Clustering Epileptiform Discharges in the Interictal Electroencephalogram with Topography Preserving Maps

DAGMAR SCOTT FRASER

MSc by Research in Pattern Analysis and Neural Networks

ASTON UNIVERSITY

September 1999

Acknowledgements

With thanks to Dr Christopher J. James who was untiring in his support.

To all at NCRG; staff and student.

ASTON UNIVERSITY

Clustering Epileptiform Discharges in the Interictal Electroencephalogram with Topography Preserving Maps

DAGMAR SCOTT FRASER

MSc by Research in Pattern Analysis and Neural Networks, 1999

Thesis Summary

Topography preserving maps have proved useful in the clustering of paroxysmal events in the Electroencephalogram (EEG) - in particular epileptiform events (EEvs) and artefacts. With the aim of enhancing performance of pre-existing systems, a novel variant of Kohonen's Self Organising Feature Map (SOFM) is considered. Realistic, synthetic EEvs have been generated using a 3-sphere head model, superimposed on true EEG. Pre-processing by means of Principal Component Analysis has allowed dimensionality reduction of the synthetic, interictal 25 channel EEG. This was clustered employing an Adaptive Subspace variant of the SOFM. The resulting clusters were interpreted to allow classification. This has permitted the development of a scheme to automatically detect and extract features from EEG traces, which offer results comparable with those in the literature over the synthetic data.

Keywords: Electroencephalography, Feature Extraction, Topography Preserving Maps, Clustering, Epileptiform Events

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Electroencephalogram

May I succeed in achieving my plan ... and create a kind of brain mirror, the electroencephalogram!

Hans Berger, from his diary of 16th November 1924.

Publishing in 1929, Hans Berger, was the first to measure scalp electrical potentials caused by underlying brain activity in humans. He introduced one of the most potent tools in the diagnosis of neurophysiological disorders; the electroencephalogram (EEG).

Berger's crude system has been succeeded by an evolving range of idiosyncratic electrode placements culminating in an international standard, the 10-20 system (see Section 2.2), which emerged in 1958 in the work of H. Jasper. Originally, these EEG traces were stored on paper, which was then laboriously interpreted by a clinical electroencephalographer (EEGer). With the advent of computer based recording, vast quantities of EEG data are now stored. However, Petsche notes, that in EEG analysis such increases in raw data can '*frequently decrease the efficiency of the work, for nothing is more difficult than to sift the chaff from the wheat*' [3]. With ever increasing amounts of data, now including video, there is a desire for automation of the processing.

Electroencephalography is a widespread diagnostic procedure, with over half of EEG referrals specifically involving the diagnosis of epilepsy. EEG is a vital tool in the diagnosis of epilepsy [4], given that the patient need not experience an obvious paroxysmal event. Diagnosis is made possible by the appearance of epileptiform patterns defined by Chatrain [5] as;

> *Epileptiform Pattern.* Interpretative term. Applies to distinctive waves or complexes, distinguished from background activity, and resembling those recorded in a proportion of human subjects suffering from epileptic disorders and in animals rendered epileptic experimentally. Epileptiform patterns include spikes and sharp waves, alone or accompanied by slow waves, occurring single or in bursts lasting at most a few seconds. Comment: (1) Term refers to interictal paroxysmal activity and not to seizure patterns. *Cf.* seizure pattern. (2) Probability of association with clinical epileptic disorders is variable.

Despite these Epileptiform Events (EEvs) evidencing stereotyped features [6] and EEG being such a widespread, largely standardised, clinical procedure difficulties remain in automating the procedure of feature extraction and classification.

Whilst the EEG trace can be construed as a background process with 'superimposed transient nonstationarities' (TNS) [7], the epileptiform transients of interest are overawed by a battery of artefacts (i.e. extracerebral sources such as eye blinks, muscle activity or electrodes), which are present in all EEG. The problem is aggravated by their similarity to the true EEv and the ultimately *subjective* definition of what constitutes an EEv.

1.2 Automated EEG Analysis

Automated EEG analysis has a substantial history in its attempt to relieve the burden on the trained electroencephalogropher. These fall into two groups; the first and simplest aims at data reduction, an event detection stage triggers a recording device thus substantially reducing the amount of data for later evaluation. The second, which is the ultimate aim of this study is, automated classification. Several techniques have been applied, including;

- Mimetic Methods; whereby parameters of the data are extracted and compared to thresholds; exceeding combinations of these thresholds signals an *event*.
- Template Matching; in which positive correlation with a template is used in classification.

- Parametric Methods; given the assumption that the background is stationary, deviation from this predicted stationarity is considered an *event*.

- Syntactic Methods, in which positives are based on structured complexes of features.

- Neural Networks; which are trained to recognise EEvs.

- Expert Systems; which attempt to use knowledge based reasoning in a fashion akin to a trained EEGer.

However, due to unacceptable rate of false positives, which result in misdiagnosis, these systems cannot be relied upon in the routine EEG setting [4].

Glover, Raghavan, Ktonas and Frost [8] note that '*Invariably, when an automated detection system's false positives are shown to an EEGer, he points to temporal and spatial (multiple channel) contextual clues to explain why the waveforms were not included in the visual scoring*'. Glover *et alia* emphasize contextual information in their knowledge based system. This reflects a growing awareness in the field of the importance of the EEv's contextual information for correct classification. Such systems often remain useful, achieving the first aim, that of data reduction e.g. Gabor and Seyal [6] considered the effectiveness of an Artificial Neural Network system as an epileptiform event detector. They concluded that the system, due to a significant false positive count, is only robust enough to be useful as a strategy for data reduction of long term EEG recordings.

Recent implementation of Topography Preserving Maps (TPM) [9], coupled with a mimetic stage and a fuzzy logic decision stage have proven favourable in comparison with results of previous automations. Consider Table 1.1.

Low values for False detections per hour, and high values for Selectivity (a measure of the system's discrimination between EEvs and artefacts) and Sensitivity (which reflects the proportion true EEvs detected) are indicators of useful performance.

The TPM utilised in System 9; Kohonen's Self Organising Feature Map (SOFM), is not considered ideal, however, for the clustering of features that have undergone

Sys	Method	Hours	% EEv	Train /Test	Sensitivity	Selectivity	Error/ hour
1	ANN + ANN	0.043	?	6/4	0.9000	0.6900	~1023
2	Mimetic	2.0	100	Blind	0.5900	0.8900	37
3	Mimetic	33.0	100	Blind	-	0.4100	117
4	Mimetic + state	33.0	100	Blind	-	0.6700	47
5	Mimetic + ANN	0.3	100	Same	0.7400	0.7400	~ 804
6	Mimetic + ANN	0.3	100	Same	0.4600	0.4600	~ 5598
7	Mimetic + Expert System	3.0	73	Same	0.5300	1.000	0
8	Mimetic + Expert System	3.2	88	Blind	0.1400	0.8100	2
9	Mimetic + SOFM + Fuzzy Logic	3.2	88	Blind	0.5500	0.8200	7

Table 1.1: A comparison of the sensitivities, selectivities and false detection rates between spike detection systems. After James *et al* [2]. For definitions of Selectivity, Sensitivity and Error Rate see Section 4.3.1 and Section 4.3.2. The Percentage EEv defines the percentage of Epiletiform Events in the Training Set.

transformation. A recent paper on the generation of synthetic EEvs [1], suggests that a common dipole morphology, which undergoes elementary transformations, can replicate the statistics of the EEG trace. As such an *invariant feature detector* should offer improved performance in the place of the SOFM.

A variant of the SOFM, inspired by the Learning Subspace Method [10], employs the property of subspaces to accurately model classes invariant of transformation of the data.

The approach considered in this Thesis involves acknowledging the data reduction methods previously researched as a first stage. The resulting reduced data will be considered as *candidate epileptiform discharges* (CEDs). These candidates (both true EEv and artefact) will remain in an appropriate window of context, spatial and temporal. These CEDs will then be clustered with an unsupervised clustering algorithm. The resultant map will then, in a principled fashion, be used to extract class discriminant information.

1.3 The Thesis in Brief

Outline of this and succeeding chapters.

Chapter 1 Has seen the introduction of the EEG, and outlined the desirability and methods which have been applied previously to the automation of EEG analysis.

Chapter 2 Here the concepts which underlie the generation of 10-20 System EEG traces will be briefly considered. These will be extended towards an explanation of the desirability and construction of the synthetic data set. Data reduction, or event detection is also addressed.

Chapter 4 This section offers brief discussion of the merits of alternative clustering methods. The novel nature of the Adaptive Subspace SOFM is considered, with thought given to the extraction of measures of performance from trained ASSOFMs. The practical steps involved in training the ASSOFM is outlined.

Chapter 5 Collated results from implementation of the ASSOFM.

Chapter 6 A discussion of the results, relating these critically to the choices made in previous sections, and offering comparisons with previous studies.

Chapter 7 Conclusions and reflections on the results discussed in Chapter 6. Avenues for related investigation. Closing comments.

Chapter 2

The EEG

2.1 Introduction

This chapter briefly considers the origins of EEG data, detailed examination is beyond the scope of this report and the reader is directed to other texts [11]. Focus is then directed to the EEvs, their morphology and the desirability of synthetic EEv. The generation of synthetic data is addressed before EEG data reduction strategies are considered.

2.2 The Neocortex

In mammals, the largest section of the brain is called the *Neocortex.* Lying immediately within the skull, it is elaborately folded to allow a large surface area to be compressed into a small volume. This small volume is considered vital to conscious process, and the electrical field it generates is reflected in the EEG.

The traces considered in this study are from electrodes placed on the scalp according to, a widespread variant of, the 10-20 system of Jasper [12]. Consider Figure 2.1. The considered implementation of the 10-20 system presented by Nuwer *et alia* [13], consists of 25 electrodes: Fp1, F3, C3, P3, O1, F7, T3, T5, Fp2, F4, C4, P4, O2, F8, T4, T6, Fz, Cz, Pz, F9, T9, P9, F10, T10 and P10. Key to regions of the *Neocortex*; F - Frontal, P - Parietal, C - Central, O - Occipital. Note that the 'central' appellation

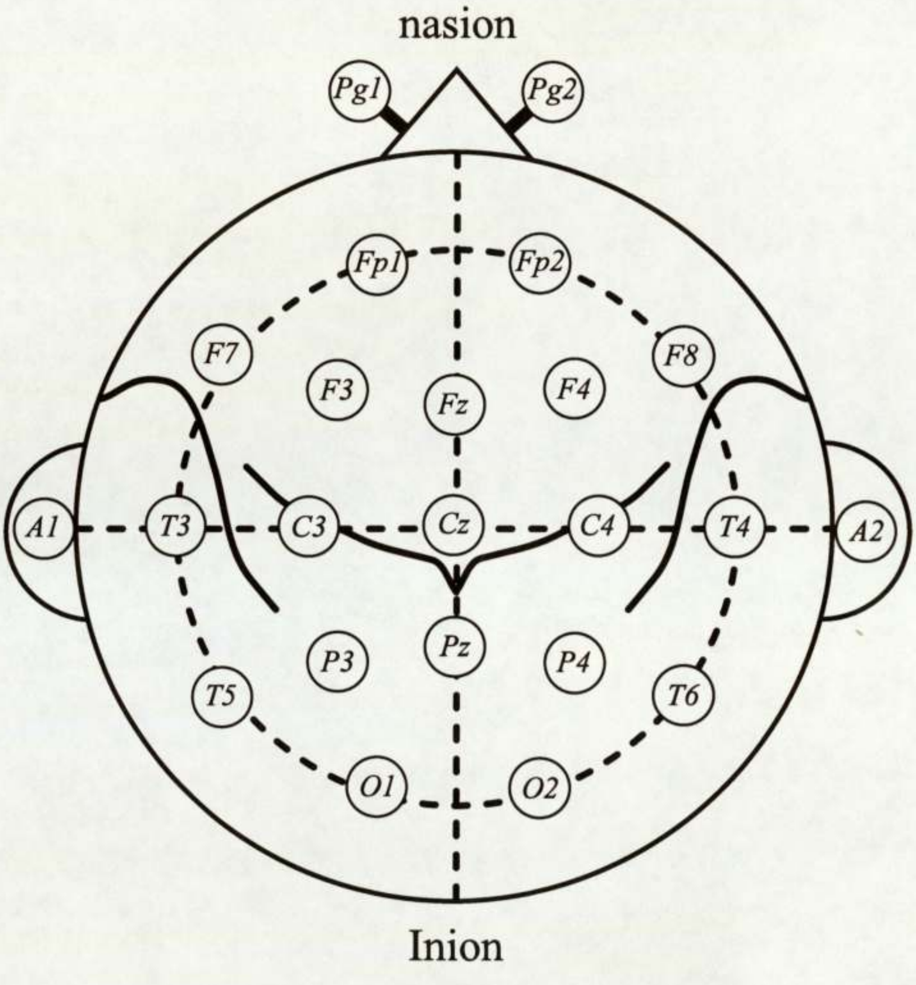

Figure 2.1: The 10-20 System for Electrode Placement, after Jasper (1958). Fpl, F3, C3, P3, O1, F7, T3, T5, Fp2, F4, C4, P4, O2, F8, T4, T6, Fz, Cz, Pz, F9, T9, P9, F10, T10 and P10. Key to regions of the *Neocortex*; F - Frontal, P - Parietal, C - Central, O - Occipital

is for convenience only, there is no such named region of the *Neocortex*. The nasion is the bridge of the nose, and the inion a raised segment of bone at the base of the skull.

2.3 Analysing the EEG

Classification of EEG is done grossly by frequency e.g. the Alpha rhythm, of 8-13Hz, which occurs in the posterior of the head, particularly in the occipital region, or in the case of localised transients such as those considered within this report, by morphology e.g. the epileptiform pattern defined by Chatrain in the Introduction (see Figure 2.2 and Figure 2.3).

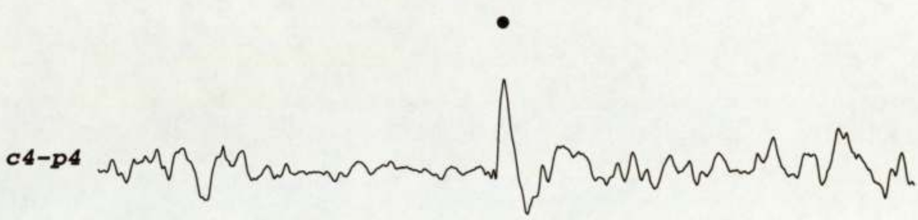

Figure 2.2: A Sharp-Wave, from a single EEG channel, clearly distinct from background.

