

Complexity Analysis of Electroencephalographic Data

Richard Germuska M.Eng

M.Sc by Research in Pattern Analysis and Neural Networks



The University of Aston in Birmingham

September 2000

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

The University of Aston in Birmingham

Complexity Analysis of Electroencephalographic Data

Richard Germuska M.Eng

M.Sc by Research in Pattern Analysis and Neural Networks
September 2000

Thesis Summary

This project investigates whether it is possible to correlate changes in the EEG structure with changes in the complexity of the original signal. Based on the assumption that the complexity of the EEG is due to the non-linear interaction of a few degrees of freedom, dynamical embedding of the EEG is performed to capture the dynamics of local sections of the underlying manifold, which are smooth non-linear fitting surfaces. Singular value decomposition (SVD) projects these sections of the manifold onto orthogonal axes that retain maximum variance, thereby identifying the degrees of freedom associated with the original EEG signal. Furthermore we assume that any change in the interaction of these degrees of freedom indicates a change in the brain state of the subject. We model this interaction by applying two measures of complexity, (i) entropy and (ii) Fisher's information content. Finally we performed experiments to see if changes in complexity corresponded to changes in the structure of the EEG and compared the performance of the two measures.

Keywords: Dynamical Embedding, Feature Extraction, Manifold, Singular Value Decomposition, Degrees of Freedom

Acknowledgements

My sincerest thanks go to my supervisor Dr. Christopher James, for all of his help and encouragement during the development of this project. I would also like to thank NCRG members; Professor David Lowe, Dr Ian Nabney, Loic L'Heveder, Olivier Dupin and Yann Brule, and Gari Clifford from the Oxford Robotics department, for all of their assistance during my year at Aston. Thanks also to Dr Helen Stone of BAE SYSTEMS for supplying the EEG data used throughout this project.

Special thanks to Venerable Geshe Kelsang Gyatso Rinpoche, Gen-la Samden, Ani-La Kunphen, Kelsang Ogmin and all at Samantabhadra centre, and to my girlfriend Eleanor Davey for her love and support. Finally I am extremely grateful to my family for their patience, love and support, and to my father for the many useful discussions throughout the duration of this project.

Contents

1	Introduction	14
1.1	Thesis overview	15
2	Electroencephalographic data	16
2.1	EEG recording	16
2.1.1	Selective sensitivity of EEG recordings	17
2.2	Brain waves	18
2.2.1	Characteristics of different frequency bands	18
2.2.2	Wave morphology	19
2.3	Experimental data	20
2.3.1	Single channel analysis	22
2.4	Summary	22
3	Modelling the underlying generator of the EEG	23
3.1	Dynamical embedding	24

3.1.1	Singular value decomposition (SVD)	25
3.2	Choice of embedding parameters	26
3.2.1	Embedding delay, τ	27
3.2.2	Embedding dimension, \hat{M}	27
3.2.3	The number of delay vectors, n	28
3.3	Experiment varying the value of n	30
3.3.1	Results	30
3.3.2	Morphology of the singular spectrum	32
3.4	Summary	34
4	Complexity analysis	35
4.1	Entropy	37
4.1.1	Entropy applied to the singular spectrum	37
4.1.2	Entropy and noise	38
4.1.3	Different normalisation	42
4.1.4	Results	45
4.1.5	Conclusions	46
4.2	Fisher's information measure	48
4.2.1	Fisher measure applied to the singular spectrum	49
4.2.2	Numerical calculation of the Fisher measure	49

4.2.3	Functional approximation	50
4.2.4	Discrete calculation	62
4.2.5	Results	63
4.2.6	Conclusions	66
4.3	Summary	67
5	Conclusions	69
6	Future Work	71
6.1	Vigilance	71
6.2	Complexity analysis of topographic systems	72
6.2.1	ICA experiment	72
A	Discrete Fisher derivation	77

List of Figures

2.1	Extracranial recording of EEG data	17
2.2	Examples of a clean section of normal wake EEG	20
2.3	Example of EMG artifact obscuring signal of interest	21
2.4	The placement of electrodes using the 10-20 system	21
3.1	GUI snapshot for $M = 150, n = 150$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.	31
3.2	GUI snapshot for $M = 150, n = 400$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.	31
3.3	GUI snapshot for $M = 150, n = 1500$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.	32
3.4	GUI snapshot for $M = 150, n = 150$ with a section of EEG corrupted with EMG noise	33

4.1	An example of the singular spectrum morphology for wake EEG. The location of the kink as suggested by previous work indicates the separation between signal and noise.	36
4.2	Weighting function introduced by entropy, $-p(x)\log p(x)$	39
4.3	Typical wake EEG singular spectrum normalised to power 1, with superimposed entropy weighting function(dotted line)	40
4.4	Typical singular spectrum for EEG corrupted with EMG, normalised to power 1, with superimposed entropy weighting function(dotted line). Notice the lack of kink structure, due to the degrees of freedom introduced by the external EMG generator.	41
4.5	Typical wake EEG singular spectrum normalised to first value 1 with superimposed entropy weighting function (dotted line)	43
4.6	Typical singular spectrum for EEG corrupted with EMG, normalised to first value 1 with superimposed entropy weighting function (dotted line). The different scaling further highlights the lack of kink structure in the singular value spectrum for data corrupted with EMG	44
4.7	Screen shots from GUI showing the two entropy measures for wake EEG. Notice how the two entropy complexity measure are almost identical when monitoring subtle changes in wake EEG	45
4.8	Screen shots from GUI showing the two entropy measures for wake EEG corrupted with EMG artifact. Notice how the entropy complexity measure with first value 1 normalisation has a far greater dynamic range. .	46

- 4.9 Polynomial interpolation plot for $n = 10$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. Notice how there are considerable errors between the original values and the interpolated function, this is because the degree of the polynomial, n , is too low. 51
- 4.10 Polynomial interpolation plot for $n = 20$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. Here we have increased the value of n to 20, but we are still under fitting the original data. 52
- 4.11 Polynomial interpolation plot for $n = 36$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. The degree of the polynomial, n , is now considerably larger (36), but although we can now approximate the first section well, we do this at the cost of introducing severe oscillations at the end. 53

- 4.12 Interpolation plot using a natural cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 13 indicating an underlying dimensionality of 6. Notice how there is a considerable error located around components 2 – 3 due to the over fitting of the data. 55
- 4.13 Interpolation plot using a cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 7 indicating an underlying dimensionality of 3. Notice how there is a considerable error located around components 2 – 3 due to the over fitting of the data. 56
- 4.14 Interpolation plot using a cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to wake EEG corrupted with EMG noise. There is no clear kink indicating the separation between signal and noise due to the extra degrees of freedom introduced to the signal by the external EMG generator. Notice how there is considerable error in the first section due to over fitting. . 57

- 4.15 Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 13 indicating an underlying dimensionality of 6. Notice how the error that was previously located around components 2 – 3 when using the natural spline, has been smoothed out resulting in no significant error. 59
- 4.16 Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 7 indicating an underlying dimensionality of 3. Notice how the error that was previously located around components 2–3 when using the natural spline, has been smoothed, but there still remains a significant notch. 60
- 4.17 Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to wake EEG corrupted with EMG noise. There is no clear kink indicating the separation between signal and noise due to the extra degrees of freedom introduced to the signal by the external EMG generator. Notice how the error that was previously located in the first section when using the natural spline, has been smoothed out resulting in no significant error. 61

4.18	Screen shot of GUI showing all three Fisher measures applied to wake EEG. From top to bottom they are: Smoothed cubic spline, forward difference, digital derivative.	64
4.19	Screen shot of GUI showing all three Fisher measures applied to wake EEG corrupted with EMG artifact. From top to bottom they are: Smoothed cubic spline, forward difference, digital derivative.	65
6.1	GUI snapshot for Fisher measure, using ICA with RMS ordering of components	73

List of Tables

- 4.1 Table summarising the relative performances of each complexity measure 67

Chapter 1

Introduction

The ability to perform non-invasive monitoring of a subject's neo-cortical brain activity has led to the formation of vast databases of high precision multi-channel electroencephalographic (EEG) data. The database at BAE SYSTEMS contains hundreds of EEGs recorded while subjects were asked to perform various tasks. These records are then used to score subject vigilance throughout the duration of these tasks.

If the scoring of subject vigilance could be accurately and robustly automated in real time, its use in monitoring those whose occupation requires a high degree of concentration and vigilance and a responsibility for the safety of others, could help in eliminating those accidents where human error is to blame.

This has motivated the search for advanced data processing techniques specifically designed to measure subject vigilance from the EEG. Our goal in this project was to try and establish if there was any interesting structure in the EEG identified by various feature extraction techniques and complexity analysis that might correlate with a clinically scored measure of subject vigilance. This project primarily explores the use of dynamical embedding and orthogonal techniques to model the underlying generator of the EEG data [1, 4, 17, 19] and looks at the use of two complexity measures, entropy

and Fisher's information content, to model the complexity of the data based on the interaction between the basis vectors of the underlying feature space [7, 15, 18]. In particular the novel approach of using single channel EEG data was explored. This approach avoids the use of standard EEG analysis techniques, such as Fourier frequency analysis, as these have been shown to be of limited practical use [16], as they do not allow us to reconstruct the underlying dynamics of the system and require the averaging of the time series over relatively long segments of EEG, whereas dynamical embedding allows us to consider much shorter segments.

1.1 Thesis overview

- **Chapter 2** gives a brief introduction to the physiology of the brain and the origin of the EEG signal, giving examples of EEG waveforms. The chapter concludes with a section on single channel analysis.
- The implementation of the dynamical embedding approach is explained in detail in **chapter 3**, and the criterion for choosing the embedding parameters are presented. The second half of this chapter gives the results obtained when dynamical embedding was performed on the EEG.
- **Chapter 4** introduces all of the theory required for the implementation of the entropy and Fisher complexity measures and presents the results of testing these measures on different waveforms of the EEG.
- **Chapter 5** presents the conclusions based on all of the key results obtained in the previous chapters.
- Recommendations for future work for the continuation of this project are made in **chapter 6**.

Chapter 2

Electroencephalographic data

The electroencephalograph records the electrical activity at the scalp, generated by underlying brain structures, using a number of non-invasive surface electrodes. First discovered by Berger in 1928, electroencephalography allows quantitative research into the dynamical function of the neocortex, from where it is believed the EEG signal originates [2].

Starting with the recording process of the EEG, this chapter takes the reader through the characteristics of the EEG and contrasts normal clean wake EEG, with noisy EEG. Finally there is a description of the data used in this project and a chapter summary.

2.1 EEG recording

The process for recording the EEG is shown in figure 2.1, which shows a cross-section of the scalp and neocortex together with the relative size of the electrode. The arrows, which represent the macro-columns of the neocortex, contain upwards of 10^6 neurons and 10^{10} synapses [2]. When these neurons fire the resulting dipole moment that is generated is recorded as a varying electrical signal by the scalp electrode. This signal,

which has a peak amplitude, before amplification, of the order of $\pm 10^3 \mu V$, is known as the EEG.

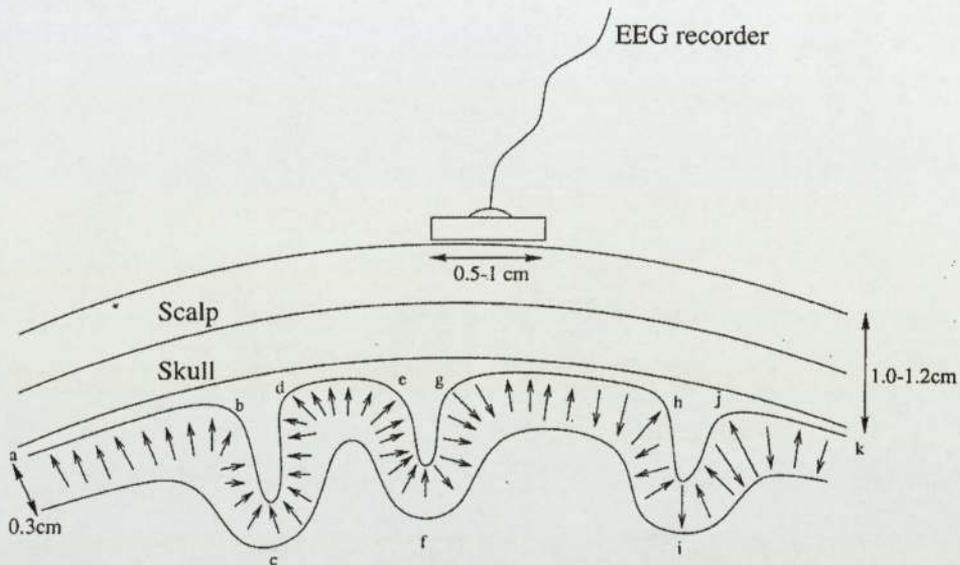


Figure 2.1: Extracranial recording of EEG data

2.1.1 Selective sensitivity of EEG recordings

There are three factors which affect the sensitivity of an EEG recording.

- The physical size of the electrode and the diffusion caused by the conductivity in the skull and scalp, which limit the surface spatial resolution of the resulting EEG to about 10 cm^2 [2].
- The proximity of the surface electrodes to the neocortex; this causes the recording to be more sensitive to sources in the neocortex than to sources deeper in the brain.
- The sensitivity of the surface electrodes to different types of dipole layer in the sulcus; they are most sensitive to correlated perpendicular dipole layers, such as

ab or *de* in figure 2.1, they are less sensitive to correlated tangential layers as in region *efg* and completely insensitive to opposing dipole layers such as *bcd*.

2.2 Brain waves

Clinical interpretation of the EEG usually involves the characterisation of constituent waveforms. This is done by observing the frequency content and to a lesser extent the morphology of these waveforms, although as previously mentioned this technique is of limited use when attempting to identify the specific characteristics of the underlying brain state, such as vigilance. The method does however provide an insight into the general characteristics of the brain state captured by the EEG and can differentiate between wake and sleep states and aid in the diagnosis of certain major brain diseases such as epilepsy.

It is generally accepted that the brain has no signal of interest with a frequency content higher than about 60Hz. Therefore the constituent waveforms are split into five frequency bands known as *alpha*, *beta*, *gamma*, *theta* and *delta*.

2.2.1 Characteristics of different frequency bands

The characteristics of each frequency band, together with brain state and type of person that exhibit them are given below.

- **Alpha**

Alpha waves are contained within the 7 to 13Hz frequency band. They are strongest over the occipital cortex (posterior region of the scalp). Alpha-activity is generally enhanced by closing the eyes and relaxation and abolished by eye opening or alerting by any mechanism *i.e.* thinking, calculating. However in

contrast it has also been strongly linked to creativity and mental work. Creative subjects show alpha activity when listening and coming to a creative solution for advanced problems. It is the major rhythm seen in normal relaxed adults - it is present during most of life especially beyond the thirteenth year when it dominates the resting trace.

- **Beta and gamma**

Beta activity is 'fast' activity. It has a frequency content of 13Hz and greater. It is usually seen on both sides in a symmetrical distribution and is most evident frontally. It may be absent or reduced in areas of cortical damage. It is generally regarded as normal rhythm and is dominant in most subjects who have their eyes open and are alert or anxious. Gamma activity corresponds to frequencies greater than 30Hz although it is sometimes included in the beta band.

- **Theta**

Theta activity has a frequency of 3 to 7Hz and is classed as 'slow' activity. It is abnormal in adults in wake state but is perfectly normal in the first stages of sleep and also in children in wake state up to 13 years old.

- **Delta**

Delta activity is 3Hz or below. It is characterised by the highest amplitude and the slowest waves. It is quite normal and is the dominant rhythm in infants up to one year and in stages 3 and 4 of sleep. It is usually most prominent frontally in adults and posteriorly in children.

2.2.2 Wave morphology

Certain waves have characteristic forms irrespective of their frequency and are recognisable by their shape: in other instances pairs or groups of waves have typical appear-

ances. The difficulty in clinical EEG scoring lies in recognising the artifacts and being able to differentiate between normal variants and abnormalities. Normal variants are waveforms that appear unusual but are not abnormal.

Normal wake EEG, a 20 second segment of which is shown in figure 2.2, is usually a linear combination of alpha, beta, theta and delta waves.

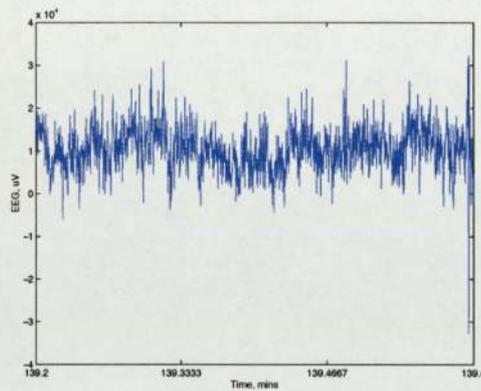


Figure 2.2: Examples of a clean section of normal wake EEG

Artifacts are disturbances in the EEG that do not originate from the neocortex and are therefore of no physiological interest. These include electrode movement and loss of contact (electrode ‘pop’) and electromyographic (EMG) noise caused by muscle movement, which includes eye blink. When electrode loss occurs there is generally no signal observed. An example of EMG artifact over a 20 second period is shown overleaf in figure 2.3, it clearly shows that the signal of interest is obscured.

