ELSEVIER

Clinical Neurophysiology 110 (1999) 2049–2063

www.elsevier.com/locate/clinph

Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages

Christopher J. James[a,*], Richard D. Jones[b,c], Philip J. Bones[a], Grant J. Carroll[d]

[a]*Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand*
[b]*Department of Medical Physics and Bioengineering, Christchurch Hospital, Christchurch, New Zealand*
[c]*Department of Medicine, Christchurch School of Medicine, Christchurch, New Zealand*
[d]*Department of Neurology, Christchurch Hospital, Christchurch, New Zealand*

Accepted 2 June 1999

Abstract

Objective: A multi-stage system for automated detection of epileptiform activity in the EEG has been developed and tested on pre-recorded data from 43 patients.

Methods: The system is centred on the use of an artificial neural network, known as the self-organising feature map (SOFM), as a novel pattern classifier. The role of the SOFM is to assign a probability value to incoming candidate epileptiform discharges (on a single channel basis). The multi-stage detection system consists of three major stages: mimetic, SOFM, and fuzzy logic. Fuzzy logic is introduced in order to incorporate spatial contextual information in the detection process. Through fuzzy logic it has been possible to develop an approximate model of the spatial reasoning performed by the electroencephalographer.

Results: The system was trained on 35 epileptiform EEGs containing over 3000 epileptiform events and tested on a different set of eight EEGs containing 190 epileptiform events (including one normal EEG). Results show that the system has a sensitivity of 55.3% and a selectivity of 82% with a false detection rate of just over seven per hour.

Conclusions: Based on these initial results the overall performance is favourable when compared with other leading systems in the literature. This encourages us to further test the system on a larger population base with the ultimate aim of introducing it into routine clinical use.

Keywords: Epileptiform discharges; Spikes; Artificial neural networks; Self-organized learning; Fuzzy logic; Bayesian

1. Introduction

The spike detection problem can be simply put: detect the presence of epileptiform discharges (EDs) in the multichannel EEG recording with high sensitivity and selectivity. That is, a high proportion of true events must be detected with a minimum number of false detections. Although desirable, it is not realistic to expect sensitivities and selectivities of 100%–if for no other reason than the imprecise definition of a spike. The lack of a proper definition of the 'ideal' spike, other than 'transients clearly distinguished from background activity with pointed peaks at conventional paper speeds and a duration from 20 to under 70 ms approximately' (for sharp-waves the duration is of 70–200 ms) (Chatrian et al., 1974), has rightly caused many researchers to ask: *What does the EEGer look for when visually detecting EDs in the EEG?*

Many researchers have attempted to answer this question by extracting features from the raw EEG which, in their opinion, best describe the ED morphology–i.e. mimetic approaches. Others have opted to use artificial neural networks (ANNs) as a means of utilising the raw EEG without having to make any decision as to what parameters are more important than others in detecting EDs are. Özdamar et al. (1991) state that the use of pre-processing to extract parameters biases the system and defeats the very purpose of a totally trainable system when utilising ANNs. Conversely, Webber et al. (1994) report better results, in terms of accuracy and speed of spike detection, through the use of

* Corresponding author. Neural Computing Research Group Aston University, Aston Triangle, Birmingham B4 7ET, UK. Tel.: +44-121-359-3611, ext: 4652; fax: +44-121-333-6215.
E-mail address: jamescj@aston.ac.uk (C.J. James)

1388-2457/99/$ - see front matter
CLINPH 99552
PII: S1388-2457(99)00168-6

parameterized EEG as opposed to raw EEG but then go on to suggest that the network may need more raw test data to abstract identifying features from EDs.

In mimetic approaches, values for peak amplitude, pre-peak slope and post-peak slope, durations, 2nd derivatives, etc., are extracted (such as Gotman and Gloor (1976); Ktonas et al. (1981); Ktonas et al. (1984)). In the parametric approach, the sharpness of the ED is used in a statistical setting where the transient or non-stationary nature of the ED is used as a criterion for detection when compared to the (assumed) stationary background (Lopes da Silva et al., 1977). The sharpness of the spike, compared to the background, should also result in a differing spectral content, with more energy being found in the 'higher' frequencies. Wavelet analysis can be utilised as a means of providing time-frequency signal analysis capabilities (Schiff et al., 1994; Kalayci and Özdamar, 1995).

It is well established that, apart from the ED itself, other contextual information is also vital to the EEGer when classifying events as ED/non-ED. These are mainly spatial information, such as *What is happening in other channels at the same time as a candidate ED?* and wide-temporal information such as *Are there similar events with similar distribution elsewhere in the EEG?* Although it is well known that EEGers use spatial information in the process of identifying epileptiform events (EVs)[1] just 'what' form the spatial cues take is not immediately identifiable. It is, therefore, somewhat surprising that most of the spike detection systems reviewed work on a channel-by-channel basis (i.e. no spatial or wide-temporal information is utilised). However, Glover et al. (1989); Dingle et al. (1993) have made use of both spatial and wide-temporal contextual information, with a high degree of success, through their use of expert systems. Özdamar et al. (1991) make use of spatial information by integrating the outputs of individual channel spike detection ANNs (from four channels) into a single ANN module trained to recognise the common spatial distributions of EDs. Webber et al. (1994) also use four channels simultaneously, including temporal contextual information of a 1.0 s long window around the ED, in the training of their ANN.

Since their inception, ANNs have been ideal candidates for classifiers, especially when little is known about the underlying statistics of the input data and there is a need to generalize to novel data. The ANN has, as mentioned above, already featured in various spike detection systems (Özdamar et al., 1991; Gabor and Seyal, 1992; Jando et al., 1993; Webber et al., 1994; Kalayci and Özdamar, 1995). However, in most cases, the ANN used has been a multi-layer feed-forward network (such as the multi-layer perceptron) trained in a supervised manner–i.e. by repetitively applying known classes of EEG segments to the ANN until it is considered as being adequately trained (Webber et al., 1994). This applies to systems that use raw EEG data as input as well as features extracted from the raw EEG. The action performed by these ANNs could be termed supervised classification learning. The drawback is the requirement of a large training set of EDs that have been labelled into classes a priori. This is not an easy task as there is often disagreement, even amongst expert EEGers, on whether candidate EVs are indeed epileptiform–i.e. lack of a gold standard (Hostetler et al., 1992; Wilson et al., 1996; Black et al., 1997; Black and Jones, 1998).

There is, however, a form of unsupervised classification learning where no *a priori* knowledge is required regarding an input's membership in a particular class. Rather, gradually detected characteristics and a history of training are used to assist the network in defining classes and possible boundaries between them. Such unsupervised classification is called clustering and the ANN is known as a self-organizing ANN. As there is no information available from the teacher on the desired classifier's responses, the similarity of incoming patterns is used as the criterion for clustering. Each cluster contains patterns of similar features.

A popular self-organising ANN is Kohonen's self-organising feature map (SOFM) (Kohonen, 1990; Kohonen, 1995). The SOFM has many applications, mainly in pattern recognition, many of which are described in Kohonen (1995). The SOFM is a single layered ANN whose weight vectors are adjusted in an unsupervised manner as inputs are presented randomly from a training set. Once training is complete, the spatial location of a neuron in the network then corresponds to a particular feature, or group of features, in the input patterns. The results achieved seem very natural, indicating that the adaptive processes at work in the SOFM may be similar to those at work in the brain. In terms of parameters, the SOFM is quite robust and through a small set of (empirically obtained) parameters the training is controlled well. A further refinement to the SOFM by Kohonen is learning vector quantization (LVQ) (Kohonen, 1990). Kohonen implements LVQ as a means of 'fine-tuning' the SOFM when it is to be used as a pattern classifier. LVQ is, however, a supervised learning technique and is intended to follow the implementation of SOFM.

Gabor et al. (1996); Gabor (1998) have applied the SOFM to the seizure detection problem.

This paper presents a novel approach to the use of the SOFM ANN in the spike detection problem. The multi-stage system relies on two fundamental issues: (1) A set of 'representative' ED/non-ED waveforms is obtained through a self-organized training process using 'raw' EEG waveforms (i.e. not parameterized ED waveforms); (2) considerable use is made of spatial cues (and limited temporal cues) in the multichannel EEG recording during the spike detection

[1] As some confusion arises from the use of the terms spike and epileptiform discharge (ED) in the literature, the convention adopted throughout this paper is to use spike and ED interchangeably to refer to epileptiform activity on a *single channel* and to refer to activity which is in evidence across *two or more channels* as an epileptiform event (EV).

