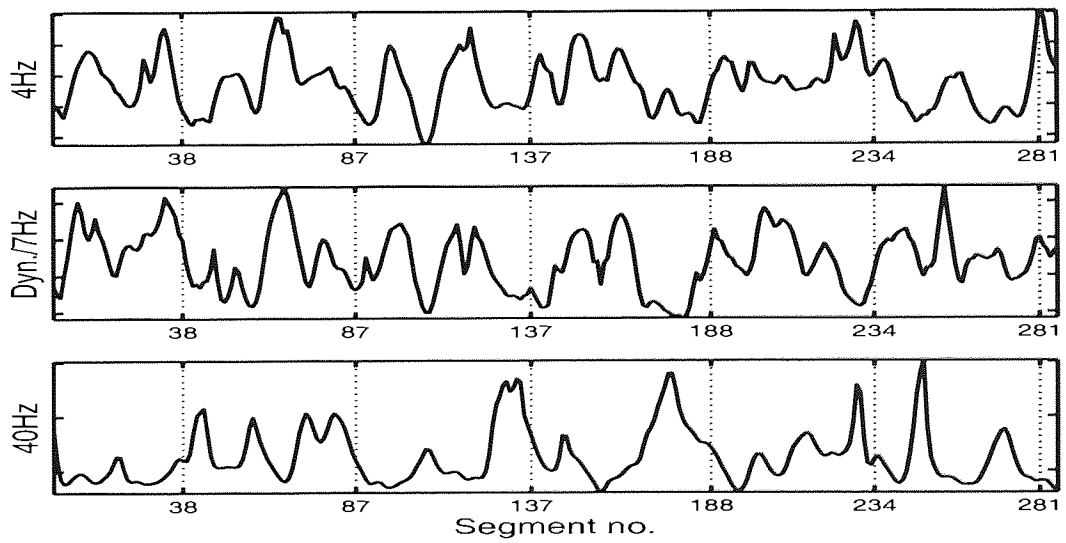


(b) Fisher information

Figure 5.13: Complexity profiles of independent subspaces from face recognition data. Dotted lines mark stimuli time instants.



(a) Entropy



(b) Fisher information

Figure 5.14: Complexity profiles of independent subspaces from natural scene recognition data. Dotted lines mark stimuli time instants.

Chapter 6

Conclusion

6.1 Discussion

6.1.1 General overview

This thesis has studied the use of a variety of novel signal processing techniques in the analysis of MEG data. We demonstrated that through the application of these techniques, it was possible to extract information, and derive results which were previously unreported in the literature.

The main feature differentiating this work from existing studies is its exclusive use of single channel, unaveraged MEG data. Analysis of MEG in this manner is completely unique and while the treatment presented in this thesis is largely preliminary in many respects, it already shows the rich promise that this direction of research holds. The traditional approach to analysing MEG data has depended largely on dividing MEG activity into frequency bins, the delineations of which are often somewhat arbitrary. This approach may encounter problems for several reasons: firstly, the MEG generators are not necessarily linear. In fact, the surrogate test results presented in chapter 3 indicate that the MEG observables are likely to consist of contributions from a combination of linear, nonlinear and stochastic components. If this is the case, frequency based analysis techniques are unlikely to be sufficient to explore and exploit the properties of the MEG recordings. An additional observation is that several classes of MEG oscillatory behaviour may possess extremely similar frequencies, which may vary between subjects as well as cognitive function explored. In cases such as this it is clearly difficult to design proper digital filters based solely on prior information about the frequency content of the sources we are interested in.

The work described in this thesis has investigated the use of methods derived from dynamical systems theory in the study of MEG recordings. The fundamental concept when working in this framework is the idea of underlying degrees of freedom which interact to give rise to the observed data. It follows that we can envisage the existence of an unobservable system manifold on which the brain dynamics evolve, and the formulation of a range of signal processing tools aimed at reconstructing and characterising this manifold. As the human brain is almost certain to contain elements with deterministic dynamics, this interpretation is likely to be accurate, at least at some fundamental level of abstraction. Therefore, the big question is not the existence of this manifold, but whether knowledge regarding its properties can be inferred given available data and resources, as well as exploited in a practical way. In the broadest sense, this thesis has set out to find an answer to this question.

The basic tool which we have depended on in this work is the delay embedding approach. Essentially, this involves the construction of a *trajectory matrix*, the columns of which consists of windows extracted from the scalar time series that we wish to study. Though simple, this procedure is of crucial importance to us as it provides a means of reconstructing the underlying system manifold given only a scalar observation, essentially serving as a link between the real world data and the hidden system dynamical properties.

The next step in our analysis was the study of individual degrees of freedom in the MEG. A raw state space reconstruction using embedding is of little practical use due to its extremely high dimensionality. We chose to partition our data linearly, using the independence condition as an optimisation criterion. This allowed us to use established ICA algorithms, which have been developed to decompose multivariate data sets into independent subspaces. In our case, because we are analysing the embedding matrices, ICA served as a means of generating customised, data adaptive filters capable of isolating and extracting underlying sources in the data. In this sense, the use of the embedding approach served a dual purpose:

1. Many signal processing techniques, including ICA, require the use of multichannel datasets as inputs. Thus, the delay embedding procedure serves as a transformation from a scalar to a multivariate data space, allowing the application of existing algorithms and techniques on single channel data. Furthermore, using the embedding matrix has the additional advantage of automatically incorporating temporal information regarding the MEG data.

2. As was demonstrated in [70], a well constructed delay trajectory matrix is an embedding of the system manifold. Hence the use of the embedding methodology raises the possibility of using tools adopted from dynamical systems theory to the analysis and characterisation of the underlying system attractor.

The application of ICA allowed the decomposition of the reconstructed MEG state space into a number of distinct component classes. Amongst these we were able to identify classes that corresponded to the traditional classes of EEG signals such as alpha and theta waves. However, far from being a justification of frequency based techniques, our observations indicated that many of the MEG signals possessed complex and multi-modal frequency spectra. Hence it is important to remember that while we made an attempt to classify the components into discrete classes, the reality is that the extracted components manifested a wide variety of dynamics and morphologies. To address this issue, we propose a method for projecting the components into a visualisation space, using Sammon Mapping. This allowed us to visualise the relationships and degree of similarity between components, which we used to subjectively group them into categories containing similar sources. To permit corroboration with existing neurophysiological knowledge, we imposed a subjective classification of these components into the familiar EEG classes. This allowed us to work with a manageable amount of data especially when studying the spatial distributions of these components. We were able to find convincing correspondence between these spatial distributions, and theoretical knowledge. An interesting example is the Alpha components. The spatial maps showed that the power of this component was clearly concentrated in the posterior of the head. It is particularly notable that we were able to extract this component from a channel in the right-temporal area, then obtain a theoretically consistent spatial distribution showing its assumed origin at the rear of the head.

However, at this preliminary stage, we do not attempt to speculate on the detailed functional or neurophysiological implications of these results, merely to offer the reader a peek at its potential significance. We note that while the subject of choosing "task-relevant" components is still highly subjective, this is a current problem which plagues all ICA applications to neurophysiological data and is not restricted to just our methods. We do feel, however, that the topographic projection of the components into the Sammon visualisation space is a positive attempt at introducing a framework in which these sources can be more effectively characterised. Nevertheless, it must be emphasised that our main aim from the start has been to draw attention to the amount of information that can be extracted from single channel MEG if one adopts the crucial paradigm shift from analysis of the time series data itself, to studying the actual hidden dynamics which we know must exist.

6.1.2 Chapter summary

As detailed in section 1.4, the aim of this thesis has been to address a number of key issues. To recap, these were:

- the analysis of the data as an unobservable dynamical system,
- the application of methods that allow for nonlinear dynamics,

- the use of single channel unaveraged MEG,
- the development of methods for noise reduction and artifact removal,
- the application of topographic projection methods for clustering and visualisation.

As far as the achievement of these aims are concerned, we may conclude that this has been satisfactorily accomplished. In general, the main task of formulating and analysing the MEG problem in a dynamical systems framework was successfully undertaken. In more detail, this thesis has targeted a number of key points:

In **Chapter 3**, we presented the problem of MEG analysis within a dynamical systems framework. The issue of embedding as a means of augmenting single channel recordings is described and we examine the issue of determining suitable embedding parameters, namely the delay lag and window length of the trajectory matrices. In addition, this chapter also details the application of a series of surrogate data tests on the MEG data. The aim was to demonstrate the presence of nonlinear generating components. The results of these tests, which indicated the presence of nonlinearities, were important as they provide support for the use of nonlinear signal processing tools.

Chapter 4 described the main components of analysis on the MEG data. The main results from ICA decomposition of MEG are reviewed in this chapter. We start by verifying the ability of ICA to separate sources from an embedding of a scalar time series. For this we generated a realistic toy MEG data set using a linear combination of sinusoidal signals in the typical alpha and theta frequency bands and clean ECG signals. This "MEG" time series was then combined with MEG background noise at varying levels to generate the final time series. ICA was then applied to the trajectory matrix formed from this and the results confirmed the ability of ICA to separate these underlying signals. Next the analysis was performed on real MEG data.

Chapter 5 presented the results of further experiments to determine the qualitative properties of the extracted ICs. One major problem when carrying out this sort of analysis is that very little is actually known about the nature of the extracted ICs. Beyond the fact that they might reflect some underlying state variable, it is unknown what some of these signals might represent. In this chapter we demonstrate the generation of spatial distribution maps for each of the ICs via the projection of the multichannel MEG recordings onto ICA weights, which we now use as digital filters to analyse recordings in other channels. A further technique which is applied is the application of measures of complexity on subsets of the ICs. The idea is that the complexities of these independent components are able to characterise the underlying manifold on which they are assumed to evolve. The ability of this analysis to distinguish between closed- and open-eye brain states, as well as to track recognition events

was tested.

6.2 Future work

As MEG is such a new and potentially powerful tool for MEG analysis, the potential for further research in this area is enormous. Similarly, there has been a corresponding explosion of interest in the general class of nonlinear analysis tools derived from the fields of neural networks, dynamical systems and statistics. In applying a subset of these tools to the analysis of MEG in a novel fashion, this work has been an attempt to bridge the gap between these two areas of research. While we feel that we have made some headway in this direction, there undoubtedly exists numerous avenues in which the research described in this thesis may be improved and extended. Possible extensions may be divided into two categories: additions and modifications to the methodology itself, and the development of applications for the proposed techniques.

6.2.1 Methodological extensions

One interesting aspect which we have been unable to examine further due to limited time is the modelling of the dynamical properties of the MEG data. The standard ICA model which we have adhered to only utilises properties of the static distribution of the data when searching for optimum directions. Fortunately, because of our use of the embedding approach, using standard ICA already incorporates some "built-in" ability to exploit temporal properties of the data. However, more effective analysis might be possible if an explicit generative model which can exploit temporal relationships, or model the actual dynamical evolution rules, is used.

Another issue which can be considered for future studies is the use of more than one channel. As has been stated a number of times in this thesis, we restrict ourselves to using just single channel data throughout this project (with the exception of the spatial topography studies) with the aim of validating the dynamical systems framework which we have chosen to adopt. By restricting ourselves in this way we intend to demonstrate the amount of information that can be extracted from just a single channel of MEG, and in this way draw attention to the redundancies inherent in enormously expensive multichannel machines. However, while in theory a single scalar observation channel is sufficient to extract a complete representation of the underlying dynamical system, the practical reality is that noise and signal attenuation due to spatial separation will of course have an effect on any method. The methods that we have described here can be extended to use with data from two or more channels, if this is

required in the actual application, if this is likely to help overcome some of the difficulties due to noise and artifacts.

6.2.2 Practical applications

The thesis has primarily focussed on the specification and demonstration of the general embedology-based framework in which MEG signals are to be analysed. As such, the development and testing of real practical applications of the methods have only been given limited attention. Nevertheless, it is easy to foresee a range of practical scenarios in which this work might be applicable, and which, given sufficient time and resources, it is likely to prove both interesting and beneficial to explore.

As this thesis focuses on the use of single channel MEG data, it is hoped that the extension and development of methods described in this thesis will eventually lead to devices based on MEG technology which contain if not one, at least a greatly reduced number of measurement channels. Such devices would undoubtedly cost a lot less to procure and maintain, as well as be a lot more appropriate for real-time, field applications. Hitherto, the drive has been for devices with ever-increasing numbers of channels, as this has been the primary way of gaining additional information regarding brain function. However, the unfortunate side effect of this trend is that MEG devices have become hugely expensive, restricting advanced research in MEG to large institutions. This is unfortunate as the advantages that MEG possesses: superior spatial resolution, as well as sensitivity to complementary areas of the cortex (relative to EEG), mean that it is certain to be a powerful addition to existing methods if only existing technical difficulties could be overcome.

Possible areas of application for single channel MEG are:

- Development of **brain computer interfaces**. A number of studies have been conducted on the use of EEG as an interface for the conversion of electrical activity recorded at the scalp to usable information for control purposes (examples are [59; 61]). Research in this area has the potential to provide exciting new tools which can help individuals with limited mobility achieve greater freedom. In view of the obvious practical limitations of MEG hardware, MEG versions of these applications have not been considered.

- Another popular application domain where EEG has been receiving a lot of attention is in **vigilance monitoring** [23; 63].The challenge in this field has been the development of methods for automatically monitoring the mental state of human operators, for example airplane pilots, for signs of fatigue and lapses in vigilance. A related application is the monitoring of brain-state during training activities to monitor the effectiveness

of training routines. As a system that provides a continuous and automated measure of this kind has the potential to greatly reduce the likelihood of serious errors, this is obviously an area which holds much promise. Unfortunately, MEG is again not considered as a practical alternative for EEG in these applications.

Appendix A

Test Images

Figure A.1: Face test images. *Courtesy: Yale faces database*

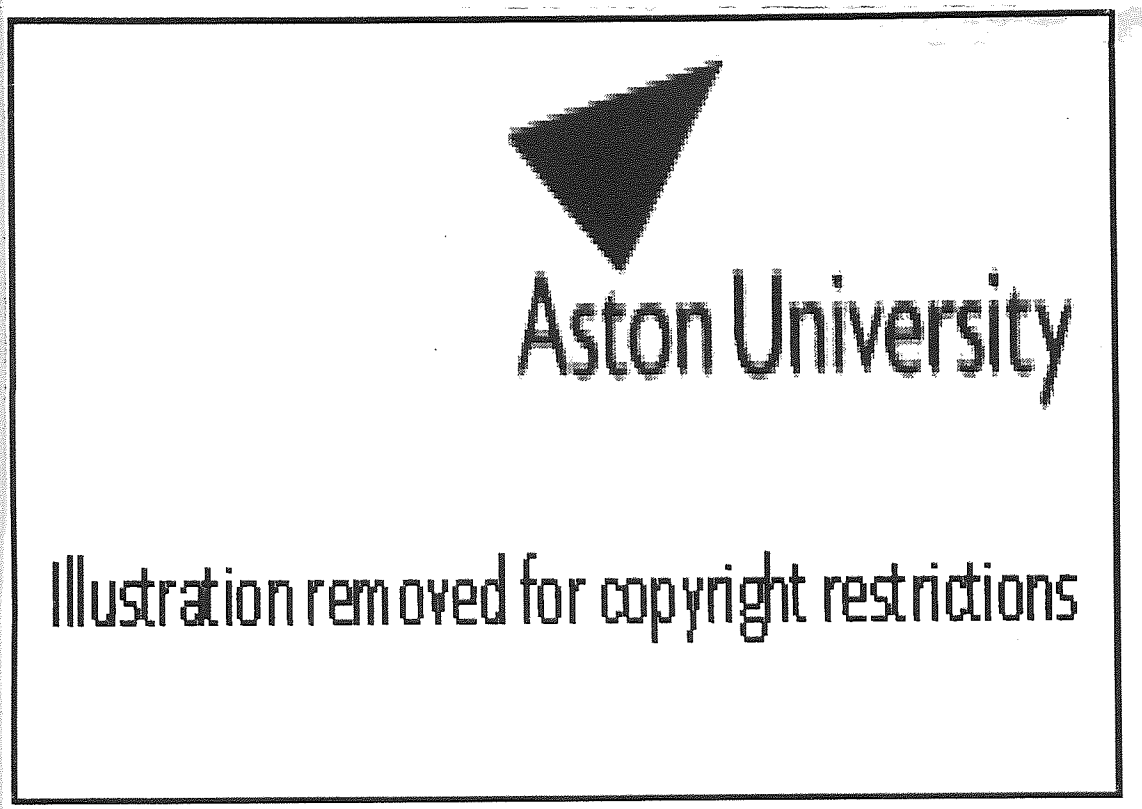


Figure A.2: Natural scenes test images. *Courtesy: Groningen natural scenes database*