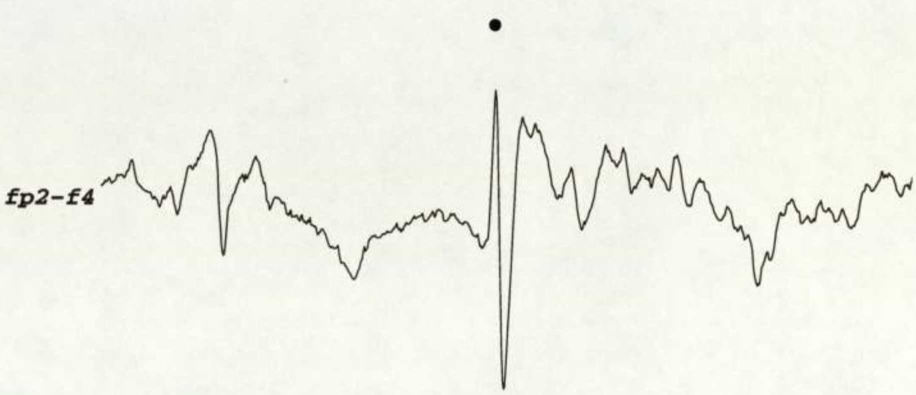

Figure 2.3: A Spike, from a single EEG channel, clearly distinct from background.

Localised transients are often extra-cranial sources, know as artefacts and many evidence a highly similar morphology to spikes and sharp-waves. Examples of these include; eye-blinks which result in frontally localised spikes, muscle contractions which generate gross electrical fields to which the EEG is sensitive, and artefacts of the recording apparatus where by movement of electrodes can cause what is termed 'electrode pop' which is evidenced by a localised spike on a single channel.

The consistent labelling of sizeable data sets takes considerable human effort and expertise, consequently they are often proprietary in nature, being commissioned and employed in the calibration of commercially available systems. Even with available data, the *subjective* nature of an EEv's definition renders these data sets inconsistent between EEGers, who given human fallibility, or even apriori knowledge of the patient's state will make differing decisions. This also results in a fundamental difficulty in the evaluation of any automated EEG analysis in comparison with previous systems employing differing test sets. This is further aggravated by the idiosyncratic levels of certainty ascribed to potential events that suffer no universal convention. For reliable, replicable results there is a requirement of objectively defined data sets, where an EEv may be labelled consistently. This is impossible with labelling by expert, or committees of experts, as there is no objective criterion for a true EEv. In an attempt to alleviate this limitation, this Thesis considers construction of a synthetic data set.

The EEG is a linear summation of electrical activity within the brain [14]. An EEv can be modelled, as superimposed TNS on an additive background noise process, if we assume that the TNS is independent from the background activity. Given a suitable model of the physiological basis of these TNS, adding it to the EEG of a normal individual would give rise to a biologically plausible, synthetic EEv. Further, normal EEG traces could be used as a source of artefact TNSs, allowing the data set to contain true artefacts, as well as the synthetic EEv superimposed on true background. There are many examples of normal EEG available. Given the previous scheme, an arbitarily sized, objectively labelled data set could be generated. As this is an exploratory report, and not an attempt to produce a diagnostic utility, it is considered that such a data set will be useful for proof of principle. It eliminates the problems associated with subjectively labelled data whilst retaining biological plausibility with the caveat that the resulting system will be proved only over the given model of EEvs.

2.4 Synthetic Data

A plausible modelling of the abnormal neuronal activity, which underlies an EEv was recently considered by Kobayashi *et alia* [1]. The field generated by the action of neurones can be modelled as a current dipole. EEvs can be synthesised using this current dipole with fixed, randomly determined, location and orientation within the brain, and a moment having a spike-like waveform. The EEG trace can then be determined at each electrode in the 10-20 system using the potential field obtained by solving the forward problem[1] for the current dipole model with a 3 sphere head abstraction. The 3 spheres referring to the brain, skull and scalp. This noiseless, synthetic EEv can then be superimposed on a background segment from a normal EEG trace giving rise to an *event* that satisfies Chatrain's subjective definition.

The synthetic EEvs were the combination of two transients, constructed in turn from the dipoles in the following Equation, split into spike transient(2.1) and slow wave transient components (2.2).

$$s_1(t) = 3000 \cdot a_{sp} \exp\{-(200 + 50 \cdot b_{sp}) \cdot \parallel t - 0.01 \cdot c_{sp} - 0.096 \parallel\} \qquad (2.1)$$

$$+1500 \cdot a_{wv} \exp\{-(40 + 15 \cdot b_{wv}) \cdot \parallel t - 0.01 \cdot c_{wv} - 0.192 \parallel\} \qquad (2.2)$$

Variables $a_{sp}, b_{sp}, c_{sp}, a_{wv}, b_{wv}$ and c_{wv} were random on the interval [0,1] to introduce jitter in time and amplitude, with the subscripts referring to the spike (Equation component 2.1) and slow wave compnents (Equation component 2.2) in turn.

A further set of variables were employed to locate and orient the current dipole within the upright skull, with the centre considered as the origin.

- ϕ, θ and f determine the orientation and distance along that orientation of the dipole. i.e. The eccentricity f can be considered the distance along the radius defined under spherical co-ordinates by ϕ and θ.

- Dx, Dy and Dz determine the current dipoles moment.

[1] The forward problem refers to classic electrostatic theory, for a useful reference see Nunez and Katznelzon, 1981 [14]. This study employs the algortithm implemented by Kobayashi *et al.* [1]

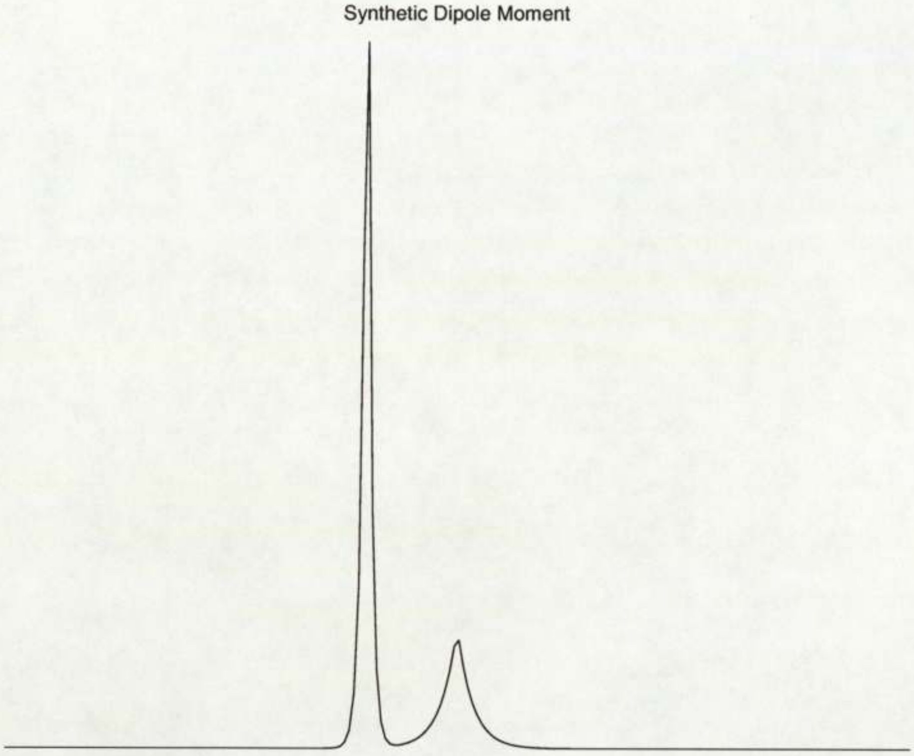

Figure 2.4: Single Spike and Slow Wave dipole as defined in Equation components 2.1 and 2.2.

As an example of the above transients (2.1 and 2.2), Figure 2.4 displays a sample spike and slow wave dipole. Figure 2.5 display the results of solving the forward model of the 10-20 system's 25 channels. Figure 2.6 show the resultant superimposition of background EEG.

The EEG, both for background and artefacts, was extracted from 4 patients. These each consisted of 2 minute readings on the hour for 48 hours to give a broad sampling of EEG activity over a resulting 6 hours and 24 minutes. The EEG data was recorded form patients in the Montreal Neurological Institute and Hospital in the long term monitoring unit using the standard 10-20 system outlined above, sampling at 200Hz.

2.5 Data Reduction

With the availability of normal EEG, attention was directed at extracting both artefact TNS, and suitable background.

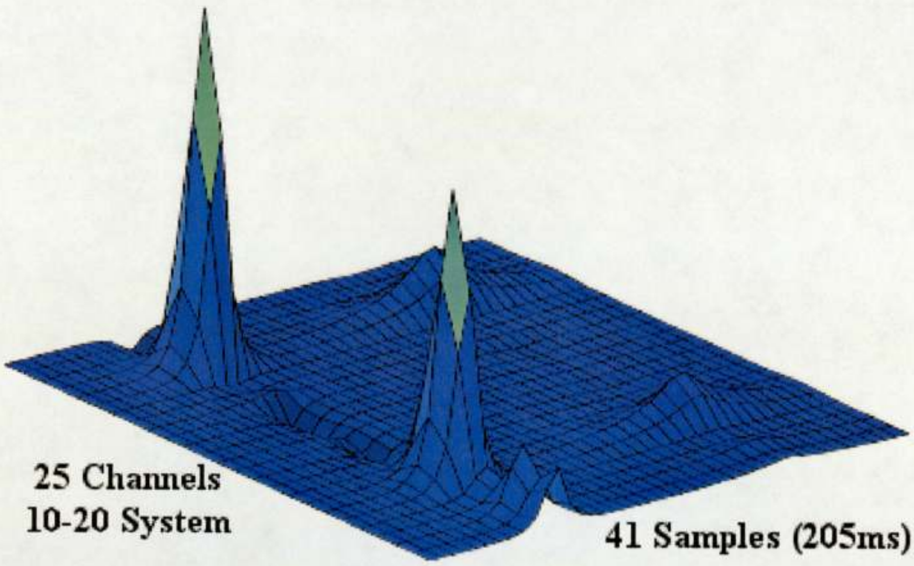

Figure 2.5: Single Spike and Slow Wave dipole propagated through 3 Sphere Model to the 25 Channels of the 10-20 System using algorithms impemented by [1].

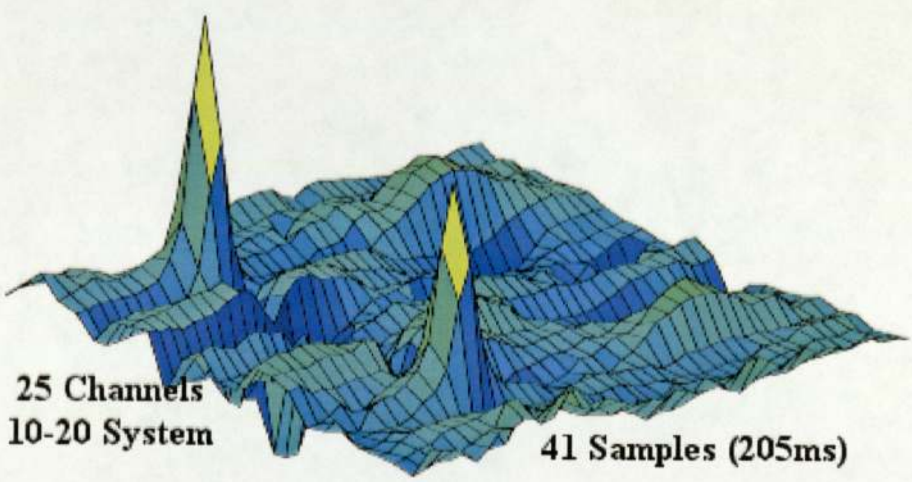

Figure 2.6: Single Spike and Slow Wave dipole propagated through 3 Sphere Model to the 25 Channels of the 10-20 System with additive noise which assumes independence of epileptogenic event from background.

The aim of this Thesis lies, not in enhancing previous methodologies for event detection, but improving the classification of EEvs. Hence, a simple implementation of the mimetic method was used in the extraction of artefact TNSs and superimposed synthetic EEvs. The mimetic approach employs a series of thresholds, detailed below, which characterise the data based on the parameters of the data. The aim of this mimetic stage is data reduction; to admit at least 100% of the synthetic EEvs and as few artefacts as possible. This study's implementation differs from previous work in the tuning of the accuracy of this mimetic stage, i.e. the data over which the parameters are chosen are definite epileptiform, without the spurious outliers of subjectively labelled data. This results in a first stage, over the chosen parameters, that can be considered optimal i.e. all synthetic EEvs are detected. However, the parameters chosen may not be the most efficient if consideration is given to the number of artefacts they also pass onto the classification stage.

The detection of events allows the extraction from the data set of windowed candidate *events*. It also ensures there is no bias in the data set in that the artefacts and synthetic EEvs both pass through the same thresholds.

The heuristic method used to extract artefact TNSs (and ultimately the superimposed synthetic EEvs to ensure parity) was as follows;

- The first order differential of each of the 25 channels of normal EEG was taken over a section of 1200 samples, or 6 seconds sampled at 200Hz.
- An increase in rate of change be followed by a decrease in rate of change or vice a versa, and the difference between indicate a certain sharpness, then a flag indicating a vertex was set.
- The true value at this flagged point was then compared to a floating average of recent values (within 50ms), which set a distinction indicator.
- The vertex and distinction indicators were tallied across channels, i.e. at a given instant in time; if there were sufficient flagged channels then it was considered a *Candidate Epileptiform Discharge* (CED) at this point.

Detecting artefacts introduced a dilemma; to extract artefacts, and suitable artefact free background EEG, required the setting of thresholds which would admit 100% of the synthetic EEvs. Without having a selection of EEvs available, it was impossible to determine these parameters without visually scanning the data for periods of suitable background onto which the synthetic EEvs could be projected. In turn, the parameters that admitted these spikes could then be used to extract artefact TNS.

These parameters were developed over the initial segments of patient A, and then tested over segments of patient B and patient C.

Empirically it was noted that the difference between initial and following slope that indicated a vertex should be 0.003, on data in the μVolts range. The prominence above the floating average was to be 0, and the tally for these flags was determined to be at least 4 for a suitable CED, i.e. to avoid isolated spikes dominating the artefact data set. EEG segments which had no CED, defined by these thresholds, were then employed in the construction of synthetic data, which were in turn tested to ensure the synthetic EEvs were detected as CEDs.

Empirical results also suggest that the synthetic EEG would rarely tally above 20 and that this could be used as a discriminant against gross artefacts such as muscle artefact. Whilst the number of artefacts admitted to any classification stage will higher it entails the misclassification of some synthetic EEvs *before* that classification stage. Introducing this, as an arbiter, is beyond the scope of the mimetic stage that is intended purely as a data reduction strategy.

The resulting CEDs were extracted embedded in suitable context. That is; 41 samples x 25 channels, in which the TNS vertex was situated at the 15th sample. This was to maintain parity with a previous study by James [9], which considered this dimension of context to retain maximal information.

Chapter 3

Clustering with Topography Preserving Maps

3.1 Introduction

Clustering is often used to aid exploratory data visualisation, or for the optimal placement of basis functions in Radial Basis function networks. Clustering entails the projection of points in a given data space $\mathcal{R}^x$ into another $\mathcal{R}^y$ (often called *feature*, or *projection* space), in which those with similar properties are grouped. Topography Preserving Maps (TPMs) attempt to ensure that data which are *close* under some *metric* (e.g. Euclidean distance) remain so after mapping, often allowing the dimension y to be less than x to aid visualisation.