Table 1
The protocols used in the routine clinical recordings of the EEG used for training and testing, including OVER5, UNDER5 and BAB

Run	Montage	Duration (s, approx.)
Adult protocol (OVER5)		
1	Longitudinal	300
2	Transverse	100
3	Long-transverse	100
4	Circumferential	100
5	Longitudinal	300
6	Longitudinal–(hyperventilation)	300
7	Longitudinal–(photic stimulation)	100
Child protocol (UNDER5)		
1	Longitudinal	200
2	Transverse	100
3	Long-transverse	100
4	Circumferential	100
5	Longitudinal	200
6	Longitudinal–(photic stimulation)	60
BABY protocol		
1	Longitudinal	500
2	Longitudinal–(photic stimulation)	60

process. These measures result in a system with a minimal false detection rate. The multi-stage system is described, stage by stage, in the next section following which the performance of the system is assessed on a 'test set' of EEGs.

2. Materials and methods

2.1. EEG recordings

The EEG was recorded by scalp electrodes placed according to the International 10-20 system (Jasper, 1958). Sixteen channels were recorded from bipolar montages. The amplified EEG was bandpass filtered between 0.5 and 70 Hz using a five-pole analogue Butterworth filter, sampled at 200 Hz and digitized to 12 bits. All data were stored for later off-line processing.

Four bipolar montages were used: longitudinal, transverse, longitudinal-transverse and circumferential. All recordings were made while the patient was awake but resting and included periods of eyes-open, eyes-closed, and extended periods of photic stimulation and hyperventilation. Three recording protocols were used, depending on the age of patients: (a) OVER5, (b) UNDER5 and (c) BABY. The details of each protocol are described in Table 1.

2.2. Overall system

The spike detection system consists of three stages (Fig. 1). The mimetic stage extracts and thresholds parameters of the EEG and presents raw candidate epileptiform discharges (CEDs) to the second stage, a trained SOFM. The SOFM has been previously trained on a large training set in a self-organized fashion and results in an ordered set of weight vectors based on the features extracted from the training data. The trained SOFM then becomes a classifier, assigning class labels to inputs based on the information held in the weight vectors of the SOFM. Both of these stages form a single-channel ED detector and are repeated (in sub-sets of 4 channels) for each channel in the multichannel EEG. The final stage incorporates the multichannel outputs of the previous stages (i.e. spatial information) to give the final EV/non-EV output of the spike detection system. The spatial combination of each of the individual single channel outputs is performed by means of a fuzzy-logic rule-based system. Through the use of fuzzy-logic it becomes possible to model the logic of an EEGer when performing spatial analysis. Each of the system's three stages is described in detail in the following sections.

The system has been developed using the MATLAB (Ver 4.2c1) package (The Math Works Inc.) with the Neural Network toolbox (Ver 2.0b) and the Signal Processing toolbox (Ver 3.0b). For testing purposes the final system was further developed using the Microsoft Visual ''C++'' (Ver 4.2) programming language. The tests were performed on a PC system with a Pentium processor running at 200 MHz. Spike-detection was performed 'off-line' on data stored on hard disk.

2.2.1. Stage 1: mimetic

The purpose of the mimetic stage is to screen the incoming EEG for CEDs, for subsequent presentation to the SOFM. Thus this stage is primarily for data reduction and, hence, it is imperative that it detects as many EDs as possible whilst rejecting most 'obvious' non-ED waveforms. The design is based on the mimetic systems of Gotman et al. (1978); Dingle et al. (1993).

All channels are scanned for a positive or negative vertex

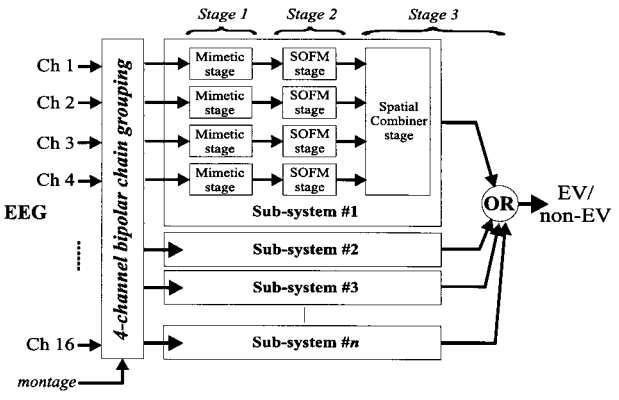

Fig. 1. A block diagram of the multi-stage spike detection system depicting each subsystem, each comprising (i) a mimetic stage, (ii) a SOFM stage and (iii) a fuzzy logic stage.

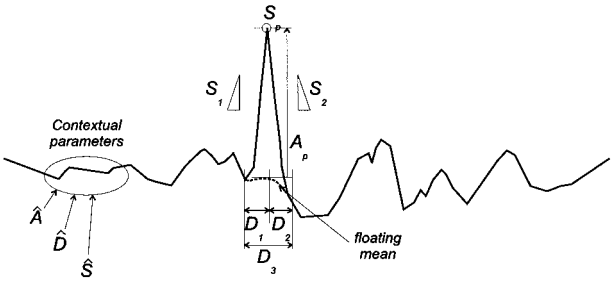

Fig. 2. The parameters extracted from around a vertex (on a single channel) by the mimetic stage that will be thresholded in order to obtain a CED that is then presented to the next (SOFM) stage. See text for definition of parameters.

using a simple peak detection algorithm. Once a vertex is found, a CED is defined as the 200 ms wave surrounding the vertex and, from this, a number of parameters are extracted (Fig. 2). We used parameters considered the most important when distinguishing EDs from non-EDs (Gotman and Gloor, 1976; Gotman et al., 1978; Gotman, 1980; Koffler and Gotman, 1985; Dingle et al., 1993; Webber et al., 1994). These are:

- amplitude A_p: The difference between the peak value and the floating mean, where the floating mean is the average value of the EEG over 75 ms centred on the peak;
- durations D_1, D_2 and D_3: D_1 and D_2 represent the durations of each half-wave before and after the vertex. D_3 represents the sum of D_1 and D_2. The half-wave durations are measured from the vertex to the point where there is more than a 60% drop in slope or a change in direction of the slope;
- slopes S_1 and S_2: Slopes are measured before and after the vertex. For waveforms of short durations (i.e. $D_1 < 20$ ms or $D_2 < 20$ ms) the peak-to-peak slope is calculated, otherwise a least-squared estimate is obtained based on four samples (excluding the peak sample);
- sharpness S_p: The sum of the magnitudes of the pre-vertex and post-vertex slopes.

The parameters are passed through a set of thresholds and a waveform that exceeds all thresholds is flagged as *passed thresholds*. The threshold values are based on values determined by discriminant analysis by Dingle (1992) and are: $A_P^{Min} = 16.8$ μV, D_1^{Min} $(= D_2^{Min}) = 10)$ ms, D_1^{Max} $(= D_2^{Max}) = 150$ ms, $D_3^{Min} = 20$ ms, $D_3^{Max} = 250$ ms, $S_P^{Min} = 1.26$ μV/ms.

The mimetic stage acts on all channels independently. Once a waveform is found which exceeds all thresholds the vertex of that waveform is called the *primary vertex*. At that point all CEDs on the remaining channels (within 50 ms of the primary vertex), *whether or not they exceed the thresholds*, are grouped so that together they make up a candidate epileptiform event (CEV) which is passed on to the next stage.

In keeping with the definition that an ED must be clearly distinguishable from background EEG, contextual parameters are extracted from the 1.0 s segment of EEG about which the vertex of each CED is centred (excluding the 200 ms of EEG describing the CED itself). The parameters are:

- average amplitude $\hat{A}$: The RMS difference between the actual EEG and a floating mean calculated over a 15 sample (75 ms) window;
- average duration $\hat{D}$: The average peak-to-peak duration of the half-waves (half-waves with a peak amplitude of less than 4.2 μV are ignored);
- average slope $\hat{S}$: The average magnitude of the slope between consecutive samples.