Appendix B

Surrogate testing: additional results

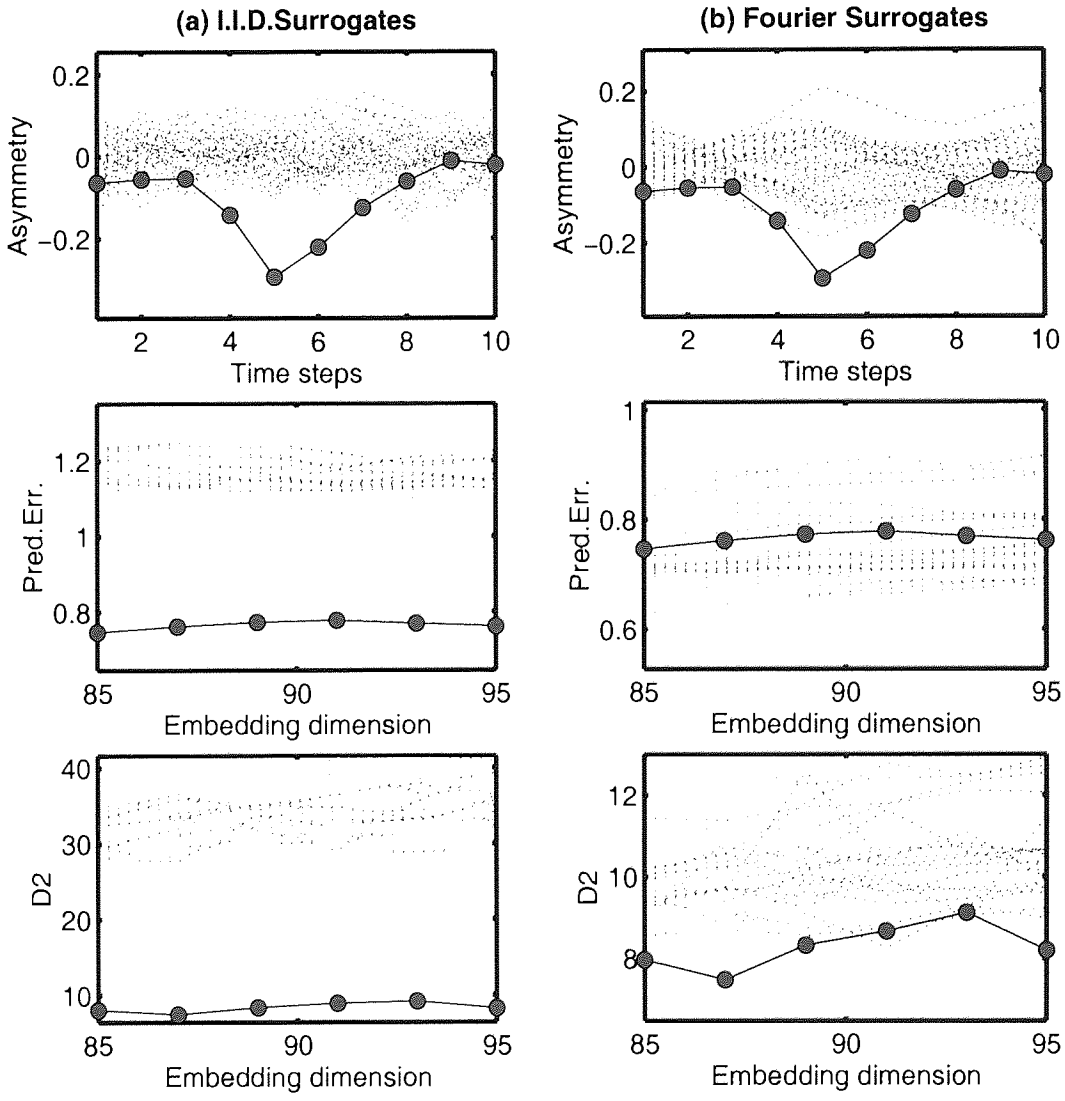


Figure B.1: Surrogate test results for MEG recordings from subject A: idle task.

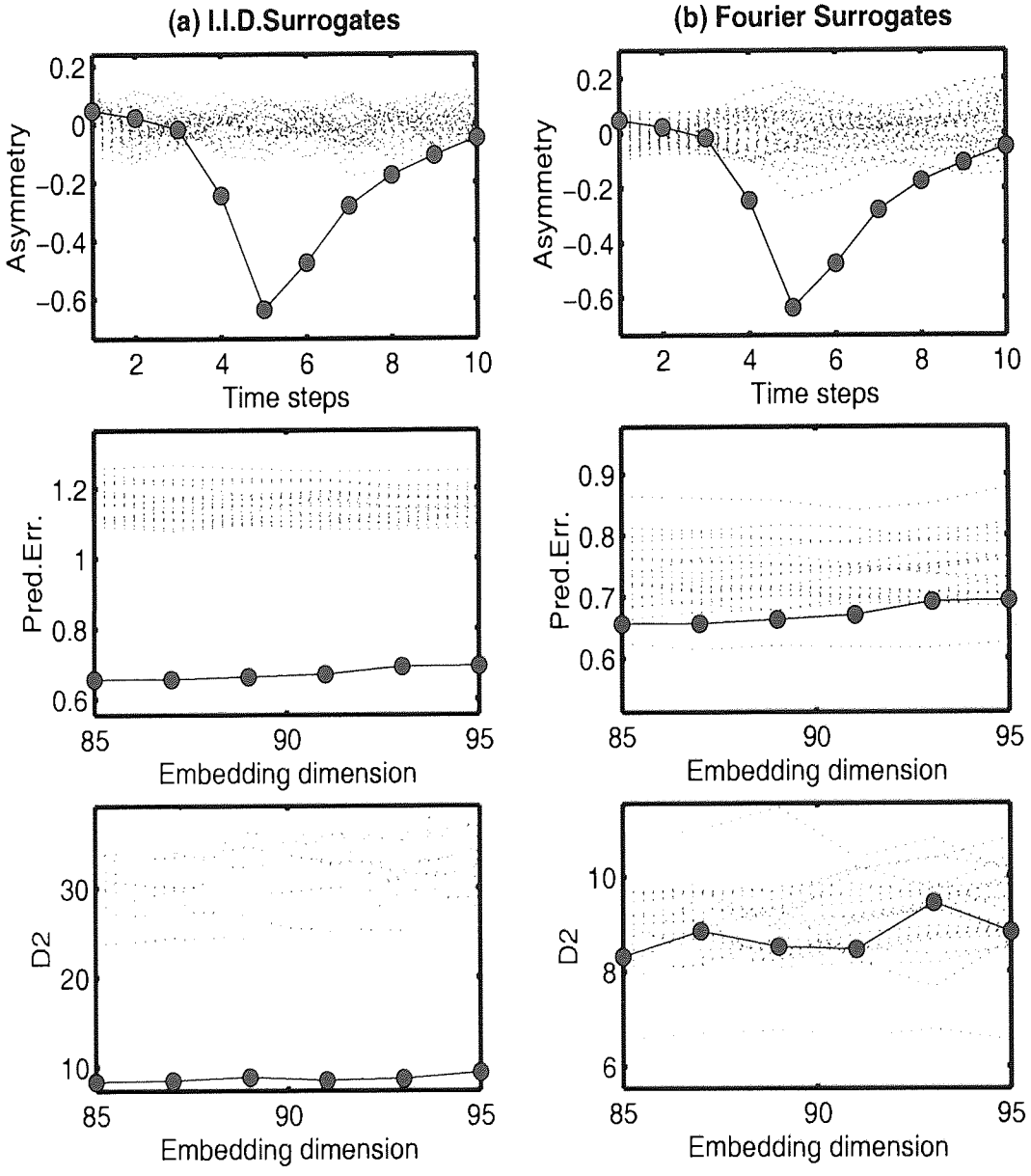


Figure B.2: Surrogate test results for MEG recordings from subject B: idle task.

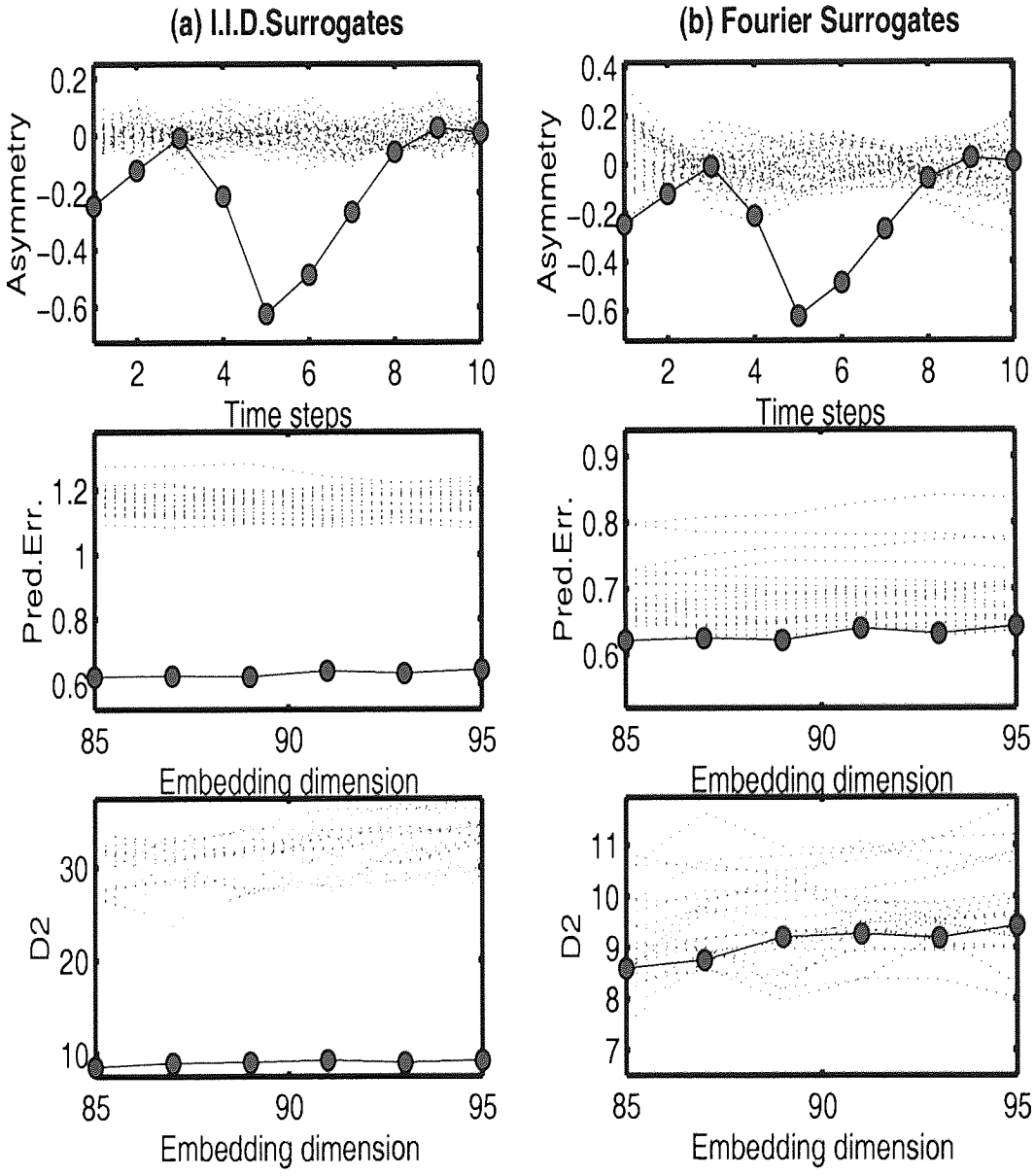


Figure B.3: Surrogate test results for facial recognition MEG recordings. Subject A

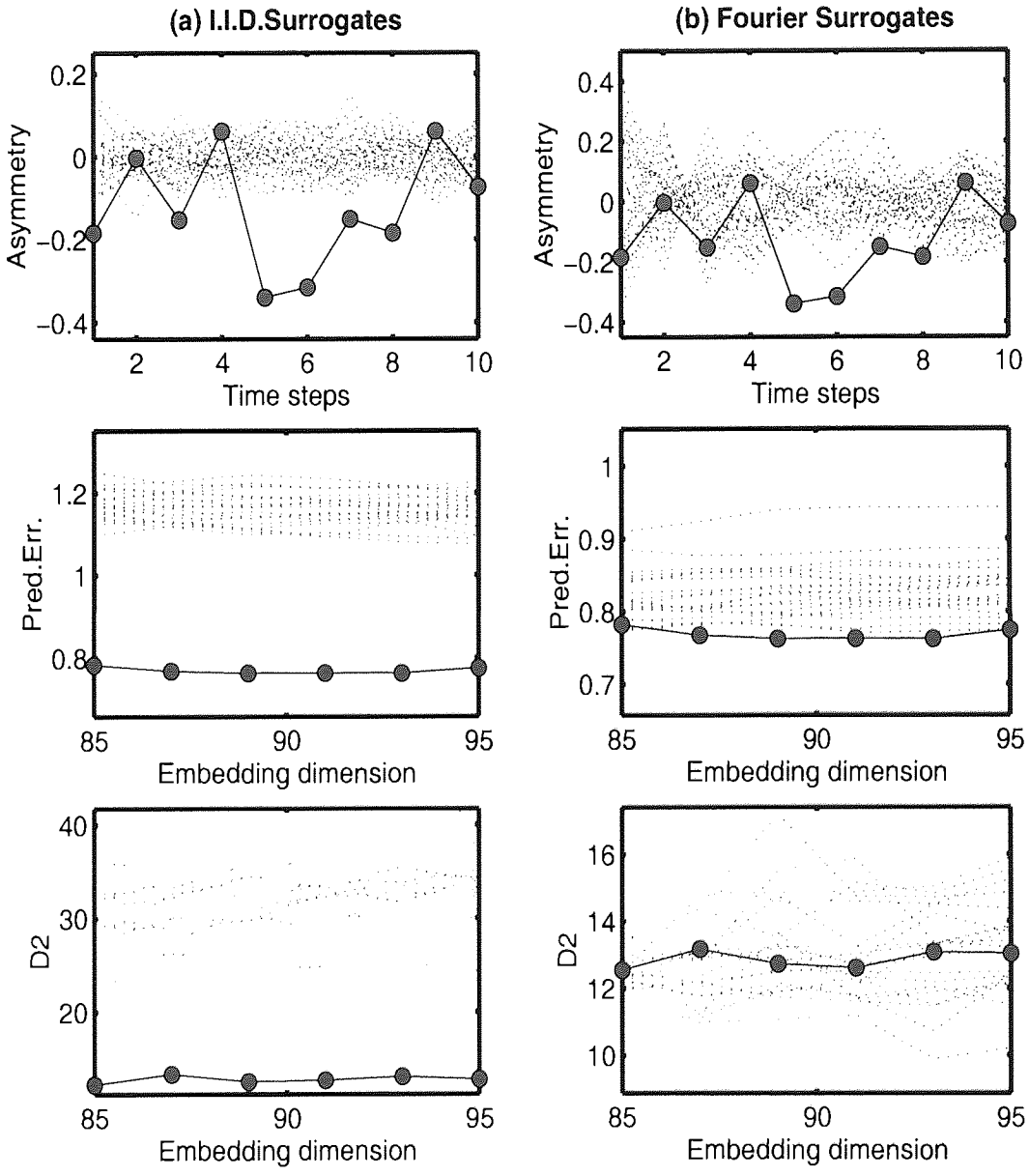


Figure B.4: Surrogate test results for facial recognition MEG recordings. Subject B