This behaviour can be usefully exploited in the construction of a classifier. Clustering, in and of itself, does not allow the classification of events. To extract a useful classifier for novel data, a further stage of calibration is required which utilises the clusters as indicators of class membership.

3.2 Selection of Topographic Mapping Algorithm

In the following passages, a brief description is given of selected TPM algorithms, before concluding with those algorithms adopted for further work. These are then given a more rigorous exposition, especially in the case of novel, or unusual variants.

Examples of TPMs include;

- Kohonen's Self-Organising Feature Map (SOFM) [10]; Between 1981 and 1982, Kohonen formulated a robust, unsupervised, globally ordering map algorithm; the SOFM. The SOFM produces an *elastic net* of vectors (nodes) that form an approximate model of the N dimensional input space's density function. This *elastic net* (commonly two dimensional), on convergence forms a grid of vectors that retains information about the *structure* of the data. When this shows smooth transition from node to node, as in the SOFM, this is termed an 'ordered' fashion. The main uses for this self-organising process lie in data visualisation and the *creation of abstractions* where trivial details are excised from the representation. The SOFM is not a classification algorithm. However, classification can be achieved by examination of the known training set and the clusters which the respectively classed data points favour. Further the SOFM generalises, i.e. it allows the clustering of novel data after training is complete.

- Bishop, Svensen and William's Generative Topographic Map (GTM); Presented in 1997, the GTM was introduced as an alternative to the SOFM [15]. Alternate in the sense that it purports to replicate the acknowledged usefulness of the SOFM without recourse to biologically inspired heuristics. The GTM employs L latent variables (again often two) to explicitly model the density function of the D dimensional input space. The algorithm uses the Expectation-Maximisation (EM) algorithm, which confers guaranteed convergence, to determine the mapping that generates the hidden variables.

- The Sammon Mapping; Published in 1969, the Sammon Mapping focuses on the structure of the data [16]. Inter-point distances are calculated, the algorithm attempts to retain these, within a margin of error, in the projection space. This is performed by the minimisation of this error function, termed *stress*. There are several implementations of this error function, perhaps the most obvious of which is gradient descent. Ultimately, this algorithm does not define a generative function that would allow novel data to be projected into the feature space, without re-calculating the inter-point distances and in turn the *stress* minimisation.

The SOFM, whilst biological and heuristic in inspiration has been employed successfully in a variety of applications, including the one at hand [9]. There is also a large body of supporting literature, which offer detailed suggestions on how to implement the SOFM. Further, it is readily amenable to modification, the neural network community suggests that the variants are too widespread to catalogue.

The GTM offers several advantages, primarily its rigorous mathematical basis. Indeed it was inspired to be a rigourous alternative to the SOFM. However it relies, ultimately, on the correct definition of several *a priori* factors; e.g. the number and placement of the basis function of the generalised linear network which underlie the mapping function are the most prominent. Further, the GTM is new, and as such literature on the calibration of these variables is minimal.

For the Sammon Mapping and related methods (e.g. multi-dimensional scaling), the lack of a functional mapping prohibits their use where we require the classification of later, novel data. A recent implementation entitled Neuroscale, which employs a feed forward network to model the Sammon Mapping, removes this restriction. Its primary advantage over the raw Sammon Mapping lies in that it allows the *stress* terms to be modified by *a priori* knowledge, ameliorating the clustering of similarly classed data points. A brief treatment of Neuroscale with regard to EEG is presented by Noel [17].

This Thesis is aimed at overcoming the limitations of previous implementations of an automated system for EEv classification [9] which employed the SOFM. The SOFM based classifier is considered sub-optimal for data that has undergone elementary transformations, i.e. Kohonen notes that the SOFM lacks the ability to detect invariant features [10]. Hence the GTM, which by inference would also demonstrate this limitation, would only be a useful candidate for a comparative study. However, as noted, the SOFM is amenable to modification. Hence this Thesis is directed towards the exploration of the viability of a novel implementation of a variant SOFM which overcomes this restriction. In this thesis, the pursuit of a novel implementation of the SOFM will be considered.

3.3 Kohonen's Self Organising Feature Map

Before considering extending the SOFM, the basic form is outlined below. The SOFM is more often than not employed in the form of a two-dimensional lattice for that is the easiest to visualise.

A simple SOFM, which defines the mapping using the Euclidean metric will be used to clarify the method by which the globally ordered mapping emerges.

The following offers the basic method of how a SOFM clusters data.

Consider now the mapping of a set of vectors; $x(t) \in \mathcal{R}^n$, to a set of i parametric reference vectors, or nodes. Define the nodes of the SOFM as; $m_i \in \mathcal{R}^n$, i.e. they have the same number of elements as the input vector $x(t)$.

These model vectors m_i will come to represent the distribution through the following algorithm, and in light of the eventual convergence properties presented below (Equation 3.2) these may be selected initially at random.

The nodes are connected in parallel to the input $x(t)$. This connection entails the comparison of the vectors in $x(t)$ and m_i, from which it can be determined under some *metric* (in this case Euclidean) which of these nodes m_i is the closest. This node is in turn designated m_c, the *winning* node;

$$c = argmin_i\{\| x - m_i \|\} \tag{3.1}$$

This *winning* node m_c, is then altered to reflect the input $x(t)$, the proportion by which this update occurs is termed the learning parameter α, where 1 would see the node replicate the data, and 0 would see no change. During this learning stage, those nodes which lie topographically near each other, i.e. nodes within a defined neighbourhood will be similarly influenced by x, or by some function of x related to distance from the *winning* node. This function is denoted $h_{ci}(t)$. This ensures a smoothing of the neighbouring nodes, from which the global ordering ensues as convergence is approached.

It is apparent from the following equation that any arbitrary value may be applied to $m_i(0)$ (as noted above), given $h_{ci}(t) \rightarrow 0$ as $t \rightarrow \infty$ i.e. the conditions for convergence;

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \tag{3.2}$$

The simplest of choices for h_{ci} is that of a so-called 'top hat' function. This sees the neighbourhood of updated nodes include those abutting m_c and in turn those abutting these. This employs a universal value of the learning rate parameter α over the neighbourhood. Both the size N of this neighbourhood (a square in two dimensions), and the value α decrease monotonically throughout training, with α decaying from 1 to 0, and N from some initial value to 1. Kohonen notes from empirical studies that if the network is not very large i.e. less than a few hundred nodes then the selection of these initial parameters is not crucial to convergence.

On the presentation of novel data, the closest node under the metric would be considered the winner. If it is assumed that some form of calibration has taken place, this node will have been assigned a class and hence classifiy the novel datum. A method of calibration is considered in Chapter 4.

3.4 The Adaptive Subspace Self Organising Feature Map (ASSOFM)

Having considered the basic SOFM, attention is now focused of the possibility of addressing the underlying cause of the SOFM inability to model data invariantly under transformation.

In certain applications, classes can be ably represented by a set of basis vectors that span a linear manifold, or subspace. This is demonstrated by the Learning Subspace Method (LSM) of Kohonen. It is possible for these basis vectors, appropriately chosen, to be able to model data which has undergone elementary transformations invariantly, i.e. linear combinations of the basis vectors allow the retention of invariant classification were mundane classifiers employing template matching, or the Euclidean metric for example might fail.

As noted SOFM itself is not an optimal solution for the classification of data that have undergone elementary transformations.

If the concept of subspace classifier and adaptive node are coupled then there is a possibility that a useful self organising visualisation tool and classifier could be constructed, which would be blind to elementary transformation between similarly labelled data. Kohonen has already considered an implementaion which combines the SOFM and adaptive basis vectors which he termed the ASSOFM. However, Kohnonen's suggested variant pursues temporal invariance, whereas in this Thesis the temporal isolation of the feature is already considered in the mimetic stage of pre-processing considered in Chapter 2.

It can be seen from the construction of the synthetic data set in Chapter 2, that the underlying epileptiform discharge has a stereotyped morphology that undergoes several transformations. This suggests that the considered EEG data classification may be susceptible to modelling with adaptive subspaces, where template matching *et alia* have given sub-optimal results.

To clarify; the nodes of the standard SOFM are to be replaced with sets of adaptable basis vectors, which in turn span an adaptive subspace. Such an implementation demands the creation of a subspace metric to allow the definition of the winning node m_c, and a means of adapting the spanning basis vectors in relation to the learning parameter α and the data, such that the distance under the new metric is reduced.

3.4.1 A New Metric and the Rotation Operator

(i) Distance Between Subspaces

After Oja [18]. To give a *distance* between two subspaces L^1 and L^2 a *metric* must be defined. Such a metric will be defined by the maximum distance from the unit sphere in one of the subspaces, to the other. Let S^1 denote the set;

$$S^1 = \{x \mid \in L^1, \| x \|= 1\} \tag{3.3}$$

i.e. the unit sphere in L^1, Let P^2 denote the projection matrix on L^2

$$P^2 = \sum_{i=1}^{p} u_i u^T, \tag{3.4}$$

where L^2 is given in terms of its orthonormal basis $\{u_i ... u_p\}$, where p is the number of linearly independent vectors spanning the subspace L^2. This gives rise to the metric;

$$\delta(L^1, L^2) = \max\{(x^T(I - P^2)x)^{\frac{1}{2}} \mid x \in S^1\} \tag{3.5}$$

$$\delta(L^1, L^2) = \max\min \| x - y \|, x \in S^1, y \in L^2. \tag{3.6}$$

Perhaps a more intuitive notion to that of *distance* would be that of angles betwixt subspaces. Watkins [19] notes that the above metric may be seen in terms of the relative orientation of two k-dimensional subspaces, which is in turn described by k *canonical angles.* The metric defined above can be seen as the sine of the largest *canonical angle* between L^1 and L^2. A MATLAB implementation is noted in the appendix.

(ii)Adaptive Basis Vectors

Having now defined a metric, some means of updating the *winning* node, and neighbourhood is required, in such a fashion that the *distance* between node and data subspace is reduced. An appeal can be made to the second interpretation of the metric, and the idea of rotation i.e. to allow reducing this angle, and hence the *distance.*

Consider the projection matrix

$$A = I - \frac{xx^T}{x^T x}, \tag{3.7}$$

with $x \in \mathcal{R}^n$ a vector. This matrix is symmetric and idempotent. A is the projection matrix onto $L^{\perp}(x)$ orthogonal to x: for any $v \in L^{\perp}(x), Av = v$, while $Ax = 0$ [for proof see Oja].

If we act on L with A, i.e. $L' = AL$, then all the basis vectors will be projected into $L^{\perp}$. All their linear combinations are therefore also in $L^{\perp}$. The transformation

$$L' = (I - \frac{xx^T}{x^T x})L \tag{3.8}$$

can thus be viewed as a *rotation* of subspace L to a direction orthogonal to vector x. If we now consider L' as one of the classification subspaces, and the vector x as our input, we see that the projection of x onto this subspace has the value zero, which entails that if we used our metric above, then perfect classification i.e. zero distance has been achieved. If we replace the projection matrix by an *elementary matrix* of the form $A = I + \alpha x x^T$ with x defined as before, and a scalar parameter α, a learning paramater lying on the interval $[0, 1]$, we have feasible way of rotating a subspace such that the projection of a given point x, can be improved as a function of the learning rate. This allows the update of a given subspace node, with respect to the data, over an arbitary value of the learning rate such that the distance from the node subspace and that of the point is reduced. A MATLAB implementation is noted in the appendix.

Coupled with the new metric, which allows the determination of the nodes to be updated, all the elements needed to construct an ASSOFM clustering algorithm are in place.

Chapter 4

Methodology

4.1 Introduction

With consideration already given to the nature of the data set, and the theory of clustering methods, this chapter relates the practical implementation and interpretation of these clustering methods as a classifier.

4.2 Probabilistic Interpretation of the SOFM used for Classification

Given the generation of clusters using the ASSOFM, a method of implementing this as a classifier is required. That is, once training is complete, the ASSOFM when presented with a CED should offer a probability of the CED being a true EEv. Each node must offer a probability between 1, for Epileptiform Discharges, and 0 for artefacts. A principled calibration of the SOFM was presented by James [9]. Some elements of this method are reproduced here.

In contrast with the method presented by Kohonen, where a label is assigned due to maximal voting, per SOFM node, over a calibration set (often the training set) an alternative approach is suggested by James [9]. Given that this study will be directly compared with this previous study that method, outlined below, is employed for parity.

The success rate for each node i is given by;

$$\theta^i = \frac{s_i}{n_i} \tag{4.1}$$

where s_i is the number of EEvs detected, and n_i is the number of detection in total for that node. By appealing to Bayes Theory, it is possible to assign probability to each node. This probability results from a weighted mixture of the *prior* probabilities and the *posterior* probabilities. This gives rise to the following estimate of the probability of any given node i;

$$\psi_i = \frac{s_i + 1}{n_i + 2} \tag{4.2}$$

Consider the following simplified example; Two nodes h and i. Node h was the winning node 10 times, 9 of which were EEvs and 1 an artefact. This gives a success rate, as defined above, of 0.9. However, employing the probabilistic estimation, this results in a value of 0.83. If node i was found to be the winning node 100 times, 90 times out of which for EEv waveforms, then the success rate remains as for node h but the probability now becomes 0.89. This reflects a greater confidence in node i as its results are over a larger sample.

Consider also, if node h was victorious for true EEvs 10 times and node i accurate for 100 classifications, this would offer a success rate 1.0. However, with the suggested scheme above of estimated probabilities, node h would have an value of 0.86 whilst node i would be assigned 0.99.

4.3 Measures of Performance

Given a method of assigning estimated probability labels to each node, we want a measure of performance of a given ASSOFM, for a given probability threshold used in classification.

4.3.1 Selectivity and Sensitivity

The final gauge of the classifier performance is not the number of EEvs detected, as there remains the likelihood that it will classify a number of artefacts as EEvs. To illuminate this a ratio of correctly and incorrectly classified events must be employed. Two new properties are be defined to allow this measure of performance; Sensitivity and Selectivity (after James [9]), which are a function of the classification threshold d_{th}, which defines those CEDs classified as EEvs.

$$sensitivity(d_{th}) = \frac{\text{correct EEv detections}(d_{th})}{\text{total number of true EEvs}}, \tag{4.3}$$

and

$$selectivity(d_{th}) = \frac{\text{correct EEv detections}(d_{th})}{\text{total number of detections}(d_{th})}. \tag{4.4}$$

Sensitivity is a measure of EEv detection, with 1 indicating only true events being detected. However, this must be contrasted with selectivity, a measure of how many unwanted artefacts enter the classification. A selectivity of 1 would indicate a discriminating classifier that resulted in no incorrect classifications. It is possible to have a system that detects all EEvs, but due to a low selectivity be useless, thus a balance between the two measure is desirable. In the case of a medical diagnosis this balance can be seen as the dilemma between missing a clinical event i.e. an EEv, which has serious consequences, or misdiagnosing an individual, which whilst serious would appear to be less important. The obvious optimal would see both as 1.