Once the primary vertex has been found and the contextual information extracted for each CED, the following information is put forward to the following SOFM stage for each channel:

1. CED: The CED for each channel is made up of a window of 'raw' EEG. As EDs are said to vary in duration from 70–200 ms, the ideal inputs to the system would consist of a window of at least 200 ms of 'raw' EEG (40 samples at 200 Hz). An important characteristic of an ED is the slow wave that often follows the spike. For a 200 ms waveform, the vertex occurs roughly during the first third and the slow wave (if present) in the remaining two thirds. Thus, a 41 sample (205 ms) window is used, such that the maximum vertex is placed at the 14th sample (at approximately one third the way across);
2. contextual information: The three contextual parameters $(\hat{A},\hat{D},\hat{S})$ for each CED;
3. flag: A flag is set for a CED which passes the thresholds and reset otherwise.

2.2.2. Stage 2: self-organizing feature map

This stage consists of Kohonen's SOFM as described earlier. The SOFM is a lattice type ANN (single layer) whose neurons become specifically tuned to various input signals through an unsupervised learning process. We adopted a square two-dimensional SOFM of order N. The synaptic weights of the SOFM are changed through a learning process known as Kohonen's learning rule which is described briefly below (for a more detailed description see Kohonen (1990)).

Consider the square two-dimensional array of neurons of Fig. 3. All the inputs are connected in parallel to all neurons i in the network (for $i = 1, 2, \ldots, N^2$). Each neuron has a corresponding weight vector given by $\mathbf{m}_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T$ (where n is the width of the input vector). The weight vectors $\mathbf{m}_i$ are initially given small random values. An input $\mathbf{x}$ ($\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$) is presented to the array and the Euclidean distance is calculated between the input $\mathbf{x}$ and each $\mathbf{m}_i$, such that:

$$|\mathbf{x} - \mathbf{m}_c| = \min_i \{|\mathbf{x} - \mathbf{m}_i|\}$$

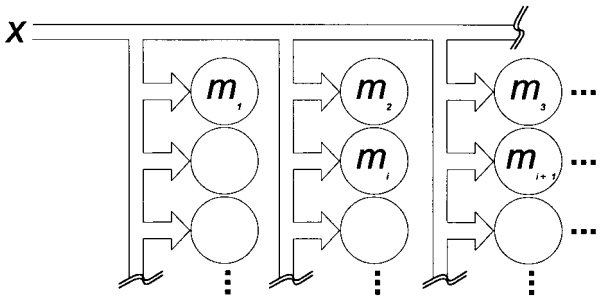

Fig. 3. The two-dimensional array of neurons in the SOFMs. All neurons are connected to the input, x, via weight vectors $\mathbf{m}_i$.

where $\mathbf{m}_c$ defines the 'winning' weight vector (i.e. that weight vector with the smallest Euclidean distance) and c is the index of the winning neuron. A neighbourhood set N_c around neuron c is then defined and all the weight vectors of the neurons in the neighbourhood N_c are updated and all others are left unchanged. The weight vectors of the SOFM are updated as follows:

$$\mathbf{m}_i(t+1) = \mathbf{m}_\mathrm{i}(t) + \alpha.(t)\left[\mathbf{x}(t) - \mathbf{m}_\mathrm{i}(t)\right] \text{ if} i \in N_\mathrm{c}(t)$$

$$\mathbf{m}_i(t+1) = \mathbf{m}_\mathrm{i}(t) \text{ if } i \notin N_\mathrm{c}(t)$$

where $\alpha(t)$ is a scalar valued learning rate, $0 \leq \alpha \leq 1$. Both the learning rate and the neighbourhood function must be gradually decreased with time. We opted for the exponential decays given by

$$\alpha(t) = \alpha_0 e^{\left(\frac{-t}{\tau_\alpha}\right)}$$

and

$$N(t) = N_0 e^{\left(\frac{-t}{\tau_\mathrm{N}}\right)}$$

although the exact form of decay proved to be not critical for training of the SOFM (James, 1997).

Once the above steps have been taken, a new input $\mathbf{x}$ is chosen randomly from the training set and presented to the SOFM and the entire process repeated. In this way the weight vectors become representative of the underlying statistics of the input vectors, the weight vectors also form topographic maps of the inputs across the (2D) lattice where similar inputs 'excite' neurons in similar regions of the SOFM. The number of iterations and the initial and final values of the learning rate, neighbourhood size and other parameters can be obtained using rules of thumb developed by Kohonen (1990). The parameters used for the SOFM shown here were obtained from extensive study of the SOFM by James (1997) and are: SOFM size ($S \times S$) $S = 20$, initial learning rate $\alpha_\mathrm{o} = 1.00$, final learning rate $\alpha_\mathrm{min} = 0.01$, initial neighbourhood size $N_\mathrm{o} = S - 1$, final neighbourhood size $N_\mathrm{min} = 1$, exponential learning rate decay, and exponential neighbourhood size decay.

The input to the SOFM is a vector made up of the 'raw' CED, contextual parameters and the passed/failed thresholds flag output by the mimetic stage. The trained SOFM is required to output a value between 0 and 1 which represents the probability of there being an ED on that channel. The SOFM is duplicated for all channels of the EEG such that the overall output of the SOFM stage is a set of 16 probabilities. Up until this point there is no interaction between channels.

Once the $\mathbf{m}_\mathrm{i}$ have become organized following training, a step known as calibration is needed. Calibration of the SOFM is a standard procedure developed by Kohonen which assigns a class label to each neuron in the SOFM (and hence to each weight vector $\mathbf{m}_\mathrm{i}$). The calibration process involves a majority voting scheme where the label assigned to each $\mathbf{m}_\mathrm{i}$ is based on the largest number of 'wins' by a particular neuron for a given class of input, when presented with the calibration data set.

It is possible to further enhance the performance of the trained and calibrated SOFM as a classifier by 'fine-tuning' the $\mathbf{m}_\mathrm{i}$ using a (supervised) learning algorithm known as Learning Vector Quantization (LVQ2) as developed by Kohonen (1990). With LVQ2 training, data (CEDs) with known class labels are used to further adjust the weights of the SOFM such that strong probabilities to true EDs are strengthened. The LVQ2 algorithm works by 'pulling' the weight vectors away from the decision surfaces to demarcate the class borders more accurately and is described in more detail in Kohonen (1990).

2.2.2.1. Preparation of EEG training data for the SOFM. The EEGs of 35 patients were obtained. All EEGs had been previously seen independently by at least two EEGers (in some cases three) and been graded as containing definite epileptiform events. The average EEG length was 24.4 min and the ages of the patients varied from 7 months to 71 years (average 19 years). Table 2 shows the characteristics of the data, which totals over 14 h of 16-channel EEG. The EEGers identified more than 2585 definite EVs and 511 questionable EVs.

The data also included a variety of background activities (e.g. alpha, delta, etc.) and most EEGs contained significant amounts of artifact, particularly eye-blinks, electrode movement, and muscle artifact. Artifacts were especially prominent during the periods of hyperventilation and photic stimulation. All of the data recorded was for routine clinical use and no segments of EEG were rejected because of excessive artifact or 'noisy' background activity.

The data was then passed through the mimetic stage and a large number of corresponding CEDs were collected. Considerably more CEDs failed the thresholds than passed (75 vs. 25%, respectively). Assuming that the values of the thresholds were optimally set such that *no* EVs were missed, then the CEDs that passed the thresholds contained the true EDs as well as many more false EDs. This means that of the large number of CEDs available for training the SOFM only a small minority had the potential of being true ED waveforms. If this was the case, only a small portion of the SOFM

Table 2
The training set comprising 35 EEGs with in excess of 2585 definite EVs and 511 questionable EVs[a]

Patient	Age	Duration	EEGers	Epileptiform events		
				Definite	Questionable	Total
1	9	27 m 7 s	2	102	11	113
2	30	25 m 2 s	2	>200	0	>200
3	49	25 m 21 s	2	1	7	8
4	2	15 m 42 s	2	>200	0	>200
5	4	15 m 26 s	2	49	0	49
6	58	25 m 46 s	2	0	4	4
7	16	26 m 27 s	2	57	41	98
8	20	33 m 20 s	2	1	4	5
9	67	27 m 50 s	2	>200	0	>200
10	6	26 m 8 s	2	2	7	9
11	5	26 m 24 s	2	201	0	201
12	25	26 m 34 s	2	0	9	9
13	6	28 m 40 s	2	2	0	2
14	11	34 m 2 s	2	230	0	230
15	7 m[b]	14 m 8 s	3	0	17	17
16	15	29 m 12 s	2	246	24	270
17	15	22 m 24 s	2	28	159	187
18	7	21 m 4 s	2	3	12	15
19	7	20 m 25 s	2	>200	0	>200
20	18	25 m 51 s	2	52	3	55
21	13	23 m 52 s	2	>200	0	>200
22	29	27 m 10 s	2	0	10	10
23	9	23 m 52 s	2	0	11	11
24	16	25 m 57 s	2	1	4	5
25	46	26 m 15 s	2	333	4	337
26	17	25 m 42 s	2	95	0	95
27*	71	18 m 49 s	2	17	40	57
28*	12	31 m 25 s	2	8	5	13
29*	8	26 m 43 s	3	3	32	35
30*	7	25 m 11 s	2	7	41	48
31*	4	14 m 58 s	2	12	17	29
32*	10 m	12 m 12 s	2	13	32	45
33*	32	24 m 55 s	2	6	4	10
34*	12	26 m 19 s	3	61	9	70
35*	17	25 m 26 s	2	55	4	59
Totals	~19	14 h 15 m 39 s		>2585	511	>3096