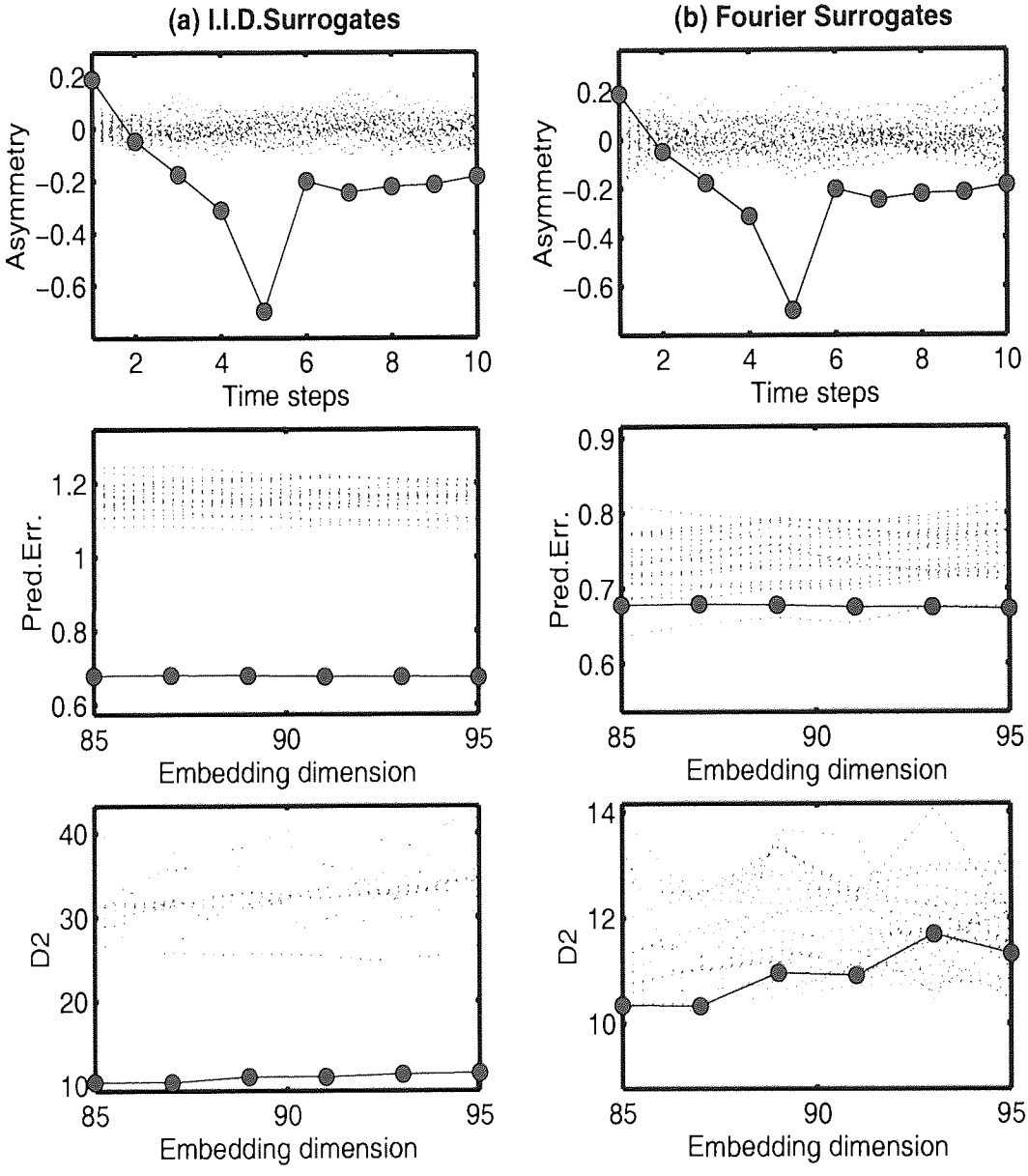


Figure B.5: Surrogate data test results for natural scene recognition MEG. Subject A

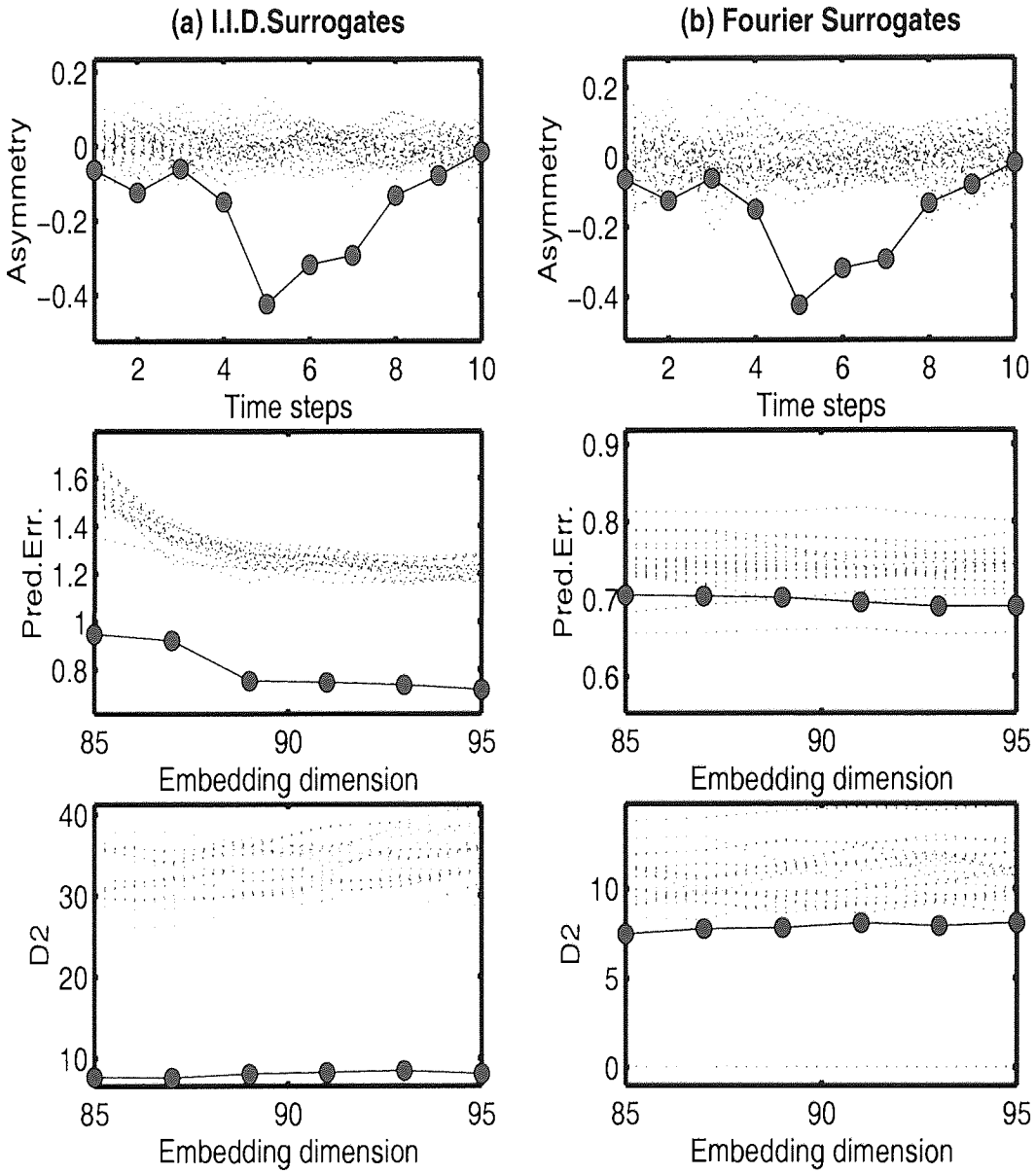


Figure B.6: Surrogate data tests performed on natural scene recognition MEG. Subject B

Appendix C

Learning algorithms

C.1 ICA contrasts

In section 4.1.3 we briefly described a number of approaches which are used to perform ICA. There exist many methods for this purpose but most of them frame the problem as a nonlinear optimisation problem. The aim then is to minimise some cost function which quantifies the redundancy of statistical dependence between variables. In this section we discuss a number of these approaches in more detail.

Maximum likelihood and Infomax

The most direct method is to optimise $P(\mathbf{x}|\mathbf{W})$. If a noisy model is assumed, the integral can be difficult to perform analytically but it is possible provided suitable assumptions are made. One notable case is *independent factor analysis* [3], in which the sources are modelled as mixtures of gaussians and the observations are assumed to have been corrupted with gaussian noise. The use of gaussians and mixtures of gaussians permits an analytical solution to be found for equation (4.5).

A method which is based on similar principles is the infomax algorithm [7]. This works by applying a nonlinear function to the output of the linear transformation:

$$\mathbf{z} = g(\mathbf{W}\mathbf{x}) \tag{C.1}$$

The nonlinear function $g(.)$ is assumed to be the cumulative distribution function (CDF) of the source distribution $p_i(s_i)$. Hence, to ensure that the linear rotation extracts components with the same distribution as the sources, ICA is performed by maximising the entropy of this output with respect to the rotation matrix. When the entropy is maximised, the outputs

$\mathbf{z}$ should have uniform distributions, indicating that $p(\mathbf{Wx}) = P(\mathrm{s})$. Further details on this, as well as proof that the infomax algorithm is equivalent to performing maximum likelihood, can be found in [14].

Mutual Information

Independence may also be quantified in terms of the mutual information between separated sources. This is defined as the difference between the joint entropy of the sources and the sum of the marginal entropies.

$$M.I.(y) = \sum_{a=1}^{n} H(y_a) - H(\mathbf{y}) \tag{C.2}$$

This expression is a direct measure of the statistical redundancy within multivariate datasets. It is also trivial to show that this is closely related to the KL-divergence between the joint and product-of-marginal densities of the variables:

$$\begin{aligned} KL(p(\mathbf{y})||\prod_{a=1}^{n} p(y_a)) &= \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{a=1}^{n} p(y_a)} dy \\ &= \int p(\mathbf{y}) \log p(\mathbf{y}) dy - \int p(\mathbf{y}) \prod_{a=1}^{n} p(y_a) dy \\ &= \sum_{a=1}^{n} H(y_a) - H(\mathbf{y}) \end{aligned} \tag{C.3}$$

Recall that from equation 4.1, full independence between the separated sources is achieved when the joint and product of marginal densities are equal, in which case the KL-divergence between these two will be zero. However, in practice this is often either not exactly true (due to noise or spurious correlations), or difficult to solve analytically. The approach that is often taken to overcome this is to use the mutual information term as a cost function and to minimise this using nonlinear optimisation tools. Further details on a method which follows this approach can be found in [2].

High order cumulants

Another popular approach for performing ICA is via maximising the high order statistics of the data. To show how this method also leads to independence between the sources, consider the expression for mutual information given in equation (C.2). The right side of the equation consists of two terms, $\sum_{a=1}^{n} H(y_a)$ and $H(\mathbf{y})$. Provided that the norm of the de-mixing matrix is constrained to one, the second term, the joint entropy of the output, is a constant since it depends only on the input (entropy is invariant under a linear rotation). This leaves us with the first term which is the sum of the marginal entropies. As our aim is to minimise the

mutual information, any solution that minimises this term will also result in independence. For a given variance, a gaussian distribution has the highest entropy compared to any other probability distribution. Hence, the independence of the output components can be obtained by forcing the marginal distributions to be as *non-gaussian* as possible. As the high order cumulants of a gaussian distribution are zero, the simplest way of achieving this aim is by finding directions in the data space that maximise these cumulants.

Examples of ICA algorithms that use the high order moments are the JADE algorithm [15] and FastICA, which is described in greater detail in the main text.

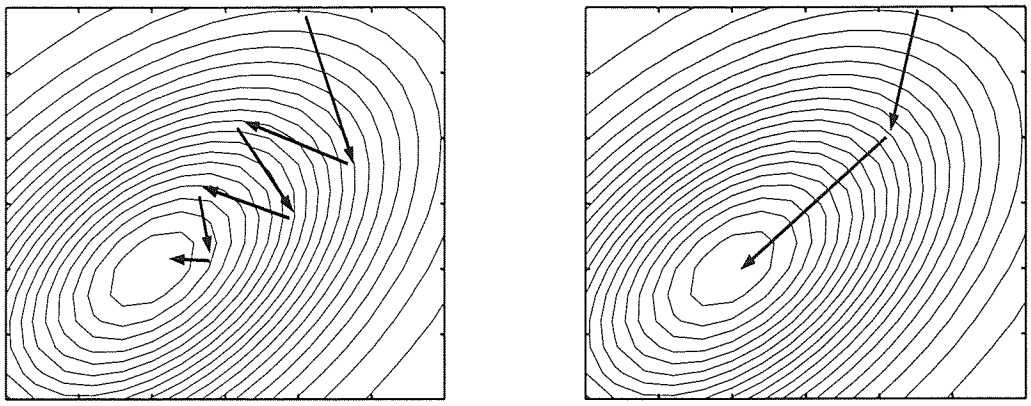
C.2 Scaled conjugate gradients algorithm

The scaled conjugate gradients algorithm (henceforth referred to as SCG) is a nonlinear optimisation algorithm designed to find the minimum of a locally quadratic error function. In this project we use it to find an optimum of the Sammon stress function, using the coordinates of the mapped points in the Sammon visualisation space as parameters. It was chosen based on its speedy convergence and stability and is suitable for a range of training scenarios. The basic idea behind its operation is the traditional gradient descent principle, but SCG attains significant performance improvements by using the principle of conjugate gradient directions.

The idea is to optimise the parameters of a nonlinear model by performing a series of successive minimisations along *conjugate* directions in the error space. Under traditional gradient descent, optimisations were performed along the direction of steepest descent, i.e.: the direction corresponding to the vector gradient of the error function. However, optimising along the gradients of the error function often leads to slow convergence as subsequent search directions often cancel out improvements made in previous steps (see figure C.1(a)). The challenge is to find an optimisation direction along which the component of the gradient parallel to the *previous* direction remains zero throughout the next optimisation step. Two search directions which are orthogonal in this fashion are called *non-interfering*, or *conjugate*. Quite fortunately, it turns out that a solution exists which allows each new search direction to be conjugate not only with respect to the previous search direction, but also to all previous directions. This is achieved by finding a set of directions which fulfil:

$$\mathbf{d}_j^T \mathbf{H} \mathbf{d}_i = 0 \quad \forall j \neq i \tag{C.4}$$

where $\mathbf{H}$ is the Hessian of the error function. Provided $\mathbf{H}$ is positive definite, it turns out that the set of vectors $\mathbf{d}_i | i = 1..W$ (W is the dimensionality of the weight space) form a complete basis set in weight space. Suppose we wish to find an optimal parameter value $\mathbf{w}^*$,



(a) Gradient descent

(b) Conjugate gradients

Figure C.1: Comparison of parameter optimisation using traditional gradient descent and a conjugate directions based method

starting from some initial value, $\mathbf{w}_1$, we can write the difference between these two as a sum of a series of steps in these conjugate directions:

$$\mathbf{w}^* - \mathbf{w}_1 = \sum_{i=1}^{W} \alpha_i \mathbf{d}_i \tag{C.5}$$

Hence, the objective in training step i is to find α_i, which defines the step length. Due to the conjugacy property of the $\mathbf{d}_i$ directions, it is possible to obtain a solution for each α_i using only terms from time step i. Thus, for time step j, the solution is given by [10]:

$$\alpha_j = -\frac{\mathbf{d}_j^T \mathbf{g}_j}{\mathbf{d}_j^T \mathbf{H} \mathbf{d}_j} \tag{C.6}$$

where $\mathbf{g}_j$ is the error gradient at time step j. In practice, a suitable procedure is to start by setting the first optimisation direction to be the negative gradient. Successive directions can be found as a linear combination of the previous direction and the gradient at the new location:

$$\mathbf{d}_{j+1} = -\mathbf{g}_{j+1} + \beta_j \mathbf{d}_j \tag{C.7}$$

where β_j is a weighting coefficient obtained by applying the conjugacy condition, giving:

$$\beta_j = \frac{\mathbf{g}_{j+1}^T \mathbf{H} \mathbf{d}_j}{\mathbf{d}_j^T \mathbf{H} \mathbf{d}_j}. \tag{C.8}$$

Thus far this has been a description of the general conjugate gradients framework. The

SCG algorithm extends and generalises this for use in a practical optimisation problem by incorporating two key adaptations:

1. In practice, the error function will most likely not be a quadratic (though the local function may still be closely approximated by one). This necessitates the calculation of the Hessian at each time step, which is computationally expensive. However, it turns out that β_j can be approximated at each time step without explicitly using the Hessian. This is achieved via the application of the *Polak-Ribiere* equation:

$$\beta_j = \frac{\mathbf{g}_{j+1}^T(\mathbf{g}_{j+1} - \mathbf{g}_j)}{\mathbf{g}_j^T \mathbf{g}_j}. \tag{C.9}$$

2. Because the local error function is not quadratic, this may lead to the Hessian being non-positive definite. SCG overcomes this problem by adding a *scaling term* to the Hessian. This is in the form of a multiple of the identity matrix, giving $\mathbf{H}^* = \mathbf{H} + \lambda\mathbf{I}$, where $\mathbf{H}^*$ is the scaled Hessian (hence, *scaled* conjugate gradients). Increasing λ causes a corresponding increase in the denominator of α (equation (C.6)) which reduces the step length. This has the effect of restricting the search to a localised region of the error function. The smaller the search region, the more likely that it can be adequately represented by a quadratic surface.

