Electroencephalography and clinical Neurophysiology 106 (1998) 488–500

An expert system for EEG monitoring in the pediatric intensive care unit

Y. Si[a], J. Gotman[a,*], A. Pasupathy[a], D. Flanagan[a], B. Rosenblatt[b], R. Gottesman[b]

[a]The Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec H3A 2B4, Canada
[b]The Montreal Children's Hospital, Montreal, Quebec, Canada

Accepted for publication: 14 November 1997

Abstract

Objectives: was to design a warning system for the pediatric intensive care unit (PICU). The system should be able to make statements at regular intervals about the level of abnormality of the EEG. The warnings are aimed at alerting an expert that the EEG may be abnormal and needs to be examined. **Methods**: A total of 188 EEG sections lasting 6 h each were obtained from 74 patients in the PICU. Features were extracted from these EEGs, and with the use of fuzzy logic and neural networks, we designed an expert system capable of imitating a trained EEGer in providing an overall judgment of abnormality about the EEG. The 188 sections were used in training and testing the system using the rotation method, thus separating training and testing data. **Results**: The EEGer and the expert system classified the EEGs in 7 levels of abnormality. There was concordance between the two in 45% of cases. The expert system was within one abnormality level of the EEGer in 91% of cases and within two levels in 97%. **Conclusions**: We were therefore able to design a system capable of providing reliably an assessment of the level of abnormality of a 6 h section of EEG. This system was validated with a large data set, and could prove useful as a warning device during long-term ICU monitoring to alert a neurophysiologist that an EEG requires attention.

Keywords: EEG; Fuzzy logic; Expert system; Intensive care unit

1. Introduction

Many neurointensivists share the opinion that EEG can become an integral part of monitoring in the ICU (Emmerson and Chiappa, 1988; Jordan, 1993). In this context, EEG recordings have been useful in the investigation of various disorders, sub-clinical seizures and coma. The technique is non-invasive, and advances in technology have made possible the collection, storage and analysis of continuously recorded EEG.

An EEG monitor should continuously record brain activity of patients in the ICU over several hours. In order to have a chance to detect cerebral dysfunction at a reversible stage, rapid interpretation of EEG is crucial. However, the presence of an EEGer throughout the entire period of recording is impractical and the complexity of EEG patterns discourages interpretation by a non-expert. In practice, there may be a considerable time lag between recording and interpretation, which may reduce significantly the effectiveness of EEG monitoring. Also, visual interpretation of long-term recordings is quite tedious and time consuming, and it is difficult to evaluate gradual changes and long-term trends. As in other monitoring situations, it is important to note that most often nothing happens: it is only in a minority of cases that interesting events take place.

In order to reduce data and simplify interpretation, quantitative techniques have been used to transform the digital EEG signals into mathematically derived parameters. In long-term monitoring, this information is displayed in a variety of formats including bar graphs, compressed spectral arrays (CSA, Bickford et al., 1972) and density spectral arrays. These methods simplify the appearance of the displays but expert interpretation is still required, particularly because many artifacts can alter the appearance of the display, thus often requiring the interpretation of the original EEG.

The aim of this project is to build a system for automated neurophysiological monitoring in the pediatric ICU. The system is to be used as a bedside EEG warning device for pediatric patients who may suffer neurological dysfunction such as those subsequent to cardiac surgery, trauma or

* Corresponding author. Tel.: +1 514 3981953; fax: +1 514 3988106; e-mail: jean@rclvax.medcor.mcgill.ca

0013-4694/98/$19.00
PII S0013-4694(97)00154-5
EEG 97106

hemorrhage. Such a monitor accepts several hours of raw EEG data as input and estimates its level of abnormality. Its purpose, however, is not to provide the intensivist with a level of EEG abnormality, a task that would require nearly perfect performance. It rather aims at determining when it is necessary that the EEG be examined by a qualified person.

Lombroso (1985) classifies EEG abnormalities in children into 3 main categories: (1) abnormalities of the background, (2) ictal abnormalities, and (3) abnormalities of states and of maturation. Among these, background abnormalities are reported as the best prognostic indicator in long-term EEG monitoring (Lombroso, 1985; Watanabe et al., 1980). An analysis of background and ictal abnormalities may give maximum prognostic information. Seizure detection methods have been described for adults and newborns (Gotman, 1990; Gotman et al., 1997a,b). We deal here with background abnormalities, which manifest themselves in several forms: inactive pattern, burst suppression, low voltage pattern through all stages, interhemispheric amplitude asymmetry, and absence of a frequency-amplitude gradient (the progressive increase in amplitude and decrease in frequency noted on the EEGs of normal children (Slater and Torres, 1979). We have selected features that represent a large proportion of these abnormalities, and have integrated them into an expert system that is able to determine a global level of abnormality from a section of EEG.