2.3 Experimental data

The EEG data used in this project was collected from healthy volunteer subjects who were asked to perform simple visual tracking tasks over a period of about 7 hours. The data was recorded using an Oxford Instruments Medilog system utilising multiple channel measurements and was supplied by BAE SYSTEMS. The electrodes were sited

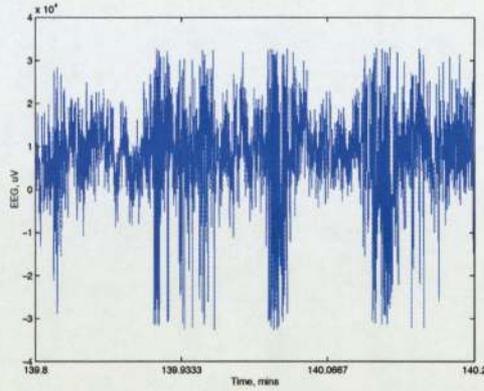


Figure 2.3: Example of EMG artifact obscuring signal of interest

according to the international 10-20 system [3] and a diagram of this set-up is shown below in figure 2.4. The sampling rate used for the recording was $\sigma_{EEG} = 256Hz$ and the data was quantised to 2 bytes per sample (16 bits).

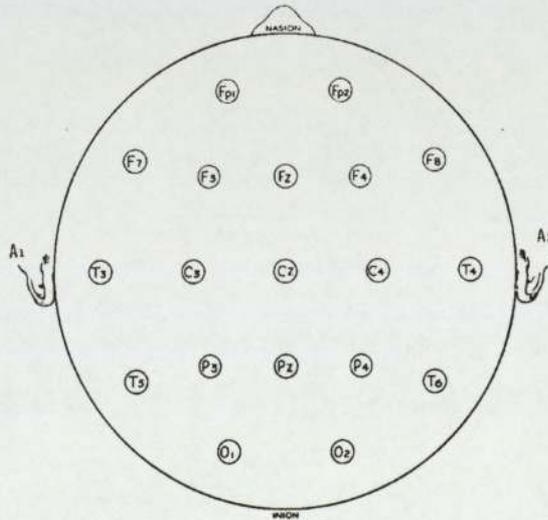


Figure 2.4: The placement of electrodes using the 10-20 system

2.3.1 Single channel analysis

There were 16 data channels available for experimental use, however we have chosen to select only one channel (O1) and to perform single channel analysis. This is because:

- We were interested to see if it was possible to perform feature extraction using single channel data.
- Since the EEG was recorded during visual tracking tasks performed by the subject, it made sense to select a single channel that was located near the occipital cortex where the visual cortex is situated. Channel O1, to the rear left of the brain was selected for use in all further analysis.

2.4 Summary

The apparently contrasting causes for alpha-activity illustrate the severe problems experienced when trying to monitor subject vigilance using simple frequency analysis. The level of alpha activity is not consistent for different types of subject; for some, creative thinking and mental work (and therefore vigilance) produces high levels of alpha activity and for others, merely opening their eyes abolishes all trace.

We need to adopt a completely different approach to EEG analysis, that is robust and subject adaptive. Subsequent chapters present a novel approach to feature extraction that allows us to model local sections of the reconstructed manifold of the underlying EEG generator. Consequently we are able to model the interaction between the degrees of freedom identified and hence the complexity of the signal.

Chapter 3

Modelling the underlying generator of the EEG

We hypothesise that the EEG signal is generated by the non-linear interaction of a few degrees of freedom as opposed to the linear interaction of many degrees of freedom [1]. A simple system (one with a few degrees of freedom), that is linear, will behave simply and will be *in equilibrium, periodic* or *quasiperiodic*. However, simple nonlinear systems can exhibit extremely complex behaviour. Such behaviour, as seen in the EEG, is characterised by its complexity and sensitivity to the initial state of the system [2]. Based on this, we can assume the existence of an unobservable deterministic generator of the observable data. Therefore from a finite number of samples of the time series, it should be possible to reconstruct the whole dynamics of the manifold that generated that time series. Sauer *et al.* [19] show that almost every smooth map from a d -dimensional smooth manifold M to \mathbb{R}^{2D+1} is a diffeomorphism on M , that is, a smooth one-to-one map which has a smooth inverse. This chapter describes how a system using dynamical embedding may be implemented to model the underlying generator of the EEG. Furthermore the choice of a set of basis vectors that span the underlying

feature space (orthogonal or independent) is made. The method presented here uses singular value decomposition (SVD) and calculates an orthogonal spanning set. Based on this decision the factors affecting the crucial choice of embedding parameters are then presented.

3.1 Dynamical embedding

Consider a continuous system governed by the set of N first-order differential equations

$$\frac{dX_i(t)}{dt} = F_i(X_1, X_2, \dots, X_N, t), i = 1, \dots, N \quad (3.1)$$

where the F_i are non-linear functions of the independent variables X_i and the system is contained in an N -dimensional vector space S . The dimension of this vector space is associated with the number of degrees of freedom of the system. Each $X(t) = (X_1, X_2, \dots, X_N, t)$ represents the state of the system at time t and is a point in the vector space S . In 1986 Broomhead and King [4] introduced SVD and embeddings to implement Takens' theorem which states that it should be possible to reconstruct the dynamics of a deterministic system [5]. The method consists of projecting each sample $x(t)$, which is the observable signal representing the state of the system $X(t)$, onto the feature space. In this space each delay vector, $X(i)$, is defined by

$$X(t) = \begin{bmatrix} x_t \\ x_{t+\tau} \\ \vdots \\ x_{t+(\hat{M}-1)\tau} \end{bmatrix} \quad (3.2)$$

which contains a section of time series data containing \hat{M} samples, sampled with a delay, or lag, of τ s between each successive sample. It is standard practice and

practical to fix τ such that there is a delay of a whole number of samples. Therefore each delay vector is a point in the embedding space $\mathbb{R}^{\hat{M}}$ and represents a window of data, of length \hat{M} , of the time series. To construct the embedding matrix \mathbf{X} that represents a discrete trajectory of the EEG data through a certain period of time, we run this window through a section of data constructing n successive delay vectors that correspond to the columns of the embedding matrix that map out this trajectory. Therefore \mathbf{X} is defined by

$$\mathbf{X} = \begin{bmatrix} x_t & x_{t+\tau} & \cdots & x_{t+n\tau} \\ x_{t+\tau} & x_{t+2\tau} & \cdots & x_{t+(n+1)\tau} \\ \vdots & \vdots & & \vdots \\ x_{t+(\hat{M}-1)\tau} & x_{t+(\hat{M})\tau} & \cdots & x_{t+(\hat{M}+n-1)\tau} \end{bmatrix} \quad (3.3)$$

We now wish to characterise this trajectory. Here, we assume an orthogonal spanning set of the feature space and calculate the eigenvectors and eigenvalues of the transformation described by the embedding matrix, \mathbf{X} , using SVD. The complementary approach involving a non-orthogonal spanning set is introduced in chapter 6.

3.1.1 Singular value decomposition (SVD)

Consider the real $M \times n$ matrix \mathbf{X} . We may decompose it as follows

$$\mathbf{X} = U * S * V^T \quad (3.4)$$

Where S is a diagonal matrix whose elements are σ_i arranged in descending order of magnitude. Then σ_i^2 is the i^{th} eigenvalue of $\mathbf{C} = \mathbf{X} * \mathbf{X}^T$. The columns of V are the eigenvectors of \mathbf{C} and the matrix U is the matrix of projections of \mathbf{X} onto the eigenvectors of \mathbf{C} [15].

3.2 Choice of embedding parameters

There are three choices that must be made when performing dynamical embedding on real data;

- Embedding delay, τ
- Embedding dimension, \hat{M}
- The number of delay vectors, n

The specific factors affecting these choices are presented in the following sections, however there are two general guiding principles. These are;

- The dimensionality of the underlying vector space S . This is unknown *a priori* and is intrinsic to the problem itself and will vary depending on the system being observed *e.g.* ECG, EEG or the stock market *etc.*
- The information content of the embedding matrix, \mathbf{X} . This is predominantly dictated by the sampling rate of the data and to a lesser extent the morphology of the data.

It is clear that each parameter is not completely independent and will have an effect on the choice of the others. However by ensuring that all of the theoretical requirements are met, we restrict the range of choice of each parameter and by analysing the results of different combinations of choices we may identify the combination that correctly captures the dynamics of the underlying system.

3.2.1 Embedding delay, τ

The first selection to be made is the delay, τ . As previously mentioned when dealing with discrete data, τ is always set to be a whole number of samples and is based on the sampling rate of the data. Nyquist's stability theorem states [6] that a signal must be resampled at at least twice the rate of the highest frequency component contained within the signal, assumed to be $60Hz$ in our case, to avoid aliasing, therefore $\sigma_{nyq} = 120Hz$. Setting τ to select every 2^{nd} data point gives a resampling rate of $\sigma_{EEG} = 128Hz$, which narrowly fulfills the Nyquist criteria. However due to the noise present in the data we choose to select every available data point, giving a resampling rate of $\sigma_{EEG} = 256Hz$, easily fulfilling the stability criteria, therefore $\tau = 1/256s$.

In our case the resampling rate is the same as the actual sampling rate of the EEG data and is therefore the highest we can choose. However for data that is sampled at a much higher frequency it is important to consider the problems that may be caused by highly correlated data, where each consecutive data point offers very little additional information about the signal due to the very high sampling rate.

Since each delay vector (which represents a discrete point on the M -dimensional reconstructed manifold) consists of a window of size n of the time series, it will be necessary to down sample when using data with a high sampling rate to avoid unnecessarily large windows that contain correlated and therefore redundant information.

3.2.2 Embedding dimension, \hat{M}

Takens' theorem states [5] that $\hat{M} \geq 2D + 1$ for the dynamics of the system to be completely captured, where D is the dimension, or the number of degrees of freedom [20], of the underlying system.

For noiseless systems we discover that the number of non-zero singular values, σ_i , in the diagonal decomposition matrix S corresponds to the embedding dimension $\hat{M} = 2D + 1$, where D is the dimension of the underlying system. The redundancies that appear indicate that the system can be completely described by a linear combination of the eigenvectors corresponding to the non-zero singular values, what we shall term *convergence of the singular spectrum* [18].

Most real systems are not noiseless and this is especially true of the EEG. What were once redundancies in noiseless data will now appear as relatively low singular values corresponding to the small contributions of the extra eigenvectors needed to reconstruct the noise which projects onto each dimension, D , of the noiseless system. Therefore for convergence of the singular spectrum we are looking for the value of M at which point the Euclidean difference between singular spectra calculated with windows of size M and $M + 1$ has converged toward zero. This value of M then corresponds to the size of the delay vector, M , required.

As discussed earlier in this chapter, the dimensionality of the EEG is assumed to be quite low, say, $D < 5$, therefore $\hat{M} \geq 11$. In practice the size of the delay vector, M , is greater than the embedding dimension \hat{M} due to correlations and the noise in the data that appears due to the EEG recording quality. This is highlighted by the value returned using convergence of the singular spectrum, and previous work in this area [7] shows that an initial value of $M = 150$ is sensible and this is confirmed by results presented in the following section.

3.2.3 The number of delay vectors, n

With a value for M and τ initially selected, we can now proceed in selecting a value for n that will fulfil all of the remaining theoretical and practical requirements.

The number of delay vectors, n , will now define the overall information content of the embedding matrix, \mathbf{X} , along with the period of time (section of time series samples) over which the corresponding trajectory, defined by \mathbf{X} in the feature space, is mapped out. The information content therefore depends on the values of M and n and the sampling rate of the data, since each embedding matrix, \mathbf{X} , contains $M + n - 1$ time series samples.

We must ensure that the information content of \mathbf{X} is large enough to contain one cycle of the lowest frequency component of interest in the EEG. This is regarded as being about $0.5Hz$, therefore $(M + n - 1) \geq 128$. We must also ensure full rank of the embedding matrix, therefore $n \geq M$.

Pseudo-stationarity

There is however a restriction placed on the upper limit of n : we must ensure pseudo-stationarity of the signal for the period of time that corresponds to the trajectory in feature space mapped out by \mathbf{X} .

When we process a large section of EEG data each consecutive non-overlapping embedding matrix will give us one corresponding singular spectrum. Clearly if we set our value of n to be too high, each embedding matrix could, for example, represent a minute's worth of wake EEG and the dynamics of the neocortex and hence the structure of the underlying system will certainly have changed over that period of time. By taking too large a section of data we are in a sense averaging the dynamics of the system over that period, instead of monitoring the subtle changes occurring within that section.

With the above requirements we performed a number of experiments described in the next section to determine a reasonable value for n based on all of these factors.

3.3 Experiment varying the value of n

Part of the remit for this project was to develop a graphical user interface (GUI) that could be used to accurately display and investigate the properties of large sections of EEG data alongside the results of the various feature extraction techniques employed (user guide and documentation are available separately). Utilising this software we were able to investigate the effect of changing the value of n , on the structure of the singular spectrum.

From the previous section we have an initial value of $M = 150$. Therefore we must ensure full rank of the embedding matrix, $n \geq M$, so we performed experiments varying n from an initial value of 150 up to 1500. Remembering that each embedding matrix, \mathbf{X} , contains $(M + n - 1)$ time series samples and that the sampling rate is $\sigma_{EEG} = 256Hz$, these values of n represent a range of approximately 1 second ($n = 150$) up to 6 seconds ($n = 1500$).

3.3.1 Results

The following pages show snapshots taken from the GUI for the different values of n . They show at the top a 3-D contour plot of the singular value spectra through time that directly corresponds to the original EEG trace shown below it.

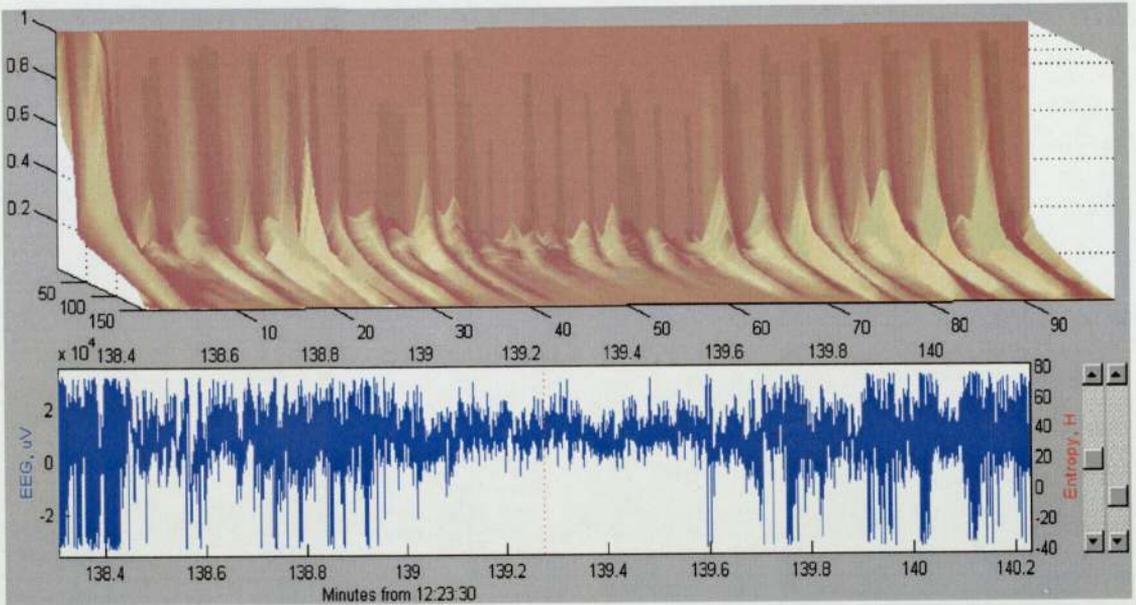


Figure 3.1: GUI snapshot for $M = 150$, $n = 150$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.

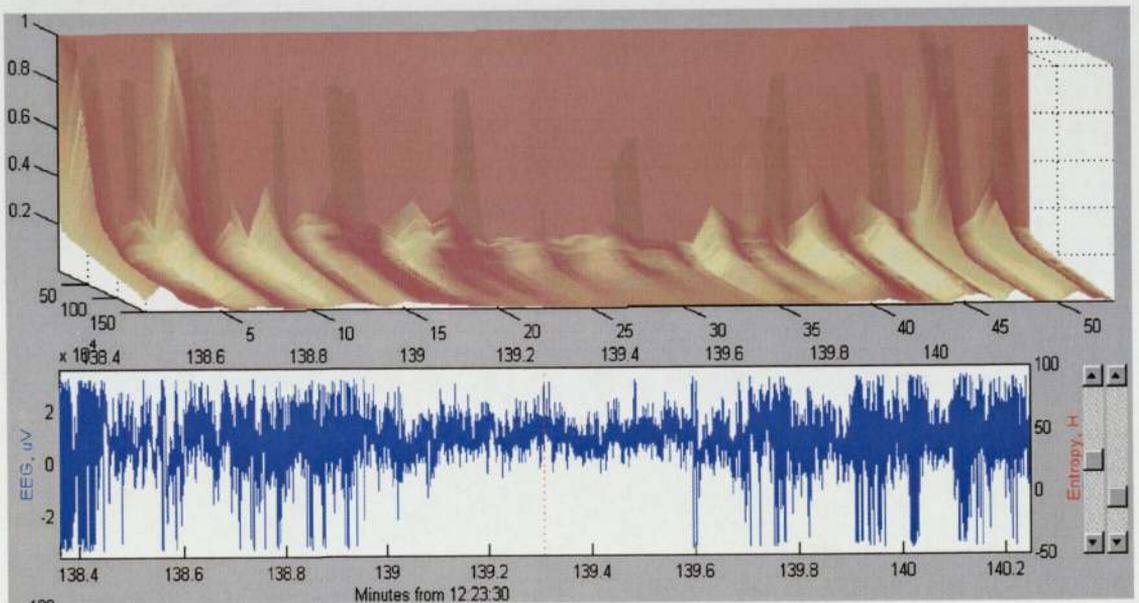


Figure 3.2: GUI snapshot for $M = 150$, $n = 400$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.