   A suitable value for the λ_j may be obtained by considering the denominator for alpha (equation (C.6)), which we will denote as δ_j:

$$\delta_j = \mathbf{d}_j^T \mathbf{H}_j \mathbf{d}_j + \lambda_j ||\mathbf{d}_j||^2. \tag{C.10}$$

   In the event that δ_j is negative, then we should increase the value of λ_j to a new value, say $\bar{\lambda_j}$. In this case, the corresponding value of δ is given by

$$\bar{\delta_j} = \mathbf{d}_j^T \mathbf{H}_j \mathbf{d}_j + (\bar{\lambda_j} + \lambda_j)||\mathbf{d}_j||^2. \tag{C.11}$$

   A common practice [10] is to set:

$$\bar{\lambda_j} = 2\left(\lambda_j - \frac{\delta_j}{||\mathbf{d}_j||^2}\right) \tag{C.12}$$

   which is then substituted back into (C.11) to give:

$$\bar{\delta_j} = -\mathbf{d}_j^T \mathbf{H}_j \mathbf{d}_j. \tag{C.13}$$

Appendix D

Additional ICA results

D.1 Infomax ICA applied to subject A recordings

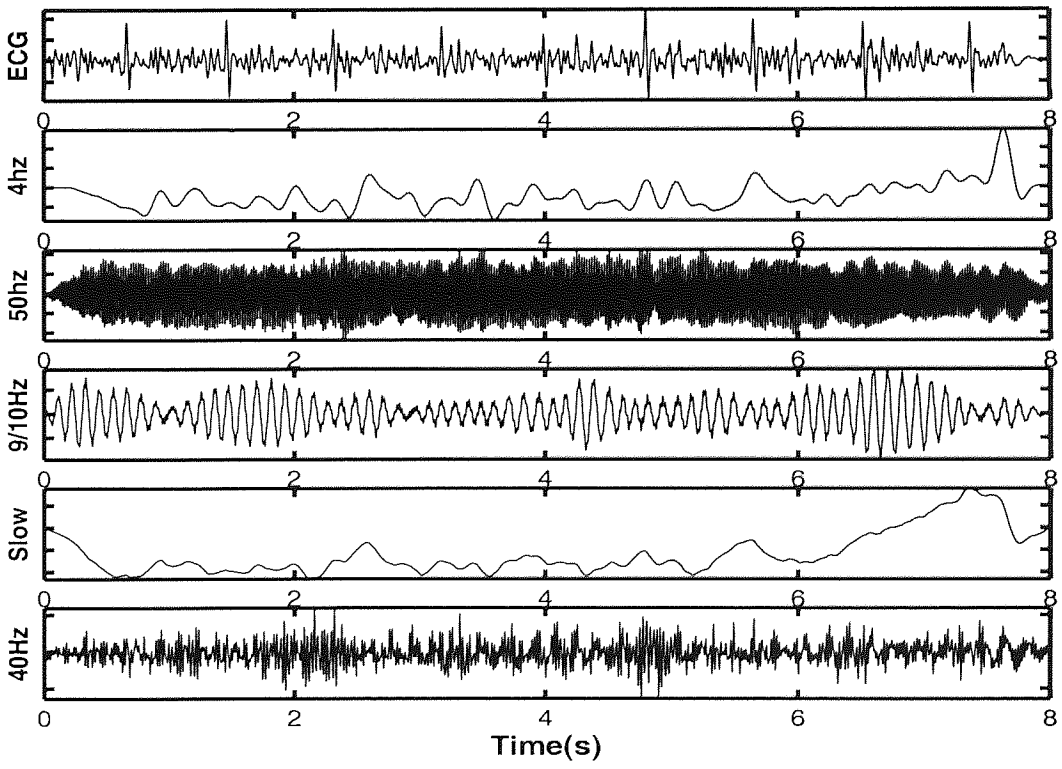


Figure D.1: Infomax ICA samples: idle task.

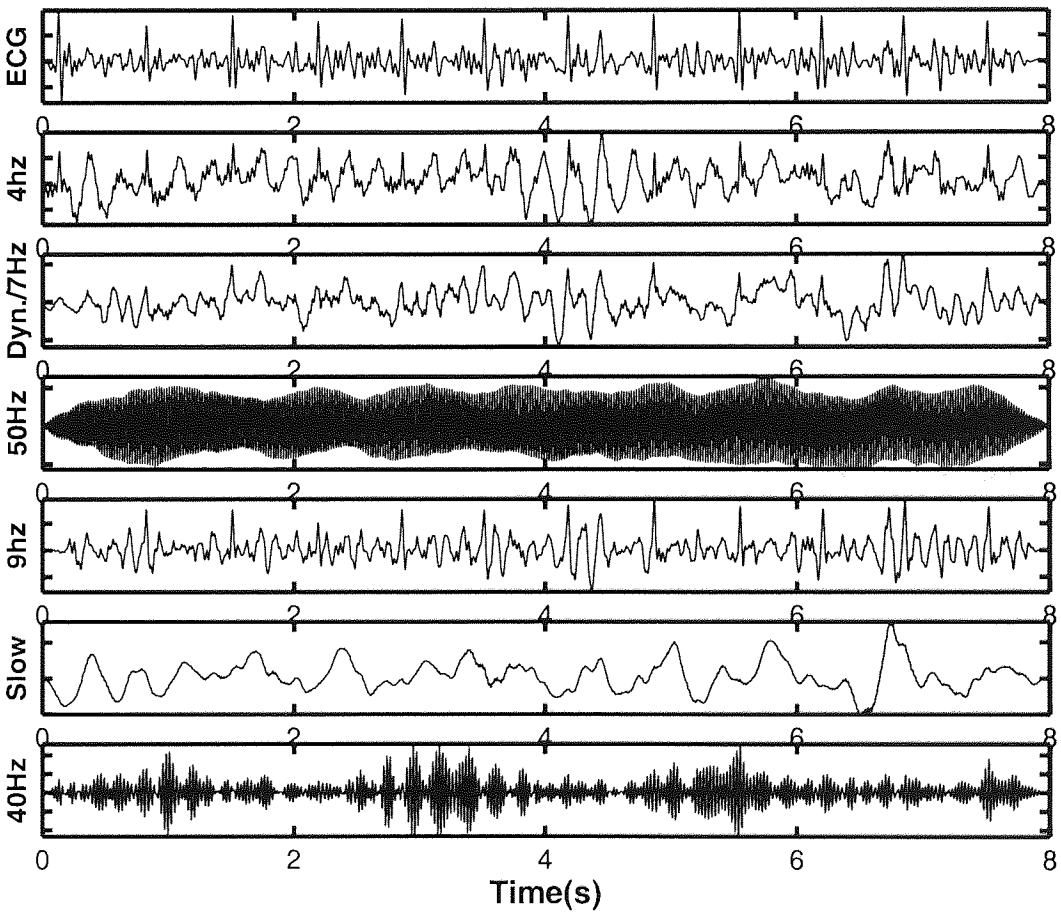


Figure D.2: Infomax ICA samples: face recognition task.

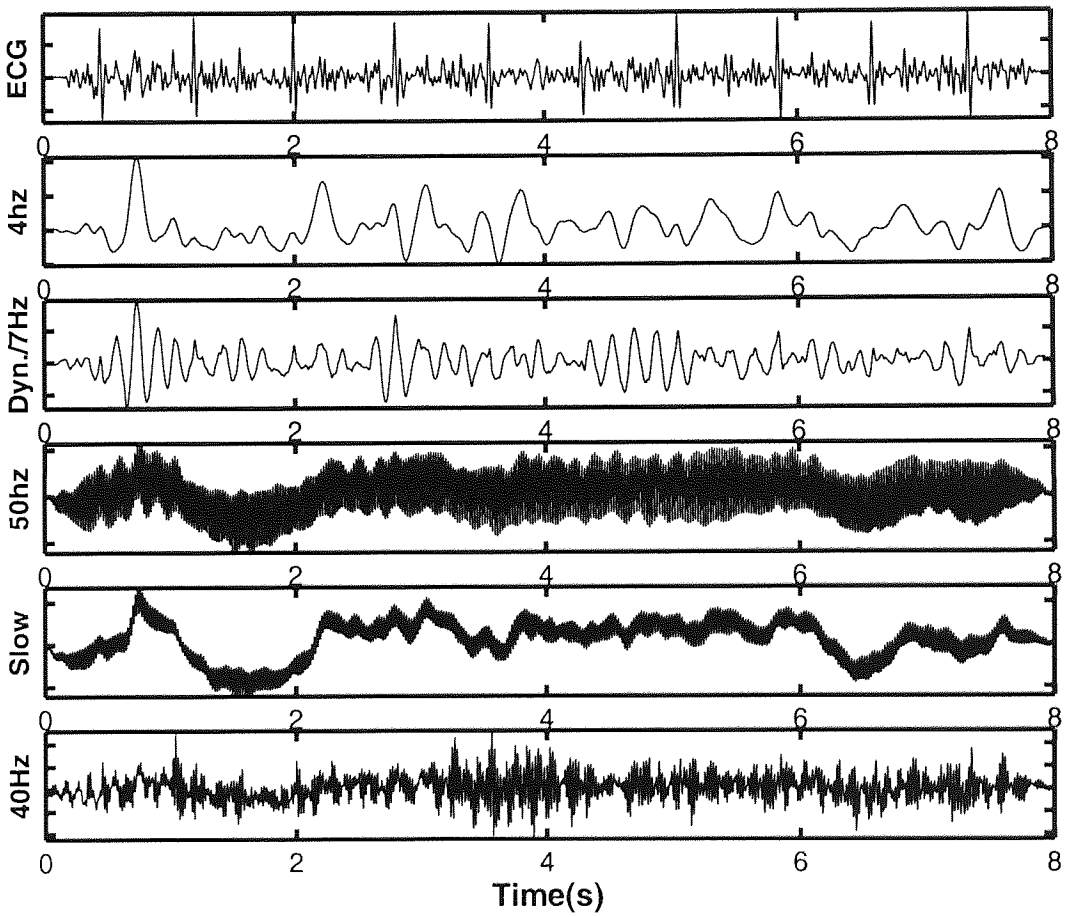
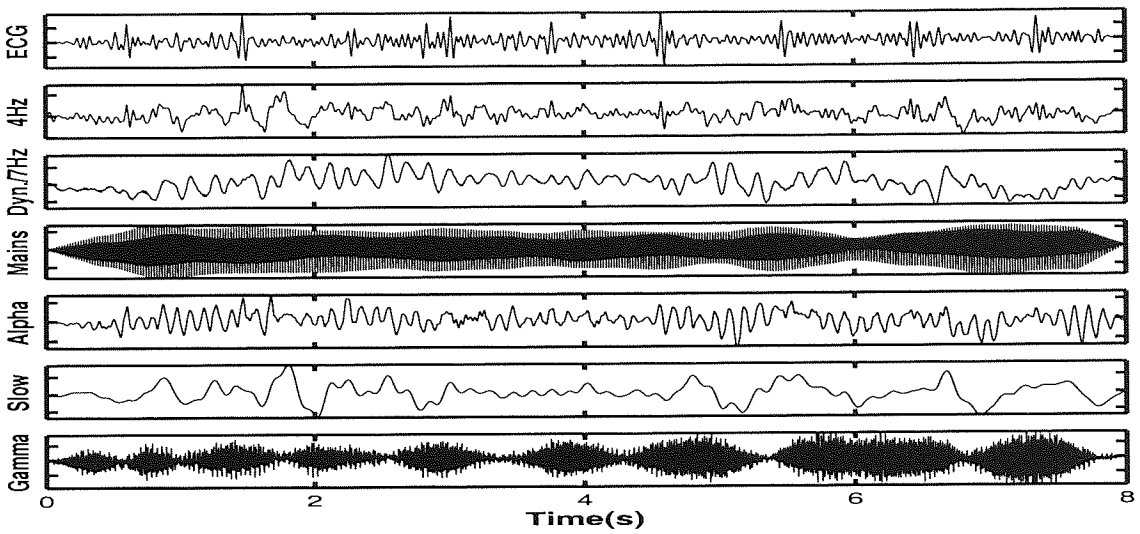
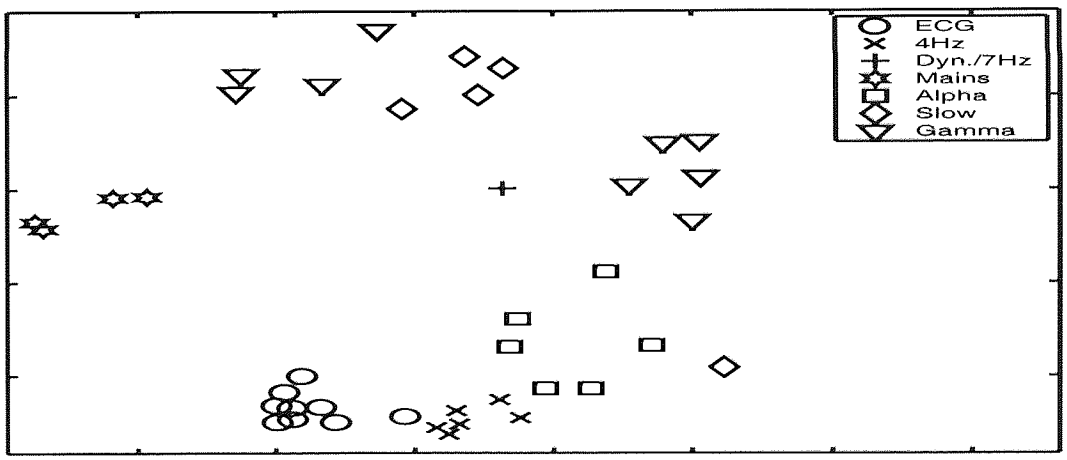


Figure D.3: Infomax ICA samples: natural scene recognition task.