[a] EEGs which contained '>200 EVs' had excessive amounts of EVs which were not individually graded by the EEGers. (EEGs indicated with an asterisk were used to form the calibration set).
[b] m, month.

would be representative of true EDs after training was complete–the actual size of the corresponding portion of SOFM would depend on the proportions of ED to non-ED in the CED data. In order to achieve a more even balance between ED and non-ED data, a number of failed-threshold CEDs selected at random were removed from the training set such that the number of passed-threshold CEDs was twice that of failed-thresholds CEDs.

From the training set, nine EEGs were chosen for calibrating the SOFM. These EEGs had an average length of 22.9 min and were recorded from children and adults (EEGs labelled with an asterisk in Table 2 represent the calibration set). The EEGers graded 182 definite EVs and 184 questionable EVs. As the calibration data was graded by more than one EEGer, the final grading assigned to each waveform was based upon a consensus amongst the two (sometimes three) EEGers. Waveforms that had widely varying labels amongst the EEGers were removed from the calibration set (but still remained part of the training set). This resulted in a calibration set of 5846 CEDs, 4574 of which passed the thresholds and 1272 of which failed. 1333 CEDs were assigned true-ED labels (definite and questionable, based on the consensus discussed above) and 4513 were assigned non-ED (200 were rejected from the calibration set due to large disagreement amongst EEGers). The calibration set data was then used to calibrate the SOFM after training was complete. The same data was also used for fine-tuning using the LVQ2 algorithm.

2.2.2.2. Training and calibrating the SOFM. The SOFM was trained by presenting the training data set in random order to the SOFM whilst adapting according to the SOFM

training algorithm as described in the section above. In a preliminary study, the size of the SOFM was varied from a [10 × 10] to a [22 × 22] SOFM in order to assess the performance of the system as a function of the SOFM size (James et al., 1996; James, 1997). It was found that as SOFM size increased, the performance increased measurably until a size of [20 × 20], when increases were negligible. The size of the SOFM needs to be large enough to capture the variability present in the inputs. However, it is important not to make the SOFM too large as this results in longer training and slower performance. For the results presented here we used an SOFM of [20 × 20].

The training data was presented randomly to the SOFM during training in such a way that all the data had been presented at least twice during the crucial initial stage of training known as the ordering stage. The ordering stage was set at 1/8th of the complete training period, which means that the entire training set was presented 16 times to the SOFM during training–this results in over half a million iterations of the training algorithm.

It is required that once a CED is presented to the SOFM after training is complete, the SOFM responds with a value indicating the probability of it being a true ED. This meant that, as part of calibration, a probability level needed to be assigned to each neuron. In operation, once a CED is presented to the SOFM, the probability assigned to the 'winning' neuron (i.e. the neuron with the closest matching $\mathbf{m}_i$) is taken to be the probability of the input CED being a true ED. We have modified the method suggested by Kohonen (1990) for calibrating the SOFM. The *success rate* for neuron i is given by:

$$\theta_i = \frac{s_i}{n_i}$$

where the number of successes s_i indicates the number of times the EEGers labelled input waveforms ED (as opposed to non-ED) and n_i is the total number of times neuron i was the 'winner'.

Using the Bayesian statistical approach, it is possible to assign a probability to each neuron. This probability results from a weighted mix of prior probabilities and the data itself–the so-called posterior probability (Schmitt, 1969; Bernardo and Smith, 1994). The prior distribution represents our uncertain knowledge of the success rate θ_i for an individual neuron. As no *a priori* knowledge of the success rate is assumed, the prior distribution is 'flat', meaning that any success rate can be applied to each neuron. Applying Bayes' theorem to the prior distribution and the Binomial likelihood function for the data, it can be shown (Bernardo and Smith, 1994) that an estimate of the success rate θ can be obtained by taking the mean value of the posterior distribution, such that

$$E = [\theta|s] = \frac{s+1}{n+2},$$

for s successes and n trials. Thus, by a Bayesian approach, a better estimate of the probability of a CED being a true ED for neuron i can be found by

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

Consider, for example, two neurons i and j. If neuron i was declared the 'winner' five times, four of which were for true ED, this gives a success rate of 0.8 (i.e. 4/5) but a probability of 0.71 using the Bayesian probabilities. If neuron j was declared the 'winner' 50 times, 40 times of which for true ED, then the success rate is 0.8 (as for neuron i) but the probability now becomes 0.79, reflecting the larger number of 'wins' for neuron j. In a similar manner, if neuron i was declared winner for true EDs all five times and neuron j declared winner for true EDs all 50 times, both would have a success rate of 1.0. However, neuron i would be assigned a Bayesian probability of 0.86 whereas neuron j a probability of 0.98.

Fig. 4 depicts the weight vectors of the [20 × 20] SOFM after training and calibration. The topological ordering of the 'ED-like' waveforms is apparent. Note that only the 'raw' EEG portion of the weight vectors is shown, the contextual parameters and passed/failed thresholds flag are not shown. Those waveforms that resulted in a probability greater than 0.5 after calibration are indicated in black whilst all others are in light grey. It can be said that all weight vectors describe 'ED-like' waveforms but at different probabilities when using the Bayesian technique described above.

2.2.3. Stage 3: spatial-combiner

The final stage of the spike detection system combines the outputs of the SOFM stage in such a way as to confirm the presence of an EV across two or more channels of EEG and, hence, report the detection of an EV. If the spatial pattern is inconsistent with the presence of an EV, it is

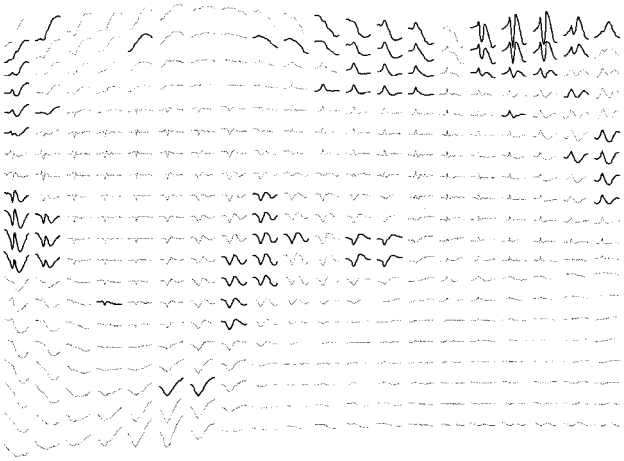

Fig. 4. The weight vectors of a [20 × 20] SOFM after training and calibration. Weight vectors which have been assigned a probability of greater than 0.5 are indicated in black whilst all others are in light grey.

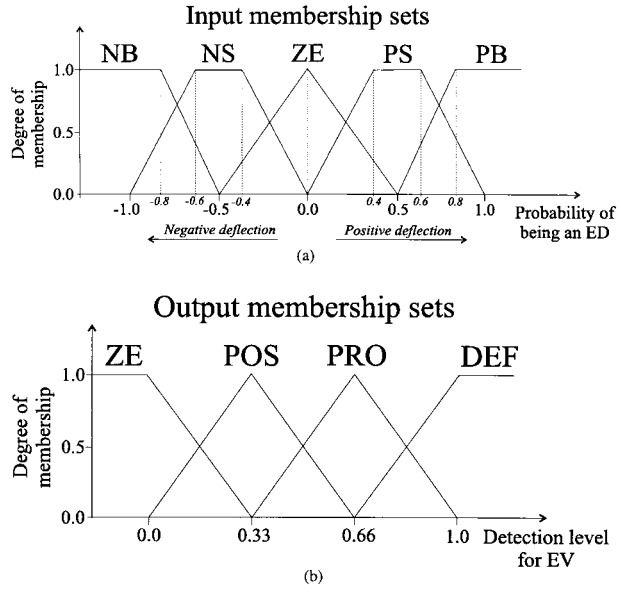

Fig. 5. The fuzzy sets used to define (a) the inputs to, and (b) the outputs of, the spatial-combiner.

rejected. This stage is dubbed the spatial-combiner (James et al., 1998).