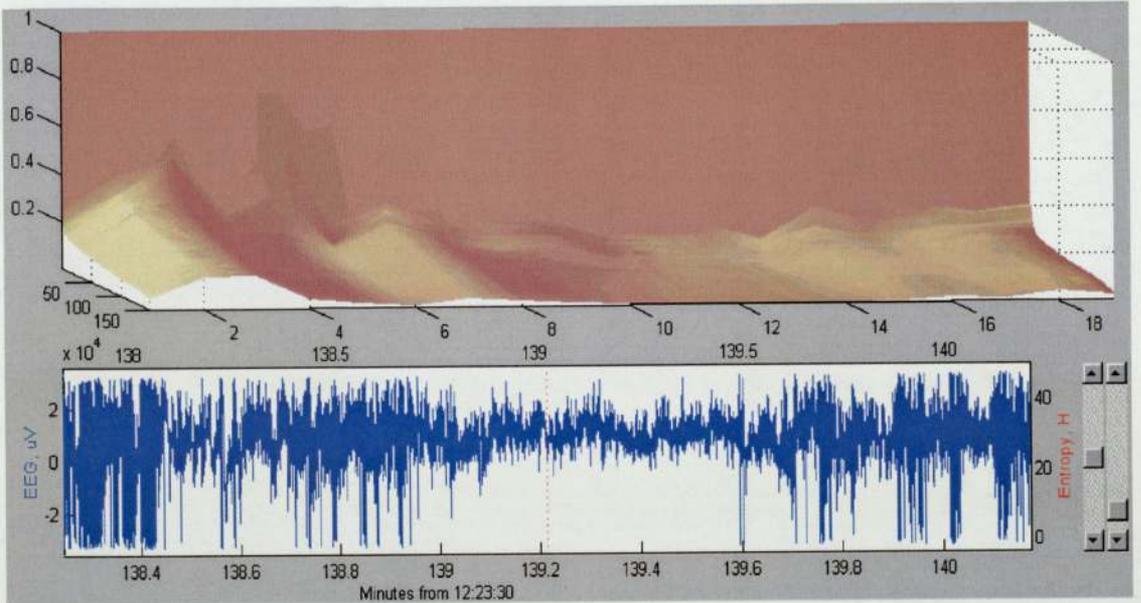


Figure 3.3: GUI snapshot for $M = 150$, $n = 1500$, showing the original EEG trace (in blue), and above it a contour plot of the singular spectrum corresponding to the section of EEG trace.

It can be clearly seen from figures 3.1-3.3 that $n = 150$ satisfies all of the conditions previously outlined: we see how the underlying system appears to change at quite a high rate and we therefore require a relatively small value of n as a result.

3.3.2 Morphology of the singular spectrum

We notice that our initial estimate of $M = 150$ guarantees convergence of the singular spectrum, indicating that we have selected the correct parameters to completely capture the dynamics of the underlying system for wake EEG based on our original hypothesis that the EEG is generated by the non-linear interaction of a few degrees of freedom. We can see that most of the power in the singular spectrum is contained within the first few singular values up to the point at which the kink occurs, usually around the 10th component. Based on previous assumptions we know that these principal components are representative of the relative contributions of the few degrees

of freedom [4] that we have previously discussed and that the relatively low singular values from the point at which that kink in the singular spectrum occurs, onwards, are a result of the intrinsic noise present in the EEG recording [17]. Furthermore we notice in figure 3.1 that the morphology of the singular spectrum changes as small bursts of externally generated noise (artifact) are encountered in the EEG. This is because there are now more degrees of freedom present in the signal due to the noise and therefore the dynamical embedding process is struggling to completely capture the dynamics of the system, highlighted by the fact that we no longer have complete convergence of the singular spectrum for those regions of data. This is further demonstrated by figure 3.4 below, which shows the singular spectrum contour map for a section of EEG data corrupted with EMG noise. It should be noticed that there is now a completely different type of singular spectrum morphology for this type of data.

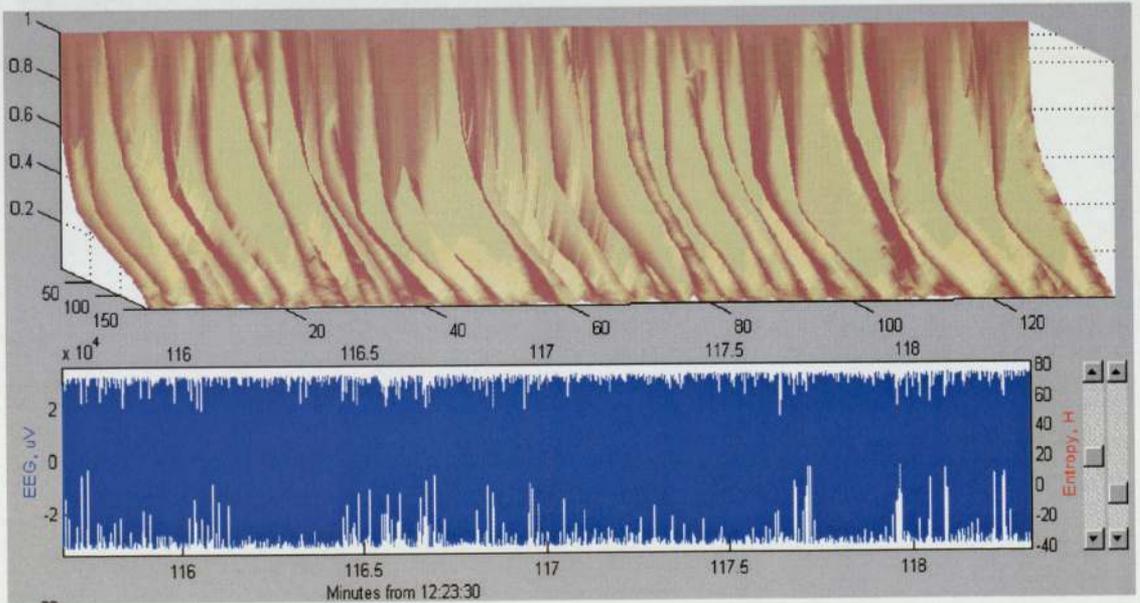


Figure 3.4: GUI snapshot for $M = 150$, $n = 150$ with a section of EEG corrupted with EMG noise

3.4 Summary

We have shown in this chapter that for wake EEG sampled at $\sigma_{EEG} = 256Hz$, the method of dynamical embedding, with parameters $\tau = 1/256$, $M = 150$ and $n = 150$, is able to capture the dynamics of the underlying system.

Using these parameters, each embedding matrix, \mathbf{X} , will represent approximately 1s of data and an advantage of being able to use such relatively small sections of data to perform the embedding is that the computational efficiency for implementation of this method will be such that real time analysis is possible, since the algorithm is exponential in the size of \mathbf{X} .

We also noticed how the morphology of the singular spectrum changed depending on the structure of the EEG and it is this that we wish to take advantage of when attempting to model the complexity of different sections of EEG data. The next chapter investigates how effective different measures of complexity are at detecting subtle changes in wake EEG.

Chapter 4

Complexity analysis

Based on our original hypothesis (chapter 3), we assume that each degree of freedom must have some physiological interpretation as to the current brain state. However although we now have an indication of the relative significance of each degree of freedom, we do not know exactly how these degrees of freedom combine to form the underlying generator, which may consist of the linear interaction of a few smaller non-linear generators [2] which will have a more obvious physiological meaning. Therefore it is possible that the structure of the EEG signal and the underlying generators can change, due to a change in brain state, without a dramatic change in the weighting of each degree of freedom, resulting in only a small change in the morphology of the singular spectrum. The general structure of the singular spectrum for normal wake EEG is shown in figure 4.1. This chapter investigates the implementation of two different types of complexity measure, (i) entropy and (ii) Fisher's complexity measure, that take advantage of this structure in the singular spectrum to highlight changes in signal complexity. Although our primary aim is to distinguish between different levels of complexity within normal wake EEG, and not to develop an EMG detector, it is vital that the complexity measure used is able to distinguish between the two types of waveform.

The first part of this chapter discusses the issues concerned with the application of the entropy measure to a function that is not inherently a probability distribution and then presents the results of using entropy to measure the complexity of the EEG. The implementation of the Fisher measure is then presented along with the many practical considerations involved in its accurate calculation. Finally there is a summary of the chapter and an evaluation of the relative performances of the two different types of complexity measure.

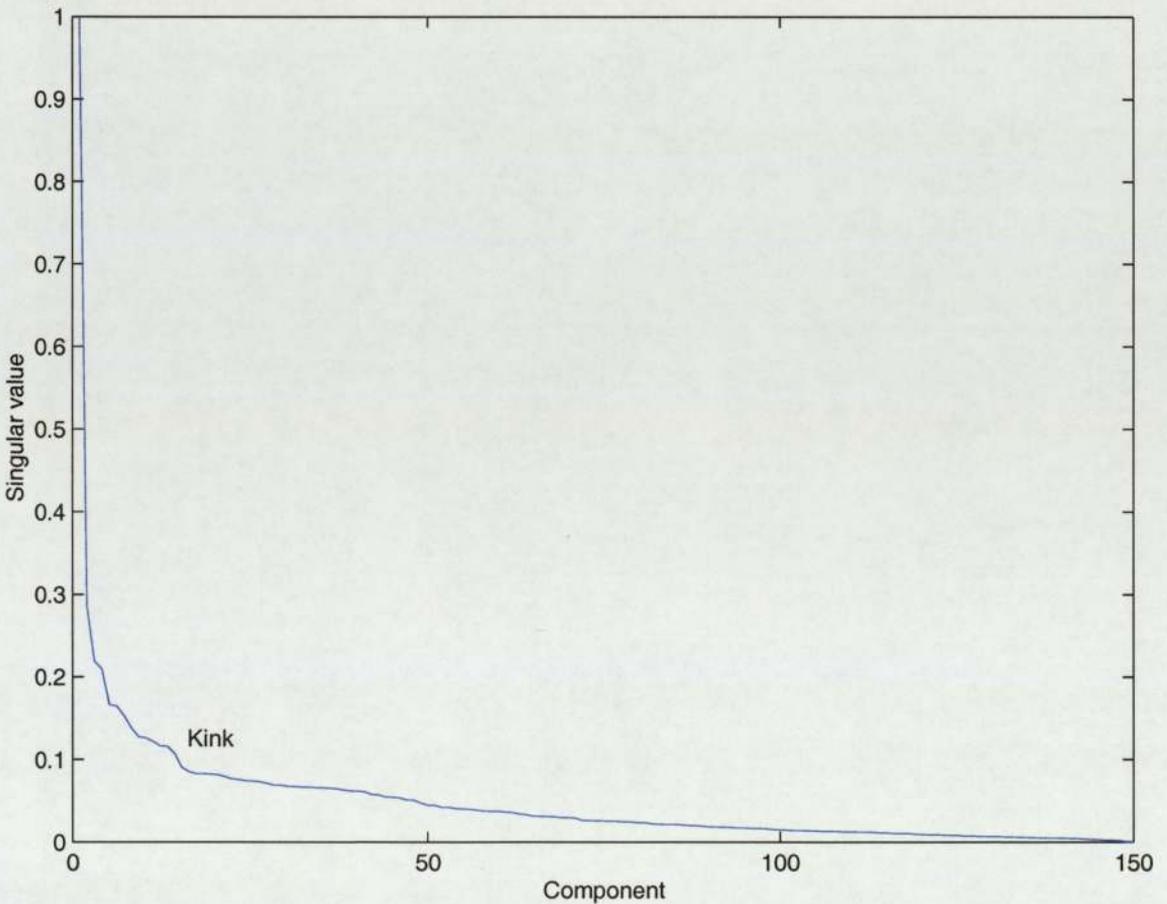


Figure 4.1: An example of the singular spectrum morphology for wake EEG. The location of the kink as suggested by previous work indicates the separation between signal and noise.

4.1 Entropy

The intuitive measure of disorder, entropy was originally developed by physicists for application in thermodynamics. It was not until 1948 that Shannon introduced entropy to information theory as a measure of uncertainty or information content in probabilistic systems. The *differential entropy* [8], which is defined for a continuous random variable x as

$$H(x) = - \int_{-\infty}^{\infty} p(x) \log p(x) dx \quad (4.1)$$

is used to measure the change in uncertainty or information content of a continuous function $p(x)$.

4.1.1 Entropy applied to the singular spectrum

We can clearly see that the structure of the singular spectrum for normal wake EEG is not random with most of the power, and therefore information content, concentrated in the first few singular values. We also notice that for different structures such as EMG noise and other artifacts there is a complete change in the morphology and therefore the spread in power distribution, leading to a change in the information content of the signal (see figure 4.4). On a smaller and more subtle scale we also notice that the structure and location of the ‘kink’ in the singular spectrum changes throughout wake EEG. As stated in section 3.2.2 our assumption is that the ‘signal’ part of the EEG is contained within the first few singular values leading up to the kink, and thereafter the smaller singular values correspond to the noise. We then hypothesise that this subtle change in the singular spectrum, due to an increase or decrease in the number

of degrees of freedom representing the signal part of the EEG, indicates a change in the brain state of the subject. We would like to monitor this change by using the most appropriate type of complexity measure.

The singular spectrum clearly lends itself to measures of information content and although it is not intrinsically a probability distribution we may model it as a probability density function (PDF) if it satisfies the requirements of a PDF (see Roberts *et al.* [15]), namely

- The probabilities must sum to 1 ie $\sum_{i=1}^M \sigma_i = 1$
- All values of $p(x)$, (σ_i) , must be positive

Clearly all singular values, σ_i , are positive and by introducing appropriate normalisation we may regard the singular spectrum as a PDF. Assuming points are sampled uniformly from σ_i , the discrete version of the formula for the differential entropy is given by

$$H = - \sum_{i=1}^M \hat{\sigma}_i \log \hat{\sigma}_i \quad (4.2)$$

where

$$\hat{\sigma}_i = \frac{\sigma_i}{\sum_{i=1}^M \sigma_i} \quad (4.3)$$

4.1.2 Entropy and noise

The graphs in figures 4.2 and 4.3 show the weighting function that is attributed to each singular value when the differential entropy is calculated and an example of a singular

spectrum that has been normalised to power 1 with superimposed weighting function.