2. Subjects

Seventy-four long-term EEGs recorded at the Montreal Children's Hospital were used as the training and testing data for the monitoring system. Among them, 69 were recorded in the intensive care unit after corrective cardiac surgery, and 5 were recorded at another monitoring unit for suspected seizures or for coma. All recordings that took place after our decision to start data collection were used (none were rejected for poor technical quality or any other reason). Thus, data selection was unbiased. The basic section of EEG used in this system lasted 6 h. Each one of the 74 EEGs yielded 1–3 6 h sections depending on the length of the recording. In all, 188 sections were available from the 74 EEGs for training and testing. The recordings were interpreted by two electroencephalographers who acted in consensus and every 6 h section was graded for each of the 3 features described below, and given an overall evaluation as one of 7 abnormality levels (see description of abnormality levels below).

3. Methods

The method of EEG analysis involves many steps, grouped in two parts: *Feature extraction* and *Knowledge-based expert system*. The first part includes the following steps: A 6 h EEG recording is first digitized and spectral band powers are computed for every 30 s epoch. An artifact rejection procedure is then followed. The 3 main features are computed for each 5 min section of EEG and their distribution for the 6 h recording is compared to the distribution obtained for control EEGs, providing statistical values representing the level of abnormality of the 3 features.

In the second part, fuzzy logic was first used to establish a correspondence between the statistical levels of abnormality and the human judgments of abnormality for each of the 3 features. A neural network then allowed the system to learn how to combine the judgments of abnormality about the separate features into an overall judgment applying to the complete recording.

The different steps and their sequence are illustrated in Fig. 1. For each step the reference is made to the heading of corresponding section.

3.1. Derivation of statistical features

3.1.1. Data acquisition and spectral analysis

The EEGs were recorded with an 8-channel bipolar montage (F3-C3, C3-P3, P3-O1, Cz-T3, and symmetrically in the right hemisphere), in the intensive care unit of the Montreal Children's Hospital. The EEG was amplified with a Grass Model 12 amplifier (Grass Instruments, Quincy, MA, USA), with an amplification factor of 10 000, and filtered by a low-pass filter at 30 Hz to prevent aliasing and a high-pass filter at 1 Hz to remove a large fraction of the artifacts due to respiration, sweating, etc. The analog signal was digitized at 200 Hz. EEGs were stored on optical disk using the Monitor software (Stellate Systems, Montréal, Québec, Canada). All subsequent analyses were performed off-line on 486 or Pentium computers. Spectral band root power was calculated for every 30 s epoch by integrating the spectra in the frequency ranges of two EEG bands, delta (1–3 Hz), and a broad band (1–14 Hz), which we called the '1-to-14' band. For each channel, the band values were then plotted against time as a 'band array' for the 6 h recording.

3.1.2. Artifact rejection

Two types of artifacts are common in EEG signals: artifacts due to patient movement and artifacts due to poor electrode contact. Since the aim of this diagnostic tool is to determine the gross state of the EEG on the basis of its frequency composition over several hours, artifact identification and rejection from the spectral band array is sufficient. In spectral band arrays, the artifacts due to the patient movement usually appear as very high amplitude rapid transients, while the artifacts generated by poor electrode contact appear as sustained high amplitude activities.

A median filter (Gabbouj et al., 1992) followed by elimination of data above a threshold were used to remove both types of artifacts from the band arrays (Fig. 2). The data were first filtered with a 5-point median filter to remove short artifacts. For sustained artifacts, we used the following: If the median of any window of 1-to-14 band value was