The spatial-combiner uses a number of rules that specify allowable combinations of EDs across channels to detect an EV. The spatial-combiner works on a four channel bipolar electrode chain basis, where the incoming (bipolar) EEG is examined based on identical sub-systems. Electrode chains with more than four channels (at most six channels) are split into two overlapping four channel chains, and shorter electrode chains are padded with 'nulls' to make up a four channel chain. The combiner relies on two pieces of information for each bipolar chain: (a) the probability assigned to each CED by the SOFM stage and (b) the polarity of each CED within the bipolar chain.

The problem of detecting allowable spatial combinations of EDs across channels is simplified by introducing fuzzy logic. With fuzzy logic, the generation of the fuzzy rules becomes easier as no explicit mathematical models are needed that describe the underlying process. Fuzzy logic was first developed by Zadeh (1965) and is based on a mathematical theory that combines elements of multi-valued logic, probability theory, and artificial intelligence. Fuzzy logic simulates aspects of human thinking by incorporating the imprecision inherent in all physical systems (Zimmermann, 1986; Klir and Folger, 1988).

The crisp inputs to the spatial-combiner (i.e. the probabilities of true ED as output by each SOFM of the previous stage) are fuzzified by using the fuzzy sets defined in Fig. 5a. The fuzzy sets are defined to be: negative big (NB), negative small (NS), zero (ZE), positive small (PS) and positive big (PB). Each sub-system produces a single output that is defuzzified using the fuzzy sets described in Fig. 5b. The four fuzzy sets are defined to be: zero (ZE), possible (POS), probable (PRO) and definite (DEF). In both cases trapezoidal membership functions are used because of their ease of implementation. The method of composite maximum was adopted for the defuzzification process such that the rule most representative of an 'allowable' EV distribution across the four inputs contributes to the fuzzy output set label and the membership value.

The fuzzy-rules are drawn-up based on our pre-defined knowledge of how an EV will manifest itself across a bipolar electrode chain. Each rule covers the possibility of a focal event at points along the bipolar electrode chain. As the probabilities assigned to each CED (by the SOFM stage) on each channel can take any value from 0 to 1, it would take a great many rules to cover every combination of polarity and probability value for each possible EV focus along a four channel chain. As there are now only five fuzzy (input) variables, for a four channel electrode chain there are a maximum of $5^4 = 625$ possible rules which cover all the possible combinations of inputs. However, a large number of these rules are meaningless and are therefore not used. This results in 127 distinct fuzzy rules describing allowable combinations of fuzzy variables for the four inputs of each sub-system. Each rule derived in this manner is assigned an outcome of either DEF, PRO or POS. Fig. 6 describes one such fuzzy-rule derived in the manner described here. (The actual fuzzy output label assigned to each particular rule was arbitrarily based on the number of PB/NB and PS/NS variables assigned to each particular rule.) The detection of an EV can then be made (a) if the crisp output exceeds a given threshold or (b) if the fuzzy output is either of POS, PRO or DEF.

Fig. 7a depicts a true EV, spanning four channels, where the CEDs presented to the SOFM were assigned high probabilities and were coupled with the right spatial distribution, resulting in a detection. Fig. 7b depicts an instance where a four channel CEV is rejected by the spatial combiner due to incompatible spatial distribution.

The underlying assumption when deriving the fuzzy spatial rules is that a focal EV is detected along the bipolar chain of electrodes. Generalized EVs, however, show no distinct focus across a number of four channel bipolar chains but will appear as a focal event at one end of each chain. Thus, the rules derived using the above method are able to detect generalized activity.

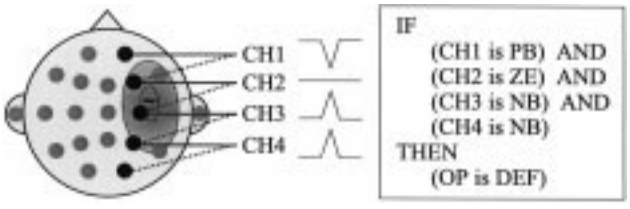

Fig. 6. An example of a fuzzy rule obtained, in this instance, for the case where a focus is assumed near the scalp with maximum negativity between electrodes 2 and 3 (i.e. F2 and C2).

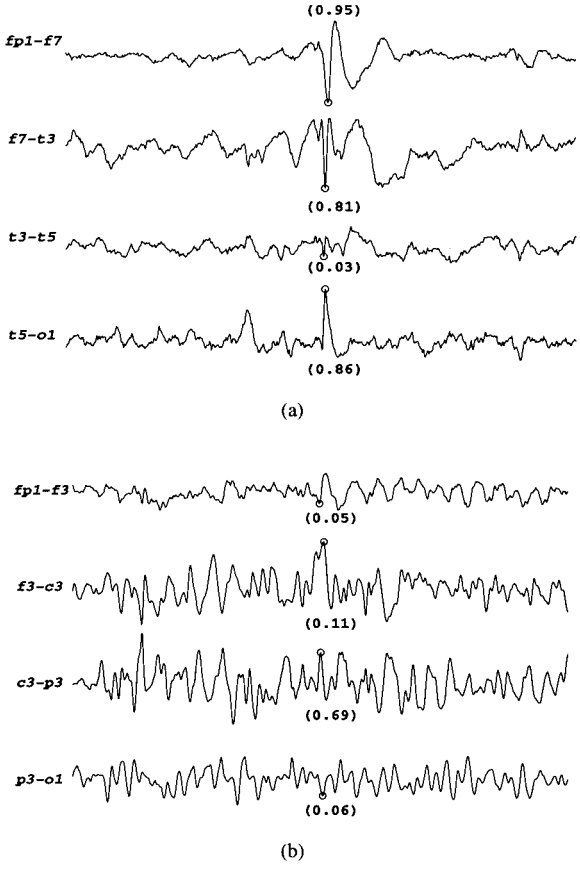

Fig. 7. Examples of actual outputs from a given four channel bipolar chain of electrodes: (a) depicts a correct detection on an EV based on the combination of the SOFM and spatial combiner outputs, (b) depicts a rejected CEV due to incompatible spatial distribution.

2.3. Grouping the inputs to the sub-systems

As stated before, the outputs from the SOFM stage needed to be grouped into four channel bipolar chains (depending on the montage in use). Each sub-system performed exactly the same operation on the four probability values input to it. For the longitudinal and longitudinal-transverse montages, the sub-systems are naturally grouped into four sub-systems of four channels. The transverse montage is divided into five sub-systems. The six channel chain A1–T3–C3–Cz–C4–T4–A2 is divided into two overlapping four channel chains such that A1–T3–C3–Cz–C4 are input to one sub-system and C3–Cz–C4–T4–A2 are input to another. The two channel chain Fz–Cz–Pz is converted to a four channel chain by padding the last two channels with 0 probabilities. The circumferential montage is divided into six sub-systems. Two non-overlapping four channel chains were formed with Fp2–F8–T4–T6–O2 and Fp1–F7–T1–T5–O1 and two overlapping four channel chains with F3–F4–C4–P4–P3 and F4–F3–C3–P3–P4. Finally two three channel chains were padded with 'null' channels (i.e. 0 probability channels) for the chains F7–Fp1–Fp2–F8 and T5–O1–O2–T6.

2.4. Artifact rejection

In order to minimize any further false detection of artifacts as EVs, outputs from the SOFM stage are eliminated if they are due to muscle contraction, eye-blinks, electrode movement or bursts of sharp-α. Elimination takes place by assigning a probability value of 0. The detection of one of the above-mentioned occurrences is made through contextual information presented with the CED and is based on work by Dingle et al. (1993). Each artifact is tackled in the following way:

- Bursts of muscle activity on any channel are detected when (a) the average background duration is short ($<$20 ms, corresponding to the high frequency waveforms characteristic of EMG) and (b) the RMS background amplitude is large ($>$12.6 μV);
- eye-blinks are detected when the floating-mean falls significantly below the baseline (<-80 μV) on a frontal channel;
- electrode movement is detected on a channel when the floating mean reaches a maximum of at least 100 μV above the baseline;
- bursts of α are detected when (a) the average background duration is 100 ms and the CED duration D_3 is similar and (b) the RMS background amplitude is comparable to the peak amplitude of the CED (A_p).