D.2 ICA results for subject B



(a) ICs extracted from each class (8 seconds)



(b) Sammon map of idle ICs

Figure D.4: ICA of subject B: idle task

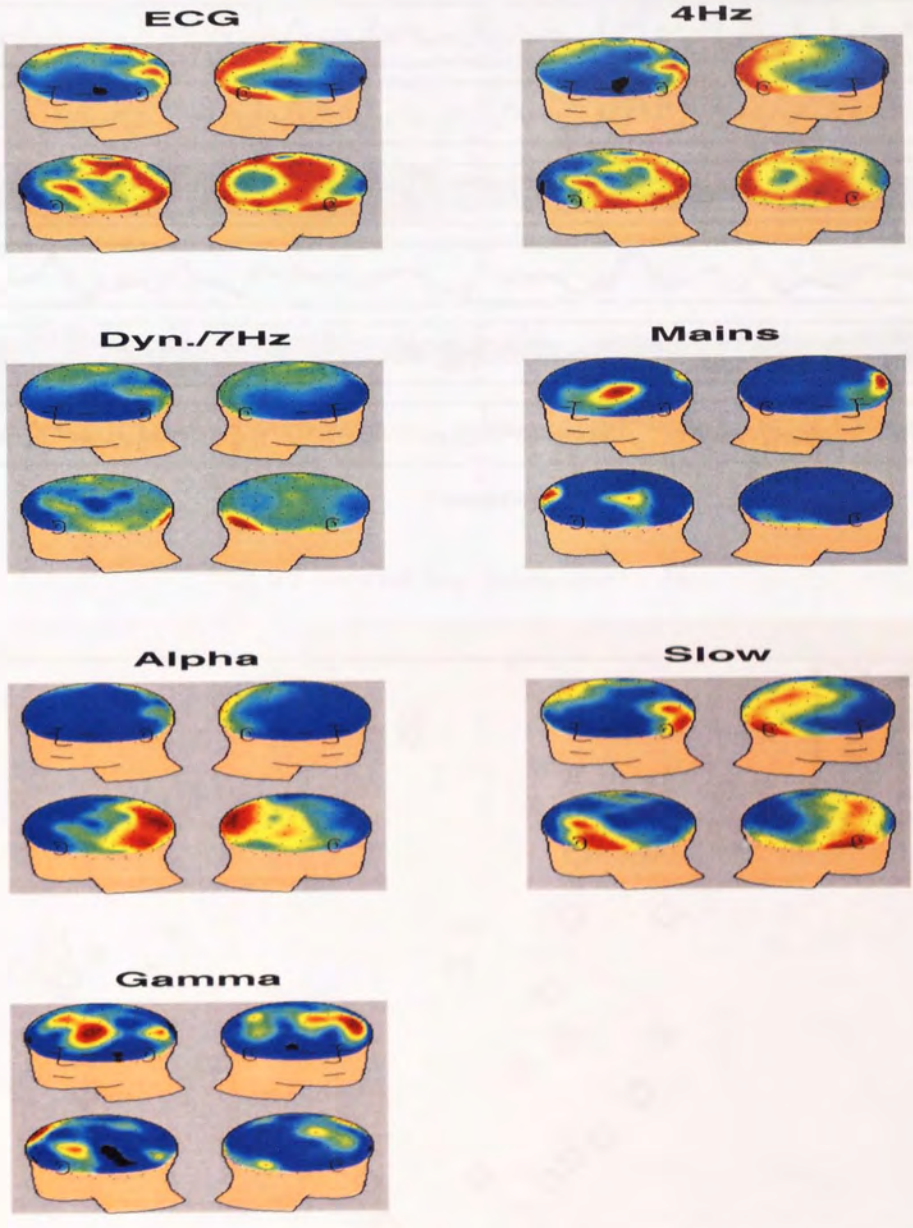
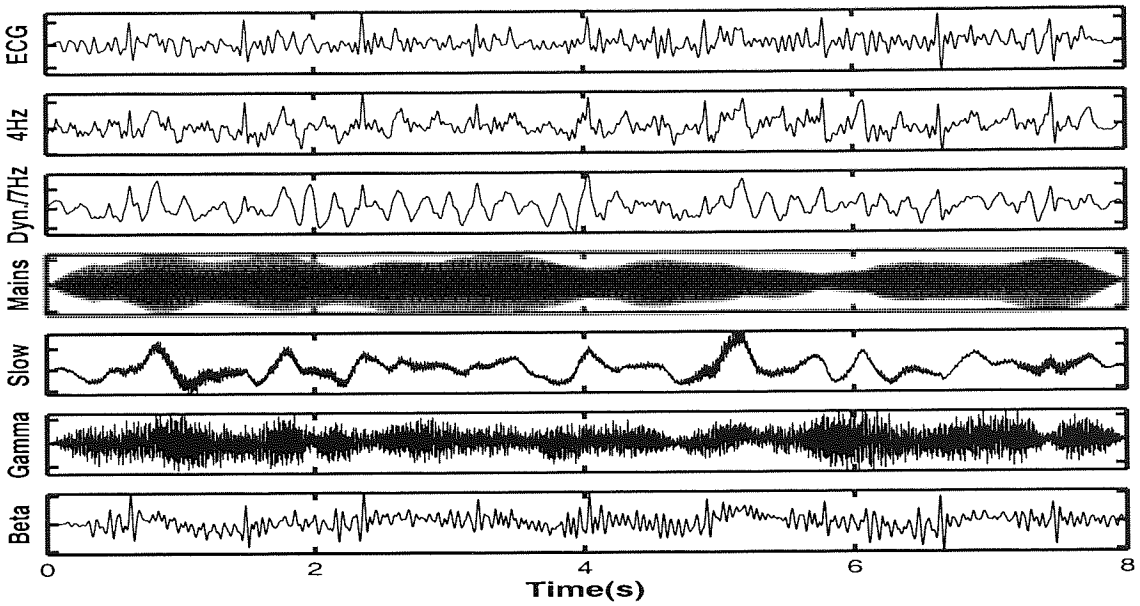
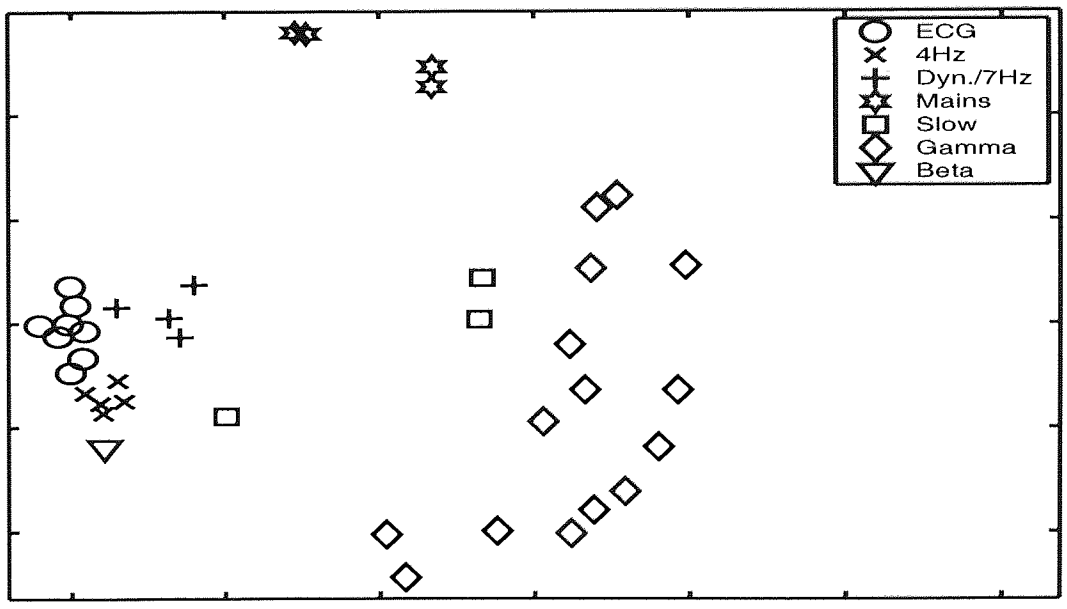


Figure D.5: Spatial maps: idle task



(a) ICs extracted from face recognition data



(b) Sammon map of face ICs

Figure D.6: ICA of subject B: faces

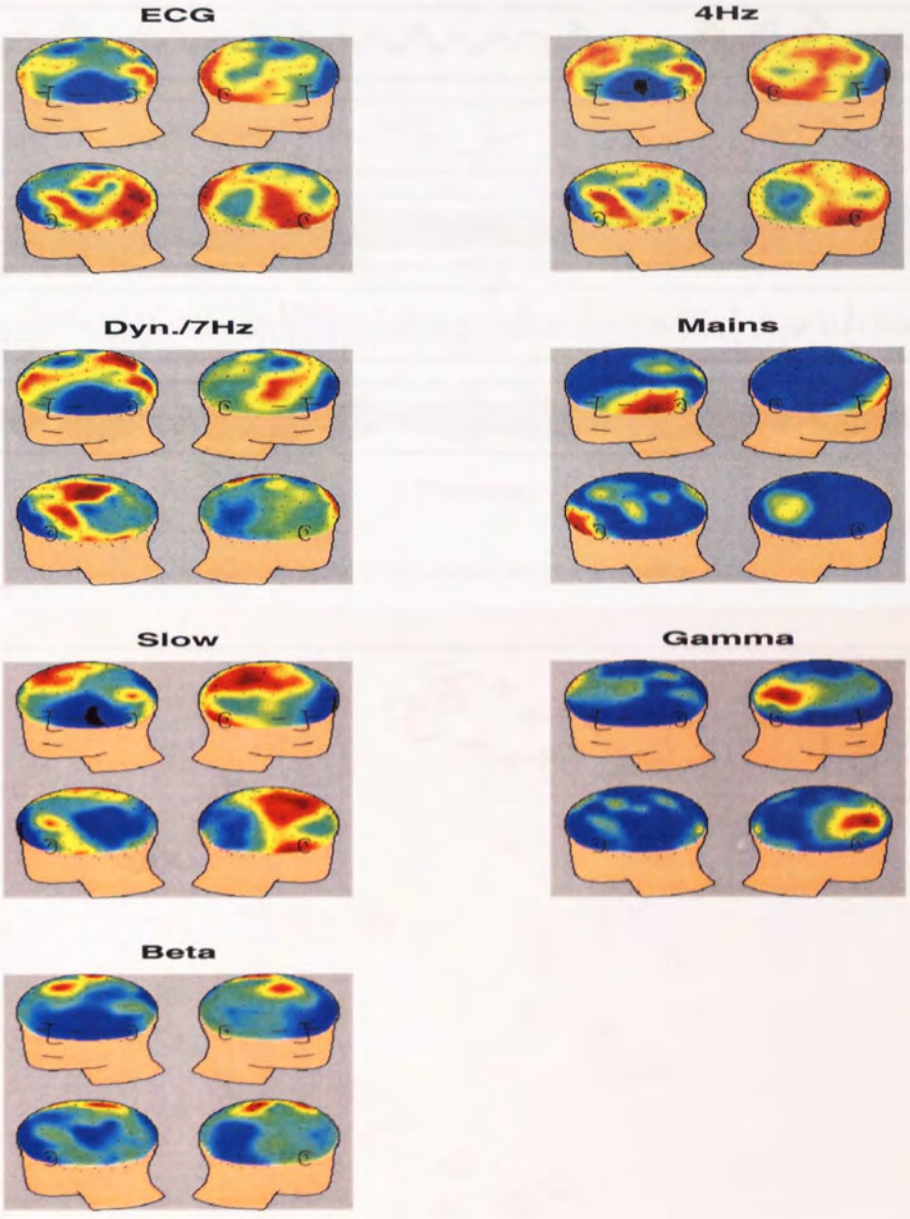
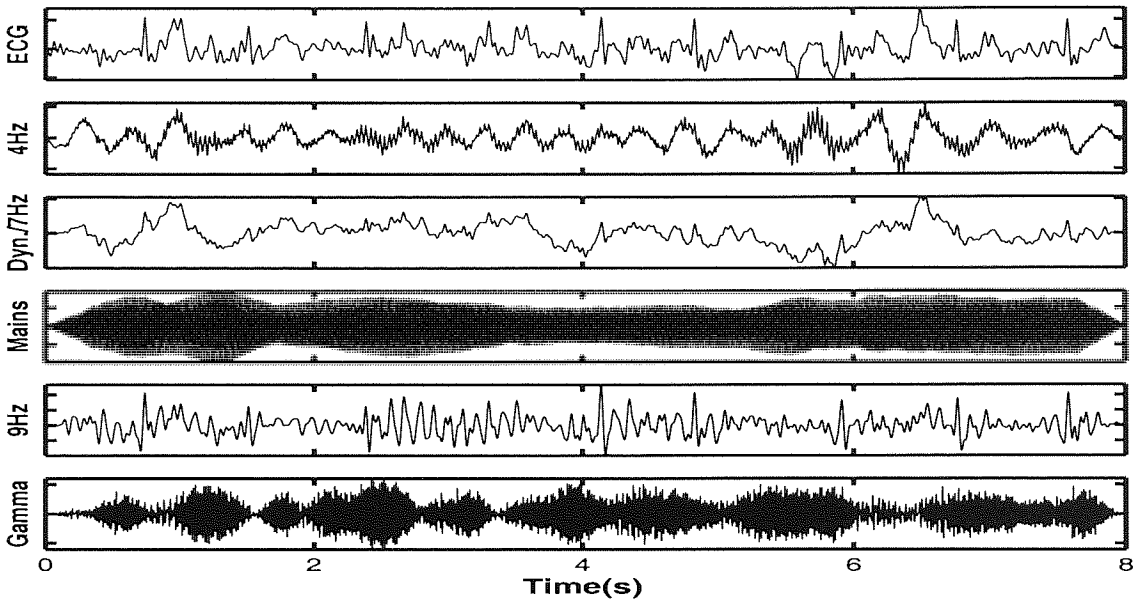
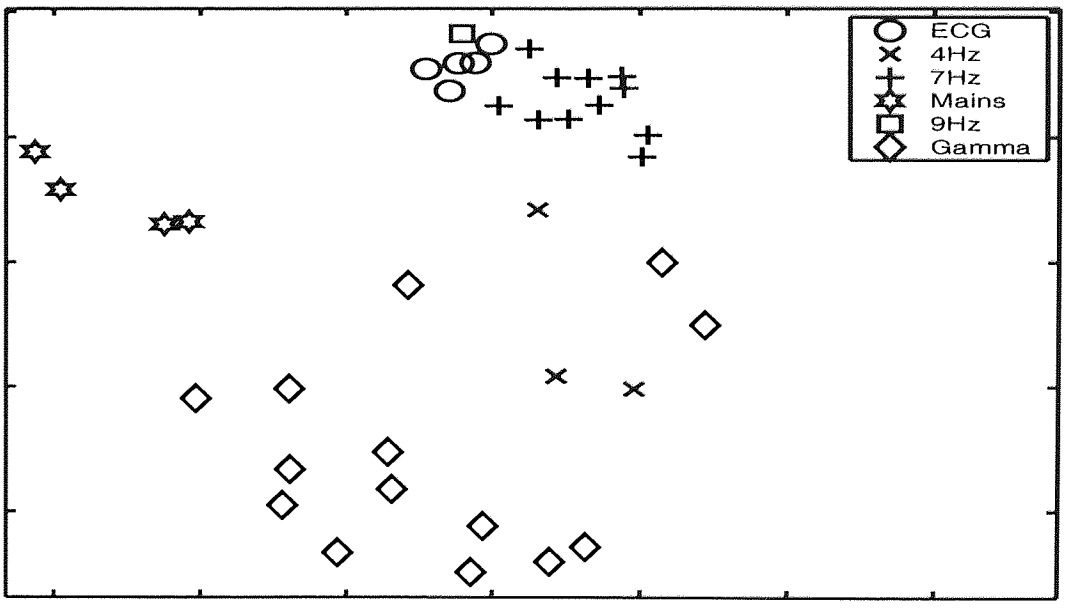