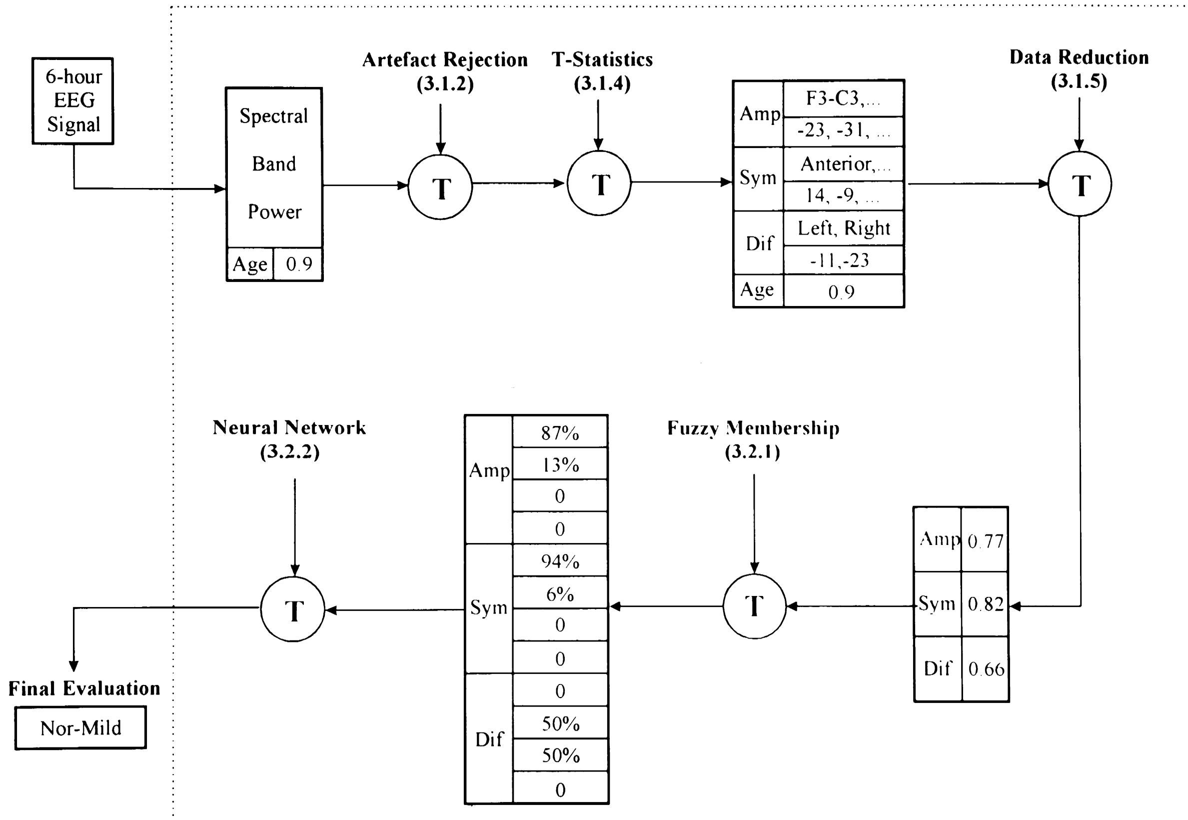

Fig. 1. Summary of the methods. The system includes 5 major steps: artifact rejection, t statistics (comparison of features to those of a control population), data reduction (taking values for all the channels and head regions and reducing them to 3 values), fuzzy membership (attributing probabilities of being normal, mildly, moderately and severely abnormal to each feature) and neural network (combining the probabilities to obtain a final evaluation). The figure illustrates with an example how data are transformed in the successive steps. The numbers above each 'T' (Transformation) refer to the headings of Section 3 where each step is described in detail.

greater than a threshold, the values in both frequency bands of that particular epoch for that channel were replaced by the respective band values averaged up to that point in time. Extensive EEG review indicated that genuine fluctuations of cerebral origin were rarely greater than 25% of the averaged background, while the amplitude of most artifacts was at least 1.5 times the average EEG background. Therefore, the threshold chosen was 1.5 times the average EEG background, ensuring that no genuine EEG activity was removed.

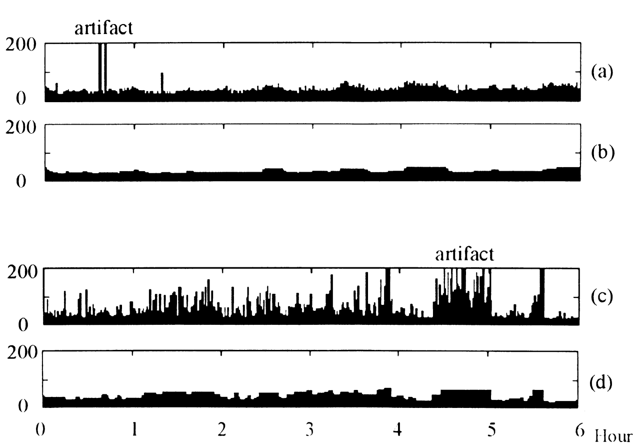

Fig. 2. Artifact-rejection with median filter and fixed threshold. (A) The 6 h band array for the total energy of an EEG containing transient artifact seen in the first and second hour. (B) Data from A passed through a 5-point median filter with a fixed threshold set at 1.5 times the average energy. (C) As in A, but for a channel with frequent sections of sustained artifact. (D) Data from C filtered with a 5-point median filter and a fixed threshold at 1.5 times the average energy.

3.1.3. Primary features

3.1.3.1. Amplitude. A depressed EEG is characterized by low amplitude values, which can be reflected in the 1-to-14 band of the EEG (Fig. 3). Since sustained rather than transient amplitude abnormalities are evaluated, the amplitude measure is defined as the logarithm of the average of the 1-to-14 band value over a 5 min period. In a band array, this represents the logarithm of the average of 10 points since basic calculations were made every 30 s. (The reason for using the logarithm of the average rather than a simple average is given below in the section on statistical analysis.) An amplitude measure extracted every 5 min over 6 h amounts to 12×6 or 72 values for each channel. A similar calculation is performed for each of the 8 channels.