2.5. The test set of EEGs

A separate test set of eight EEGs recorded in a similar manner to the training set described earlier was used to test the performance of the trained system. The test set consisted of seven EEGs containing epileptiform activity (as graded by two or three EEGers) and one normal EEG (i.e. no epileptiform activity as graded by all three EEGers). The data (Table 3) consisted of routine clinical recordings of an average length of 24.4 min (total EEG length 3.2 h) with 190 epileptiform events (65 definite and 125 questionable) as marked by two or three EEGers using the convention as described before for the calibration set. The EEGs contained significant artifact such as EMG, electrode movement and bursts of alpha waves and no segment was discarded because of artifact. The presence of artifact was most significant during photic stimulation and hyperventilation, which made up over 30% of the recorded EEG.

2.6. Performance measures

The performance of the spike detection system has been assessed (a) after the mimetic stage, (b) after the SOFM stage and (c) at the output of the spatial combiner. For both (a) and (b), no spatial contextual information is present, whereas (c) uses spatial information to act on the outputs of (b) making the final EV/non-EV decision. A detection was considered to have taken place at each stage as follows:

Table 3
The test set comprising seven definite epileptiform EEGs (with a total of 65 definite and 125 questionable EVs) and one normal EEG (i.e. no epileptiform activity)

Patient	Age	Duration	EEGers	Epileptiform events		
				Definite	Questionable	Total
1	5	27 m 24 s	3	9	3	12
2	71	25 m 08 s	2	15	17	32
3	24	26 m 16 s	2	25	20	45
4	11	23 m 34 s	2	2	1	3
5	3	16 m 03 s	2	8	21	29
6	50	21 m 13 s	3	6	58	64
7	84	26 m 50 s	3	0	5	5
8	28	25 m 15 s	3	0	0	0
Totals	~34.5	3 h 11 m 43 s		65	125	190

- mimetic stage: if at least one CED on any of the 16 channels passed the thresholds;
- SOFM stage: if the probability assigned to at least one CED on any of the 16 channels exceeded 0.84. (A detection threshold of 0.84 results in a sensitivity similar to that measured at the spatial combiner stage);
- spatial combiner: if an output was assigned any one of the fuzzy variables POS, PRO or DEF.

In each case the performance measures are the sensitivity and selectivity of the system to EVs (similar to Webber et al., 1994). The sensitivity and selectivity are given by

$$\text{Sensitivity} = \frac{\text{correct detections}}{\text{total number of true events}} \times 100\%$$

and

$$\text{Selectivity} = \frac{\text{correct detections}}{\text{total number of detections}} \times 100\%.$$

In addition, the number of false detections per hour has been calculated. The false detection rate is considered an important measure of performance of a system as it gives an indication of how useful a system will be in routine clinical practice. Furthermore, the measure of number of false detections per hour of EEG can be used to place the reported performance of the system into context when considering the length of EEGs used in the test sets.

In this study, EVs rated by the EEGers as definite or questionable were both treated as *true* EVs. If some of the questionable EVs were in fact non-epileptiform, this would lead to a slight underestimation of sensitivity and overestimation of selectivity. The effect of using 'questionable's along with 'definite's should be the subject of another study.

3. Results

Our hybrid system runs at approximately 20 × real-time (i.e. a 20 min EEG takes about 1 min to process off-line) but can take longer for EEGs with 'noisy' backgrounds due to the excessively large number of CEDs put forward by the mimetic stage.

Table 4 gives the performance at each stage of the system for each patient in the test set for a [20 × 20] SOFM. For each stage measures for the sensitivity, selectivity and false detection rate are indicated for each EEG separately and for the eight EEGs overall. It can be seen that the mimetic stage results in a high sensitivity for almost every EEG with an overall sensitivity of 91.1%, coupled with a very low selectivity of 0.8% and a false detection rate of 6796/h. This shows how the mimetic stage is 'screening' the incoming EEG with very little selectivity – as expected.

At the output of the SOFM stage, the sensitivity has been reduced to 58.9%. The selectivity improved slightly to 7.9% with a marked improvement in the false detection rate to 410/h. Although the false detection rate shows a marked improvement over the mimetic stage, it is still unacceptable–approximately one false detection every 7 s of EEG. (A lower detection threshold at the SOFM stage could, of course, give a higher sensitivity for this stage, but at the expense of a lower selectivity and much higher false detection rate. The threshold of 0.84 was chosen so that comparisons could be made between the SOFM stage and output stage with a similar sensitivity for both). This notwithstanding, the performance at the SOFM stage indicates that a large quantity of the CEDs were assigned a reasonably high probability by each channel SOFM based on their single-channel morphology.

The performance at the final stage indicates a similar sensitivity of 55.3% but a massive increase in the selectivity to 82.0% coupled with an equally impressive 7.2 false detections/h–one false detection every 500 s or 8.3 min of EEG. This false detection rate reflects almost a 60 fold reduction from the SOFM stage and almost a 1000 fold reduction from the mimetic stage. These results support our contention of the crucial importance of spatial analysis in the spike detection process. Whilst the improvements in the selectivity and false detection rate were substantial, it was at the expense of reducing the sensitivity to 55.3%. This is not, however, seen as a massive drawback as although not ideal (i.e. 100%) the sensitivity complemented the low false detection rate acceptably well.

Of particular note are the results of patient 7: the 5 questionable EVs marked by the EEGers were heavily contaminated by EMG and resulted in a poor sensitivity of 20% and selectivity of 33.3%. However, the sensitivity and selectivity at the SOFM stage for this patient were 0%. This indicates that on an individual ED level (i.e. single channel) the relevant CEDs were assigned a low probability (a value less than d_{th})–which was not surprising when the CEDs in question were observed. This again highlights the importance of spatial analysis.

Also of note are the results of patient 8 (the normal EEG): the mimetic and SOFM stages resulted in unacceptably high false detection rates of 8678.7/h and 261.3/h, respectively.

Following spatial analysis the false detection rate was correctly reduced to 0/h.

4. Discussion

This paper has presented a new and innovative approach for the detection of epileptiform discharges in the EEG. Many attempts have been made previously in the literature to solve this problem but all with limited success. The main problem lies with the extreme difficulty met in attempting to eliminate considerable numbers of false detections due to spike-like artifacts and sharp background activity in the EEG. The EEGer makes considerable use of spatial and temporal information when visually performing spike detection to influence his decision in the process. Surprisingly, few systems incorporate any aspects of such reasoning in their spike detection algorithms. In the approach described here, the ability of ANNs to be trained to solve problems and their ability to generalize to novel data once training is complete are drawn upon to the advantage of the spike detection system.

The spike detector/classifier is based around the SOFM. The capability of the SOFM to extract identifying features from large amounts of input data is drawn upon heavily in this stage. This is particularly useful in the spike detection problem where there is a great difficulty in accurately grading large numbers of candidate epileptiform discharges (CEDs) due to the amount of disagreement between EEGers. Using the self-organising abilities of the SOFM it is possible to train an ANN with a large number of EEGs known to contain EDs and then use only a selected sub-set of EDs (which have been graded with a high degree of agreement amongst EEGers) to accurately label each weight abstracted from the input data by the SOFM. The resulting spike-like waveforms abstracted from the large training set as shown in Fig. 4 indicates the large variability in spike (and non-spike) morphology. Although it would be possible to label the weight vectors of the trained SOFM as ED/non-ED manually, this would be subjective and the further requirement of a probability level attached to each waveform would make the task an extremely subjective one. A lot of the subjectivity is removed through the novel use of Bayesian statistics to assign a probability value to each SOFM weight during the calibration stage. The use of EDs in the calibration set labelled based on a consensus of 2/3 EEGers also helps reduce the subjectivity of the labelling process.

The SOFM stage is preceded by a mimetic stage in order to screen the incoming EEG and reduce the number of CEDs presented to the SOFM stage. Importantly, the mimetic stage also gives a time reference to each CED such that it is presented to the SOFM stage with its vertex positioned at the same point in each case. Each mimetic/SOFM stage works independently on each channel of EEG (i.e. utilising no spatial information). The results given in Table 4 show that the mimetic stage performed its function well; the sensitivity of 91.1% is acceptable and the low selectivity of 0.8% was expected.