Figure D.7: Spatial maps: facial recognition



(a) ICs extracted from natural scene recognition data



(b) Sammon map of natural scene ICs

Figure D.8: ICA of subject B: natural scene recognition

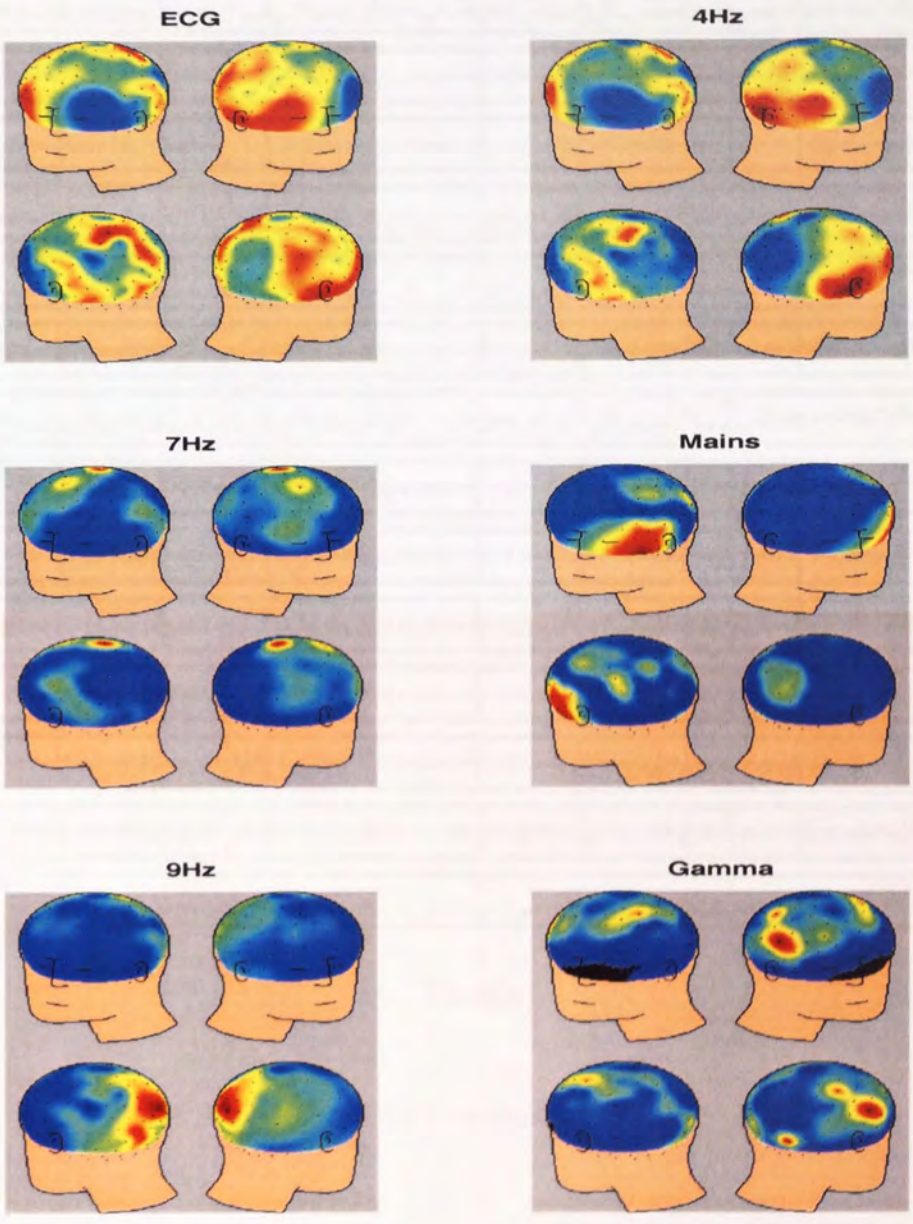


Figure D.9: Spatial maps: natural scene recognition

D.3 Full subject A decomposition results

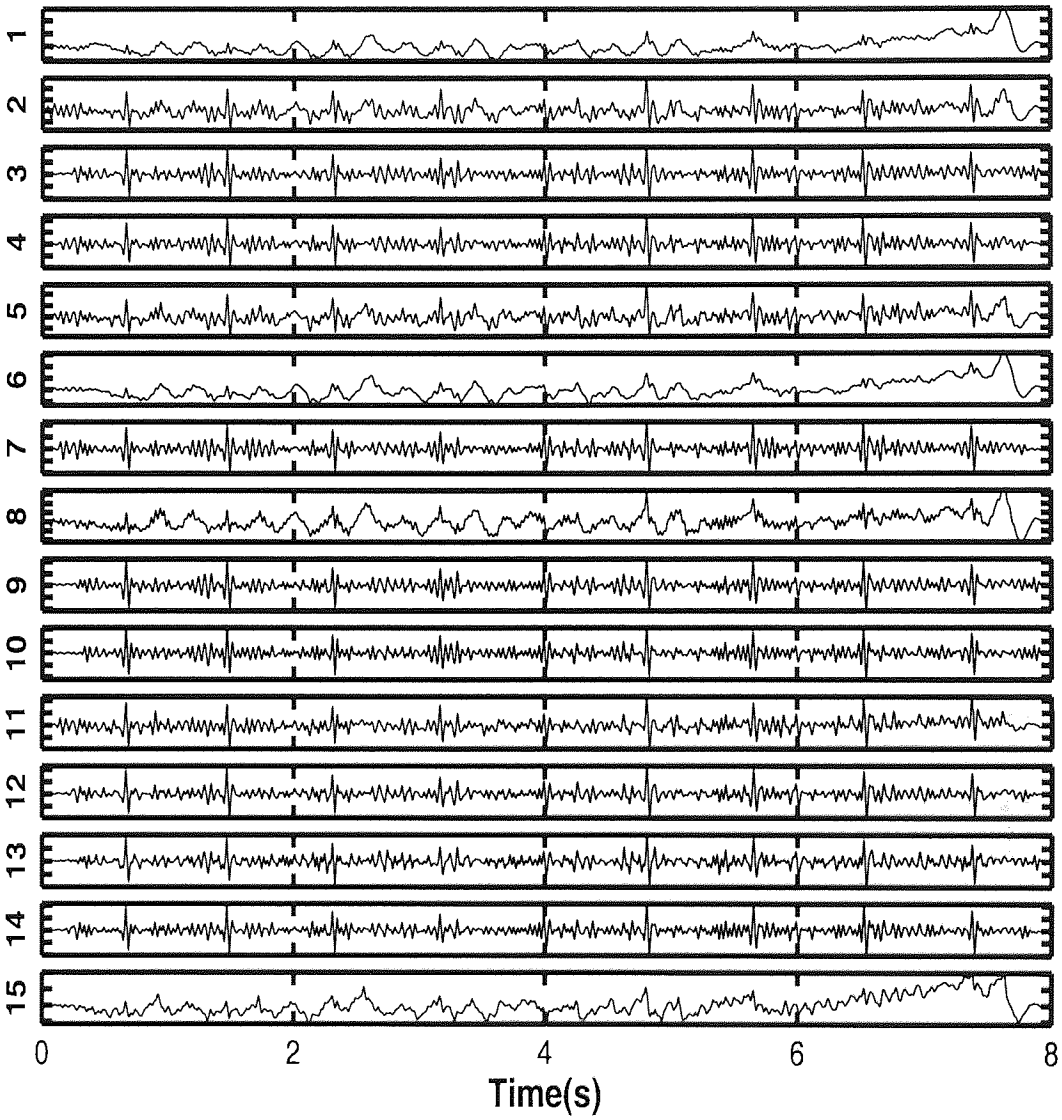


Figure D.10: ICs 1-15 from the idle data set

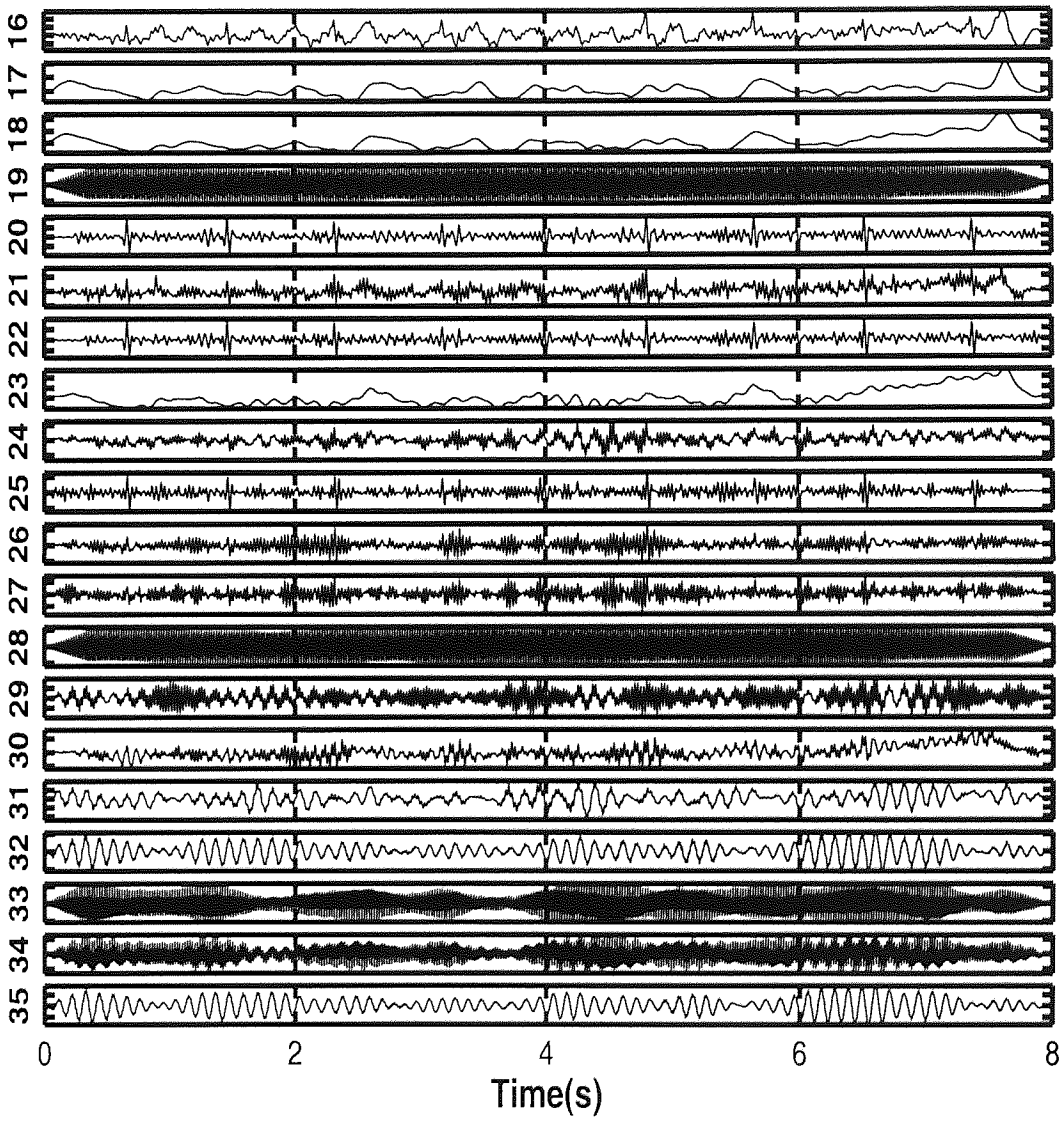


Figure D.11: ICs 16-35 from the idle data set

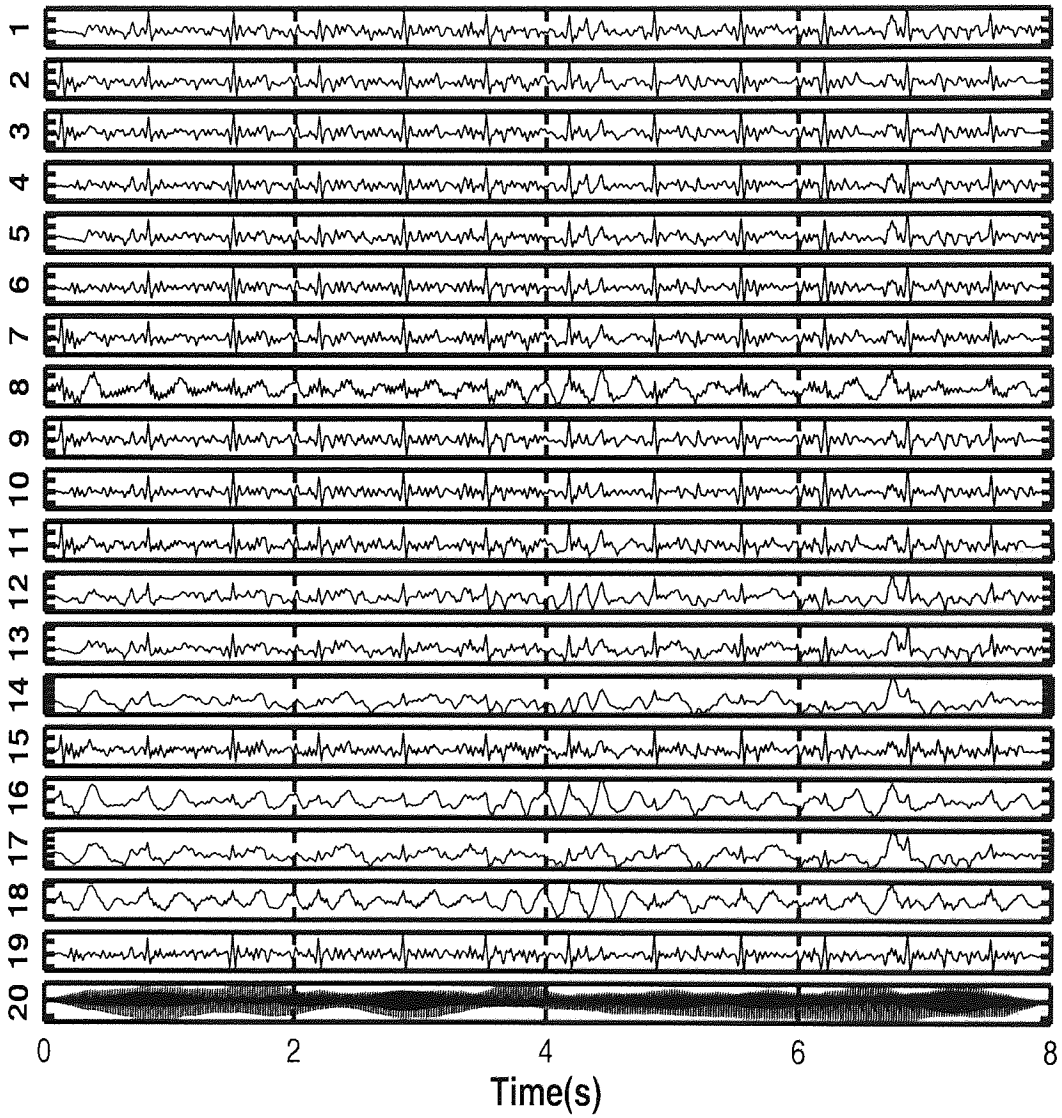


Figure D.12: ICs 1-20 from the face recognition data set

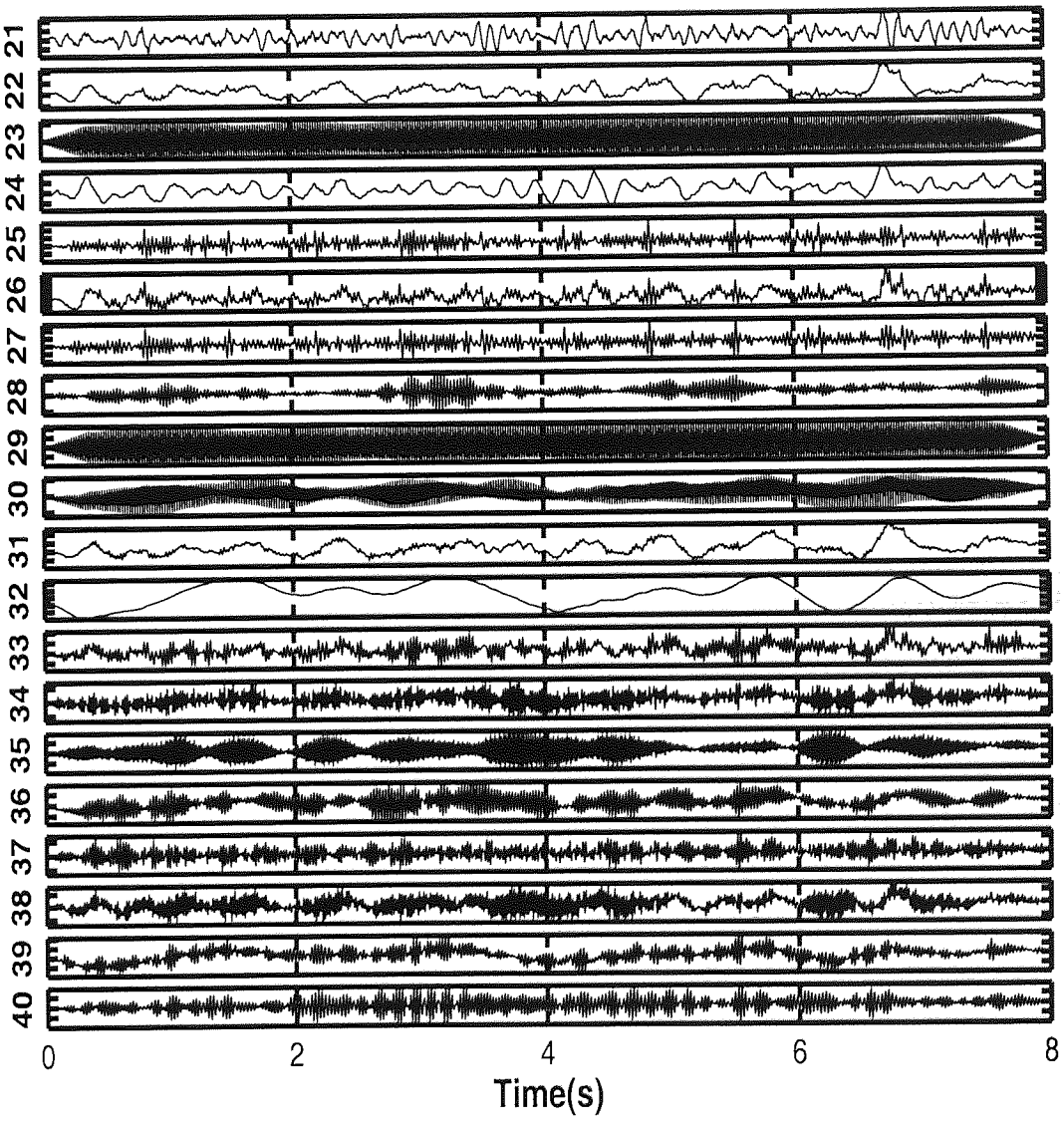


Figure D.13: ICs 21-40 from the face recognition data set

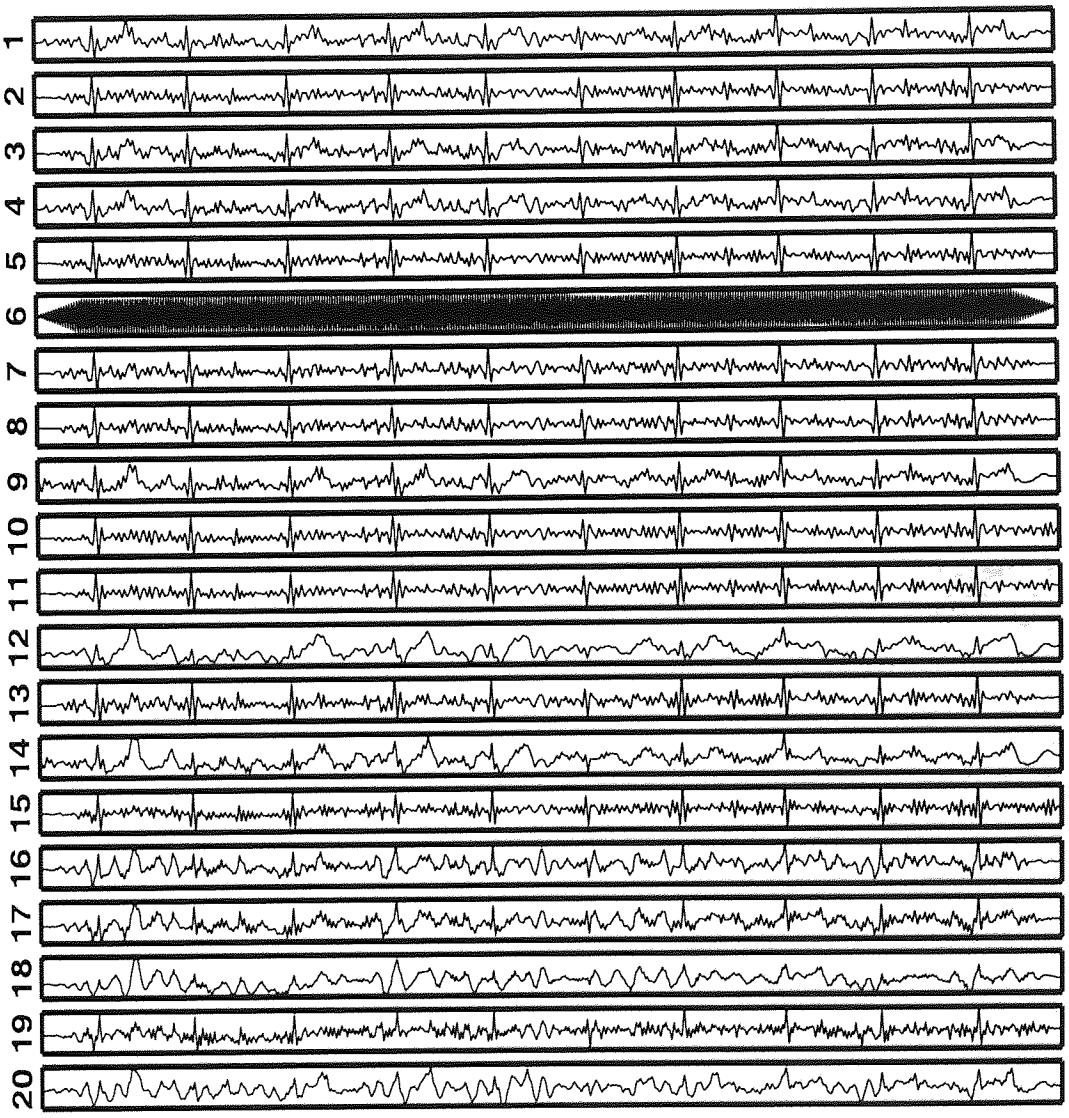


Figure D.14: ICs 1-20 from the image recognition data set