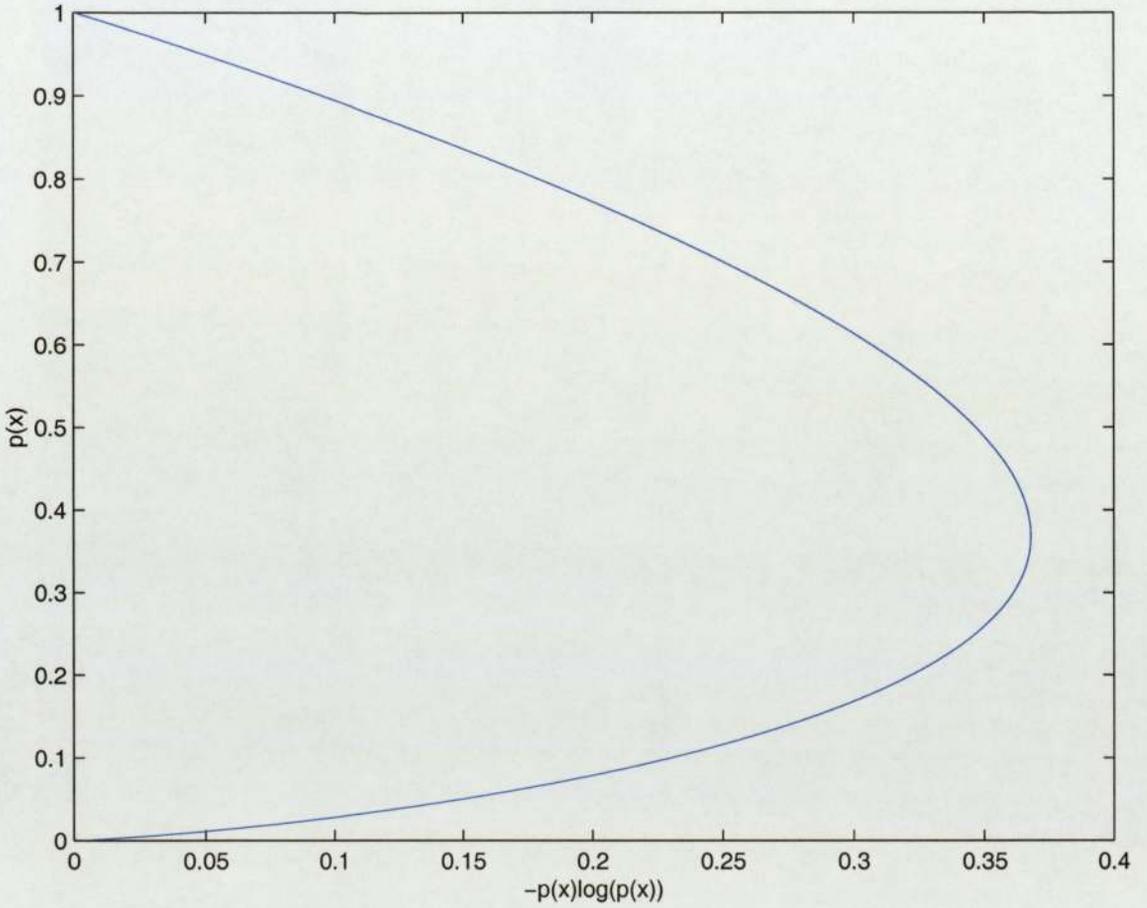


Figure 4.2: Weighting function introduced by entropy, $-p(x)\log p(x)$

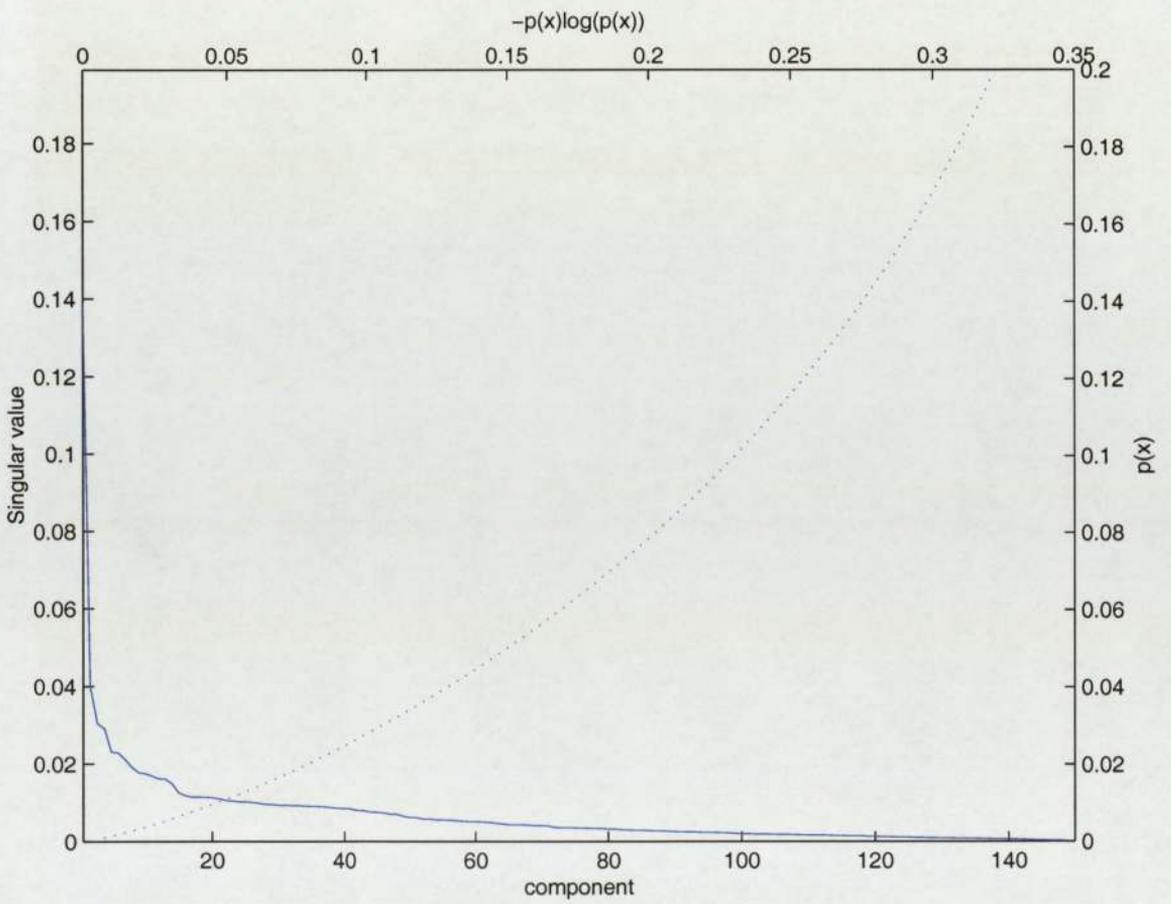


Figure 4.3: Typical wake EEG singular spectrum normalised to power 1, with superimposed entropy weighting function(dotted line)

We can see that the superimposed weighting function happens to intersect the singular spectrum at roughly the point at which the kink occurs. It increases or decreases very sharply either side of it and therefore we would expect to detect subtle changes in the morphology of the singular spectrum such as the movement of the kink which defines the boundary between signal and noise. We also notice that due to the normalisation of the signal, the ability to distinguish between clean wake EEG and sections corrupted with EMG, an example of which is shown below in figure 4.4, could be compromised as the singular spectrum for EMG now appears to be very similar to that for wake EEG.

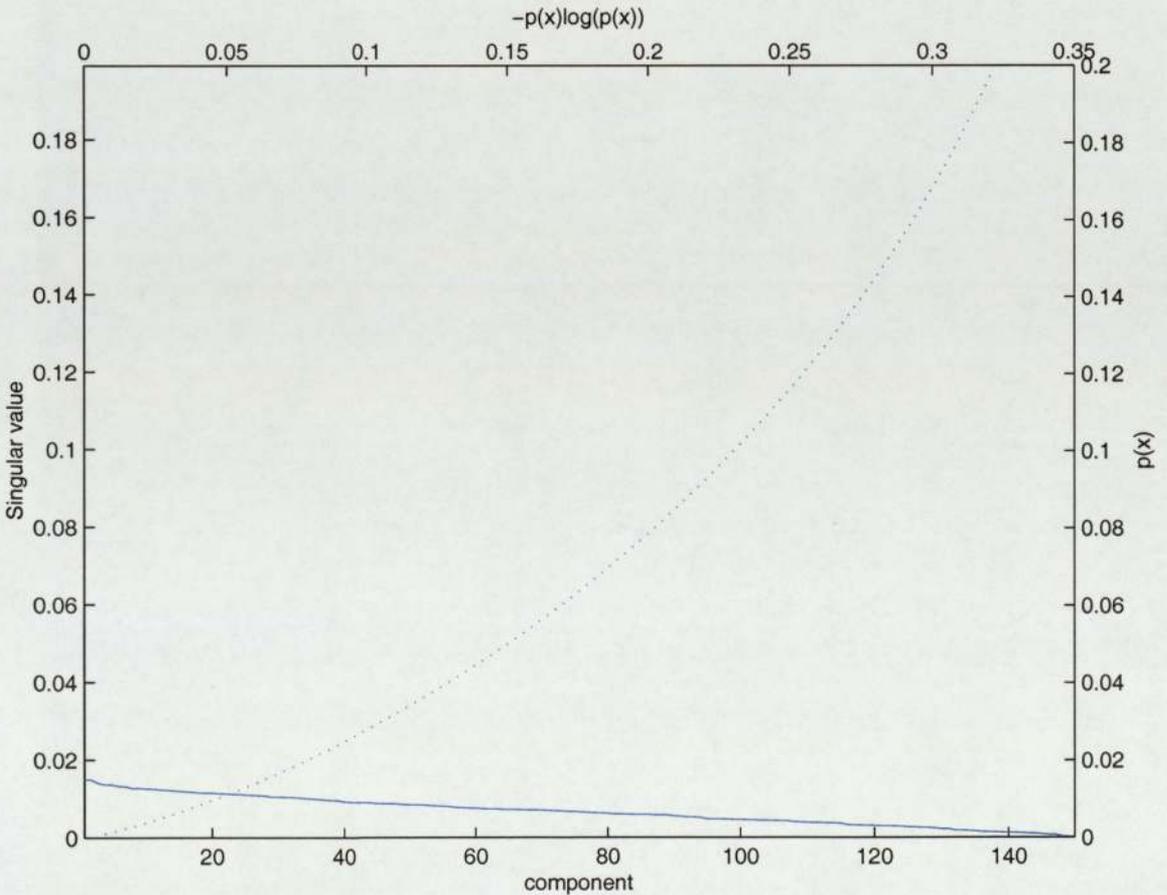


Figure 4.4: Typical singular spectrum for EEG corrupted with EMG, normalised to power 1, with superimposed entropy weighting function(dotted line). Notice the lack of kink structure, due to the degrees of freedom introduced by the external EMG generator.

4.1.3 Different normalisation

Although we would no longer fulfil the theoretical requirements for modelling the singular spectrum as a PDF, we could emphasise the change in the structure of the singular spectrum for the two extreme types of signal by using a more appropriate normalisation.

By normalising the singular spectrum so that the first singular value, σ_1 , has a magnitude of 1, we would take full advantage of the structure of the differential entropy weighting function, $-p(x)\log(p(x))$ where $p(x) = \sigma_i$, to highlight a change in signal complexity. Figures 4.5 and 4.6 show the different normalisations of the singular spectra corresponding to the same sections of normal wake EEG and EEG corrupted with EMG shown previously with the superimposed weighting function. We can now clearly see the difference in the singular spectrum between normal wake EEG and EEG corrupted with EMG, and would still expect to be able to detect any changes in the kink structure.

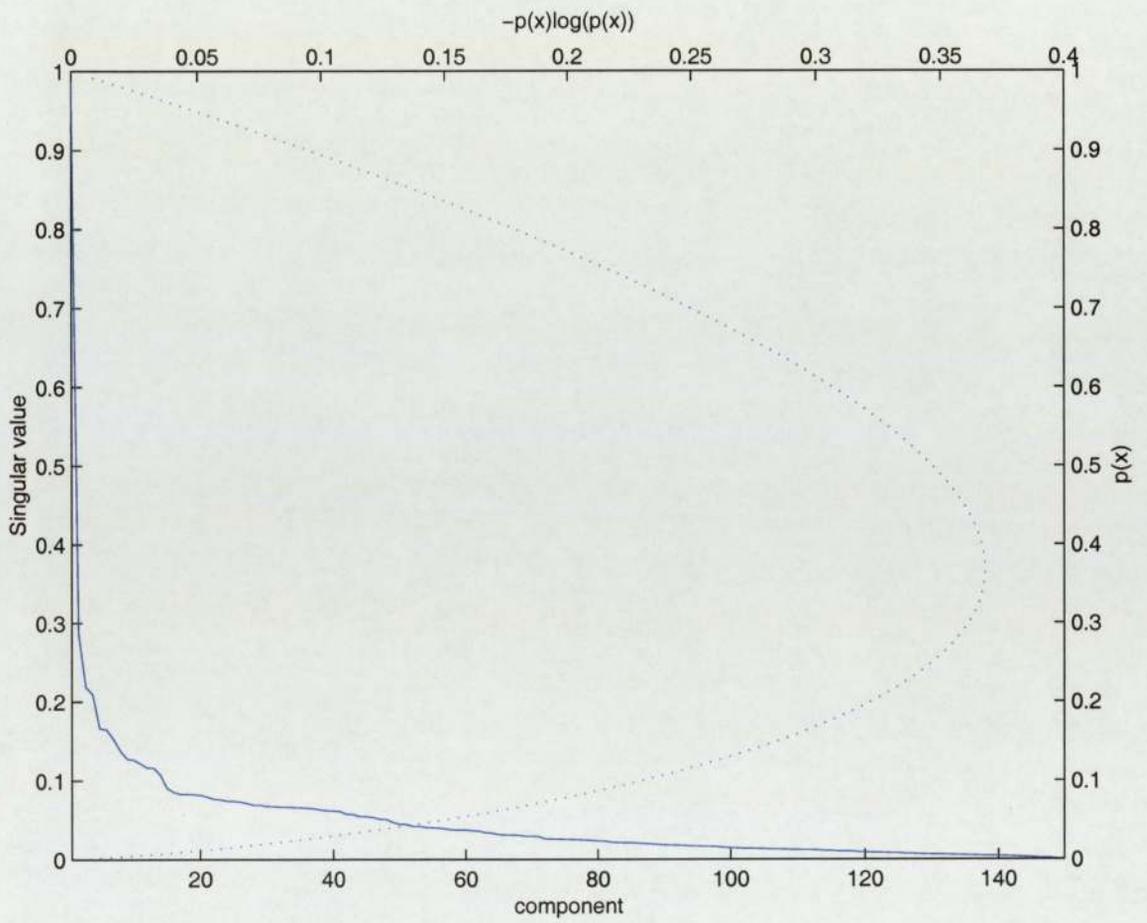


Figure 4.5: Typical wake EEG singular spectrum normalised to first value 1 with superimposed entropy weighting function (dotted line)

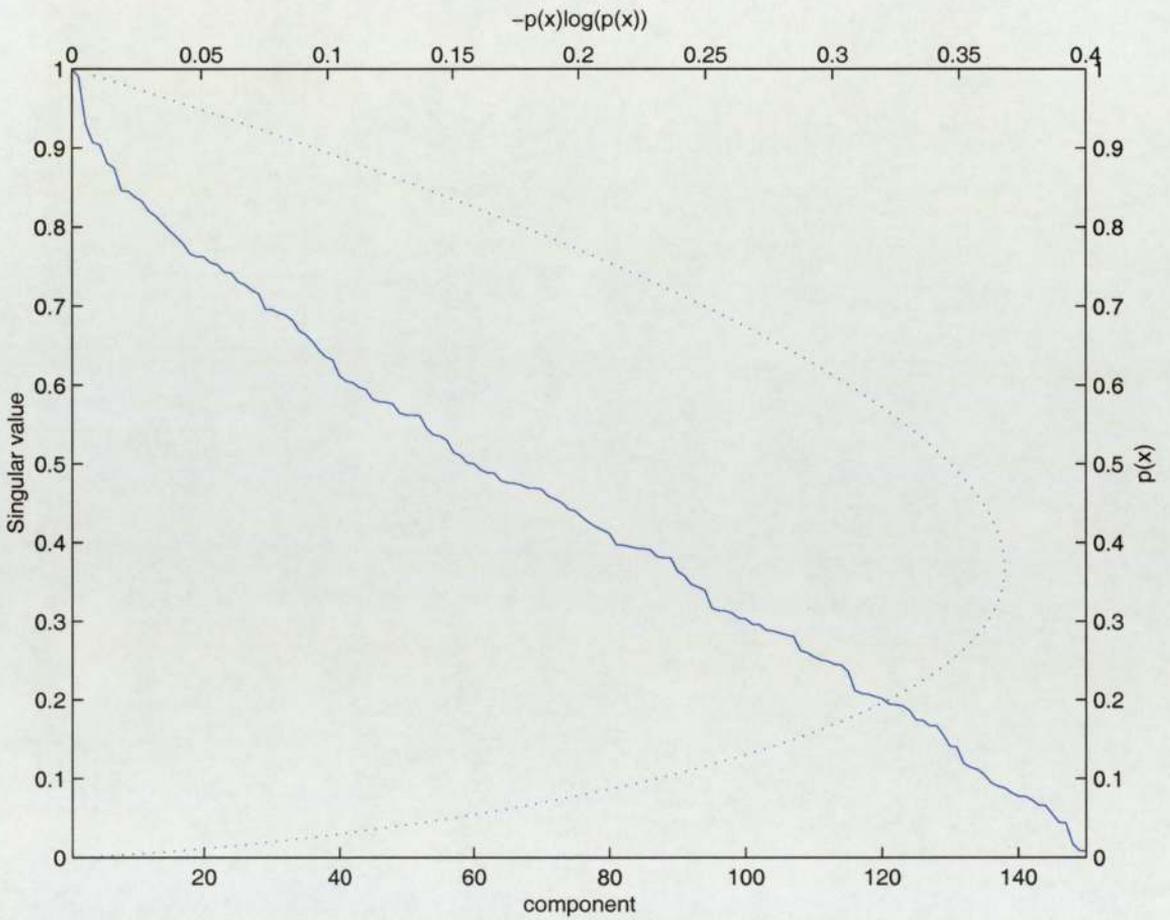


Figure 4.6: Typical singular spectrum for EEG corrupted with EMG, normalised to first value 1 with superimposed entropy weighting function (dotted line). The different scaling further highlights the lack of kink structure in the singular value spectrum for data corrupted with EMG

4.1.4 Results

The following page shows screen shots of sections of the GUI that was designed as part of the project. Figures 4.7 and 4.8 show the original EEG in blue with the superimposed entropy complexity measure in red. Figure 4.7 compares the two different types of entropy measure, with different normalisations, when applied to normal wake EEG, and figure 4.8 compares the same two measures over normal wake EEG corrupted with EMG.