The performance of the system following the SOFM stage (at $d_{th} = 0.84$) was a sensitivity of 58.9%, a selectivity of 7.9% and a false detection rate of 410.6/h. The threshold is arbitrary and is used only to give an indication of the performance of the system. Nonetheless an almost 17-fold reduction is achieved in the false detection rate over the previous stage. However, apart from increasing the performance of the system per se, the SOFM stage performs the more important function of assigning a probability level to each CED detected by the mimetic stage to the best advantage of the spatial combiner.

Finally, a spatial-combiner stage, based on fuzzy logic, is presented which groups the single-channel probabilities according to the (bipolar) montage in use and forms a decision on the detection of an EV based on the spatial distribution of the CEDs. Following spatial analysis, the performance of the system was measured at a sensitivity

Table 4
The sensitivities, selectivities and false detection rates of the system at each stage to definite and questionable EVs for each patient[a]

Patient	Mimetic stage			SOFM stage ($d_{th} = 0.84$)			Spatial combiner stage		
	Sensitivity (%)	Selectivity (%)	FD rate/h	Sensitivity (%)	Selectivity (%)	FD rate/h	Sensitivity (%)	Selectivity (%)	FD rate/h
1	100.0	0.7	4840.8	83.3	10.0	270.3	83.3	58.8	21.0
2	93.8	1.9	4615.6	50.0	17.6	225.2	71.9	100.0	0.0
3	100.0	1.5	9099.1	66.7	13.6	573.6	44.4	95.2	3.0
4	100.0	0.4	2354.4	66.7	2.8	207.2	66.7	40.0	9.0
5	96.6	2.9	2840.8	58.6	13.2	336.3	65.5	79.2	15.0
6	79.7	2.9	5153.2	57.8	18.0	507.5	46.9	85.7	15.0
7	80.0	0.0	27 627.6	0.0	0.0	1558.6	20.0	33.3	6.0
8	–	–	8678.7	–	–	261.3	–	–	0.0
Totals	91.1	0.8	6796.0	58.9	7.9	410.6	55.3	82.0	7.2

[a] The SOFM stage values are for a [20 × 20] SOFM followed by fine-tuning with LVQ2). For the SOFM stage the detection threshold, d_{th}, was set at 0.84 so that the measured sensitivity was similar to that obtained by the spatial combiner stage, in order to make a justifiable comparison between selectivities and false detection rates for each stage.

of 55.3%, a selectivity of 82.0% and a false detection rate of 7.2/h. The selectivity and false detection rate depict the immense gains achieved through the use of spatial analysis in the spike-detection process. A sensitivity of 55.3% may be considered on the low side but we contend that, as long as the sensitivity is adequate (i.e. as long as a reasonable proportion of EVs are detected in the overall EEG), the selectivity and more importantly the false detection rate are the measures that will be used to 'judge' a system's utility in routine clinical use. As is always the case due to the large variability between EEGs, the sensitivity was high for four out of the seven epileptiform EEGs with an average sensitivity of 71%. It should also be noted that our calculation of sensitivity included detection of questionable events; many of these could, in fact, have been non-events, resulting in an incorrect estimation of our system's sensitivity (e.g. patient 7's 20% may have been invalid). Of the 85 EVs missed by the system, 4% were missed due to the wrong CEDs being chosen by the mimetic stage when grouping across a four channel bipolar chain, resulting in CEVs with incompatible spatial distribution. The mimetic stage completely failed to pick up 20% of the missed EVs and the remaining 76% of the missed EVs were due to low probabilities being assigned to one or more CEDs for a given event, resulting in incompatible spatial distributions. The low probabilities were mainly due to EDs obscured by artifact or background EEG. The appreciable fall in sensitivity following the SOFM stage indicates that the major cause of these missed events is indeed due to incompatible spatial distributions. In contrast, the 17 false detections were almost entirely due to large focal artifact present at the beginning or end of an electrode chain (mainly as sharp alpha waves measured by longitudinal montage) or, less frequently, electrode 'pop'.

Comparisons between different spike detection systems are made difficult by the wide range of measures used for evaluating their performance. The greatest differences in assessing the performance occur when obtaining measures for false detections and missed detections. Gotman and Wang (1991, 1992) define false detections as detections which are obviously artifact, whereas other methods include defining a false detection as an event not marked by any of a panel of six EEGers (Eberhart et al., 1989) or marked by fewer than six of seven EEGers (Fischer et al., 1980). In a similar way, Gotman and Wang (1991); Gotman and Wang (1992) state that missed detections are events falsely rejected by the system as non-epileptiform, but they ignore EVs/EDs missed by the system altogether. In contrast, Eberhart et al. (1989) identify missed detections as those not detected by the system but marked by at least four of six EEGers.

The performance measures used for this system, where possible, will be used to compare between systems. Table 5 shows a number of spike detection systems found in the literature and compares their performance with that of the SOFM based spike detection system. In addition it was possible to directly compare our system with that of Dingle et al. (1993) using the same test set of EEGs as in the current study; this is listed as system ''8'' in Table 5.

The results given in Table 5 show a great variability in the measures of performance, especially false detection rate. Of the systems described, only Gotman and Wang (1992) and our system use a totally new set of test data to validate performance. Of all of the systems, only Dingle et al. (1993) and our system have included normal EEGs in their tests. Webber et al. (1994) tested their system on the EEGs obtained from ten patients (the same EEGs were used for training), and report satisfactory sensitivity and selectivity for the mimetic + ANN (parameters) case (both 74%) but at a cost of around 804 false detections/h. For their performance using 'raw' EEG instead of parameters, both the sensitivity and selectivity were relatively low (both 46%), with an even higher false detection rate (over 5000/h). Özdamar et al. (1991) report similarly good results for sensitivity and selectivity but a similarly high false detection rate (~1023/h). Hostetler et al. (1992) carried out an independent evaluation of the Gotman et al. (1978) system and reported a much lower false detection rate (37 false detections/h). Gotman and Wang (1992) report a similar false detection rate for their system (which estimates the state of a subject during recording to improve performance). In contrast, a false detection rate of zero is reported by Dingle et al. (1993) with a reasonable sensitivity (53%) on data from 11 patients including three with normal EEGs. However, the system used the training data to test the system. When the same system was used to assess the novel test set used by our system, the false detection rate rose slightly to two detections/h, the selectivity fell to 81%, and the sensitivity fell to 14%. The system of Dingle et al. (1993) has also been evaluated blindly on 521 consecutive routine EEGs (173 h) compared with two or three EEGers. The system detected 36 of 38 EEGs containing definite EVs–a global sensitivity of 95%–and had an average false detection rate in the non-epileptiform EEGs of 0.29/h (Jones et al., 1996), thus confirming the low false detection rate of their system.

In relation to the other spike detection systems of Table 5 our system compares well. The importance of the false detection rate as a measure of performance of such a system is immediately apparent when one considers the performance of systems that have outwardly good measures of sensitivity and selectivity, such as Özdamar et al. (1991); Webber et al. (1994). It was not possible to determine measures of the sensitivity for the systems of Gotman and Wang (1992) so it was not possible to measure at what cost (in terms of missed events) the values of selectivity and false detection rate were obtained. Irrespective, the current system resulted in superior selectivities and false detection rates. Certainly results from the present study and those from Dingle et al. (1993); Jones et al. (1996) confirm the crucial importance of incorporating spatial and wide-temporal contextual information in the spike detection

Table 5
A comparison of the sensitivities, selectivities and false detection rates between the SOFM based spike detection system (9) and others found in the literature (systems 1–8)

System		Methods	EEGs	H	Epileptic (%)	EEGers	Train/test	Sensitivity (%)	Selectivity (%)	False/h
1	Özdamar et al., 1991	ANN + ANN	10	0.043	?	4	6/4	90	69	~1023
2	Hostetler et al., 1992	Mimetic	5	2.0	100	5	Blind	59	89	37
3	Gotman and Wang, 1992	Mimetic	20	33.0	100	2	Blind	–	41	117
4	Gotman and Wang, 1992	Mimetic + state	20	33.0	100	2	Blind	–	67	47
5	Webber et al., 1994	Mimetic + ANN (Params)	10	0.3	100	1	Same	74	74	~804
6	Webber et al., 1994	Mimetic + ANN (Raw EEG)	10	0.3	100	1	Same	46	46	~5598
7	Dingle et al., 1993	Mimetic + expert system	11	3.0	73	1	Same	53	100	0
8	Dingle et al., 1993[a]	Mimetic + expert system	8	3.2	88	2/3	Blind	14	81	2
9	*James et al.*	*Mimetic + SOFM + Fuzzy logic*	*8*	*3.2*	*88*	*2/3*	*Blind*	*55*	*82*	*7*

[a] System '8' used the method of Dingle et al., 1993 on the current test data.

process. At this point it is worth exercising caution in that the results obtained from the current system only represent those extracted from a small set of novel routine EEGs. Although the results look extremely promising, based on novel EEGs as they are, testing on a much larger population of routine and long term EEGs is warranted before the system can be implemented in routine clinical use. This is the next step in our study.