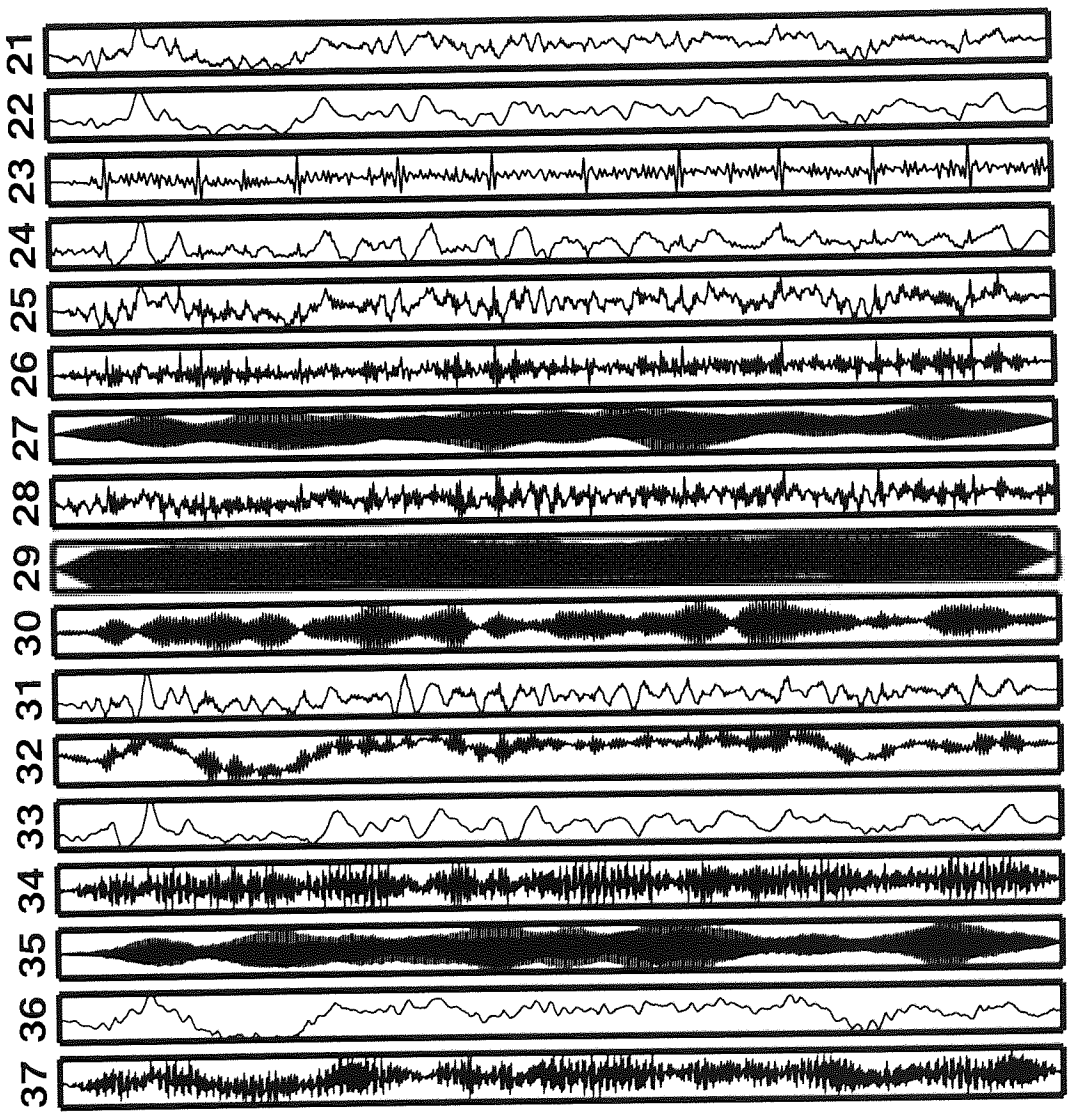
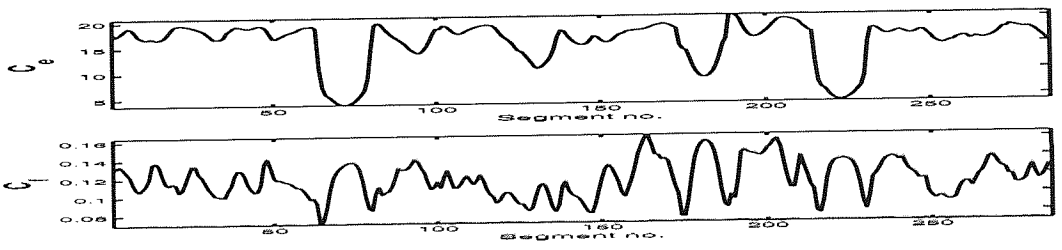


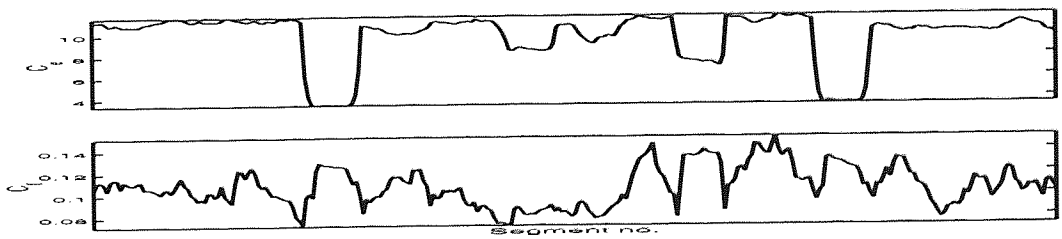
Figure D.15: ICs 21-37 from the image recognition data set

Appendix E

Windowing parameters: effect on C_f/C_e

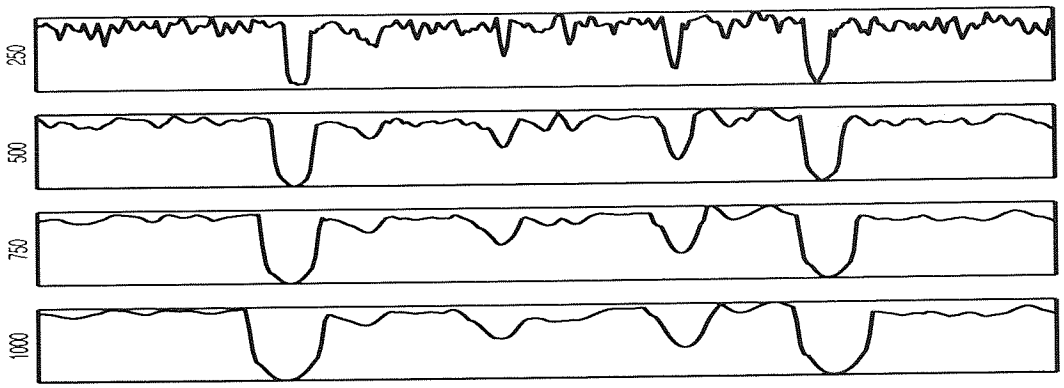


(a) Hamming window

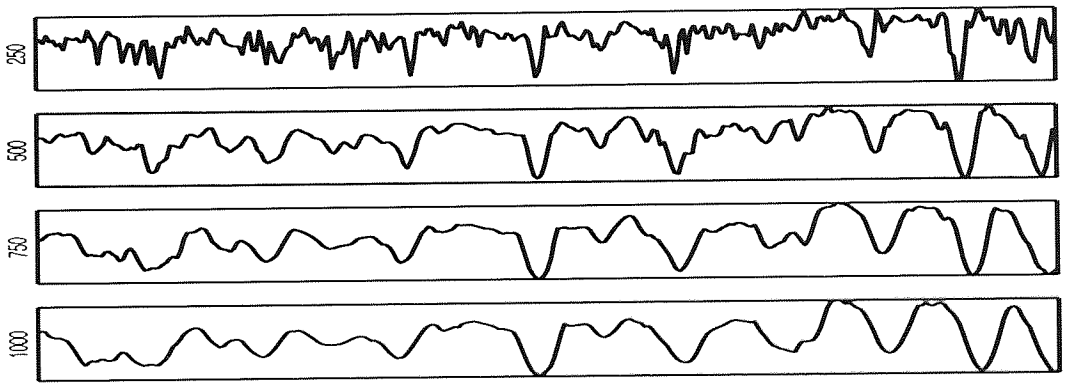


(b) Rectangular window

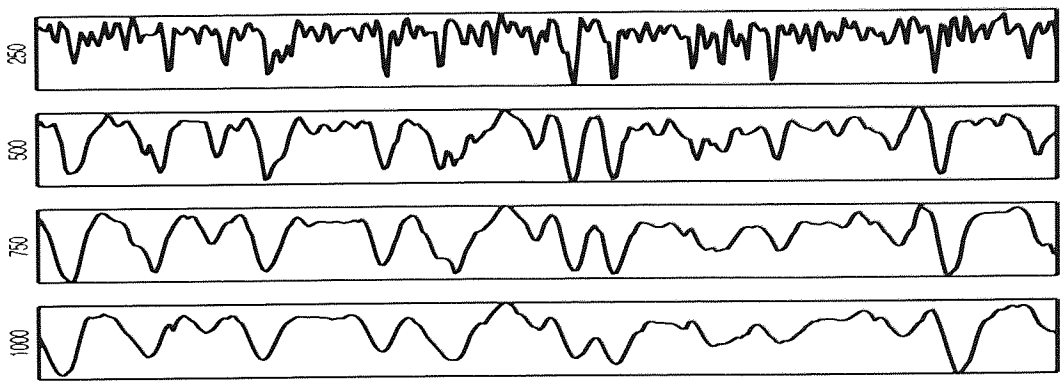
Figure E.1: Effect of windowing function: it can clearly be seen that the use of the Hamming window helps to smooth the signal by removing discontinuities. Note, for example, the initial section of C_f, where the Hamming window version clearly shows smooth curves which reflect the structure of the singular spectra. Where the rectangular window has been used, this structure has been obliterated