3.1.3.2. Asymmetry. A simple ratio of band values between

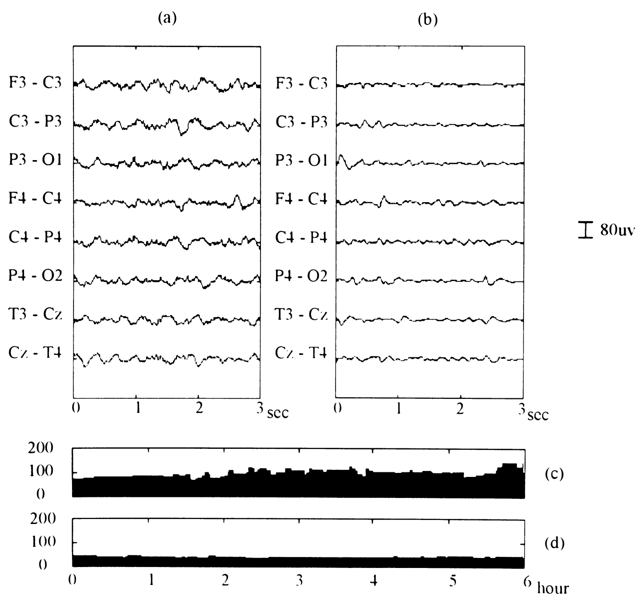

Fig. 3. Samples of 8-channel EEGs and their representation in a band array. (A) Normal amplitude. (B) Depressed amplitude. (C) Band array corresponding to EEG in A, for channel P3-O1, showing normal EEG amplitude over the 6 h interval. D. Band array for EEG in B, showing depressed amplitude sustained over the whole 6 h recording.

the corresponding channels of the left and right hemispheres can quantify the level of asymmetry. The formula for the Left/Right symmetry is:

$$R = \log\left(\frac{\text{Average band value in left}}{\text{Average band value in right}}\right) \tag{1}$$

Here, the averaged band value represents the averaged 1-to-14 band value over a 5 min period. Four Left/Right asymmetry measures corresponding to the 4 pairs of channels symmetrically located in the two hemispheres are extracted every 5 min. A 6 h EEG record is thus associated with 4 Left/Right symmetry measures, each of which contains a set of 72 values.

3.1.3.3. Front/back differentiation. The ratio of the delta band value of the posterior channel to that of the anterior channel of the same hemisphere reflects the extent of their differentiation. Two such measures are extracted, one for each hemisphere. The parameter is derived using the following formula:

$$R = \log\left(\frac{\text{Average band value in posterior}}{\text{Average band value in anterior}}\right) \tag{2}$$

where the average band value corresponds to an average of delta over a 5 min period. For each hemisphere, the front/back differentiation extracted from every 6 h record is once again described by a set of 72 values.

3.1.4. Statistical analysis

Following the above calculation, 14 quantitative measures are extracted from a 6 h EEG, 8 for amplitude, 4 for symmetry and two for front/back differentiation, containing 72 values. Interpretation of these quantitative measures is performed by comparison to values obtained from a control population.

Since it is impractical to get real ‘normal’ subjects to undergo long-term EEG monitoring in the ICU, a group of ten post-cardiac surgery patients varying in age from 6 months to 12 years, with normal post-operative long-term EEG recordings and normal neurological examination on discharge from the ICU, were carefully chosen to form the control population. Normality of the EEG was assessed by clinical experience in the context of the monitoring situation (gradual elimination of anesthetic effects and effects of sedation).

Data from the 10 patients of the control population were used to construct 14 frequency distributions of amplitude, asymmetry and front/back differentiation for each of the 8 channels and the 6 ratios (Fig. 4A). These distributions can also be constructed for a new 6 h EEG section. For a particular feature (e.g. symmetry in posterior head regions, Fig. 4B), a measure of similarity between the distribution of the control population and the distribution of the EEG being analyzed gives an estimate of the normality of that feature. We use a statistical comparison of the distributions rather than a simple comparison of mean values because the distributions for the different features have different shapes (Fig. 4), and therefore comparison of the whole distributions is a more representative measure.

The *t* statistic was used as measure of similarity. As indicated earlier, all the features extracted from the frequency band arrays were defined as the logarithm of the variable rather than the variable itself. This renders the distributions of the estimates closer to a normal distribution (Gasser et al., 1982), which facilitates the *t* statistic comparison.

For each pair of distributions, a *t* statistic value was calculated using the formula:

$$t = \frac{x_1 - x_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \tag{3}$$

where x_1, s_1, n_1 and x_2, s_2, n_2 represent the mean, the standard deviation and the sample size of the distributions from the control population and from the new EEG being analyzed. Each *t* value represents the level of normality of the corresponding measure.

The range of *t* statistic values extends from $-\infty$ to $+\infty$. For easier interpretation, they were mapped into the range of 0 to 1, with 0 indicating severe abnormality and 1 indicating normality. For amplitude, all *t* statistic values greater than zero imply amplitude normality (we only consider abnormal a *decrease* in amplitude) and are mapped to the value 1. We have observed that a *t* statistic value less than −100 indicates a very severe depression and it is therefore mapped to the value 0. The transformation is formalized as:

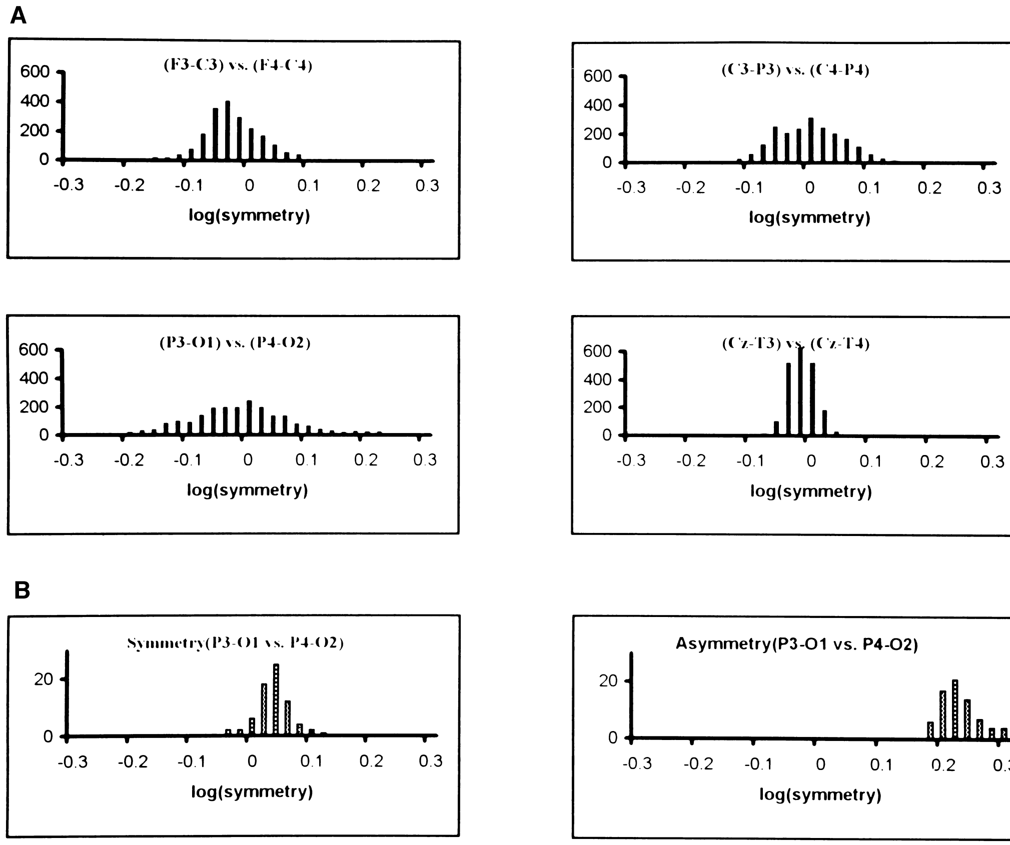

Fig. 4. (A) Distributions of the symmetry measures for the control population in the 4 pairs of channels. Note that the level of symmetry is different in the different regions; the distribution is particularly narrow in Cz-T3 vs. Cz-T4 because of the common electrode Cz. (B) Distribution of symmetry values in P3-O1 vs. P4-O2 in two patients: in the first, the distribution is quite similar to that of the control population; the second is obviously very different from the control population.

$$P_{\mathrm{Amp}} = \begin{cases} 1.0, & t \geq 0.0 \\ (100+t)/100, & -100 < t < 0 \\ 0.0, & t \leq -100 \end{cases} \tag{4}$$

Similarly, we observed that t statistic values of symmetry larger than 80 or smaller than –80 represented very important asymmetry and therefore calculated the asymmetry by using the following formula:

$$P_{\mathrm{Sym}} = \begin{cases} 0.0, & t \leq -80 \\ (80-|t|)/80, & -80 < t < 80 \\ 0.0, & t \geq 80 \end{cases} \tag{5}$$

For front/back differentiation, we found that values beyond –50 represented a total lack of differentiation and therefore used the following transformation formula:

$$P_{\mathrm{Dif}} = \begin{cases} 1.0, & t \geq 0.0 \\ (50+t)/50, & -50 < t < 0 \\ 0.0, & t \leq -50 \end{cases} \tag{6}$$

3.1.5. Reduction in the number of statistical features

For each new EEG analyzed, 14 *probability values* of normality were calculated as indicated above, 8 for amplitude, 4 for symmetry and two for front/back differentiation. These values were further reduced to 3. For amplitude, since depression in several channels is worse than in only one channel, the 8 probability values were averaged to indicate the overall normality of amplitude. Similarly, the front/back differentiation measures corresponding to the two hemispheres were averaged to evaluate the normality of front/back differentiation. For symmetry, ordinarily, the EEGer thinks an EEG with one severely asymmetrical pair of channels is more abnormal than one with several mildly asymmetrical pairs of channels. The simple average may not be enough to represent the overall estimate of symmetry. Here, the estimation of symmetry normality is achieved by

$$P_{\mathrm{Overall}} = \frac{1}{2}\left(P_{\mathrm{Sev}} + \frac{1}{3}\sum P_{\mathrm{Other}}\right) \tag{7}$$

where P_{Sev} is the probability value of the most asymmetrical channel pair and P_{Other} are the values of the other 3 pairs.