In conclusion, a new spike detection system has been developed which makes considerable use of spatial and limited use of temporal information in the EEG whilst aiming to emulate the EEGer's approach to the spike detection problem. A modular approach was used in the system's development and it makes particular use of the attributes inherent in ANNs (i.e. nonlinear, adaptive, etc.). In a clinical setting, such a system should prove to be an important tool in the automatic and real-time detection of epileptiform activity in routine EEG recordings and in long-term EEG monitoring.

Acknowledgements

The authors gratefully acknowledge the help and comments of Dr Jean Gotman of the Montreal Neurological Institute in the preparation of this manuscript. Funding in the form of a Commonwealth Scholarship to Dr Christopher J. James by the New Zealand Vice-Chancellor's Committee is gratefully acknowledged.

References

Bernardo J, Smith F. Bayesian theory. New York: Wiley, 1994.

Black MA, Jones RD. Sensitivity and selectivity for continuous perception values–a comment. Electroenceph clin Neurophysiol 1998;106(5):457–459.

Black MA, Jones RD, Carroll GJ, Smith MH. Bayesian evaluation of automated EEG analysis system in the absence of a gold standard. New Zealand Med J 1997;110:444.

Chatrian G, Bergamini L, Dondey M, Klass D, Peterson I, Lennox-Buchtall M. A glossary of terms most commonly used by clinical electroencephalographers. Electroenceph clin Neurophysiol 1974;37:538–548.

Dingle, A.A., Engineering in brain research: processing electroencephalograms and chaos in neural networks. PhD Thesis. University of Canterbury, Christchurch, New Zealand, 1992.

Dingle AA, Jones RD, Carroll GJ, Fright WR. A multistage system to detect epileptiform activity in the EEG. IEEE Trans Biomed Eng 1993;40:1260–1268.

Eberhart R, Dobbins R, Webber WRS. Neural network design considerations for EEG spike detection. Proc. 15th Northeast Bioeng. Conf., Boston 1989. pp. 97–98.

Fischer, G., Mars, N., Lopes da Silva, F., Pattern recognition of epileptiform transients in the encephalogram. Report #7, Institute of Medical Physics, Da Costakade, Utrecht, The Netherlands, 1980.

Gabor AJ. Seizure detection using a self-organizing neural network: validation and comparison with other detection strategies. Electroenceph clin Neurophysiol 1998;107:27–32.

Gabor AJ, Seyal M. Automated interictal EEG spike detection using artificial neural networks. Electroenceph clin Neurophysiol 1992;83:271–280.

Gabor AJ, Leach RR, Dowla FU. Automated seizure detection using a self-organizing neural network. Electroenceph clin Neurophysiol 1996;99(3):257–266.

Glover JJ, Raghavan N, Ktonas P, Frost JJ. Context-based automated detection of epileptogenic sharp transients in the EEG: elimination of false positives. IEEE Trans Biomed Eng 1989;36:519–527.

Gotman J. Quantitative measurements of epileptic spike morphology in the human EEG. Electroenceph clin Neurophysiol 1980;48:551–557.

Gotman J, Gloor P. Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. Electroenceph clin Neurophysiol 1976;41:513–529.

Gotman J, Wang L. State-dependent spike detection: concepts and preliminary results. Electroenceph clin Neurophysiol 1991;79:11–19.

Gotman J, Wang L. State-dependent spike detection: validation. Electroenceph clin Neurophysiol 1992;83:12–18.

Gotman J, Gloor P, Schaul N. Comparison of traditional reading of the EEG and automatic recognition of interictal epileptic activity. Electroenceph clin Neurophysiol 1978;44:48–60.

Hostetler W, Doller H, Homan W. Assessment of a computer program to detect epileptiform spikes. Electroenceph clin Neurophysiol 1992;83:1–11.

James CJ The detection of epileptiform activity in the electroencephalogram using artificial neural networks. PhD Thesis. University of Canterbury, Christchurch, New Zealand, 1997. pp. 287

James CJ, Jones RD, Bones PJ, Carroll GJ. The self-organising feature map in the detection of epileptiform transients in the EEG. Proc. 18th Int. Conf. IEEE Eng Med Biol Soc, Vol. 18, Amsterdam, 1996, pp. 2 (CD-ROM).

James CJ, Jones RD, Bones PJ, Carroll GJ. Spatial analysis of multi-channel EEG recordings through a fuzzy-rule based system in the detection of epileptiform events. Proc 20th Int Conf IEEE Eng Med Biol Soc, Hong Kong 1998. 4 pp (CD-ROM).

Jando G, Siegel RM, Horvath Z, Buzsaki G. Pattern recognition of the electroencephalogram by artificial neural networks. Electroenceph. clin. Neurophysiol. 1993;86(2):100–109.

Jasper H. Report of the committee of methods of clinical examination in electroencephalography. EEG J. 1958;10:370.

Jones RD, Dingle AA, Carroll GJ, Green R, Black MA, Donaldson L, Parkin P, Bones P, Burgess K. A system for detecting epileptiform discharges in the EEG: real-time operation and clinical trial. Proc. 18th Int. Conf. IEEE Eng. Med. Biol. Soc., Amsterdam, 1996, 2 pp. (CD-ROM).

Kalayci T, Özdamar Ö. Wavelet preprocessing for automated neural network detection of EEG spikes. IEEE Eng Med Biol Mag 1995;14:160–166.

Klir G, Folger T. Fuzzy sets, uncertainty and information, Englewood Cliffs, NJ: Prentice-Hall, 1988.

Koffler DJ, Gotman J. Automatic detection of spike-and-wave bursts in ambulatory EEG recordings. Electroenceph clin Neurophysiol 1985;61:165–180.

Kohonen T. The self-organizing map. Proc. IEEE 1990;78:1464–1480.

Kohonen T. Self-organizing maps, Berlin: Springer, 1995.

Ktonas P, Luoh W, Kejariwal M, Seward M. Computer-aided quantification of EEG spike and sharp characteristics. Electroenceph clin Neurophysiol 1981;51:237–243.

Ktonas P, Glover J, Webster L, Antonthanasap R, Velamuri S, Urunuela J, Reilly E. Automatic detection of epileptogenic sharp EEG transients. Electroenceph clin Neurophysiol 1984;58:38.

Lopes da Silva F, Van Hulten K, Lommen J, Van Leeuwen W, Van Veelen C, Vliegenthart W. Automatic detection and localisation of epileptic foci. Electroenceph clin Neurophysiol 1977;43:1–13.

Özdamar Ö, Lopez C, Yaylali I. Multilevel neural network system for EEG spike detection. In: Bankman I, Tsitlik J, editors. Computer-based medical systems, New York: IEEE Computer Society Press, 1991. pp. 272–279.

Schiff S, Aldroubi A, Unser M, Sato S. Fast wavelet transformation of EEG. Electroenceph clin Neurophysiol 1994;91:442–455.

Schmitt S. Measuring uncertainty: an elementary introduction to Bayesian statistics. Reading, MA: Addison-Wesley, 1969.

Webber WRS, Litt B, Wilson K, Lesser R. Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized data. Electroenceph clin Neurophysiol 1994;91:194–204.

Wilson SB, Harner RN, Duffy FH, Tharp BR, Nuwer MR, Sperling MR. Spike detection. I. Correlation and reliability of human experts. Electroenceph clin Neurophysiol 1996;98:186–198.

Zadeh LA. Fuzzy sets. Inform Contrib 1965;8:338–353.

Zimmermann H. Fuzzy sets, decision making and expert systems. Boston, MA: Kluwer Academic Publishers, 1986.