(a) Idle

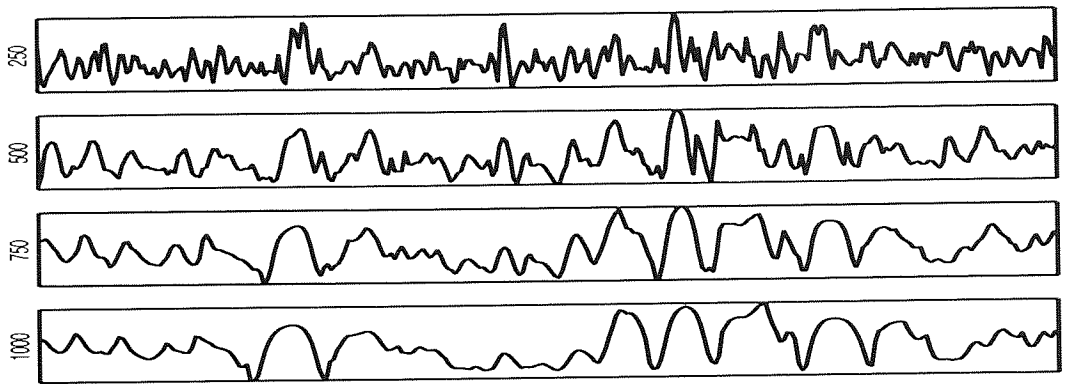


(b) Face recognition

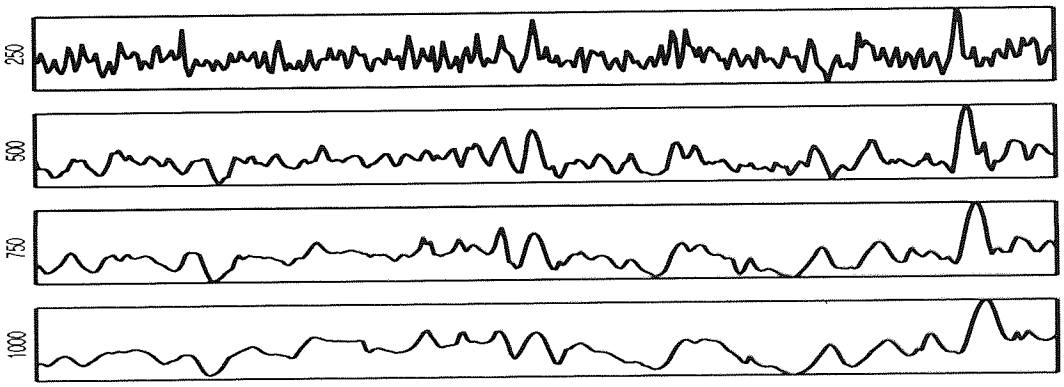


(c) Scene recognition

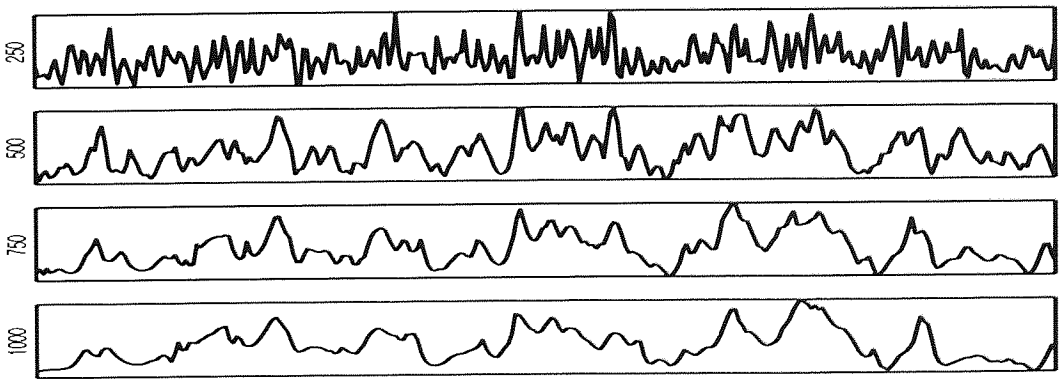
Figure E.2: Figures showing the effect of varying window lengths on C_e measurements. The labels on the y-axes denote the window lengths which were used. Note the gradual increase in smoothness of the curves as the window size is increased



(a) Idle



(b) Face recognition



(c) Scene recognition

Figure E.3: C_f curves calculated using varying window lengths

Appendix F

Publically available resources

The following publically available datasets and toolkits were used during this project:

Toolkits

- FastICA - *http://www.cis.hut.fi/projects/ica/fastica/*
- Netlab - *http://www.ncrg.aston.ac.uk/netlab/*

Datasets

- Groningen Natural Images database - *http://hlab.phys.rug.nl/archive.html*
- Physionet - *http://www.physionet.org/*
- Yale Face Database - *http://cvc.yale.edu/projects/yalefaces/yalefaces.html*

Bibliography

[1] P. Akritas, P. G. Akishin, I. Antoniou, A. Y. Bonushkina, I. Drossinos, V. V. Ivanov, Y. L. Kalinovsky, V. V. Korenkov, and P. V. Zrelov. Nonlinear analysis of network traffic. *Chaos, Solitons and Fractals (in press)*, 2001.

[2] S. Amari, A. Cichoki, and H. Yang. A new learning algirthm for blind signal separation. *Advances in Neural Information Processing Systems*, 8:752–763, 1996.

[3] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, 1999.

[4] E. Başar, C. Başar-Eroglu, and S. K. M. Schürmann. Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? *Neuroscience Letters*, 259:165–168, 1999.

[5] E. Başar, M. Schürmann, T. Demiralp, C. Başar-Eroglu, and A. Ademoglu. Event-related oscillations are real brain responses wavelet analysis and new strategies. *International Journal of Psychophysiology*, 39:91–127, 2001.

[6] C. Basar-Eroglu, D. Struber, M. Schomenn, M. Stodler, and E. Basar. Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. *International Journal of Psychophysiology*, 24:101–112, 1996.

[7] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1004–1034, 1995.

[8] O. Bertrand and C. Baudry. Oscillatory gamma activity in humans: a possible role for object representation. *International Journal of Psychophysiology*, 38:211–233, 2000.

[9] O. Bertrand and C. Tallon-Baudry. Oscillatory gamma activity in humans: a possible role for object representation. *International Journal of Psychophysiology*, 38:211–223, 2000.

[10] C. Bishop. *Neural networks for pattern recognition.* Clarendon Press, Oxford, 1995.

[11] B. Bower. Faces of perception. *Science news*, 160(1), July 2001.

[12] D. S. Broomhead and G. P. King. Extracting qualitative dynamics from experimental data. *Physica D*, 20:217–236, 1986.

[13] J. P. C. Tallon-Baudry, O. Bertrand. A ring-shaped distribution of dipoles as a source model of induced gamma-band activity. *Clinical Neurophysiology*, 110:660–665, 1999.

[14] J. F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 4(4):112–114, April 1997.

[15] J. F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[16] M. C. Casdagli, L. D. Iasemidis, J. C. Sackellares, S. N. Roper, R. L. Gilmore, and R. S. Savit. Characterizing nonlinearity in invasive EEG recordings from temporal lobe epilepsy. *Physica D*, 99:381–399, 1996.

[17] M. C. Casdagli, L. D. Iasemidis, R. S. Savit, R. L. Gilmore, S. Roper, and J. C. Sackellares. Non-linearity in invasive EEG recordings from patients with temporal lobe epilepsy. *Electroencenphalogr. Clin. Neurophysiol.*, 102:98, 1997.

[18] D. Cheyne, L. Roberts, W. Gaetz, D. Bosnyak, H. Weinberg, B. Johnson, C. Nahmias, and L. Deecke. EEG and MEG source analysis of somatosensory evoked responses to mechanical stimulation of the fingers. In *10th International Conference on Biomagnetism*, Santa Fe, New Mexico, 1996.

[19] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[20] D. Cox and D. Hinkley. *Theoretical Statistics*. Chapman and hall, London, 1974.

[21] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 62:845–848, 1987.

[22] R. Germuska. Complexity analysis of electroencephalographic data. Master's thesis, NCRG, Aston University, 2000.

[23] A. Gevins, H. Leong, R. Du, M. E. Smith, J. Le, D. DuRousseau, J. Zhang, and J. Libove. Towards measurement of brain function in operational environments. *Biological Psychology*, 40:169–186, 1995.

[24] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley. Physiobank, physiotoolkit and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, June 2000.

[25] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman and Hall, 2001.

[26] R. Govindan, K. Narayanan, and M. Gopinathan. The evidence of deterministic chaos in ECG: Surrogate and predictability analysis. *Chaos*, 8(2):495 – 502, 1998.

[27] A. Haig, E. Gordon, J. Wright, R. Meares, and H. Bahramali. Synchronous cortical gamma band activity in task-relevant cognition. *Computational Neuroscience*, 11(4):669–675, 2000.

[28] M. Hämäläinen, R. Hari, J. Ilmoniemi, R. Knuutila, and J. Lounasmaa. Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413 – 497, 1993.

[29] K. L. Hansen, J. Palva, M. Sams, J. Hietanen, H. Aronen, and R. Ilmoniemi. Face selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, 253:147–150, 1998.

[30] S. Haykin. *Neural Networks: a Comprehensive Foundation*. Prentice hall, 2 edition, 1999.

[31] A. Hyvarinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[32] A. Hyvrinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[33] S. Ikeda and K. Toyama. Independent component analysis for noisy data - MEG data analysis. *Neural Networks*, 13:1063–1074, 2000.

[34] C. James, J. Kobayashi, and D. Lowe. Isolating epileptiform discharges in the unaveraged EEG using independent component analysis. In *Medical Applications of Signal Processing colloquium (IEE)*, 1999.

[35] C. James and D. Lowe. Using dynamical embedding to isolate seizure components in the ictal EEG. *IEE Proc. Sci. Meas. Technol.*, 146(6), November 2000.

[36] C. James and D. Lowe. Information dynamics view of brain processing function. *23rd Annual International Conference IEEE EMBS*, October 2001.

[37] C. James and D. Lowe. Single channel analysis of electromagnetic brain signals through ICA in a dynamical systems framework. *23rd Annual International Conference IEEE EMBS*, October 2001.

[38] O. Jensen and J. Lisman. An oscillatory short-term memory buffer model can account for data on the sternberg task. *The Journal of Neuroscience*, 18(24):10688–10699, 1998.

[39] O. Jensen and C. D. Tesche. Ongoing theta activity (6-8 hz) over the frontal midline increases with memory load in a short-term memory task: a parametric MEG study. *Society for Neuroscience Abstracts*, 2000.

[40] T. Jung, S. Makeig, M. Wester, J. Townsend, E. Courchesnea, and T. Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111:1745–1758, 2000.

[41] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. Analysis and visualization of single trial event related potentials. *Human Brain Mapping*, 14:166–185, 2001.

[42] N. Kanwisher. Domain specificity in face perception. *Nature Neuroscience*, 3:759–763, 2000.

[43] W. Klimesch, M. Doppelmayr, A. Yonelinas, N. Kroll, M. Lazzara, D. Rohm, and W. Gruber. Theta synchronization during episodic retrieval: neural correlates of conscious awareness. *Cognitive Brain Research*, 12:33–38, 2001.

[44] W. Klonowski, W. Jernajczyk, K. Niedzielska, A. Rydz, and R. Stepien. Quantitative measure of complexity of EEG signal dynamics. *Acta Neurobiol. Exper.*, 59:315–321, 1999.

[45] T. Kohonen. The self-organizing map. *Proceedings of the Institute of IEEE*, 78:1464–1480, 1990.

[46] D. Kugiumtzis. State space reconstruction parameters in the analysis of chaotic time series - the role of the time window length. *Physica D*, 95:13–28, 1996.

[47] D. Kugiumtzis. Lack of robustness of the surrogate data test. *Physica D.*, December 1999.

[48] D. Kugiumtzis. Test your surrogate data before you test for nonlinearity. *Phys. Rev. E*, 60(3):2808–2816, 1999.

[49] D. Kugiumtzis and N. Christophersen. State space reconstruction: Method of delays vs singular spectrum approach. Research report 236, Department of Informatics, University of Oslo, 1997.

[50] R. Leahy, J. Mosher, M. Spencer, M. Huang, and J. Lewine. A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Clinical Neurophysiology*, 107(2):159–173, 1998.

[51] S. Little, S. Ellner, M. Pascual, M. Neubert, D. Kaplan, T. Sauer, H. Caswell, and A. Solow. Detecting nonlinear dynamics in spatio-temporal systems, examples from ecological models. *Physica D: Nonlinear Phenomena*, 96:321–333, 1996.

[52] D. Lowe, C. James, and R. Germuska. Tracking complexity characteristics of the wake brain state. *4th International Conference on Neural Networks and Expert Systems in Medicine and Healthcare - NNESMED 2001*, June 2001.

[53] D. Lowe and M. Tipping. Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications*, 4:83–95, 1996.

[54] S. Makeig, M. Westerfield, J. Townsend, T. Jung, E. Courchesne, and T. Sejnowski. Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society: Biological Sciences*, 354:1135–1144, 1999.

[55] J. Malmivuo, V. Suihko, and H. Eskola. Sensitivity distributions of EEG and MEG measurements. *IEEE Trans. Biomed. Eng.*, 44(3):196–208, 1997.

[56] P. Nunez. Neocortical dynamics and human EEG rhythms, 1995.

[57] M. Palus. Nonlinearity in normal human EEG: Cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics*, 75(5):389–396, 1996.

[58] G. Pfurtscheller and F. L. da Silva. Event-related EEG/MEG synchronization and desynchronization:basic principles. *Clinical Neurophysiology*, 110:1842–1857, 1999.

[59] G. Pfurtscheller, C. Guger, G. M. ller, G. Krausz, and C. Neuper. Brain oscillations control hand orthosis in a tetraplegic. *Neuroscience Letters*, 292:211–214, 2000.

[60] I. Rezek and S. Roberts. Stochastic complexity measures for physiological signal analysis. *IEEE Transactions on Biomedical Engineering*, 45:1186, September 1998.

[61] S. Roberts, W. Penny, and L. Rezek. Temporal and spatial complexity measures for EEG-based brain computer interfacing. *Medical and Biological Engineering and Computing*, 1998.

[62] C. Robinson. *Dynamical Systems*. CRC Press, 1995.

[63] S. Rokicki. Psychophysiological measures applied to operational test and evaluation. *Biological Psychology*, 40:223–228, 1995.

[64] M. S, J. T-P, A. J. Bell, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Science, USA*, 94:10979–10984, 1997.

[65] J. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, C-18:401–409, 1969.

[66] T. Schreiber and A. Schmitz. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett*, 77:635, 1996.

[67] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142:346–382, 2000.

[68] K. Singh. Functional imaging of the brain using superconducting magnetometry. *Endeavour*, 19(1):39–44, 1995.

[69] M. Small, D. Yua, J. Simonottoa, R. G. Harrisona, N. Grubbb, and K. Fox. Uncovering non-linear structure in human ECG recordings. *Chaos, Solitons and Fractals*, 13:1755–1762, 2001.

[70] F. Takens. Detecting strange attractors in turbulence. *Lecture notes in mathematics (dynamical systems and turbulence, Warwick 1980)*, 898:366–381, 1981.

[71] J. Theiler. On the evidence for low-dimensional chaos in an epileptic electroencephalogram. *Phys. Lett. A*, 196:335–341, 1994.

[72] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58, 1992.

[73] J. Theiler and P. E. Rapp. Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalogr. Clin. Neurophysiol.*, 98:213, 1996.

[74] A. Treisman. The binding problem. *Current Opinion in Neurobiology*, 6:171–178, 1996.

[75] R. Vigario. Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. Clin. Neurophysiol.*, 103:395–404, 1997.

[76] R. Vigario, V. Jousmaki, M. Hamalainen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems*, volume 10, pages 229–235. The MIT Press, 1998.

[77] J. Vrba. Magnetoencephalography: The art of finding a needle in a haystack. *Physica C (In Press)*, 2001.

[78] C. Wade and C. Tavris. *Psychology*. HarperCollins College Publishers, 1993.

[79] W. L. Woon and D. Lowe. A nonlinear dynamical systems approach to MEG signal processing: an example of brain source elicitation. *4th International Conference 'Neural Networks and Expert Systems in Medicine and Healthcare' (NNESMED)*, June 2001.