Selectivity and sensitivity are functions of the probability threshold d_{th}, above which it is considered classification of an EEv to be accurate. The definition of an actual EEv could be altered by raising (or lowering) this probability threshold, this would have an effect of increasing selectivity and decreasing sensitivity (or vice versa). To exploit this a range of thresholds must be examined.

4.3.2 False Detection Rate

The False detection rate, often measured per hour, is simply the number of artefacts erroneously classified as EEvs by the system over the period of one hour of EEG. This is important in measuring the practical utility of the system (in tandem with the abstracted measures of sensitivity and selectivity). If this rate is zero, then there is certainty over any positive classification. Given the suspicion of a high false positive rate, this entails a manual re-examination of the data, defeating the aim and reducing the system to a data reduction strategy.

4.4 The Dimensionality of the Data

In the Introduction, the importance of the retention of *context* into which the CED was embedded to visual, and hence must be in turn to human like automatic classification was noted. However, the dimensionality of the data offered by the mimetic stage requires some further investigation. It would be desirable if the data dimensionality could be reduced, whilst retaining the implicit information, even from a purely computational point of view. Preprocessing's primary aim is the improvement of the Signal to Noise ratio, allowing the enhancement of the performance of any next stage, including those considered here.

There are a variety of methods for dimensionality reduction including; the ubiquitous Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Independent Component Analysis (ICA, a form of Blind Source Separation), Projection Pursuit, Canonical Variables and the SOFM itself.

Previous research suggests that EEG data has a low underlying dimensionality [1], i.e. a dimensionality less than the 25 channels of the 10-20 EEG. In this study it is known that two generators underly the synthetic data without noise. Anecdotal application of ICA to the extraction of the underlying generator in EEvs has offered a dimension of 1. It is therefore considered that dimensionality reduction can be employed in the expectation of the retention of information.

The aim of this Thesis is not to produce the optimal utility but rather exploit the theorised qualities of the ASSOFM. As such, an ideal pre-processing stage has the requirement of not affecting classification, but merely aiding it by rendering the data in a concise form.

4.4.1 Principal Component Analysis

Initially PCA was considered, also called the Karhunen-Loève transform. PCA is a linear method of dimensionality reduction that results in a data re-expression that is a linear combination of the original variables, thus retaining the same dimension. However, it returns information about the relative importance of these components in the terms of variance explained. This allows specific elements, which have least associated variance to be discarded, and hence the dimensionality reduced in a fashion that attempts to retain maximal information.

The Principal Components may be extracted practically using this algorithm;

1. First compute the mean of data point and then subtract this.

2. The covariance matrix of the resultant is then calculated from which the Eigenvalues and Eigenvectors are then found.

3. These Eigenvalues are ordered by variance explained. The data is then projected onto the Eigenvectors, thus giving rise to the new variables each of which has a related variance.

4. By discarding those of lowest variance, the dimensionality is reduced whilst retaining maximum information.

Principal Component Analysis is not without limitations. The Principal Components are not invariant to scaling in the data, though this can be alleviated by the use of normalisation post subtracting the mean. PCA evidences sensitivity to outliers, and being a linear technique it may return sub-optimal components that do not reflect the implicit dimensionality of non-linear data.

4.4.2 Results and Conclusions of PCA/SVD applied to the Candidate Epileptiform Discharges

The CEDs had dimensions of 41 samples x 25 channels. Spatial information, i.e. has already been used to judge the candidates, so this information is explicit in each of the

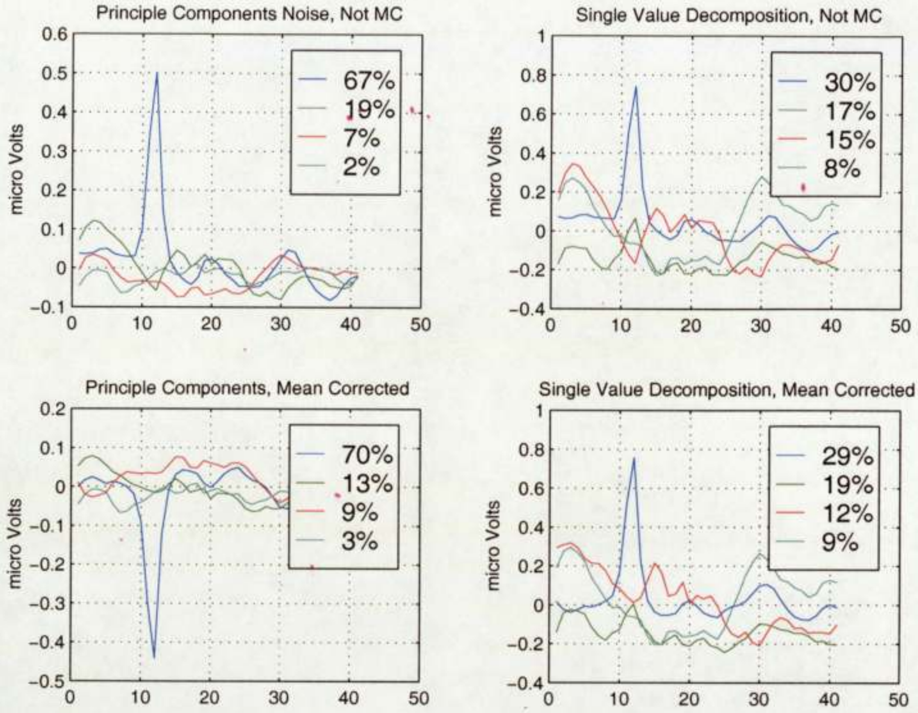

Figure 4.1: Comparison of PCA and SVD over the example dipole, using mean and non mean corrected data of 41 samples x 25 channels.

data points. The spatial dimension is therefore the primary candidate for reduction.

SVD and from that in turn PCA were applied to both synthetic and artefact candidate epileptiform discharges with the following results. Figure 4.1 shows the results of PCA and SVD on the example 3 Sphere Model propagated dipole with additive noise (Figure 2.4) considered in Chapter 2. This results conform with the expectation of a low dimensionality, given the assumption that variance is a useful arbiter of dimensionality of the EEG, and the fact that it is know that two generators underly the noiseless data. PCA consistently accounts for 85% of the data variance with 5 principal components. However this does not entail that there are 5 underlying components, or that the first 85% of the variance is the signal of interest. It could reflect a non-linearity within the data that PCA is struggling to model. However, it is felt that the arbitrariness of a specific variance target given that noise is a significant and varying component, accounts for such a high number of dimensions. Empirically, visual comparison of the generative dipole and the dominant components suggest that PCA is extracting useful data. Consider again Figure 4.1 in contrast with that of the dipole; Figure 2.4.

It was decided to employ the first 5 principal components, given the variance over these 5, in tandem with the fact that two generators underly the synthetic data without noise. Whereas, had this been unknown, further components would have been included to explain 95% of the variance. This gives rise to candidate event data points of 41 temporal steps x 5 principal components.

4.5 Initial Implementation of the ASSOFM

A cautious strategy was employed in the testing of the variant adaptive subspace SOFM algorithm. A trial data set employing a jittered, fixed dipole morphology for the generation of the synthetic candidate epileptiform discharges was implemented. This would allow evaluation both, of the practical implementation of the ASSOFM from a computational point of view, and offer preliminary clustering results. As a secondary goal these initial ASSOFM would also allow the testing of the classification procedure and invariance to dipole transformations. This invariance is tested by examining the performance of the trial ASSOFM over the data sets that have dipoles of a random nature, as opposed to the fixed nature of the training set.

Ensuing ASSOFMs would be trained on the fully random data sets to evaluate whether this would result in any alteration in performance. Further, these ensuing ASSOFMs would explore modelling ability of the subspace in situation where there was no 1 to 1 correspondence between data and adaptive basis vectors. These will then be evaluated over the training set and novel data.

Trial Method A and B

The first trial algorithm, termed Method iA (where the i denotes initial), employs the complete CED, i.e. 5 Principal Components by 41 times steps, to determine the rotational operator.

Employing empirical parameters, previously developed for the standard SOFM implementation, several sizes of network were implemented to discover the optimum size. The nodes themselves employed five basis vectors, again of dimension 41. Starting with 6 by 6 nodes, followed by an 8 by 8, 10 by 10 and ultimately 15 by 15 nets. In all instances the initial neighbourhood was half that of the net size (7 in the case of the 15 by 15), and decayed monotonically to one. The learning rate parameter, α, was initialised as one and decayed monotonically to zero.

Given a data set of 7000 CEDs, alternating between artefacts and synthetic, it was estimated that 15 passes through the data set to allow useful convergence. To facilitate analysis of this projection the map was displayed and saved after each complete pass through the data to allow analysis. The nodes were populated at random with equal numbers of artefact and synthetic candidates.

The second exploratory implementation, Method iB, followed a differing form, primarily to explore the usefulness of smaller nodes, and an alternate winner strategy. The nodes in this instance employed three basis vectors, again of dimension 41. Further, each data point was considered as an *event*, composed of the five components. The rotation operator, in contrast to Method A, was not defined by the whole *event*, but by a single winning component. This imposes a five fold increase in computational cost, as each component is compared to every node individually to determine both the winning component and winning node. Further, the distance was then weighted, with the primary principal component considered twice as important as the next and so on. Other parameters remained as for Method iA.

As noted in Section 4.2, it is possible to extract a probabilistic interpretation of the trained SOFM, and by extension the ASSOFM. In all cases, this calibration is done over the training set.

These procedures were then repeated but with a data set constructed using 2000 random dipoles and a further 2000 artefacts extracted from 3 patients, which were ordered randomly by each pass of the ASSOFM algorithm. Method A was extended to test the modelling ability of the ASSOFM by examining a range of subspace spanning basis vectors.

Chapter 5

Results

5.1 Introduction

This section collates the results of the ASSOFM.

5.1.1 ASSOFM Labelling Convention

To aid swift recall of the ASSOFM employed this convention is employed throughout the results; (i)ANxNbv$M$. Where the leading 'i' reflects the initial examination over the fixed dipole data set defined below. The first upper case letter refers to the method employed, either A or B. The variable N refers to the ASSOFM's size, whilst M is the number of basis vectors employed in each node. For example iB6x6bv3, refers to the initial implementation of Method B, over a 6 by 6 network of nodes, each of which defines a subspace spanned by 3 subspace vectors.

5.1.2 Review of Data Types

Definition of Data Types employed. As noted previously the synthetic data is generated with the aid of 6 variables and incorporating jitter. The initial implementations employed data from a fixed dipole, whereas ensuing results were obtained employing a data set composed of randomised dipoles. The following define the labels employed in the tables.

- Fixed data - Data which is composed of single jittered reference dipole, which has a constant set of variables. The data set annotated with an asterisk, Fixed* was composed of 7500 CEDs, equally composed of synthetic EEvs and artefact TNS.

- Type A - Data composed of a dipole which was jittered, and had random eccentricity f.

- Type B - Data composed of a dipole which was jittered, and had random eccentricity and orientation ϕ and θ.

- Type C - Data composed of dipoles that were generated with all variables random. The data set annotated with an asterisk, Type C was composed of 4000 CEDs, again equally balanced.

- An asterisk further indicates that the data set was the Training Set for that particular ASSOFM

5.2 Example Estimated Probability Map

All the following results are derived from the Probability Maps defined under Section 4.2. An example, A10x10bv5 is given here to make clearer the process involved in extracting these results.

The results from a converged ASSOFM take the following form displayed in Figure 5.1.

The training set, in this case Type C*, is used to calibrate the map i.e. to calculate the EEv probability. This results in a Probability Map of the nodes. This is presented numerically in Table 5.1 and in Figure 5.2.

Node	1	2	3	4	5	6	7	8	9	10
1	0.76	0.90	0.79	0.64	0.61	0.54	0.53	0.54	0.41	0.30
2	0.92	0.85	0.80	0.77	0.90	0.75	0.73	0.60	0.52	0.38
3	0.72	0.83	0.85	0.89	0.79	0.72	0.65	0.60	0.63	0.23
4	0.76	0.75	0.90	0.80	0.76	0.77	0.70	0.54	0.44	0.13
5	0.63	0.81	0.83	0.77	0.68	0.78	0.71	0.46	0.25	0.15
6	0.58	0.79	0.66	0.80	0.68	0.81	0.54	0.56	0.28	0.05
7	0.58	0.51	0.65	0.63	0.64	0.66	0.56	0.26	0.19	0.09
8	0.45	0.41	0.41	0.29	0.38	0.43	0.28	0.20	0.19	0.063
9	0.37	0.27	0.27	0.27	0.21	0.11	0.03	0.17	0.16	0.07
10	0.20	0.22	0.22	0.16	0.04	0.09	0.03	0.05	0.15	0.09

Table 5.1: Numerical estimated EEv Probability Map from trained ASSOFM: A10x10bv5

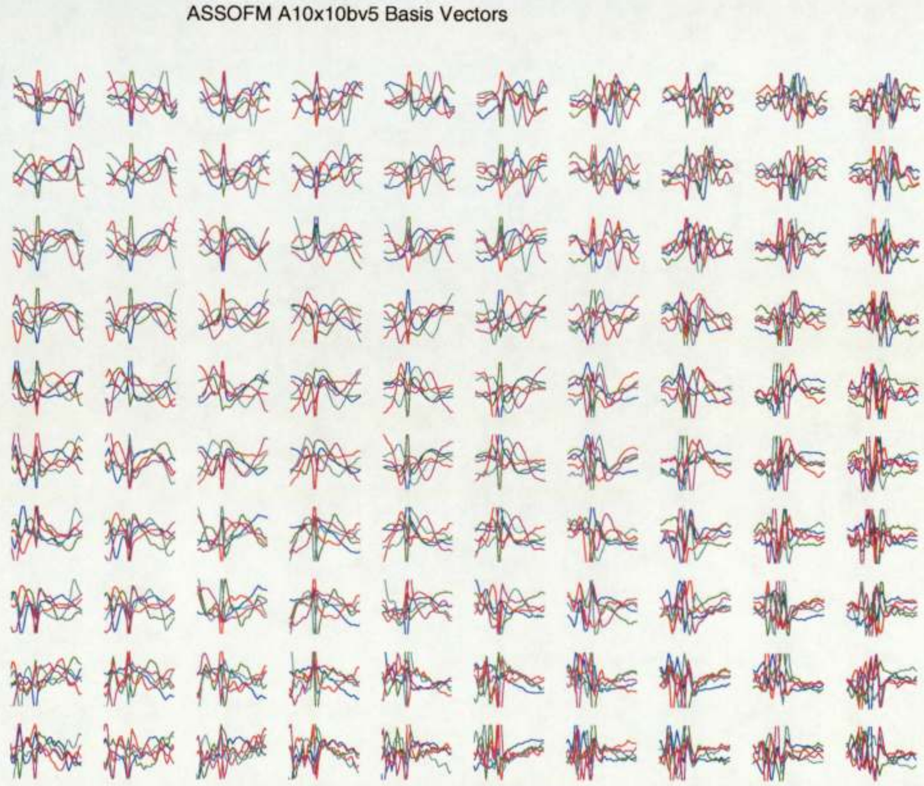

Figure 5.1: Converged ASSOFM Nodes. This displays the 41x5 Adaptive Basis Vectors in a 10x10 ASSOFM: A10x10bv5.