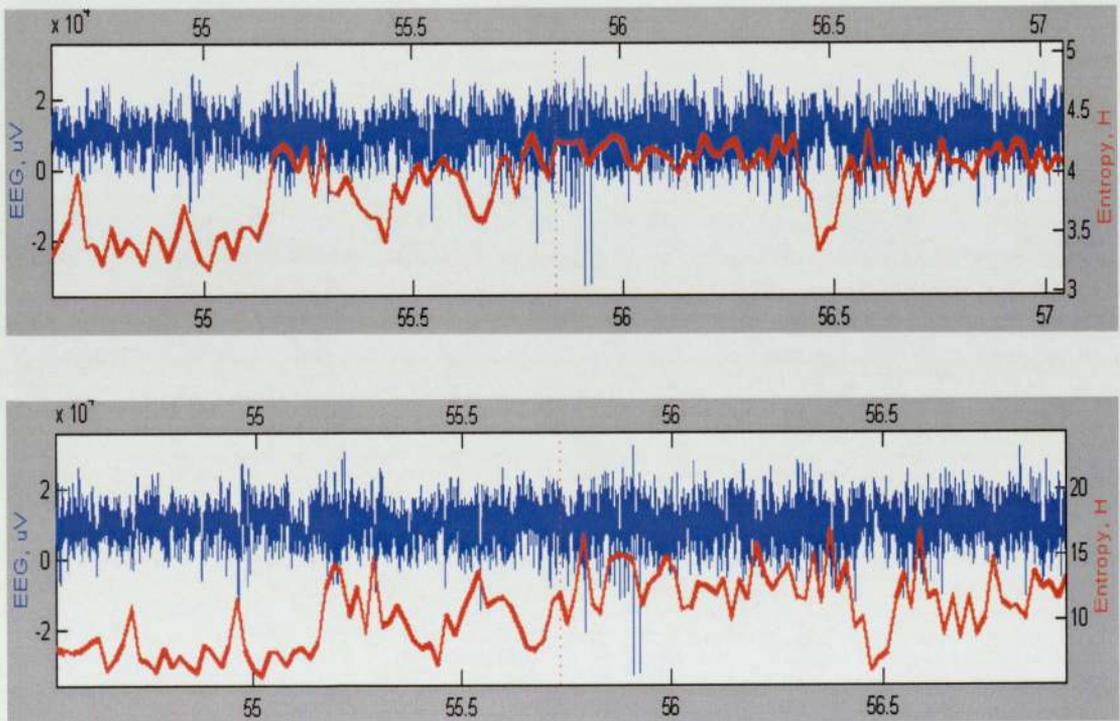


Figure 4.7: Screen shots from GUI showing the two entropy measures for wake EEG. Notice how the two entropy complexity measure are almost identical when monitoring subtle changes in wake EEG

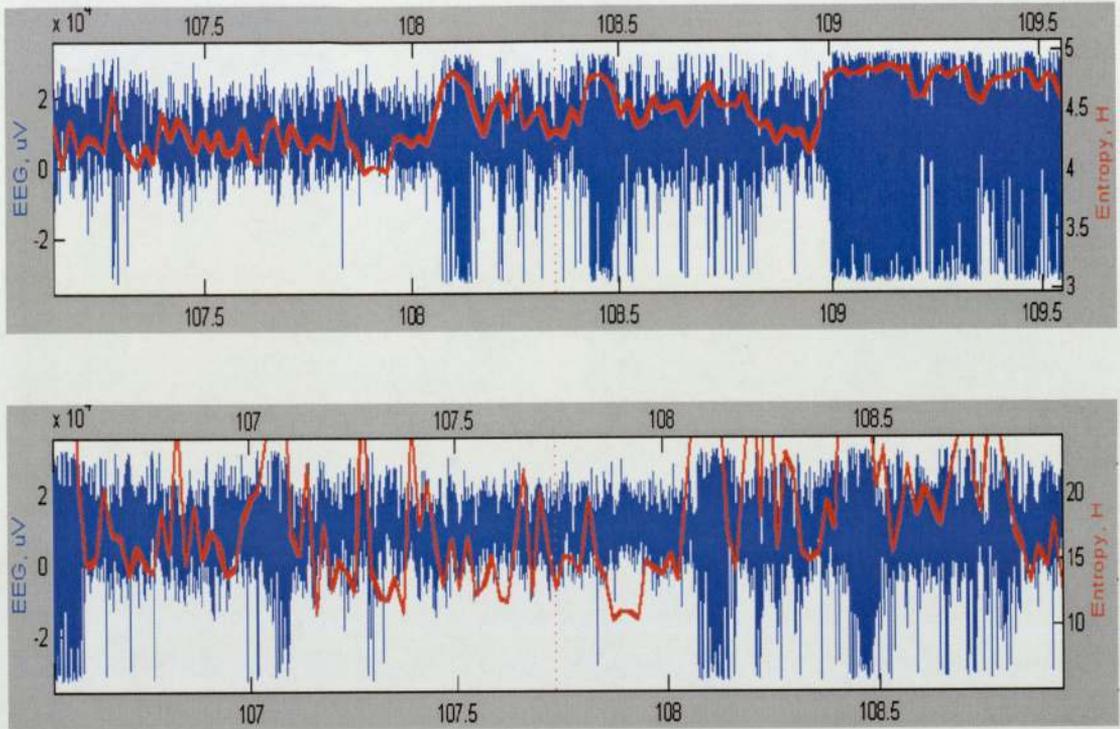


Figure 4.8: Screen shots from GUI showing the two entropy measures for wake EEG corrupted with EMG artifact. Notice how the entropy complexity measure with first value 1 normalisation has a far greater dynamic range.

4.1.5 Conclusions

The results confirm our hypothesis, that different normalisation will result in two entropy measures that are selectively sensitive to changes in different sections of the singular spectrum.

Sensitivity to subtle changes in kink structure

From figure 4.7 we can see that the entropy measure is highly variable corresponding to very subtle changes in the EEG waveform and the singular spectrum. Therefore

we can say that the entropy measures normalised to power 1 and first value 1, are sensitive to subtle changes in kink structure corresponding to a change in the number of degrees of freedom.

Sensitivity to EMG artifact

Figure 4.8 confirms our hypothesis that the entropy measure with power 1 normalisation is unable to highlight the difference in structure of EEG corrupted with EMG artifact. The entropy values corresponding to the corrupted section are very similar to those in normal wake EEG sections. However for the first value 1 normalisation the entropy values for the corrupted section are off the scale, peaking at approximately $H = 47$, indicating a much higher relative signal complexity. Therefore the normalisation of the singular spectrum, resulting in the best entropy measure, is given by equation 4.4 below and requires the first singular value, σ_1 , to have a magnitude of 1. This normalisation allows us to clearly highlight sections of EEG that are contaminated with EMG noise (which are of no interest to us) whilst retaining the sensitivity to subtle changes in the morphology of the singular spectrum.

$$\hat{\sigma}_i = \frac{\sigma_i}{\sigma_1} \tag{4.4}$$

4.2 Fisher's information measure

Proposed by Fisher in 1934, Fisher's information content, $I(\theta)$ is defined as the information about θ in a sample of n independent observations [9] and is given by

$$I(\theta) = E \left[\left(\frac{\partial(\log P(\theta))}{\partial \theta} \right)^2 \right] \quad (4.5)$$

Where $P(\theta)$ is the likelihood of the sample, and the expectation E is over n samples of $P(\theta)$.

If we sample uniformly from θ we find that the discrete version of the Fisher measure is given by equation 4.6. The full derivation is given in appendix A.

$$I_n(\theta) = \sum_{i=1}^n \frac{(P'_i(\theta))^2}{P_i(\theta)} \quad (4.6)$$

Where

$$P'_i(\theta) = \frac{\partial P_i(\theta)}{\partial \theta} \quad (4.7)$$

This is equivalent to the sum of the partial derivatives at each point squared divided by the value of the function at each point.

4.2.1 Fisher measure applied to the singular spectrum

The structure of the singular spectrum for normal wake EEG is such that the gradient varies greatly; being very large over the first few singular values, then gradually flattening off with a stationary point of inflection near the kink. Any movement in the position of this kink will result in a change in the gradient to the region either side of it. Therefore we expect the Fisher measure to be able to detect subtle changes in the location of the kink.

As was shown earlier in this chapter the singular spectrum corresponding to EMG artifact is completely different, therefore we also expect the Fisher measure to be able to distinguish between normal wake EEG and EEG corrupted with EMG artifact.

4.2.2 Numerical calculation of the Fisher measure

By modelling the singular spectrum as a PDF, as before, where the $\hat{\sigma}_i$ now correspond to the likelihood $P_i(\theta)$, we may monitor the change in Fisher information content of the singular spectrum. Computationally, the most difficult part of the Fisher measure calculation is the evaluation of the derivative at each point of the singular spectrum. There are two obvious ways of doing this:

- Functional approximation of the singular spectrum by a continuous function leading to a functional evaluation of the derivative at each point.
- The use of existing singular values to perform a discrete calculation of the Fisher measure, by numerical implementation of difference equations.

4.2.3 Functional approximation

Based on our hypothesis that the structure of the singular spectrum and specifically that around the kink is an indication of the underlying brain state, it is essential that none of the shape in the structure is lost when the functional approximation is performed. This section describes two techniques that were used to try and model the singular spectrum.

Polynomial interpolation

Interpolation of the singular spectrum using least squares polynomial curve fitting [10], enables us to accurately calculate the derivative at each data point, since this only requires functional differentiation and the evaluation of a known n^{th} order polynomial, which is trivial. However it proved impossible to retain the structure of the sharp initial gradient and the kink without introducing large errors in other parts of the function. Figures 4.9-4.11 show plots of the evaluated interpolation function superimposed on the original singular spectrum for values of n between 10 and 36. We can clearly see that the interpolation resulted in either oscillations due to over fitting, or smoothing of the kink structure when lower order polynomials were used.

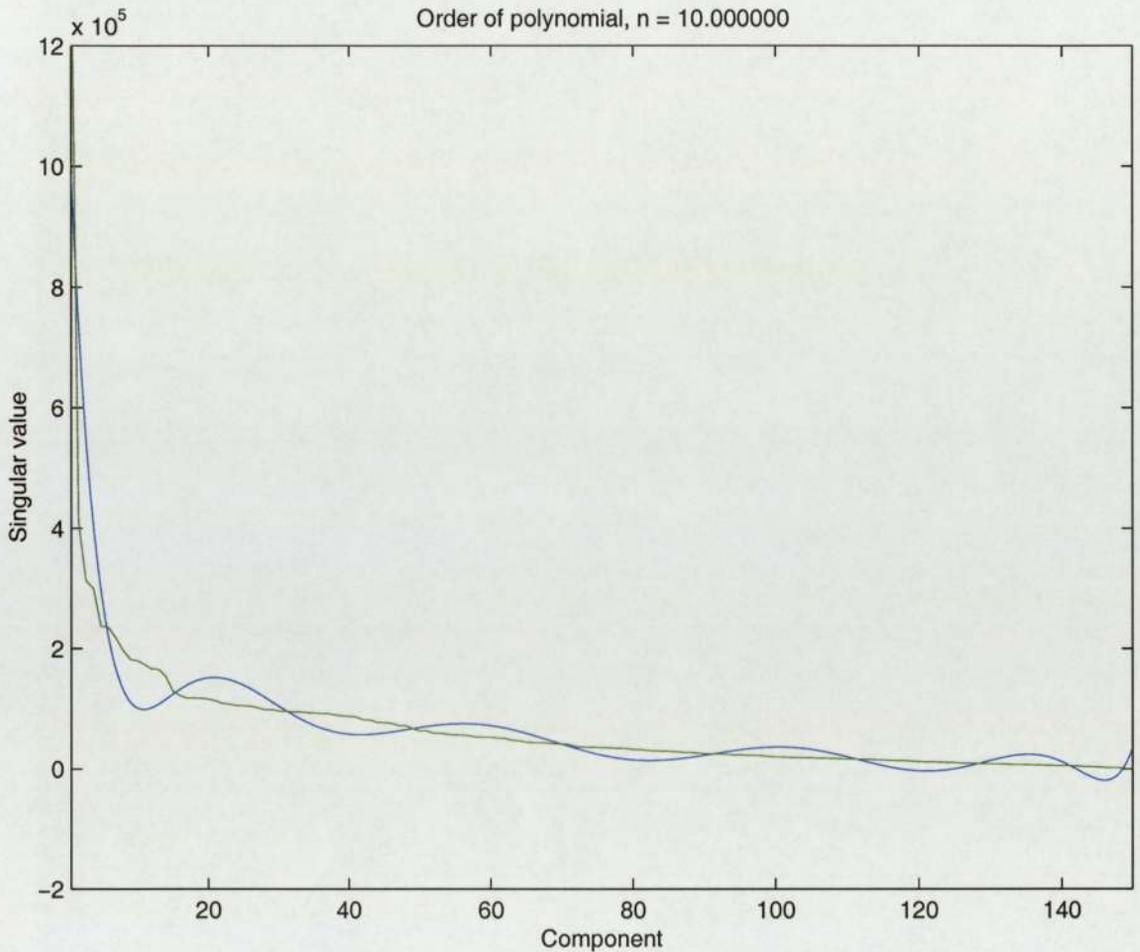


Figure 4.9: Polynomial interpolation plot for $n = 10$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. Notice how there are considerable errors between the original values and the interpolated function, this is because the degree of the polynomial, n , is too low.

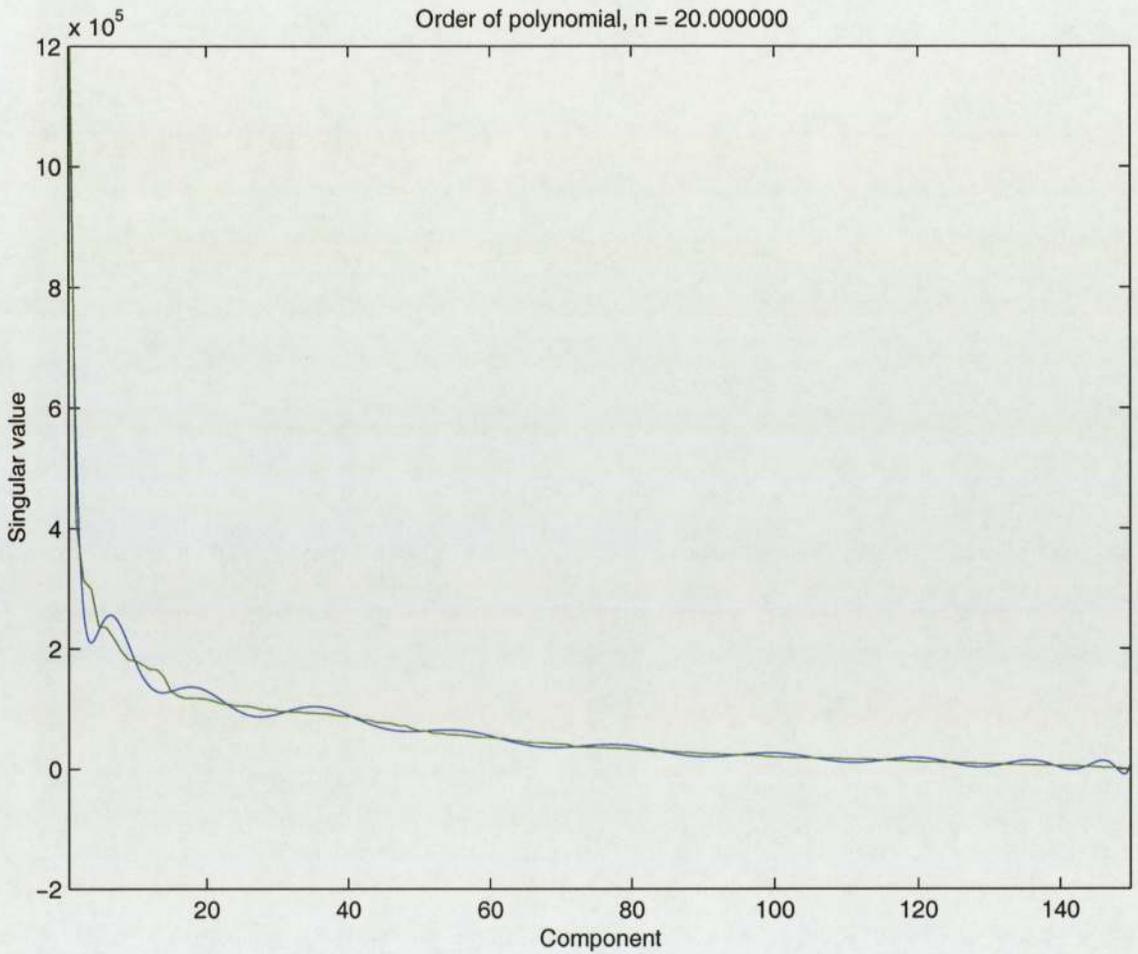


Figure 4.10: Polynomial interpolation plot for $n = 20$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. Here we have increased the value of n to 20, but we are still under fitting the original data.