Front/back differentiation is a feature largely influenced by the patient's age. It is, therefore, necessary to transform its probability value based on the patient's age. Fig. 5 illustrates a function of front/back differentiation, S(age), which is established by the interpretation of published data by an

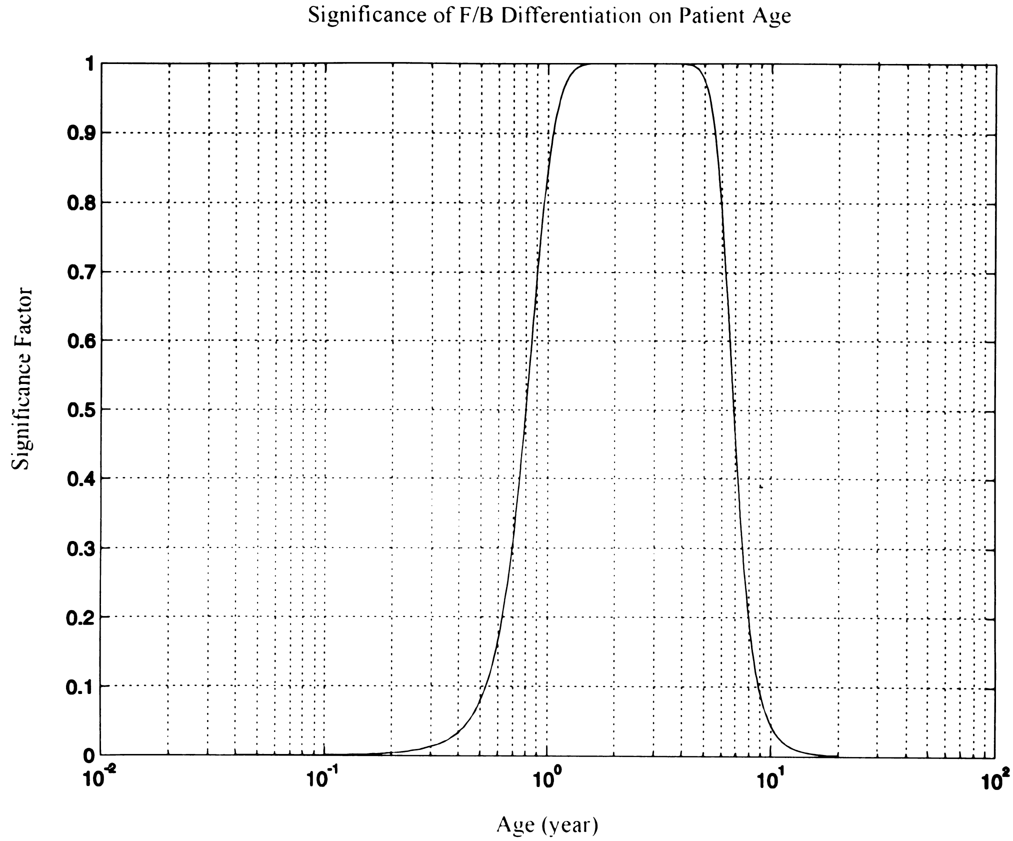

Fig. 5. Curve representing the importance attached to the presence of a front/back differentiation as a function of age. 10^{-1} is 1/10th of a year or just over 1 month; 10^{0} is one year; 10^{1} is 10 years. For young patients (below about 3 or 4 months, and for older patients (above about 10 years), the front/back differentiation feature is thus not used. Between these ages, its importance is weighed by this curve.

experienced EEGer (Slater and Torres, 1979). The transformation is described by:

$$P_{\text{New}} = \begin{cases} 1.0, & \text{age} < 4 \text{ month} \\ \lceil P_{\text{Old}} / S(\text{age}) \rceil_{1.0}, & 4 \text{ month} \leq \text{age} \leq 10 \text{ year} \\ 1.0, \quad 1.0, & \text{age} > 10 \text{ year} \end{cases} \tag{8}$$

where $\lceil\ \rceil_{1.0}$ means that the value inside $\lceil\ \rceil$ would be set to 1.0 if greater than 1.0.

3.2. Knowledge-based expert system

The knowledge-based monitor system should accept as input the 3 indices, amplitude, asymmetry and front/back differentiation for an EEG, and classify that EEG into one of several abnormality levels. It should learn the knowledge embedded in the EEGer's decision-making process on the level of EEG abnormality, organize the learned knowledge and perform the classification task under a suitable inference engine.

Six-hour EEG samples were used to train this system. Each sample was quantified by the probability values of normality varying from 0 to 1 calculated for the 3 features. The EEGer was asked to classify *each of the 3 features* and *the overall EEG* into one of 4 categories: normal, mildly abnormal, moderately abnormal and severely abnormal. As illustrated in Fig. 6, the entire learning system included two parts. The first part was designed to transform the 3 features' quantitative measures into the EEGer's interpretation of these features (normal, mildly abnormal...); it uses fuzzy logic. The second part was implemented to evaluate how to combine the different attributes to obtain a final classification: for instance what is the final classification of an EEG having normal amplitude, mildly abnormal asymmetry and moderately abnormal front/back differentiation? This second part uses neural networks.

When giving their overall assessment of an EEG, EEGers often have difficulty in deciding to which of two adjacent classes it belongs. They may give an interpretation such as 'the overall EEG is mildly to moderately abnormal.' After discussion with the EEGers, we decided to accept the 3 intermediate classes of normal-mild, mild-moderate and moderate-severe. We were therefore dealing with a total of 7 classes of abnormality.