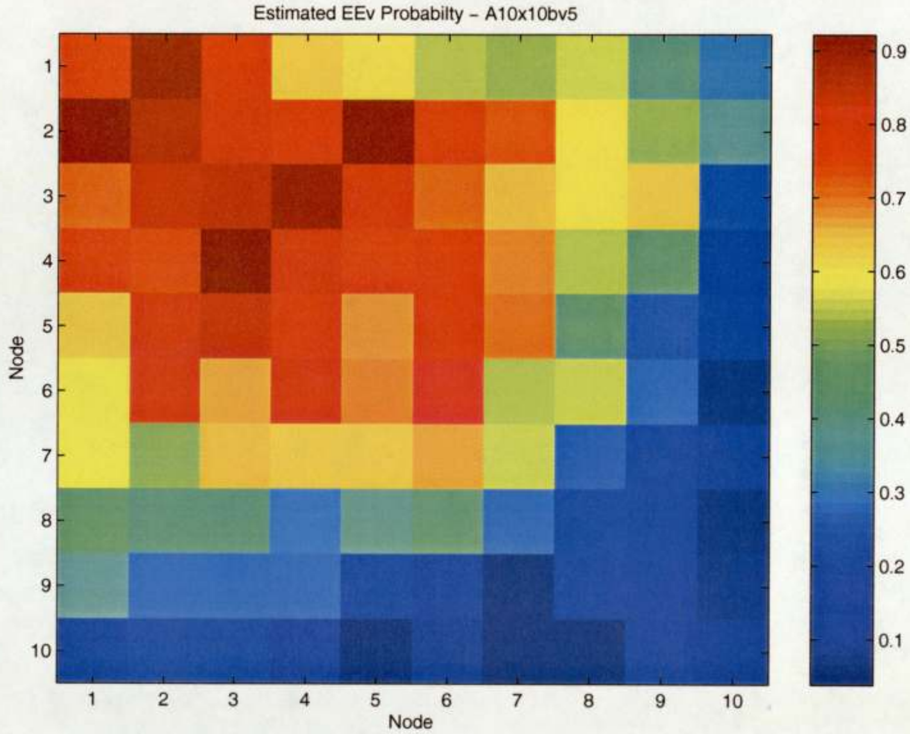

Figure 5.2: This displays the estimated EEv probabilities over the 10x10 ASSOFM, A10x10bv5. The estimated probability values above 0.5 suggest with increasing confidence that any novel data which is assigned that node can be classified as EEv, and that those below 0.5 are not.

5.3 Results for both Method iA and A, employing nodes spanned by five basis vectors. ASSOFM: (i)ANxNbv5.

The ASSOFM implementation iANxNbv5; results are presented graphically in Figure 5.3 and Figure 5.4. Sensitivity and Selectivity over the Fixed*, i.e. training data, and Type C (random dipole data set) are presented in Tables 5.3 and 5.2. Error rate is tabulated in Table 5.4.

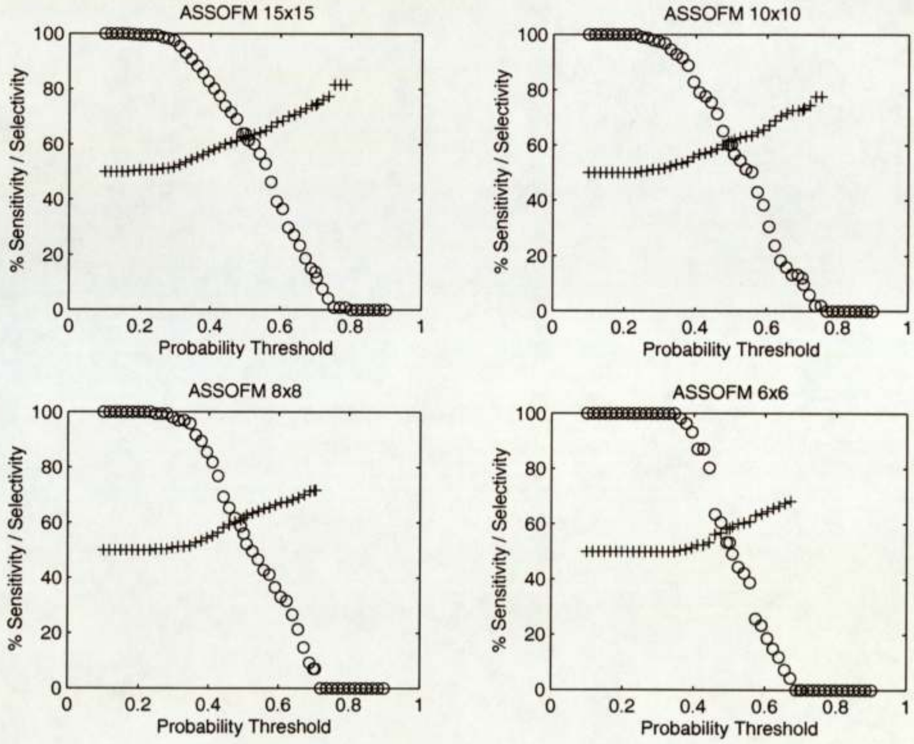

Figure 5.3: Selectivity and Sensitivity from ASSOFM for iANxNbv5 over Fixed* data set
KEY + Selectivity o Sensitivity

The ASSOFM implementation ANxNbv5; results are presented graphically in Figure 5.5. Sensitivity and Selectivity over Type C* (random dipole data set) are presented in Tables 5.6 and 5.5. Error rate is tabulated in Table 5.7.

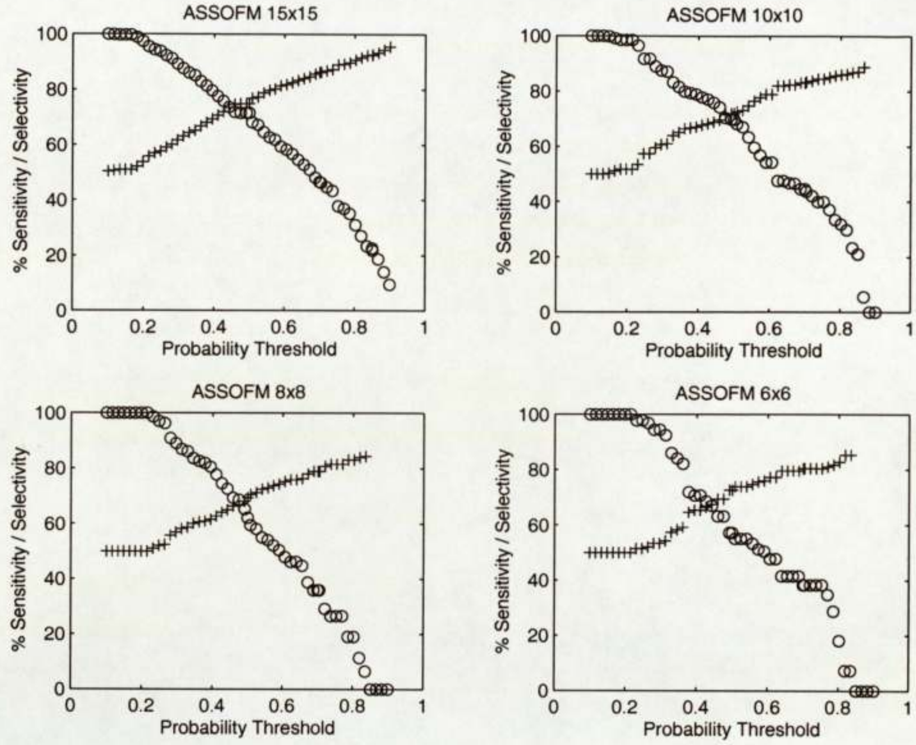

Figure 5.4: Selectivity and Sensitivity from ASSOFM for iANxNbv5 over Type C.
KEY + Selectivity o Sensitivity

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Selectivity			
15 x 15	Fixed*	0.5000	0.6258	0.7434	-
	Type C	0.5093	0.7492	0.8622	0.9242
10 x 10	Fixed*	0.5000	0.6126	0.7284	-
	Type C	0.5000	0.7194	0.8313	0.8696
8 x 8	Fixed*	0.5000	0.6163	0.7168	-
	Type C	0.5000	0.6939	0.7899	-
6 x 6	Fixed*	0.5000	0.5824	-	-
	Type C	0.5000	0.7263	0.8055	-

Table 5.2: Results for ASSOFM iANxNbv5- Selectivity. Measured over the Fixed* data set, and calculated over novel data of Type C.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
15 x 15	Fixed*	1.0000	0.6383	0.1374	0
	Fixed	1.0000	0.8000	0.2667	0
	Type A	1.0000	1.0000	0.6000	0
	Type B	1.0000	1.0000	0.5000	0
	Type C	0.9990	0.7155	0.4660	0.2260
10 x 10	Fixed*	1.0000	0.6014	0.1203	0
	Fixed	1.0000	1.0000	0.3667	0
	Type A	1.0000	0.9000	0.3000	0
	Type B	1.0000	1.0000	0.2000	0
	Type C	1.0000	0.6985	0.4460	0.2100
8 x 8	Fixed*	1.0000	0.5594	0.0709	0
	Fixed	1.0000	0.9000	0.2333	0
	Type A	1.0000	1.0000	0.1000	0
	Type B	1.0000	1.0000	0.3000	0
	Type C	1.0000	0.6210	0.3590	0
6 x 6	Fixed*	1.0000	0.5331	0	0
	Fixed	1.0000	1.0000	0	0
	Type A	1.0000	1.0000	0	0
	Type B	1.0000	1.0000	0	0
	Type C	1.0000	0.5720	0.3830	0

Table 5.3: Results for ASSOFM iANxNbv5- Sensitivity over the Fixed* data set, and over data sets; Fixed, Type A, Type B and Type C

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
15x15	Fixed*	3500	1336	166	0
10x10		3500	1331	157	0
8x8		3500	1219	98	0
6x6		3500	1338	0	0
15x15	Type C	1967	479	149	37
10x10		2000	545	181	63
8x8		2000	548	191	0
6x6		2000	431	185	0

Table 5.4: Results for ASSOFM iANxNbv5- Error Rate over the Fixed* data set (3500 possible artefacts) and over Type C(2000 possible artefacts).

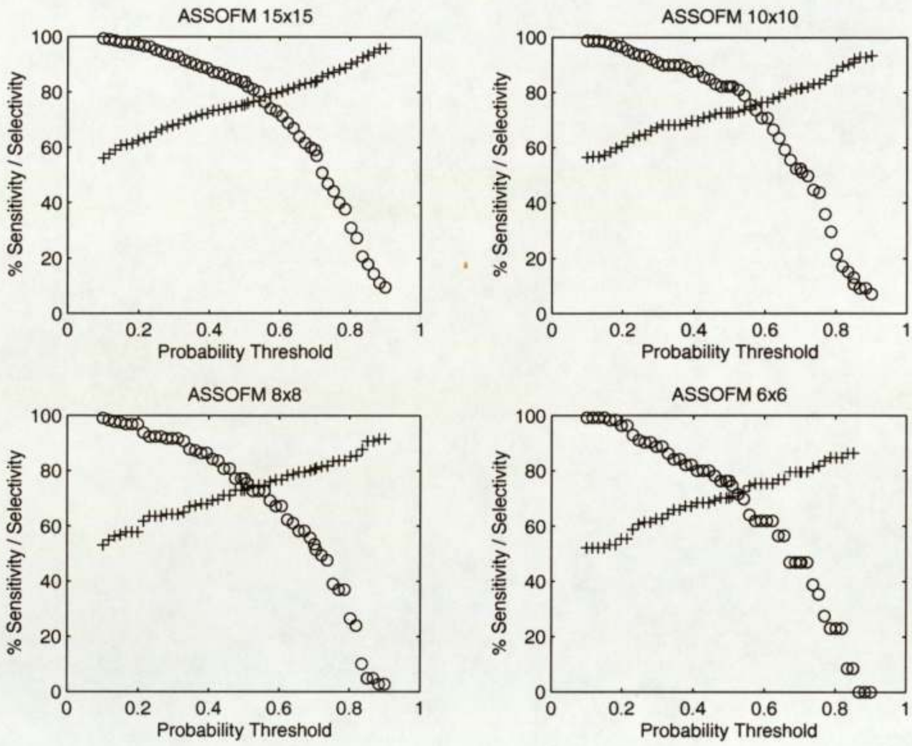

Figure 5.5: Selectivity and Sensitivity from ASSOFM for ANxNbv5 over Type C*. KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
15 x 15	0.5605	0.7525	0.8371	0.9344
10 x 10	0.5643	0.7272	0.8143	0.9097
8 x 8	0.5307	0.7293	0.8066	0.9074
6 x 6	0.5211	0.7013	0.7963	0.8636

Table 5.5: Results for ASSOFM ANxNbv5- Selectivity over data set Type C*.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
15 x 15	Type C*	0.9935	0.8375	0.5910	0.1790
	Fixed	1.0000	0.9000	0.8667	0.2333
	Type C	1.0000	1.0000	0.6667	0.3000
10 x 10	Type C*	0.9875	0.8225	0.5240	0.1310
	Fixed	1.0000	1.0000	0.9667	0.4667
	Type C	1.0000	1.0000	0.9000	0.4333
8 x 8	Type C*	0.9910	0.7720	0.5340	0.0490
	Fixed	1.0000	1.0000	0.9667	0.1000
	Type C	1.0000	0.8667	0.8667	0.0667
6 x 6	Type C*	0.9925	0.7630	0.4690	0.0855
	Fixed	1.0000	1.0000	0.8000	0.2667
	Type C	1.0000	1.0000	0.9000	0.6000

Table 5.6: Results for ASSOFM ANxNbv5- Sensitivity over data sets Type C*, Fixed and Type C

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
15x15	Type C*	1558	551	230	25
10x10		1525	617	239	26
8x8		1753	573	256	10
6x6		1824	650	240	27

Table 5.7: Results for ASSOFM ANxNbv5- Error rate over Type C* (2000 possible artefacts)

5.4 Results for Method A, employing nodes spanned by four basis vectors. ASSOFM: ANxNbv4.

The ASSOFM implementation ANxNbv4; results are presented graphically in Figure 5.6. Sensitivity and Selectivity are presented in Tables 5.9 and 5.8. Error rate is tabulated in Table 5.10.