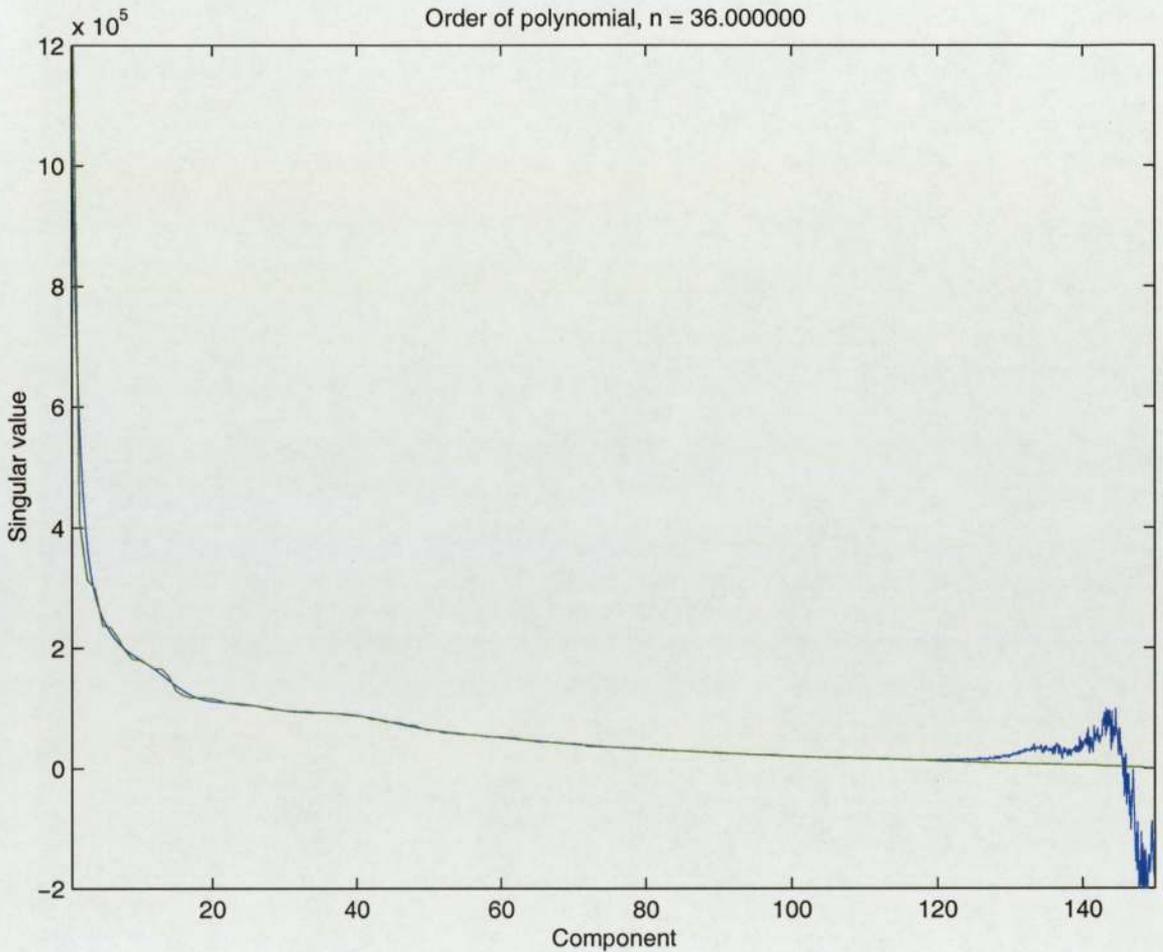


Figure 4.11: Polynomial interpolation plot for $n = 36$, the original function, which is the singular value spectrum, is shown in green with the interpolated function superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 15 indicating an underlying dimensionality of 7. The degree of the polynomial, n , is now considerably larger (36), but although we can now approximate the first section well, we do this at the cost of introducing severe oscillations at the end.

Spline interpolation

It is well known that interpolation using high-order polynomials often produces ill-behaved results [12]. There are numerous approaches to eliminating this poor behaviour. Of these approaches, cubic splines are very popular. In cubic splines, cubic polynomials are found to approximate the curve between each pair of data points, called abscissae. Since an infinite number of cubic polynomials can be used to approximate a curve between two points, additional constraints are placed on the cubic polynomials to make the result unique. By constraining the first and second derivatives of each cubic polynomial to match at the abscissae, all internal cubic polynomials are well defined, with the slope and curvature of the approximating polynomials continuous across the abscissae. Since the first and last polynomials do not have adjoining cubic polynomials additional constraints must be introduced. The most common approach is to adopt a *not-a-knot* condition. This condition forces the third derivative of the first and second cubic polynomials to be identical, and likewise for the last and second to last cubic polynomials. The derivative of a function described by splines is then straightforward. Since the k^{th} cubic polynomial is given by

$$s_k(x) = a_k(x - x_k)^3 + b_k(x - x_k)^2 + c_k(x - x_k) + d_k \quad (4.8)$$
$$x_k \leq x \leq x_{k+1}$$

the derivative of $s_k(x)$ is over the same section and is written as

$$\frac{ds_k(x)}{dx} = 3a_k(x - x_k)^2 + 2b_k(x - x_k) + c_k \quad (4.9)$$

Figures 4.12-4.14 overleaf show three typical singular spectra, corresponding to two sections of normal wake EEG, and a section with significant EMG artifact, that have been interpolated using natural cubic splines with *not-a-knot* end conditions.

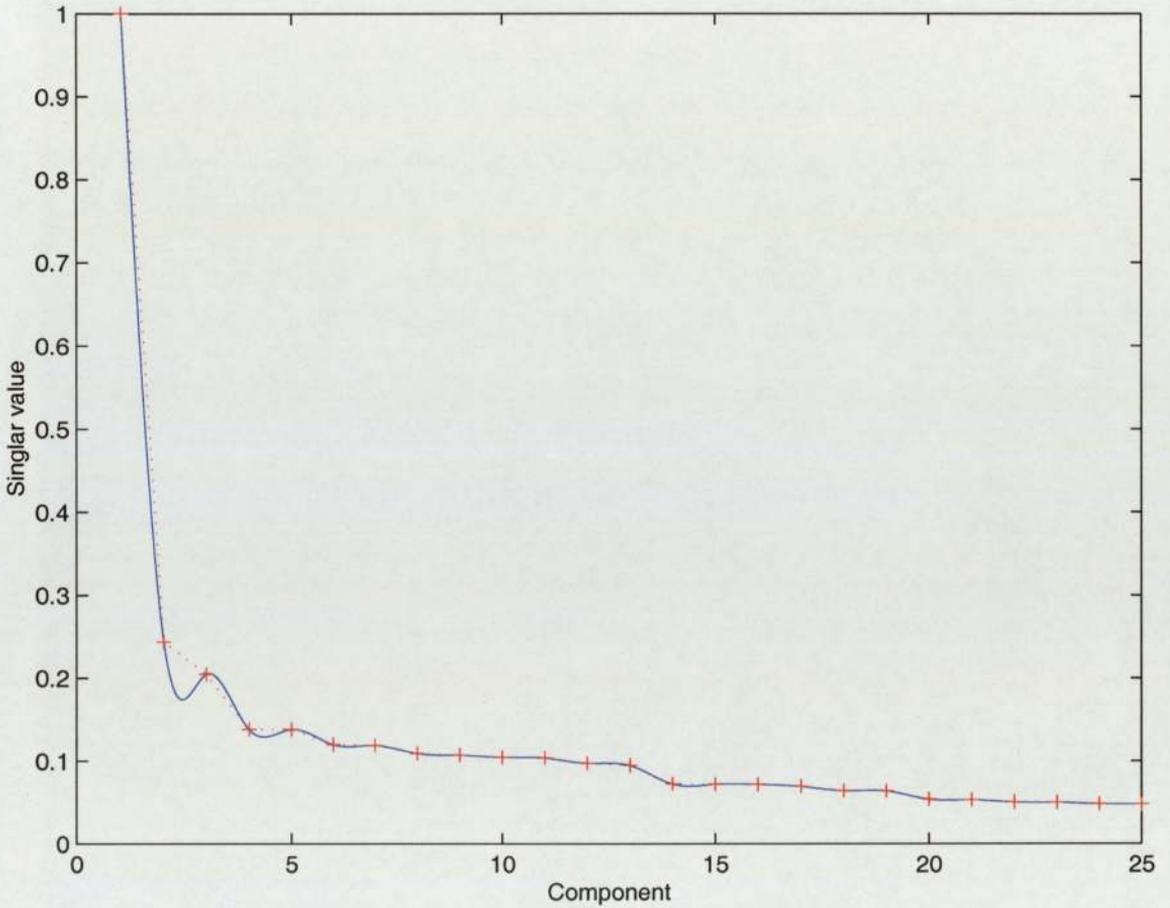


Figure 4.12: Interpolation plot using a natural cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 13 indicating an underlying dimensionality of 6. Notice how there is a considerable error located around components 2 – 3 due to the over fitting of the data.

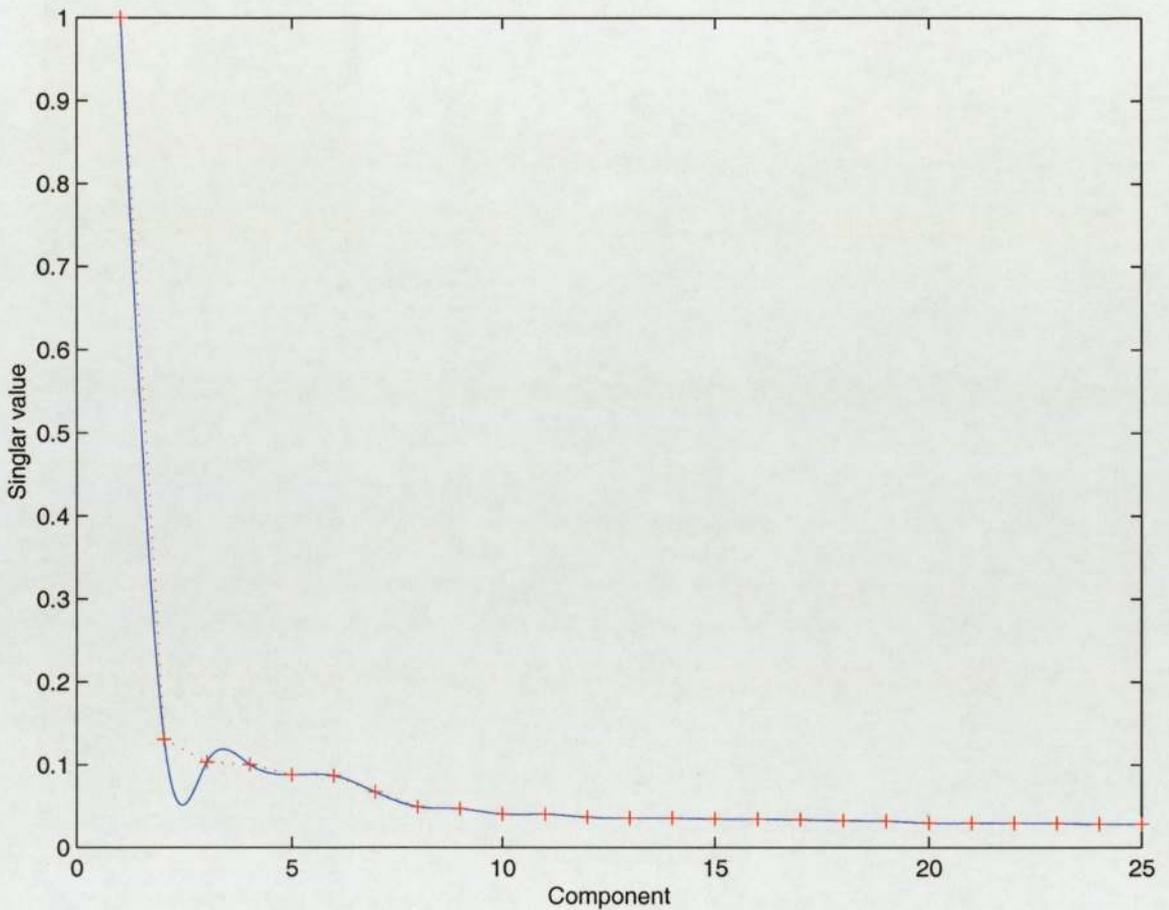


Figure 4.13: Interpolation plot using a cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 7 indicating an underlying dimensionality of 3. Notice how there is a considerable error located around components 2 – 3 due to the over fitting of the data.

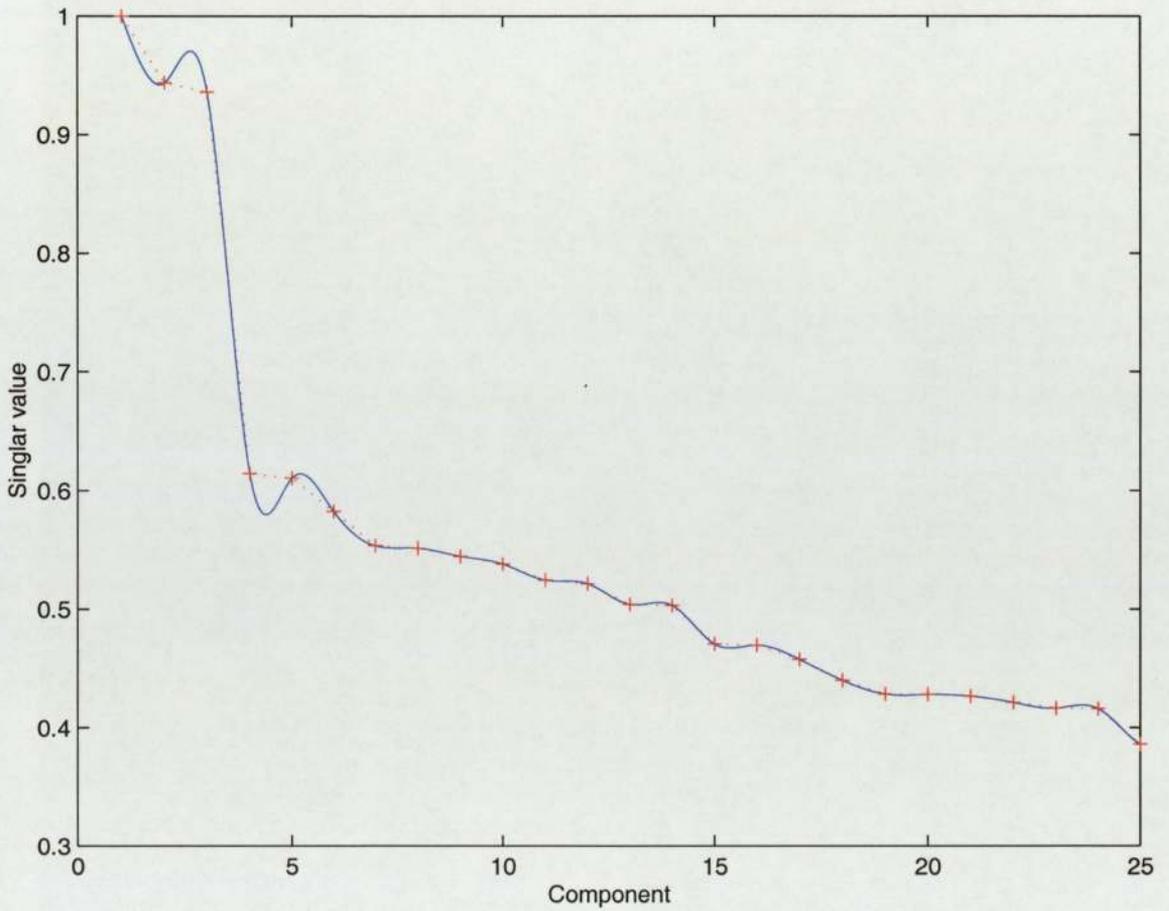


Figure 4.14: Interpolation plot using a cubic spline with *not-a-knot* end conditions. The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to wake EEG corrupted with EMG noise. There is no clear kink indicating the separation between signal and noise due to the extra degrees of freedom introduced to the signal by the external EMG generator. Notice how there is considerable error in the first section due to over fitting.

These results indicate a vast improvement over polynomial interpolation. There is very little noise in the interpolated function and the structure of the kink remains intact. However there still remains a slight error in the first section of the spline function which can be seen as a notch, which will unfortunately introduce noise into the Fisher measure. We can try and eradicate this notch by employing the more advanced technique of smoothing splines.

A smoothing spline is a cubic spline which more or less follows the presumed underlying trend in noisy data using a smoothing parameter P which can vary between 0 and 1. For $P = 0$, this is the least-squares straight line fit to the data, while, on the other extreme, *ie*, for $P = 1$, this is the ‘natural’ or variational cubic spline interpolant. The transition region between these two extremes is usually only a rather small range of values for P and its location strongly depends on the data. It has been shown [13] that the value of P at which this transition occurs, and for which the ‘best’ fit is expected, is given by

$$P = \frac{1}{1 - \epsilon} \tag{4.10}$$

$$\epsilon = h^3/16$$

where h is the average distance between the given abscissae. Figures 4.15 - 4.17 show the improvement in the interpolation of the singular spectra shown previously, although there still remains a slight notch. The results for the calculation of the Fisher measure using smoothed splines are displayed alongside those using discrete difference equations (presented in the following section) at the end of the chapter.