3.2.1. Fuzzy membership learning

As described above, each of the 3 features was quantified by a value from 0 to 1. These values are associated with the EEGer's assessment, 0 implying severe abnormality and 1 implying normality. However, the boundaries between the EEGer's normality levels are not defined precisely: we can-

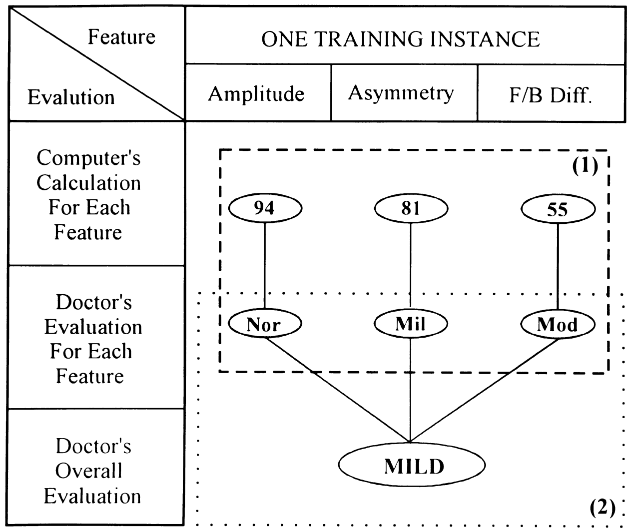

Fig. 6. Structure of the examples that are used in training the expert system. From the computer's calculations of features and the EEGer's independent evaluation of that aspect of the EEG, the fuzzy logic membership functions were obtained. From the EEGer's evaluation of each aspect of the EEG and his overall evaluation, the neural network was trained.

not assume that amplitude values from 0 to 0.25 correspond to severely abnormal, from 0.25 to 0.5 correspond to moderately abnormal, etc. We must find the process allowing the numerical values to correspond to the human classification levels. We use fuzzy logic to establish this correspondence, by learning from our data set of 150 6 h sections how the EEGer performed the classification. The details of this procedure are given in Appendix A. The procedure can be summarized with an example: assume that out of the 150 sections, there are 20 for which the amplitude feature is between 0.8 and 0.9. If the EEGer classified 19 of these 20 cases as having a normal amplitude and 1 as having a mildly abnormal amplitude, we conclude that an amplitude feature between 0.8 and 0.9 has a probability of 95% of corresponding to a normal amplitude and a probability of 5% of corresponding to a mildly abnormal amplitude.

3.2.2. Neural network learning

In the second part of the expert system, we use neural network techniques to learn how the 3 features, amplitude, symmetry and front/back differentiation are combined to obtain the EEGer's assessment of overall abnormality of the EEG.

A single layer network, as shown in Fig. 7, was established. In this network, we use 12 inputs instead of 3, each representing a different abnormality level of the 3 features. This is based on the consideration that different classifications for one feature may influence the final evaluation nonlinearly and having 12 inputs therefore reduces the complexity of the training. The Widrow-Hoff method was used to train this network. The detailed algorithm is as follows:

1. Choose a small positive value for ρ (= 0.05), the step size, and assign randomly selected small initial weights $\{W_i \in [-0.2,\ 0.2]\}$ to the 12 inputs, where i is the input index.
2. Repeat steps 3 to 6 until changes in the mean squared error, ϵ, are less than 0.001 for every 1000 iterations.
3. Randomly choose the next training examples, E, with correct output C.
4. Calculate the weighted sum,