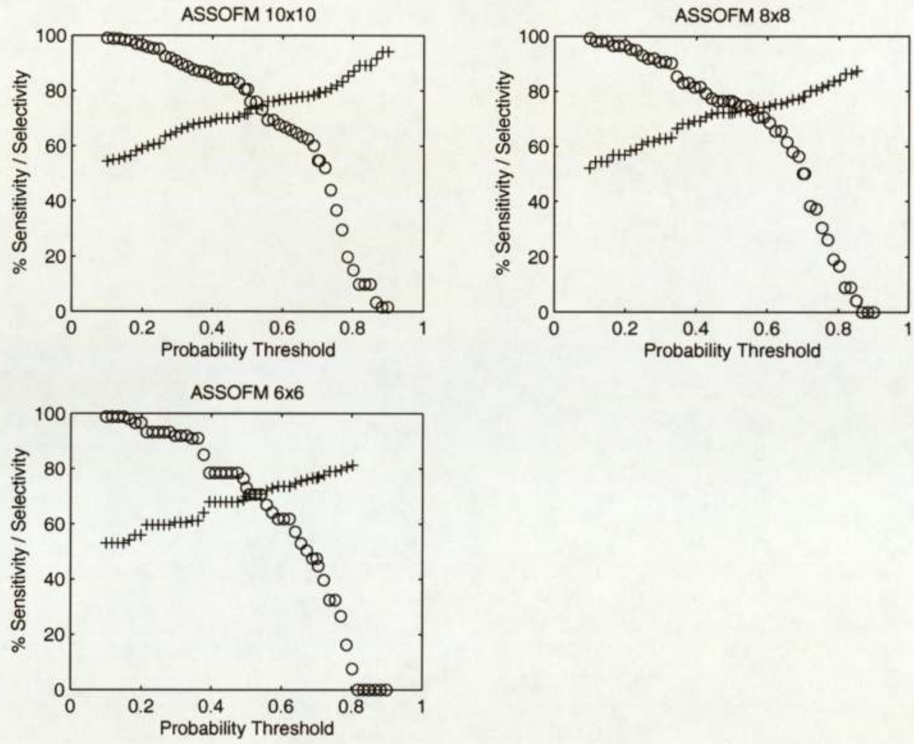

Figure 5.6: Selectivity and Sensitivity from ASSOFM for ANxNbv4 over Type C*. KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5439	0.7162	0.7926	0.8919
8 x 8	0.5211	0.7224	0.7812	0.8737
6 x 6	0.5306	0.6999	0.7672	-

Table 5.8: Results for ASSOFM ANxNbv4- Selectivity over data set Type C*.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Type C*	0.9875	0.9905	0.5465	0.0990
	Fixed	1.0000	0.9333	0.9000	0.4333
	Type C	1.0000	0.9667	0.8000	0.1000
8 x 8	Type C*	0.9910	0.9920	0.5015	0.0415
	Fixed	1.0000	1.0000	0.9667	0
	Type C	1.0000	1.0000	0.8000	0
6 x 6	Type C*	0.9925	0.9880	0.4745	0
	Fixed	1.0000	1.0000	0.6000	0
	Type C	1.0000	1.0000	0.6000	0

Table 5.9: Results for ASSOFM ANxNbv4- Sensitivity over data sets Type C*, Fixed and Type C.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
10x10	Type C*	1661	638	286	24
8x8		1823	588	281	12
6x6		1748	627	288	0

Table 5.10: Results for ASSOFM ANxNbv4- Error rate over Type C* (2000 possible artefacts).

5.5 Results for Method A, employing nodes spanned by three basis vectors. ASSOFM: ANxNbv3.

The ASSOFM implementation ANxNbv3; results are presented graphically in Figure 5.7. Sensitivity and Selectivity are presented in Tables 5.12 and 5.5. Error rate is tabulated in Table 5.13.

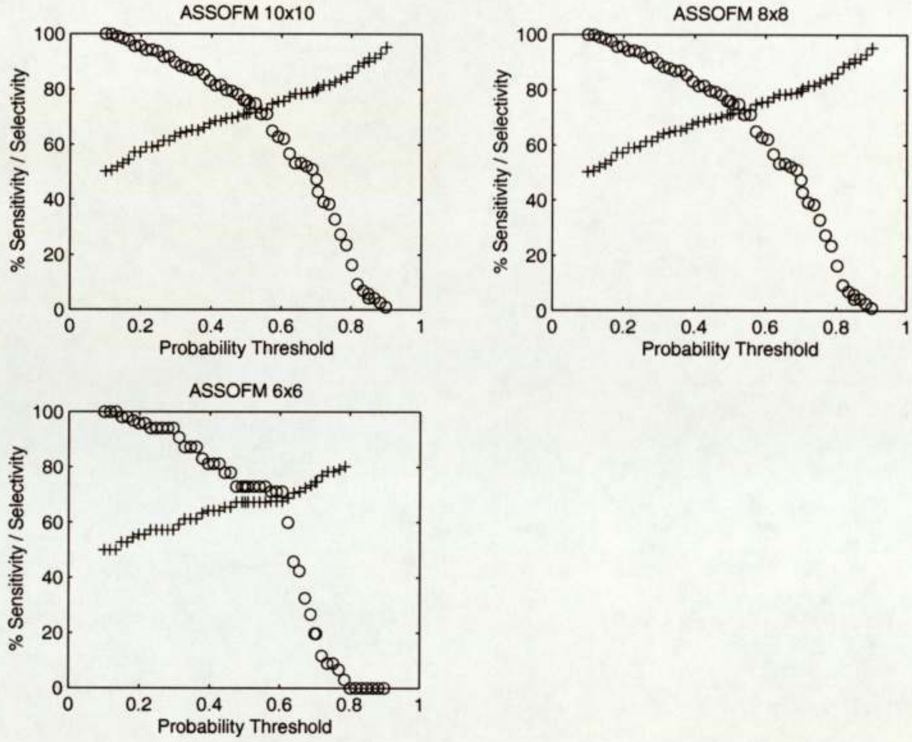

Figure 5.7: Selectivity and Sensitivity from ASSOFM for ANxNbv3 over Type C*
KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5029	0.7103	0.7965	0.9000
8 x 8	0.5029	0.7103	0.7965	0.9000
6 x 6	0.5000	0.6730	0.7476	-

Table 5.11: Results for ASSOFM ANxNbv3- Selectivity over data set Type C*

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Type C*	0.9995	0.7600	0.4735	0.0585
	Fixed	1.0000	1.0000	0.5667	0.1333
	Type C	1.0000	0.7667	0.7000	0.0667
8 x 8	Type C*	0.9995	0.7600	0.4735	0.0585
	Fixed	1.0000	1.0000	0.7000	0.1667
	Type C	1.0000	0.8667	0.7000	0.0333
6 x 6	Type C*	1.0000	0.7295	0.1970	0
	Fixed	1.0000	0.9333	0.3333	0
	Type C	1.0000	1.0000	0.1333	0

Table 5.12: Results for ASSOFM ANxNbv3- Sensitivity over data sets Type C*, Fixed and Type C.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
10x10	Type C*	1976	620	242	13
8x8		1976	620	242	13
6x6		2000	709	288	0

Table 5.13: Results for ASSOFM ANxNbv3- Error rate over Type C* (2000 possible artefacts)

5.6 Results for Method A, employing nodes spanned by two basis vectors. ASSOFM: ANxNbv2.

The ASSOFM implementation ANxNbv2; results are presented graphically in Figure 5.8. Sensitivity and Selectivity are presented in Tables 5.15 and 5.6. Error rate is tabulated in Table 5.16.

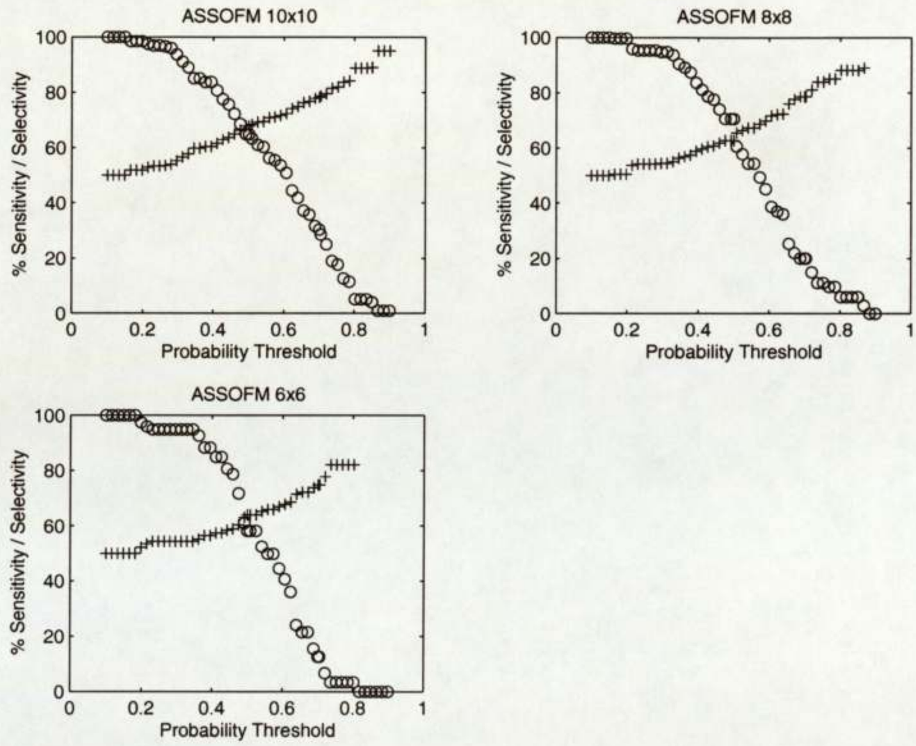

Figure 5.8: Selectivity and Sensitivity from ASSOFM for ANxNbv2 over data set Type C*.
KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5000	0.6785	0.7822	0.8901
8 x 8	0.500	0.6276	0.7874	0.8824
6 x 6	0.5000	0.6394	0.7478	-

Table 5.14: Results for ASSOFM ANxNbv2- Selectivity over data set Type C*.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Type C*	1.0000	0.6480	0.3035	0.0405
	Fixed	1.0000	0.6333	0.1000	0
	Type C	1.0000	0.8333	0.3667	0
8 x 8	Type C*	1.0000	0.7060	0.2000	0.0600
	Fixed	1.0000	0.9000	0.3333	0
	Type C	1.0000	0.6333	0.1333	0
6 x 6	Type C*	1.0000	0.5815	0.1260	0
	Fixed	1.0000	0.8000	0.3333	0
	Type C	1.0000	0.9000	0.1333	0

Table 5.15: Results for ASSOFM ANxNbv2- Sensitivity over data sets, Type C*, Fixed and Type C.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
10x10	Type C*	2000	614	169	10
8x8		2000	838	108	16
6x6		2000	656	85	0

Table 5.16: Results for ASSOFM ANxNbv2- Error rate over Type C* (2000 possible artefacts)

5.7 Results for Method A, employing nodes spanned by one basis vector. ASSOFM: ANxNbv1.

The ASSOFM implementation ANxNbv1; results are presented graphically in Figure 5.9. Sensitivity and Selectivity are presented in Tables 5.18 and 5.17. Error rate is tabulated in Table 5.19.

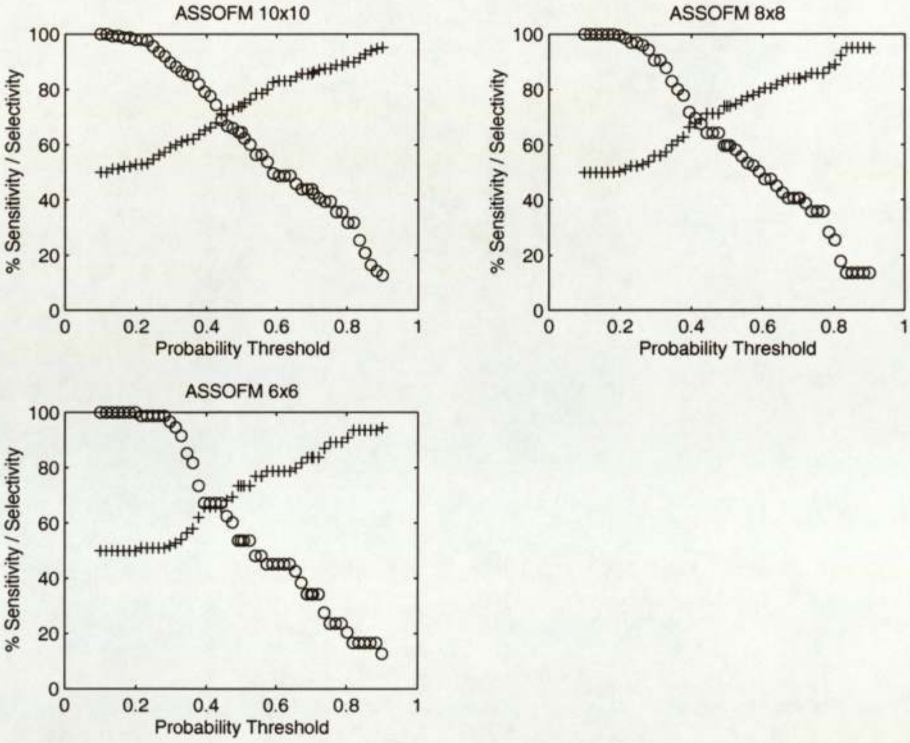

Figure 5.9: Selectivity and Sensitivity from ASSOFM for ANxNbv1 Type C*
KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5000	0.7381	0.8566	0.9265
8 x 8	0.500	0.7392	0.8395	0.9507
6 x 6	0.5000	0.7339	0.8374	0.9352

Table 5.17: Results for ASSOFM ANxNbv1- Selectivity over data set Type C*.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Type C*	1.0000	0.6455	0.4390	0.2080
	Fixed	1.0000	0.8333	0.4667	0.3667
	Type C	1.0000	1.0000	1.0000	0.3667
8 x 8	Type C*	1.0000	0.5965	0.4055	0.1350
	Fixed	1.0000	0.8000	0.5000	0.3667
	Type C	1.0000	1.0000	1.0000	0.5000
6 x 6	Type C*	1.0000	0.5365	0.3425	0.1660
	Fixed	1.0000	0.8000	0.5000	0.3000
	Type C	1.0000	1.0000	0.9667	0.1667

Table 5.18: Results for ASSOFM ANxNbv1- Sensitivity over data sets Type C*, Fixed and Type C.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Error Rate			
10x10	Type C*	2000	458	147	33
8x8		2000	421	155	14
6x6		2000	389	133	23

Table 5.19: Results for ASSOFM ANxNbv2- Error Rate over Type C* (2000 possible artefacts).

5.8 Results for both Method iB and B, employing nodes spanned by three basis vectors. ASSOFM: (i)BNxNbv3.

The ASSOFM implementation iBNxNbv3; results are presented graphically in Figure 5.10. Sensitivity and Selectivity over the Fixed*, i.e. training data are presented in Tables 5.21 and 5.20.