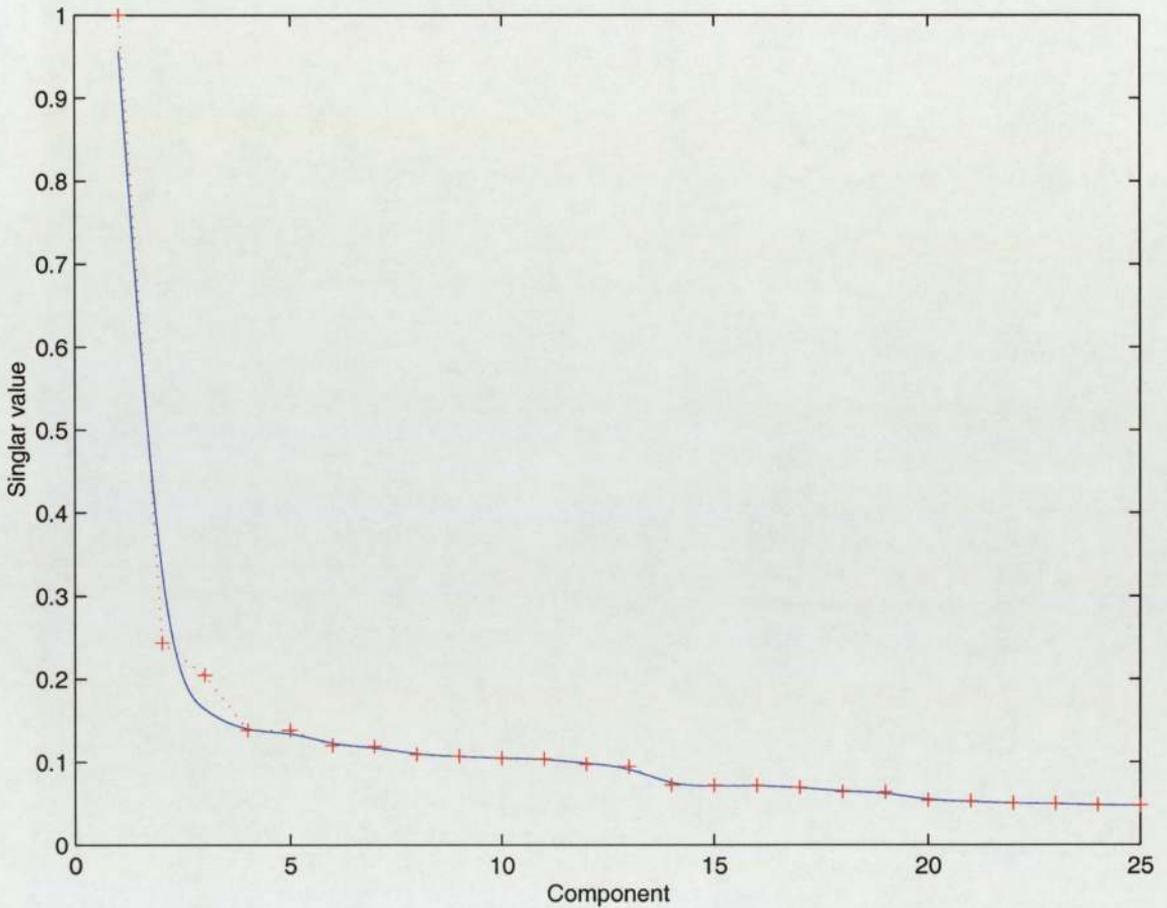


Figure 4.15: Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 13 indicating an underlying dimensionality of 6. Notice how the error that was previously located around components 2 – 3 when using the natural spline, has been smoothed out resulting in no significant error.

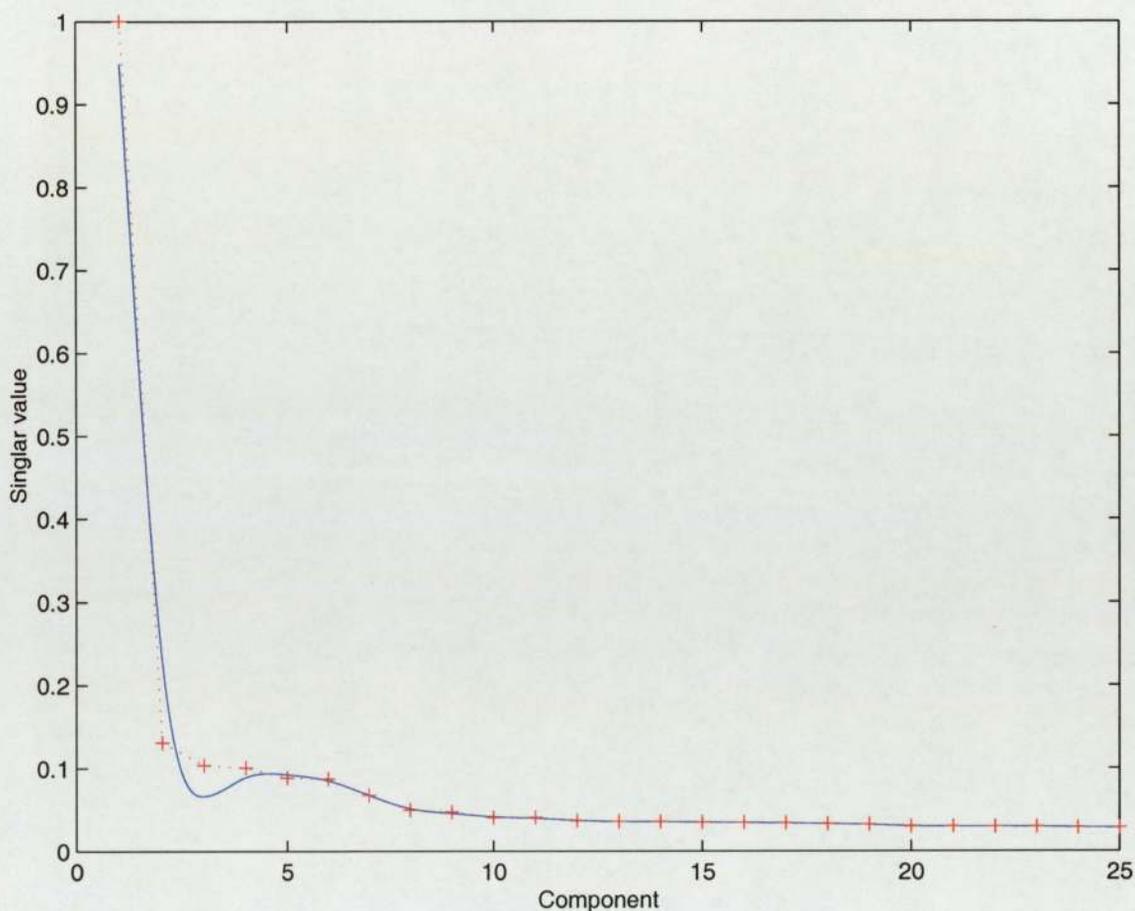


Figure 4.16: Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to normal wake EEG, with the kink located around component 7 indicating an underlying dimensionality of 3. Notice how the error that was previously located around components 2 – 3 when using the natural spline, has been smoothed, but there still remains a significant notch.

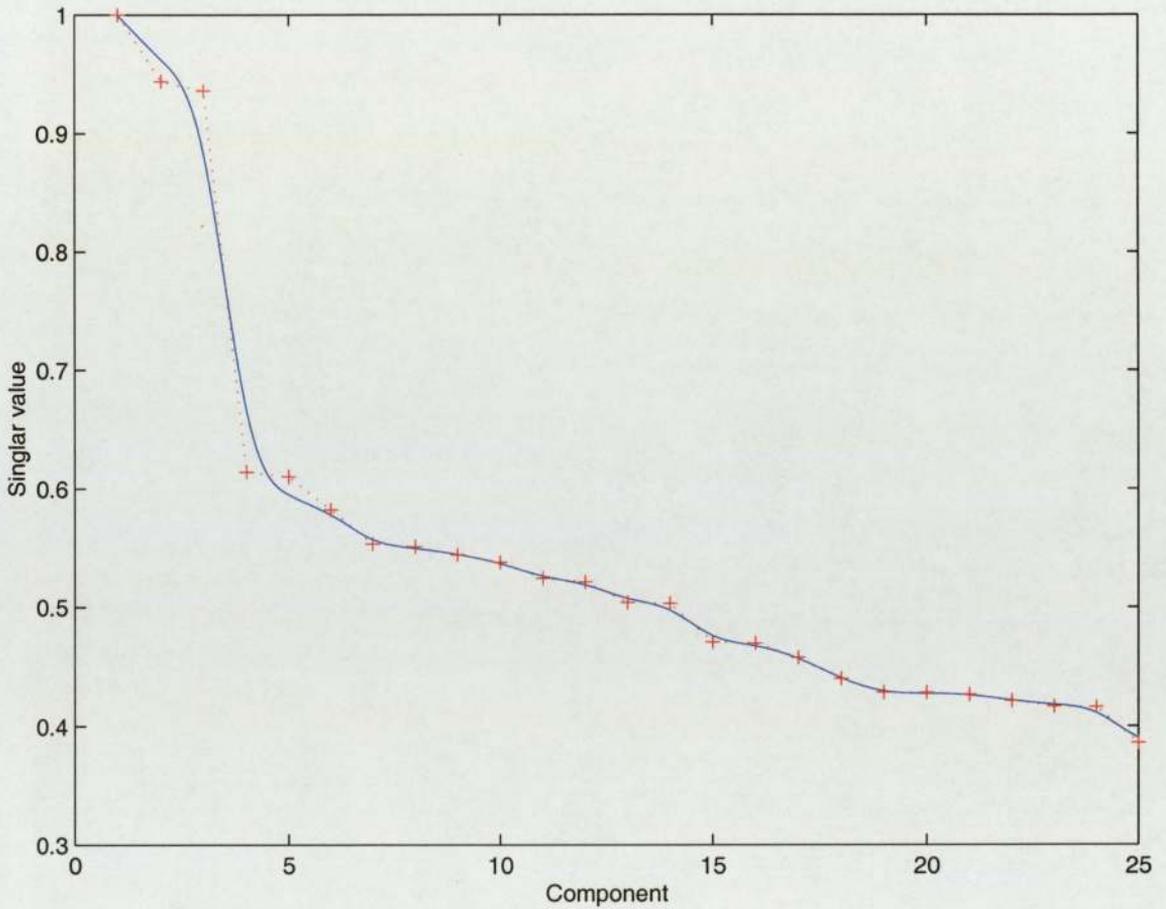


Figure 4.17: Interpolation plot using a smoothed cubic spline with *not-a-knot* end conditions, with optimal smoothing parameter P . The original data points are in red and the spline interpolation function is superimposed in blue. The singular value spectrum corresponds to wake EEG corrupted with EMG noise. There is no clear kink indicating the separation between signal and noise due to the extra degrees of freedom introduced to the signal by the external EMG generator. Notice how the error that was previously located in the first section when using the natural spline, has been smoothed out resulting in no significant error.

4.2.4 Discrete calculation

Differentiation describes the slope of a function at a point, which is a microscopic property of a function. As a result numerical differentiation is avoided where possible due to the extreme sensitivity of the derivative to minor changes in the shape of a function, especially where the data is obtained experimentally. The standard approach is to perform a least squares polynomial fit or cubic spline interpolation and differentiate the resulting polynomials. However we have already seen that these techniques are unable to perfectly fit the singular spectrum, and so we shall perform numerical differentiation of the singular values in order to compare the two different approaches. The derivative of $y = f(x)$ can be approximated by

$$\frac{dy}{dx} \approx \frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{h} \quad (4.11)$$

which is the forward finite difference of y divided by the forward finite difference of x . The corresponding difference equation for the derivative at each point of the singular spectrum is given below by equation 4.12, where we know that $h = 1$.

$$\Delta(\sigma_n) = \sigma_n - \sigma_{n+1} \quad (4.12)$$

We know that that the simple forward difference is extremely sensitive to slight changes in the shape of the function [10], and will result in a very noisy Fisher measure. We may improve the numerical stability by utilising more data points in the calculation of the derivative to average out the sudden changes, and act as a low pass filter. However we must ensure that the averaging does not destroy the information contained within the kink structure. Pan and Tompkins [11] present an implementation of a derivative filter used to calculate the derivative at each point of an electrocardiogram (ECG). The filter is designed to retain the spike structure of the QRS complex and that of the slower T-

wave, seen during each heartbeat for patients who exhibit normal sinus rhythm (NSR). Since the structure in the singular spectrum that we wish to preserve is also apparent in the ECG we shall implement a modified version of the Pan and Tompkins filter [21], that takes advantage of modern processing power giving an improved response, to enable a cleaner Fisher measure to be derived. The difference equation (4.13) for the digital differentiator is given below, the sample delay is 2.

$$\Delta(\sigma_n) = \frac{2\sigma_n + \sigma_{n+1} - \sigma_{n+3} - 2\sigma_{n+4}}{8} \quad (4.13)$$

4.2.5 Results

The following pages show the Fisher measure results of the same 4 hour section of EEG data previously used for the entropy analysis. They show the original EEG trace, with the Fisher complexity measure superimposed. Figure 4.18 shows the three different Fisher measures calculated using smoothed cubic splines, forward difference equation and digital derivative filter, applied to wake EEG. Figure 4.19 shows the three different Fisher measures applied to EEG corrupted with EMG artifact.

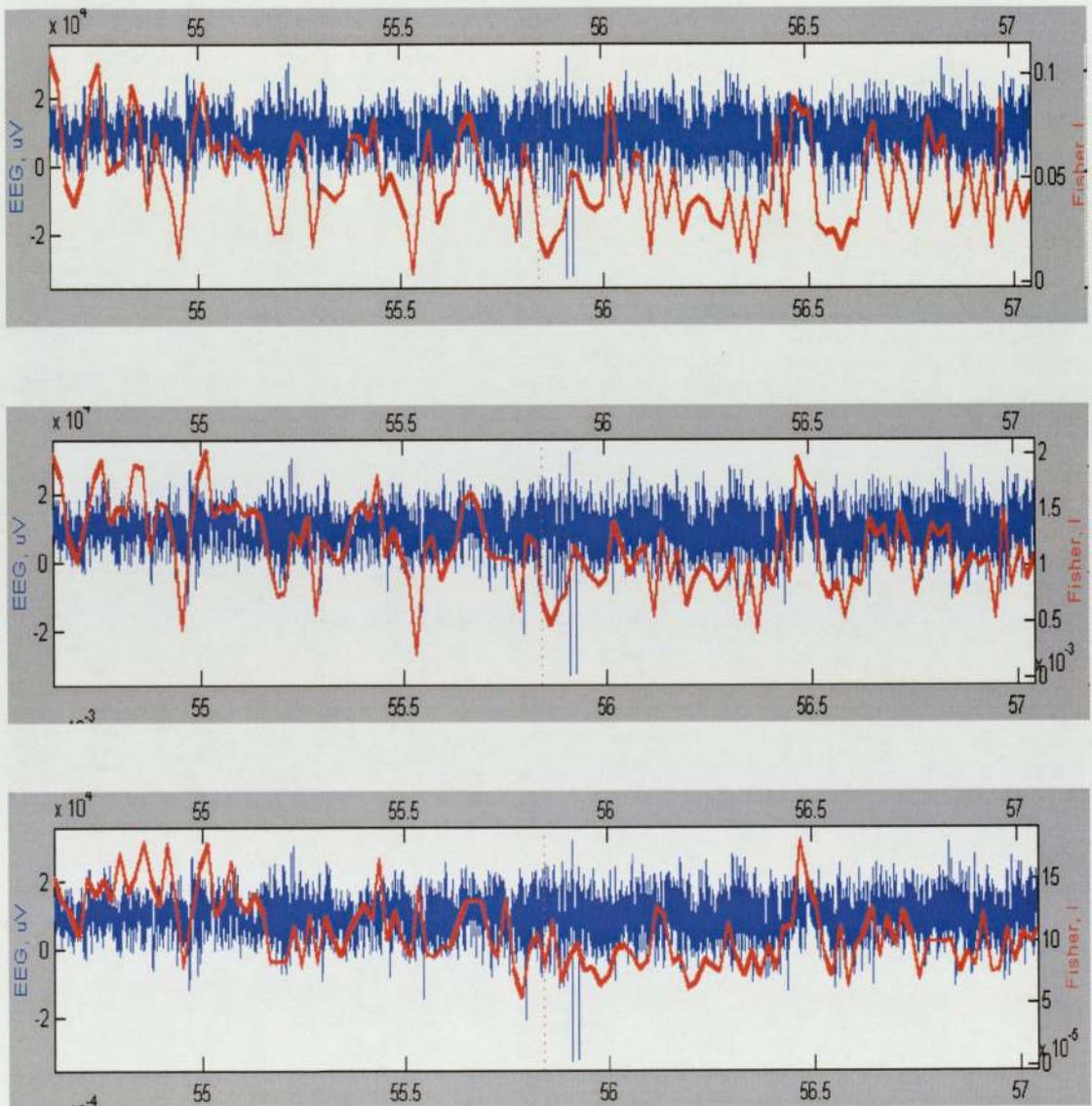


Figure 4.18: Screen shot of GUI showing all three Fisher measures applied to wake EEG. From top to bottom they are: Smoothed cubic spline, forward difference, digital derivative.