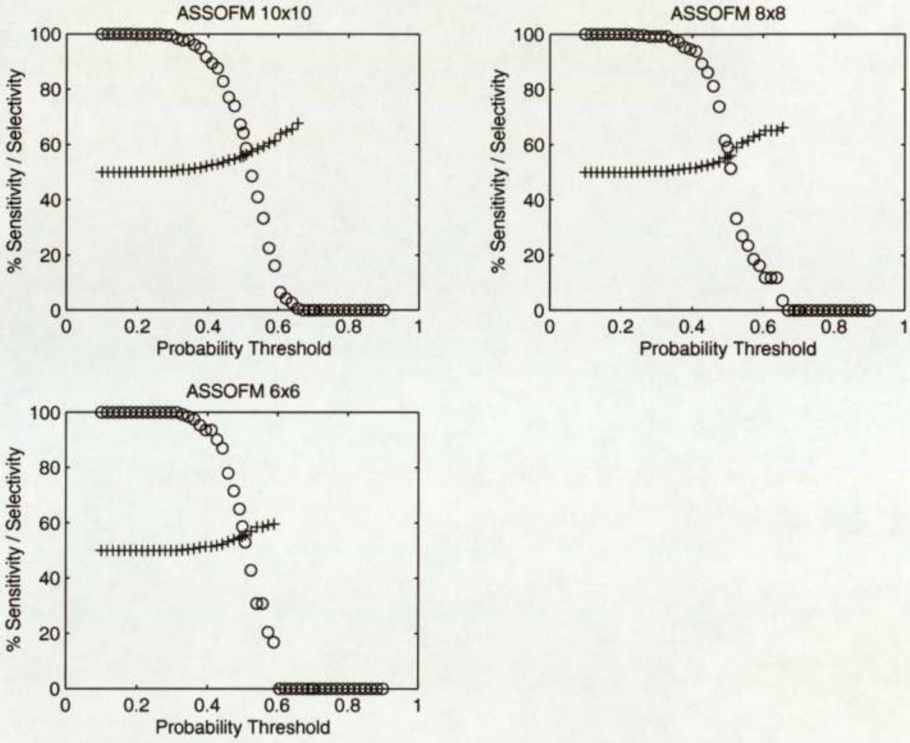

Figure 5.10: Selectivity and Sensitivity from ASSOFM for iBNxNbv3 over Fixed* data set.
KEY + Selectivity o Sensitivity

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5000	0.5580	-	-
8 x 8	0.5000	0.5528	-	-
6 x 6	0.5000	0.5516	-	-

Table 5.20: Results for ASSOFM iBNxNbv3- Selectivity over Fixed* data set.

The ASSOFM implementation BNxNbv3; results are presented graphically in Figure 5.11. Sensitivity and Selectivity over Type C* (random dipole data set) are presented in Tables 5.23 and 5.22.

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Fixed*	1.0000	0.6417	0	0
	Fixed	1.0000	0.7667	0	0
	Type A	1.0000	0	0	0
	Type B	1.0000	0.1000	0	0
	Type C	1.0000	0.5000	0	0
8 x 8	Fixed*	1.0000	0.5903	0	0
	Fixed	1.0000	0.5000	0	0
	Type A	1.0000	0	0	0
	Type B	1.0000	0.4000	0	0
	Type C	1.0000	0.6000	0	0
6 x 6	Fixed*	1.0000	0.3866	0	0
	Fixed	1.0000	0.4333	0	0
	Type A	1.0000	0.8000	0	0
	Type B	1.0000	0.8000	0	0
	Type C	1.0000	0.8000	0	0

Table 5.21: Results for ASSOFM iBNxNbv3- Sensitivity over Fixed* data set

Probability Threshold	0.10	0.50	0.70	0.85
Size N	Selectivity			
10 x 10	0.5000	0.6580	0.7344	-
8 x 8	0.5000	0.5894	0.7174	-
6 x 6	0.5000	0.6856	0.8148	-

Table 5.22: Results for ASSOFM BNxNbv3- Selectivity over data set Type C*.

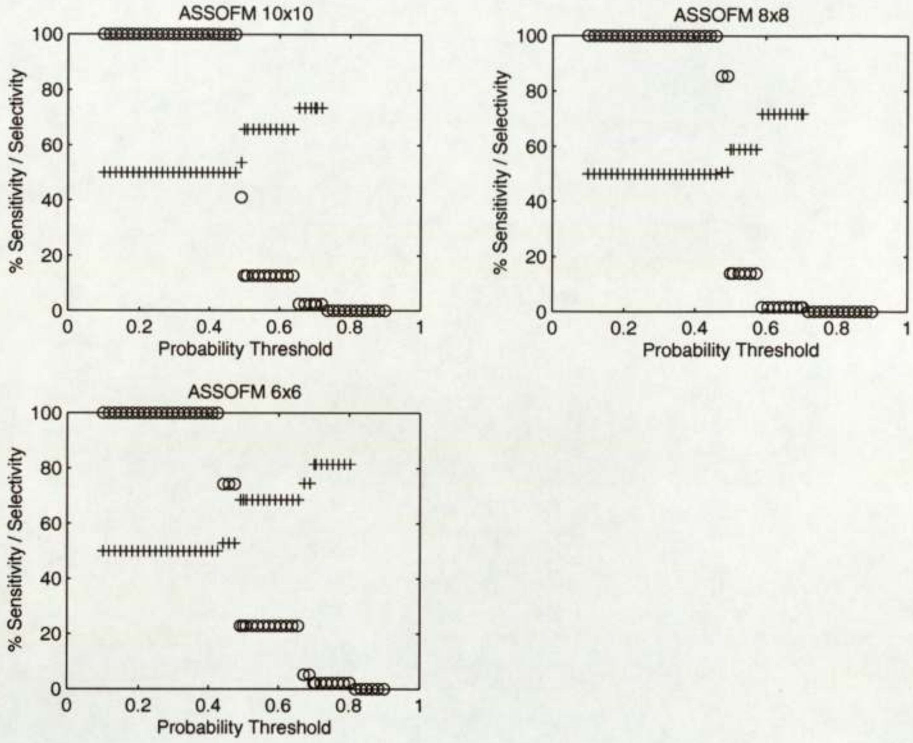

Figure 5.11: Selectivity and Sensitivity from ASSOFM for BNxNbv3 over Type C*. KEY + Selectivity o Sensitivity

Probability Threshold		0.10	0.50	0.70	0.85
Size N	Data Type	Sensitivity			
10 x 10	Type C*	1.0000	0.1260	0.0235	0
	Fixed	1.0000	0.8333	0.4667	0.3667
	Type C	1.0000	1.0000	1.0000	0.3667
8 x 8	Type C*	1.0000	0.1385	0.0165	0
	Fixed	1.0000	0.8000	0.5000	0.1667
	Type C	1.0000	1.0000	1.0000	0.5000
6 x 6	Type C*	1.0000	0.2290	0.0220	0
	Fixed	1.0000	0.8000	0.5000	0.3000
	Type C	1.0000	1.0000	0.9667	0.1667

Table 5.23: Results for ASSOFM BNxNbv3- Sensitivity over data sets Type C*, Fixed and Type C.

Chapter 6

Discussion

6.1 Introduction

This section offers a subjective evaluation of the ASSOFM clustering, followed by an objective analysis of the results. It then continues with the comparison of the measures of performance obtained in this study with a range of results from the literature in the field of automated EEG analysis.

6.2 The Viability of the ASSOFM as a Topography Preserving Map

Before considering the resulting classification, attention must be directed at the clustering ability of the ASSOFM.

Table 5.1 and in Figure 5.2 reflect example results from the ASSOFM A10x10bv5 (i.e. Method A, 10x10 ASSOFM with Adaptive Subspaces spanned by 5 x 41 dimensional basis vectors, trained and calibrated over a Type C* data set).

Consider especially the visual representation, Figure 5.2. For a topographically ordered map (of sufficient size to avoid sharp quantisation) it is expected to observe a smooth transition from one node to its neighbours in this case, in terms of estimated probability.

This topological ordering behaviour can also be inferred from the sensitivity curves, where plateaux as evidenced in ANxNbv1's Selectivity/Sensitivity (Figure 6.1 repeated in this section) represent sharp quantisation of the topographic ordering. This sharper quantisation here is a direct result of there being fewer nodes.

Consider the graph of the A10x10bv1; as the Probability Threshold d_{th} increases, those nodes assigned that probability or greater are included in the scoring as EEv detections. It can be determined from A6x6bv1 (which has 36 nodes as opposed to the 100 or 64) that ranges of thresholds exist over which the performance is not altered. This is representative behaviour of topographically ordered maps, and in the next section discussion will include the ramifications of this quantisation.

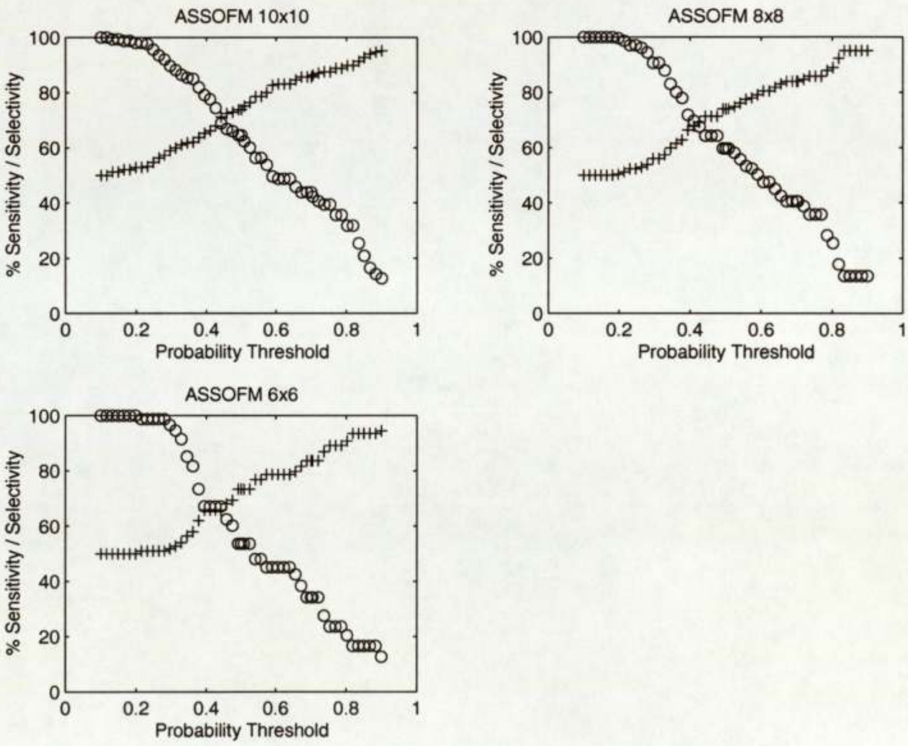

Figure 6.1: Selectivity and Sensitivity from ASSOFM for ANxNbv1 over data set Type C*. Demonstrating the degradation of the Sensitivity curve.
KEY + Selectivity o Sensitivity

An extreme example of this quantisation can be noted in the results of BNxNbv3 (Figure 5.11), where it is considered that a useful topographically ordered map has not evolved. This cannot be considered a direct product of ASSOFM size as it is evident across all three sizes considered. In this case the winning strategy, i.e. the method of determining the rotational operator, involved only a single principal component, as opposed to the results considered above for ANxNbv1. However this quantised result contrasts sharply with the initial trial, iBNxNbv3, which differs in that it has a fixed dipole data set as opposed to the random dipoles in training set Type C*, but as is evident in Figure 5.10 does not suffer this catastrophic stepping. It is considered that this winning strategy was unable to generalise given differing forms of the dipole where it had shown success over a fixed dipole.

The results of the iBNxNbv3 implementation remain valuable as they demonstrated that a topographically ordered map could evolve in a situation in which there was not a one-to-one correspondence between basis vectors and principal components. This spurred the exploration of ASSOFMs with a variety of nodes which allowed the production of useful maps. A further analysis of the modified winning strategy, not coupled with the node alteration was omitted, this implies that this strategy cannot be dismissed without further inquiry.

6.3 Classification

Previous studies (Table 6.1) have offered selectivities of less than the minimum 0.5 noted in this study. The Selectivity and Sensitivity curves differ from the stereotypical relationship. James *et alia* [9], show the classic 'X' shape; with Selectivity initially 0, and Sensitivity at 1. As the threshold increases they converge and cross, tending towards 1 and 0 respectively. This selectivity occurs at low thresholds (e.g. trivially 0) which entails classification of all CEDs (both synthetic and artefact) as EEvs. Recalling that selectivity is defined as the total of true EEvs detected, divided by the total number of detections it becomes obvious that the value of 0.5 is to be expected. This is a product of the balanced data set (i.e. 50% EEv, 50% artefact) which is unique amongst those trials considered. Presented in Table 6.1 below are the collated results of previous automated systems. Overleaf on Table 6.2 we can see selected results from this study tabulated in a simialr form

6.4 Comparison with Previous Studies

Table 6.2 offers results collated from Chapter 5 to be compared with Table 6.1. However, caution must be employed in any direct comparison. Consider the highlighted result for A15x15bv5 and the final result in Table 6.1. Whilst The latter suggests a lower error rate than A15x15bv5, whilst offering reduced Selectivity and Sensitivity this exposes the arbitrariness of the Error Rate unless it is considered across standard data.

With this caveat, the following have been selected to offer a representative overview of the ASSOFM implementations.

It is immediately obvious that there is a great range of performance in previous studies. If Error Rate is taken as the arbiter it suggest that this study is comparable with the systems exhibiting the lowest rates i.e. 0 - 117. As noted Error Rate is an arbitrary measure related to the number of artefacts happen to lie within a given hour in the data set employed by those studies. As these data sets were not available for this study (again a primary motivation for synthetic data sets, or at least standardised data sets) Error Rate cannot be considered a useful measure alone. Hence, Sensitivity and Selectivity must be considered the primary measures of classification performance.

Sys	Method	Hours	% EEv	Train /Test	Sensitivity	Selectivity	Error/ hour
1	ANN + ANN	0.043	?	6/4	0.9000	0.6900	~1023
2	Mimetic	2.0	100	Blind	0.5900	0.8900	37
3	Mimetic	33.0	100	Blind	-	0.4100	117
4	Mimetic + state	33.0	100	Blind	-	0.6700	47
5	Mimetic + ANN	0.3	100	Same	0.7400	0.7400	~804
6	Mimetic + ANN	0.3	100	Same	0.4600	0.4600	~5598
7	Mimetic + Expert System	3.0	73	Same	0.5300	1.000	0
8	Mimetic + Expert System	3.2	88	Blind	0.1400	0.8100	2
9	Mimetic + SOFM + Fuzzy Logic	3.2	88	Blind	0.5500	0.8200	7

Table 6.1: A comparison of the sensitivities, selectivities and false detection rates between spike detection systems. After James *et al* (1999). For definitions of Selectivity, Sensitivity and Error Rate see Section 4.3.1 and Section 4.3.2

Tallying the proffered results offers a rough measure of performance. This should be done with reference to the secondary measure, Error rate, else the results from low thresholds d_{th} (i.e. 0.1) which offer Selectivities of 1 and Sensitivities 0.5 would be classified as useful. The highest tally offered in the previous literature is 1.59 for System 1, though this has an Error rate of an order above those evidenced by the best results in this study. For those of a lower Error rate, for example System 7 (which uniquely offers 0 errors; though given that this study continues the search for an optimal EEv classification system we must suggest that this system has not yet been implemented for novel data), a tally of 1.53 measured across the training data is noted. The equivalent systems developed here, with results across their training sets offer 1.58 in the case of A15x15bv5 to 0.90 for iA15x15bv5, with both offering Error rates of ~ 125.

A more useful test of the systems is across novel data, i.e. Blind tests. Consider

System	Method	d_{th}	Hours	Train /Test	Sensitivity	Selectivity	False/ hour
iA15x15bv5		0.70	4.48	Blind	0.4660	0.8622	37
iA15x15bv5	Mimetic	0.70	1.60	Same	0.1374	0.7434	104
iA6x6bv5	+ASSOFM	0.70	4.48	Blind	0.3830	0.8055	41
A15x15bv5		0.10	4.48	Same	0.9935	0.5605	348
A15x15bv5		0.50	4.48	Same	0.8375	0.7525	122
A15x15bv5	#	0.70	4.48	Same	0.5910	0.8371	51
A15x15bv5		0.85	4.48	Same	0.1790	0.9344	6
A8x8bv5		0.50	4.48	Same	0.7720	0.7293	128
A8x8bv4		0.70	4.48	Same	0.5015	0.7182	62
A10x10bv3		0.70	4.48	Same	0.4735	0.7965	63
A10x10bv2		0.70	4.48	Same	0.3035	0.7822	38
A6x6bv1		0.70	4.48	Same	0.3425	0.8374	30

Table 6.2: Collated ASSOFM performance measures offered for comparison with previous studies in Table 6.1. Training set Type C* was composed of three patient EEGs. These were recorded for 2 minutes per hour for 48 hours, resulting in 4 hours and 48 minutes of EEG which spanned the full gamut of human, and hence artefact activity. 2144 artefacts were extracted from this, of which a random subset of 2000 were employed, giving rise to an adjusted time 4 hours 28 minutes. The Fixed* training set employs 3500 artefacts, extracted from a single patient over a single 48 hour period offering 1 hour 36 minutes. The error rate presented here may be obtained by dividing the values presented previously by 4.48 for Type C*, or 1.6 for Fixed*. The artefacts were matched with a similar number of synthetic EEvs giving rise to an % EEv of 50. Applying Type C* to the initial trials (iA and iB) results in a Blind trial, applying this set to the secondary implementations results in a trial noted as Same (A and B).

system 2 and 9, both that exhibit desirably low Error rates, and a tallies of 1.48 and 1.35 in turn. ASSOFMs iA15x15bv5 and iA6x6bv5 considered here, offer commensurate Error results, with comparable tallies of 1.32 and 1.18.

If it is considered that a high Selectivity is most desirable i.e. false diagnosis must be minimised, or that Epileptiform Spikes are prolific in Epileptic individuals hence allowing a low Sensitivity to still be useful in diagnosis. System 8, which offers the desirable Error Rate of 2 can be construed as an example of such a system. Such a system can be constructed easily by raising the threshold above which EEvs are classified in the ASSOFM's considered here. ASSOFM A15x15bv5, which is shown at a range of thresholds highlighted by a # in Table 6.2, shows a directly comparable result to System 8 at threshold of 0.85.

Chapter 7

Conclusions

7.1 Evaluation of the ASSOFM System

It is apparent from the results collated in Chapter 5, and discussed in Chapter 6, that the system involving the novel variant ASSOFM (Method A) has the ability to form clusters in a self organising fashion.

There is an awareness however, that clustering algorithms, when presented with random data will exhibit clustering behaviour, and that clustering itself is not an arbiter of utility. The converged ASSOFMs were further examined to give rise to a classifier, calibrated with an appeal to Bayesian methods (after James [9]). It is noted that this calibration reveals a topological ordering of these clusters.

These estimated probabilities were used to extract measure of performance over the training set and novel data. These measures show parity with previous research in the field of automated EEG analysis. As such, it is considered that the clusters formed by the ASSOFM can be interpreted in such a way as to form a classifier, given the assumption that the synthetic data is representative of clinical data.

Hence, it is considered that the subspaces employed have an ability to model artefacts and synthetic EEvs. As a consequence the Metric (defined in Section 3.4.1) to determine the distance from a candidate node subspace from the subspace in which a data point was embedded is considered to be a valid measure of *difference* between these two classes. Further the Rotational Operator (defined in Section 3.4.1) is thus an appropriate tool in the modification of a given subspace's basis vectors.

However, consideration must be given to the nature of the data set employed in this study. Whilst it is established that the system presented can classify to a degree commensurate with systems offered elsewhere, the lack of examination over novel, clinical data means the results can only be considered for proof of the system as a classifier over this data set. Any comparison between these systems is dependent on the representative nature of this synthetic data. It may be suggested that previous results also suffer from data set specific results, consider especially the arbitrary nature of the Errors per hour over differing patients. In the absence of a verification of the accuracy of the EEv model employed, either by the processing of novel clinical data or further study of the model, this study can only suggest the potential utility of this ASSOFM system.

This utility may be exposed with a consideration of the results of the preliminary examination of the ASSOFM algorithm. It may be noted that iANxNbv$M$ was trained and calibrated using a fixed dipole, i.e. defined by a fixed eccentricity f, ϕ, θ and by moment Dx, Dy and Dz. Measures of performance were obtained over this data set, and over a data set composed of random dipoles. The results tabulated in Tables 5.2, 5.3 and 5.4 demonstrate the system's ability to model novel data out-with that considered in the training set in line with the performance inferred by the training set. This, it is suggested, is evidence of the system's ability to model an underlying generator (the jittered synthetic dipole) which has undergone different transformations.

This demonstration of invariant classification is an evaluation of the total system. Thus, it may be held that the Principal Component Analysis employed in dimensionality reduction was vital in presenting the clustering stage with data in which underlying generators may have been rendered prominent, and thus rendered trivial the differing clustering of artifact and synthetic data. Thus, without further examination, it cannot be concluded that the ASSOFM subspace nodes confer this behaviour on the system, even given that this behaviour in the Learning Subspace Method of Kohonen directly inspired their use. Without results for PCA across genuine EEvs, it cannot be asserted that the application of PCA was not taking advantage of the *apriori* knowledge that the synthetic EEvs had a low underlying dimensionality. This is a re -expression of the caveat that the system is proved only for the data set considered. However, even given that this may not be a distinguishing feature of true EEvs and TNSs, it can be said that the clustering algorithm ordered elements which had undergone differing transformations, the synthetic EEvs, similarly.

7.2 Optimal ASSOFM Parameters

It is apparent that performance, measured as Error Rate (which between results in this study is an accurate measure of relative performance over the same data) improves marginally between 6x6, to 8x8, to 10x10 and 15x15 maps. The performance of the differing numbers of basis vectors spanning the subspaces, again considering Error Rate as an arbiter shows that for a given d_{th} that two basis vectors is optimal, with a single subspace offering 50% larger error rates and 3,4 and 5 offering at least 100% greater. The Selectivity remains similar across all ASSOFM sizes and basis vectors, the improved Error Rate being accounted for by a reduction in Sensitivity, i.e. proportionally less of both artefacts and synthetic EEvs are classified as EEvs. This is considered a degradation in performance, as such it is concluded that 5 basis vectors, i.e. one to one correspondence with the data, offers the best overall performance.

7.3 Expanding the Study

The desirability of synthetic data, as discussed previously, must be contrasted with the fact that the system cannot be realistically evaluated without clinically labelled data. Whilst an accurate, widely accepted, synthetic model could improve the comparability of systems by creating a standard method of a implementing a data set, the proof of utility arises only with real data. As such, the first suggested course is the application of the presented system to real data, this would offer a true measure of performance.

This would not answer the questions about the influence of the pre-processing i.e.

PCA, on the systems performance. A comparable study using this study's mimetic and PCA stages, but employing the basic version SOFM, would potentially reveal any performance gains in employing the ASSOFM.

It is noted that the SOFM and in turn the variant ASSOFM are heuristically inspired, albeit useful, algorithms. This suggest that the parameters involved are arbitrary, and given the novel nature of the ASSOFM, the SOFM heuristics which were employed need to be re-examined, or a principled alternative examined. EXPAND!!

The Generative Topographic Mapping is presented as a principled alternative to the SOFM. Considering that the aim of this study is to develop a system element which can surpass the SOFM in System 9, in Table 6.1, an examination of this alternative is desirable given its claim to replicate the results of the SOFM, but without the heuristic nature of parameters such as network size, or in the ASSOFM number of basis vectors. It is suggested that future work consider such a principled TPM.

7.4 Closing Comments

In synopsis; a system is presented that, given accurate synthetic modelling of EEv, can be compared favourably to present systems extant in the literature. This system employs a novel implementation of the SOFM that by modelling classes with adaptive subspaces, coupled with PCA, confers invariant classification to transformations in the data. It is considered that the ASSOFM element of the system will bear further examination, especially in contrast to principled TPMs such as the GTM.

Bibliography

[1] K. Kobayashi, C.J. James, T. Nakahori, T. Akiyama, and J. Gotman. Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. To appear in Electroenceph. And Clin. Neurophys.

[2] C.J. James, R.D. Jones, B.J. Bones, and G.J. Caroll. Detection of epileptiform discharges in the eeg by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic states. *Clin. Neurophys.*, 110:2049–2063, 1999.

[3] H. Petsche. Topography of the EEG: survey and porspects. *Clin. Neurol. Neuroseurg.*, 79:15–28, 1976.

[4] A.A. Dingle, R.D. Jones, G.J. Carroll, and W.R. Fright. A multistage system to detect epileptiform activity in the EEG. *IEEE Trans. Biomed. Eng.*, BME-40:1260–1268, 1993.

[5] G.E. Chatrian, L. Bergamini, M. Dondey, D.W. Klass, I. Peterson, and M. Lennox-Buchtall. A glossary of terms most commonly used by clinical electroencephalographers. *Electroencephalography and Clinical Neurophysiology*, 37:538–548, 1974.

[6] A.J. Gabor and M. Seyal. Automated interictal EEG spike detection using artificial neural networks. *Electroenceph. Clin. Neurophysiol.*, 83:271–280, 1992.

[7] C.J. James, M.T. Hagan, P.J. Jones, R.D. Bones, and G.J. Carroll. Multireference adaptive noise canceling applied to the EEG. *IEEE Trans. on Biomed. Eng*, 44:775–779, 8 1997.

[8] J.R. Glover, Jr., N. Raghavan, P.Y. Ktonas, and J.D. Frost, Jr. Context-based automated detection of epileptogenic sharp transients in the EEG: Elimination of false positives. *IEEE Trans. Biomed. Eng.*, BME-36:519–527, 1989.

[9] C.J. James. *Detection of epileptiform activity in the electroencephalogram using artificial neural networks.* PhD thesis, University of Canterbury, Christchurch, New Zealand, 1997.

[10] T. Kohonen. *Self-Organizing Maps.* Springer-Verlag, Berlin, Heidelberg, Germany, 1995.

[11] R.H.S. Carpenter. *Neurophysiology.* Edward Arnold, London, 1990.

[12] H.H. Jasper. Report of the committee of methods of clinical examination in electroencephalography. *EEG Journal*, 10:370, 1958.

[13] M.R. Nuwer, G. Comi, A. Emerson, R. Fuglsang-Frederiksen, J.M. Gurit, H. Hinrichs, A. Ikeda, and R. Luccas, F.J.C. Rappelsburger. IFCN standards for digital recording of clinical EEG. *Electroenceph. Clin. Neurophys.*, 106:259–261, 1998.

[14] P.L. Nunez and R.D. Katznelson. *Electric Fielda of the Brain : The Neurophysics of EEG.* Oxford University Press, New York, 1981.

[15] C.M. Bishop, M. Svensen, and C.K.I. Williams. GTM: A principled alternative to the self organising map. *Advances in Neural Info. Proc. Sys*, 9, 1997.

[16] P.L. Nunez and R.D. Katznelson. Electric fielda of the brain : The neurophysics of eeg. *IEEE Trans. On Comp.*, C-18(5):401–409, 1997.

[17] N.C. Noel. Dynamical embedding and feature extraction of electroencephalographic data. Master's thesis, Aston University, Birmingham - United Kingdom, 1998.

[18] E. Oja. *Subspace Methods of Pattern Recognition.* Research Studies Press, Letchworth Hertfordshire England, 1983.

[19] D.S. Watkins. Understanding the QR algorithm. *IEEE Trans. On Comp.*, 24(4):427–440, October 1997.

Appendix A

Appendix - Selected MATLAB Functions

A.1 Subspace Distance Function

MATLAB 5.2 Function for determining the distance between the data, and nodes subspace (*cf.* Section 3.4.1.

```
function [subspace_delta] = subspace_distance_function(X,U)
%subspace distance function
%
%Dagmar Scott Fraser May 1999
%
%    After Oja (1983) Subspace Methods of Pattern Recognition
%    & Watkins (1982) Understanding the QR Algorithm
%
% Define a metric for subspace distance [difference].
%
% X - input data matrix - assumes ||X|| = 1.
%
% U - ASSOFM node subspace basis vectors -
```

```
% must be orthonormal.  Though a treatment
% exists in Oja for a non-orthogonal basis.
%
% Construct the projection matrix P from the candidate
% subspace L_assofm, spanned by U.
%
%     p
% P = E u_i u'_i . (2.8 Oja)
%    i=1
%
% where U is the orthonormal basis {u_1, ... u_p} of L_assofm where
% p is the no. of linearly independent vectors spanning the subspace,
% where u_i is a vector of dimension (in this case) 41.
%
%More simply P = UU', the projection matrix of X onto subspace L_assofm
%


P = U * U'; %Projection matrix for L_assofm
I = eye(size(P));
argument = I - P;
% Determine distance delta
subspace_delta = max( max( sqrtm( X' * argument * X )));
%end file
```

A.2 Rotation Operator

MATLAB 5.2 Function for determining the Rotation Operator (*cf.* Section 3.4.1.

```
function [rotation_operator] = subspace_rotation_operator(alpha,X)
%subspace rotation operator
```

```
%
%Dagmar Scott Fraser May 1999
%
%     After Oja (1983) Subspace Methods of Pattern Recognition
%     & Kohonen (1995) Self Organising Maps
%
%     alpha - 0 < alpha < 1, learning parameter of the rotation operator.
%
%     X - input matrix, assumes that ||X|| = 1.
%
%     R_d = I + alpha XX' / ||X||^2 --[ denominator ignored as ||X||==1]

argument = X * X';
[m,n] = size(argument);
rotation_operator = zeros([m,n]);
rotation_operator = eye([m,n]) +  ( alpha * argument );
%end file
```