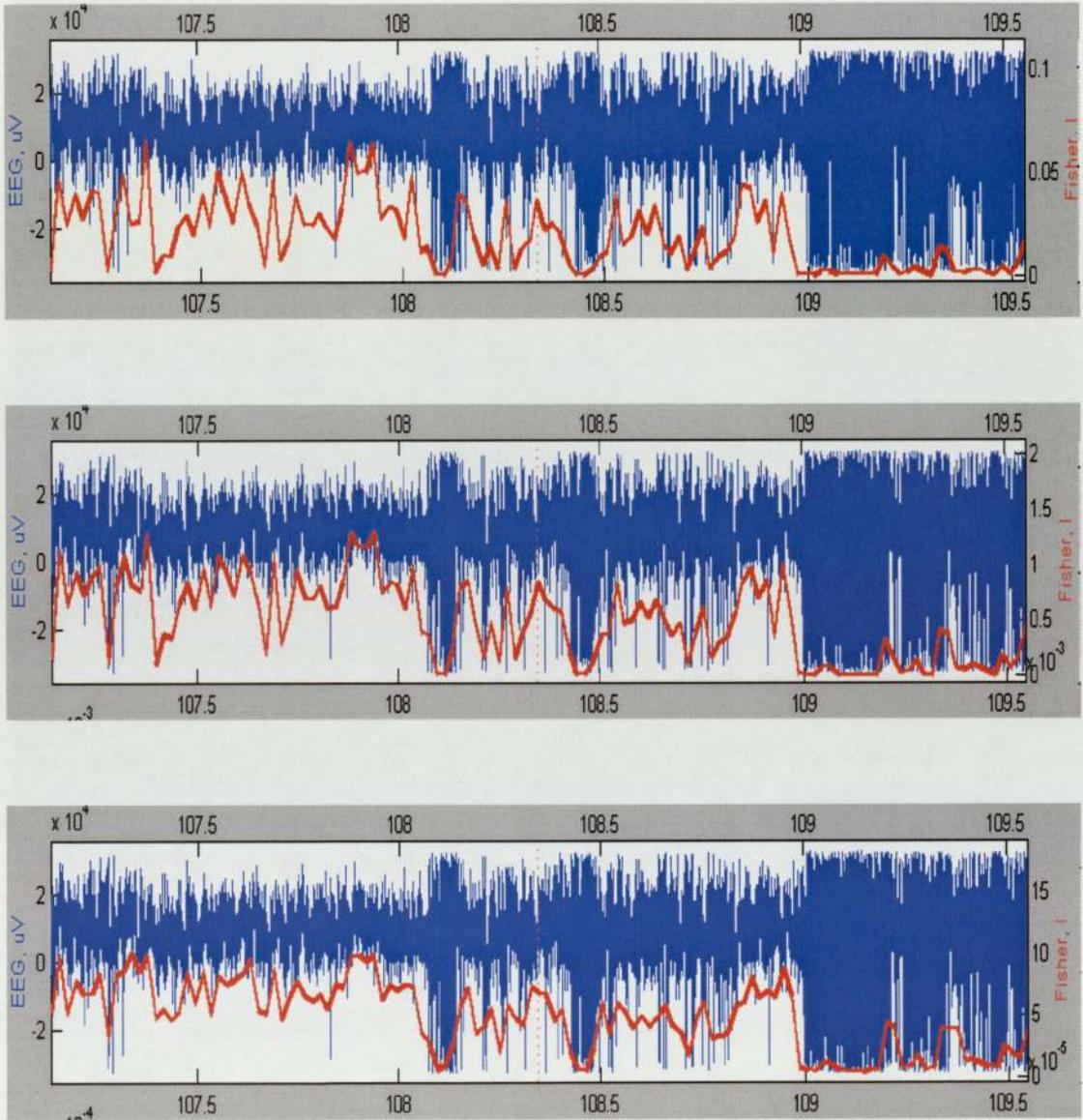


Figure 4.19: Screen shot of GUI showing all three Fisher measures applied to wake EEG corrupted with EMG artifact. From top to bottom they are: Smoothed cubic spline, forward difference, digital derivative.

4.2.6 Conclusions

The results confirm our hypothesis, that different methods used to calculate the derivative at each point in the singular spectrum will result in Fisher measures that have different noise profiles, and are selectively sensitive to changes in different sections of the singular spectrum.

Sensitivity to subtle changes in kink structure

As expected we find the forward difference Fisher measure to be extremely noisy with random spikes occurring throughout the section shown in the middle plot of figure 4.18. Surprisingly the smoothed cubic spline Fisher measure is almost identical (top plot), and no real trend can be observed in either of these measures over the clean wake EEG section. However the digital derivative Fisher measure is somewhat smoother, and clear DC trends can be observed, along with spikes that clearly correspond to changes in the EEG structure (see bottom plot on figure 4.18 at ~ 55.4 mins and ~ 56.5 mins corresponding to the digital derivative Fisher measure)

Sensitivity to EMG artifact

Again both the smoothed cubic spline and forward difference Fisher measures are noisy and the dynamic range is poor, with values corresponding to sections of wake EEG occasionally equal to sections corrupted with EMG artifact (see ~ 107.7 mins on top and middle plots of figure 4.19). We see a marked improvement in the digital derivative Fisher measure (bottom plot of figure 4.19), and can clearly distinguish between wake EEG and EMG artifact. Therefore the method used to calculate the derivative at each point of the singular spectrum, resulting in the best Fisher measure, is the digital

derivative filter for which the difference equation (4.14) is given below

$$\Delta(\sigma_n) = \frac{2\sigma_n + \sigma_{n+1} - \sigma_{n+3} - 2\sigma_{n+4}}{8} \quad (4.14)$$

4.3 Summary

Table 4.1 below gives a summary of the relative performances of each complexity measure. When comparing the best Fisher measure (digital derivative) with the best entropy measure (first value 1) we see that the entropy measure is smoother, with a superior dynamic range, and we therefore conclude that although both measures are able to monitor the changes in EEG signal complexity the first value 1 entropy measure is a slightly better measure of complexity to use in this instance.

Complexity measure	Kink sensitivity	EMG sensitivity
Entropy (Power 1)	Good	Poor
Entropy (First value 1)	Good	Good
Fisher (Smoothed cubic spline)	Poor	Poor
Fisher (Forward difference)	Poor	Poor
Fisher (Digital derivative)	Fair	Fair

Table 4.1: Table summarising the relative performances of each complexity measure

However the entropy and the Fisher measures are fundamentally different. Entropy is a *global* measure that will remain constant irrespective of the ordering of components, and Fisher is a *local* measure that is highly dependent on the local structure (*i.e.* gradient) which is determined by the ordering of components. When using singular value decomposition, the ordering of components is not explicitly determined, but is intrinsic to the method itself, whereby components are rank ordered by variance between eigenvectors. However for *topographic* systems where neighbourhood relation-

ships matter, and component ordering is not obvious or predetermined, such as for independent component analysis (ICA), the Fisher measure could be of considerable use.

Chapter 5

Conclusions

We have presented in this thesis a new application of entropy and the Fisher measure for monitoring the complexity of the EEG.

In chapter 3 a technique known as dynamical embedding was introduced. The technique allows us to reconstruct the dynamics of the underlying generator by modelling the interaction of the degrees of freedom associated with local sections of the manifold, which are non-linear fitting surfaces, using an arbitrary basis, which for this thesis was an orthogonal spanning set.

The principal challenge in using this method was choosing the correct values for the embedding parameters. We have shown that a number of criterion intrinsic to the data and the problem itself dictate the values of the embedding delay τ and the embedding dimension, M . The final selection for the number of delay vectors, n , was based on the convergence of the singular spectrum and experiments showed that parameters chosen based on these criterion were correct and captured the dynamics of the system.

The complexity measures implemented in chapter 4 showed us that although entropy is a global measure, by appropriately normalising the singular spectrum, we made the

entropy measure sensitive to changes in the local structure of the singular spectrum which resulted in an excellent complexity measure. Furthermore we showed that careful discrete calculation of the gradient at each point of the singular spectrum using a derivative filter, reduced the noise in the Fisher measure making it of significant practical use for the analysis of topographic systems where neighbourhood relationships between components matter and entropy is of little use in highlighting this.

The testing of the complexity measures was performed using only one EEG record. The next task should be to test their performance on other EEG records whose waveform morphology differs from the record used in this project. Also single channel analysis should be applied to other EEG channels over much longer time scales, in order to highlight any possible longer term variability in the complexity of the signal.

Chapter 6

Future Work

6.1 Vigilance

With the GUI in place, the first task for any future work would be to attempt to correlate changes in complexity with a scored measure of ‘vigilance’. However, the existing approach of attempting to detect subtle changes in subject ‘vigilance’, over a long section of EEG data, without knowing *a priori* the expected change in complexity is unrealistic. We should first limit ourselves to being able to discriminate between gross levels of ‘vigilance’, and for this I suggest we would require the following two types of EEG data:

- A section of EEG recorded whilst the subject is given every opportunity to perform at an optimal level of concentration. This would mean plenty of sleep and rest in the days leading up to the experiment together with a sound nutritional program and a moderate degree of physical exercise and plenty of mental relaxation, since it has been shown that the state of the mind and suppleness of the body are highly interdependent. The surroundings in which the experiment takes place are also important and should be quiet and comfortable to induce a

peaceful state of mind and minimise any distractions.

- An EEG recorded whilst the same subject has been deprived of the opportunity to cultivate a mind capable of single-pointed continuous concentration. Therefore sleep and rest should be deprived leading up to, and immediately prior to, the experiment and any form of exercise should be stopped and the diet adjusted accordingly. The surroundings can be less comfortable and distracting.

Whilst these are clearly two rather unrealistic and extreme cases, and we are changing many variables at the same time, the initial aim should be to analyse ‘vigilant’ and ‘non-vigilant’ data and investigate the structure of the complexity measures.

6.2 Complexity analysis of topographic systems

As mentioned in chapter 5, by choosing a different basis for modelling the local manifold, and corresponding technique where the ordering of components is not predetermined (such as ICA), we can investigate the effects of various ordering criterion on the Fisher complexity measure.

6.2.1 ICA experiment

We performed initial experiments using Hyvärinen and Oja’s fast-fixed point ICA algorithm [14] to calculate 150 non-orthogonal source components to correspond to the same number of components needed to model the local manifold when an orthogonal basis was used. Therefore we may write

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{6.1}$$

where \mathbf{x} is the matrix of n observed random vectors, \mathbf{A} is the $n \times M$ mixing matrix and \mathbf{s} is the source vector. We then ordered each component according to its total RMS power contribution to the original observed random vectors contained by \mathbf{x} . To do this we rank ordered the RMS values of the columns of the mixing matrix, \mathbf{A} . Figure 6.1 shows the results obtained for the same EEG record over the first minute, with the Fisher measure superimposed over the EEG trace. It can be seen that the structure of the function from which the complexity measure is derived is now very different to the singular spectrum, and the Fisher measure also exhibits some interesting structure. Any further work in this area is expected to prove worthwhile.

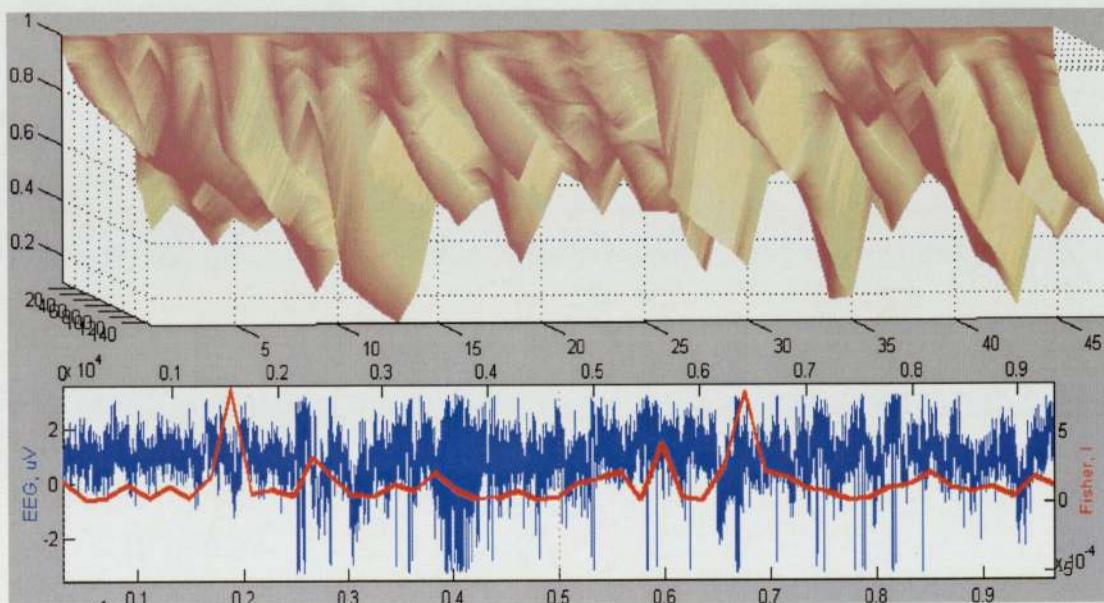


Figure 6.1: GUI snapshot for Fisher measure, using ICA with RMS ordering of components

Bibliography

- [1] Lowe D. *Extracting Structure From Wake EEG Using Neural Networks*. Aston University, 1996.
- [2] Nunez P. *Neo-cortical Dynamics and Human EEG Rhythms*. OUP. 1995.
- [3] Nuwer MR, Comi G, Emerson R, Fuglsang-Frederiksen A, Guérit JM, Hinrichs H, Ikeda A, LuccasFJC, Rappelsburger R. *IFCN Standards for Sigital Recording of Clinical EEG*. *Electroenceph. clin. Neurophysiol.*, 106:259-261. 1998.
- [4] Broomhead DS, King GP. *Extracting Qualitative Dynamics from Experimental Data*. *Physica D*, 20:217-236. 1986.
- [5] Takens F. *Detecting Strange Attractors in Turbulances*. *Lecture Notes in Mathematics*, 898:366-381. 1981.
- [6] Oppenheim AV, Willsky AS, Young IT. *Signals and Systems*. Prentice Hall. 1983.
- [7] Marzio P. *Complexity Measures of Elecroencephalographic Data*. M.Sc. Thesis, Aston University. 1999.
- [8] Bishop C. *Neural Networks and Pattern Analysis*. OUP. 1995.
- [9] Saporta G. *Probabilités Analyse des Données et Statistique*. Volume I. Editions Technip. 1990.
- [10] Hanselman D, Littlefield B. *Mastering Matlab 5*. Prentice Hall. 1998.

- [11] Pan J, Tompkins W. *A Real-Time QRS Detection Algorithm*. IEEE Transactions on Biomedical Engineering, vol. BME-32, No.3, 233p. 1985.
- [12] Kreyzig E. *Advanced Engineering Mathematics*. Wiley, New York, 7th ed. 1993.
- [13] De Boor. *A Practical Guide to Splines*. Applied Math. Sciences Vol. 27, xxiv+392p. Springer V. 1978.
- [14] Hyvärinen A, Oja A. *A Fast Fixed Point Algorithm for Independent Component Analysis*. Neural Computation, 9:1483-1492. 1997.
- [15] Roberts SJ, Penny W, Rezek I. *Temporal and Spatial Complexity Measures for EEG-Based Brain-Computer Interfacing*. Medical and Biological Engineering and computing. 3-4p. 1998.
- [16] Roberts SJ, Rezek I, Penny W, Everson RM. *The Use of Advanced Information Processing Methods in EEG Analysis*. Proceedings of IEE colloquium on Intelligent Decision Support in Clinical Practice, June 1998. Imperial College, London.
- [17] James CJ, Lowe D. *Using Independent Component Analysis and Dynamical Embedding to Isolate Seizure Activity in the EEG*. World Congress on Medical Physics and Biomedical Engineering, Chicago, 23-27 July 2000.
- [18] Noël N. *Dynamical Embedding and Feature Extraction of Electroencephalographic Data*. Aston University. M.Sc. Thesis, 1998.
- [19] Sauer T, Yorke J, Casdagli M. *Embedology*. Journal of statistical physics. Vol. 65. Nos. 3/4. 1991.
- [20] Meiss J. *The Nonlinear Science Frequently Asked Questions Document*. Colorado, sci.nonlinear newsgroup, Version 1.0.4.; 4-5. 1995.

- [21] Germuska, R. *Neural Networks for the Detection of Abnormal Beats in the ECG*. Oxford University, Honour School of Engineering and Computing Science, Technical Report, 1999.

Appendix A

Discrete Fisher derivation

Fisher's information content, $I(\theta)$, is defined by

$$I(\theta) = \int_{\theta} \left(\frac{\partial}{\partial \theta} \log(P(\theta)) \right)^2 P(\theta) d\theta \quad (\text{A.1})$$

where $P(\theta)$ is the *likelihood*. In the case of the singular spectra, the *prior* probability distribution is uniform. Therefore, since the *likelihood* is given by

$$\text{posterior} \propto \text{prior} \times \text{likelihood} \quad (\text{A.2})$$

and the *posterior* distribution is assumed to be the singular spectrum, $P(\theta)$, we may regard $P(\theta)$ as the *likelihood*. Therefore we may write

$$I(\theta) = \int_{\theta} \left(\frac{P'(\theta)}{P(\theta)} \right)^2 P(\theta) d\theta \quad (\text{A.3})$$

$$I(\theta) = \int_{\theta} \frac{P'(\theta)^2}{P(\theta)} d\theta \quad (\text{A.4})$$

So we may write the discrete version of the Fisher measure, over M samples, as

$$I_M(\theta) = \sum_{i=1}^M \frac{(P'_i(\theta))^2}{P_i(\theta)} \quad (\text{A.5})$$

where

$$P'_i(\theta) = \frac{\partial P_i(\theta)}{\partial \theta} \quad (\text{A.6})$$

and

$$P_i(\theta) = \hat{\sigma}_i \quad (\text{A.7})$$