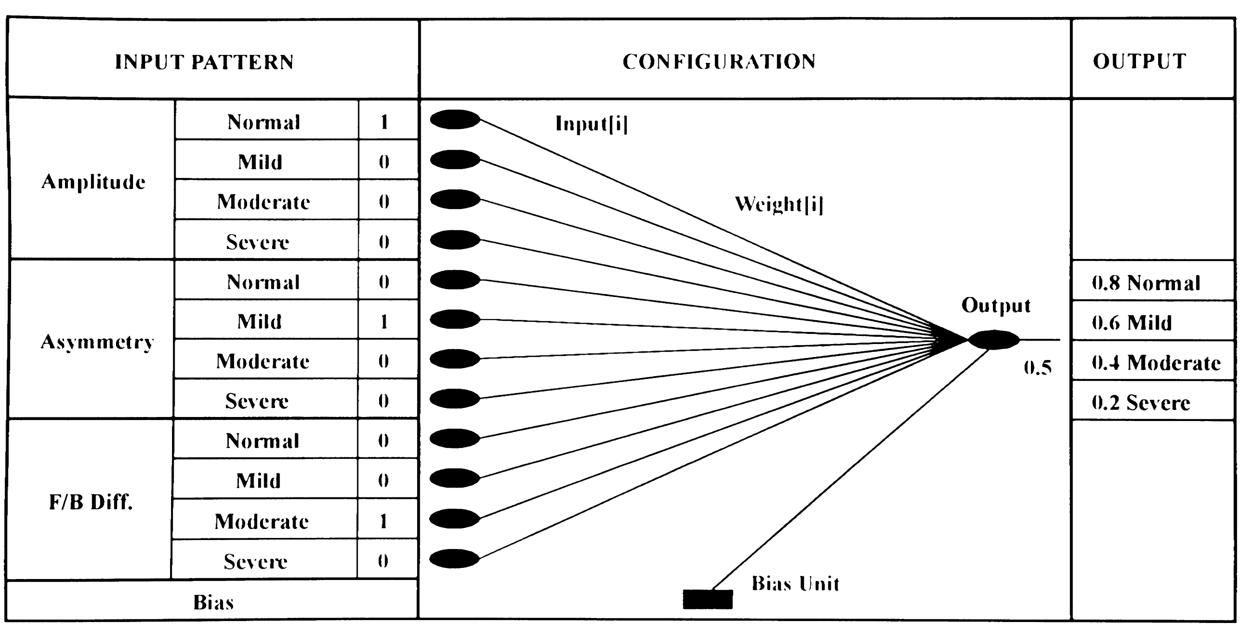

Fig. 7. Structure of the single layer network, with an example of input (the EEGer's assessment for amplitude, asymmetry and front/back differentiation are normal, mildly abnormal and moderately abnormal). With this input, the network's output is 0.5. By comparing its result with the EEGer's *overall* evaluation (mildly abnormal, 0.6), the network learns the classification by revising its weights.

$$S = W_0 + \sum_{i=1}^{12} W_i E_i$$

5. Update weights:

$$W_0^* = W_0 + \rho(C - S)$$

$$W_i^* = W_i + \rho(C - S)E_i$$

6. Update step size: $\rho^* = \rho * \zeta$, where $\zeta = 0.999$.

Several two-layer networks with 12 inputs, one output, and different numbers of intermediate cells were also implemented. They were trained by using the standard back-propagation algorithm and their generalization ability was compared with that of the single-layer network. We wrote all software for neural networks and fuzzy logic.

3.3. Training and testing data

For the purpose of training and testing the expert system, the 188 EEG sections were grouped into training and testing subsets according to the rotation method (Devijver, 1982). This method makes optimum use of the available data. A more traditional approach would have been to divide the data in two equal sets, and use one for training and one for testing. This is done when training is not automatic (when the training includes the selection of features and the development of the method itself). In our case, the method was developed with data independent from these 188 sections and its training was automatic. We could therefore use the rotation method, which makes optimum use of the data by allowing all the recordings to be used in training and all to be used in testing. The 188 sections were grouped into 6 subsets with 31 or 32 sections each (subset 1 includes records 1 to 32, subset 2 includes records 33 to 64...). For each training and testing run, 5 subsets were used to train the system (approximately 150 sections) and the remaining subset was used to test, thus maintaining the independence between training and testing data. Six training and testing runs could thus be carried out and the overall performance was the summation of the 6 runs (for instance, the first run used subsets 1, 2, 3, 4, 5 for training and 6 for testing; the second run used subsets 1, 2, 3, 4, 6 for training and 5 for testing).

For the actual implementation of this method, one could use a detector trained with *all* subjects. Its performance should be very similar to that of each detector, since each is trained with over 150 of the 188 sections. We cannot, of course, formally evaluate such a detector at present since all available data would have been used in its training.

4. Results

4.1. Fuzzy memberships

Fig. 8 illustrates the membership functions obtained for amplitude in the training series 1. The first 4 graphs delineate the estimated frequency probabilities for 12 bins ('o') and the fitted exponential functions for the 4 classes: normal, mild, moderate and severe; the fifth graph outlines the combination of the 4 fitted curves which partitions the total value space into 4 successive but overlapping subspaces. It is clear from Fig. 8 that the subspaces are not equal and could not have been guessed.

4.2. Neural networks

The single-layer and two-layer networks discussed above were trained with the Least Mean Square and Back-propagation algorithms, respectively. The generalization ability of different network models for the training and the testing data were evaluated as follows: the EEGer's interpretation for the 3 features served as input, and the output of the network was compared with the EEGer's assessment of overall EEG abnormality. Two-layer models included 2 to 32 nodes in the middle layer. Comparing the two-layer models with the single-layer model, most of the models achieve 70% correct classification for both training and testing data, and over 95% of results are within one-half level distance from the EEGer's outputs. However, the single-layer model required only 10 000 iterations for training, one-tenth of that required by the two-layer model, which indicates that the single-layer model converges much faster than two-layer models. We conclude that the single-layer model is more suitable for this learning problem.

4.3. Entire system

The classification capability of the entire system was evaluated as the cumulative result of the test data of the 6 runs corresponding to the 6 partitions of the 188 6 h EEG sections (Table 1), in which the quantitative measures of the features were used as the input of the lower part of Fig. 1, and the output of the system was compared with the EEGer's global interpretation. The single layer model was used as the final evaluation network.

In the 7×7 matrix presented in Table 1, rows represent the system's classification for the testing instances and columns correspond to the EEGer's evaluation. The boldface cells in the main diagonal represent the cases of perfect concordance between the two classifications, and the italic-face cells stand for the instances where the system's outputs were half a level away from the EEGer's classification. The results are very encouraging and most of the instances within these areas. Forty-five percent of the total testing instances are classified as the same class as the EEGers'. If we add the cases with a half-level deviation, the total percentage of instances which are considered as having acceptable results reaches 91%. In contrast, the recordings having a classification error exceeding one class are fewer than 3%.

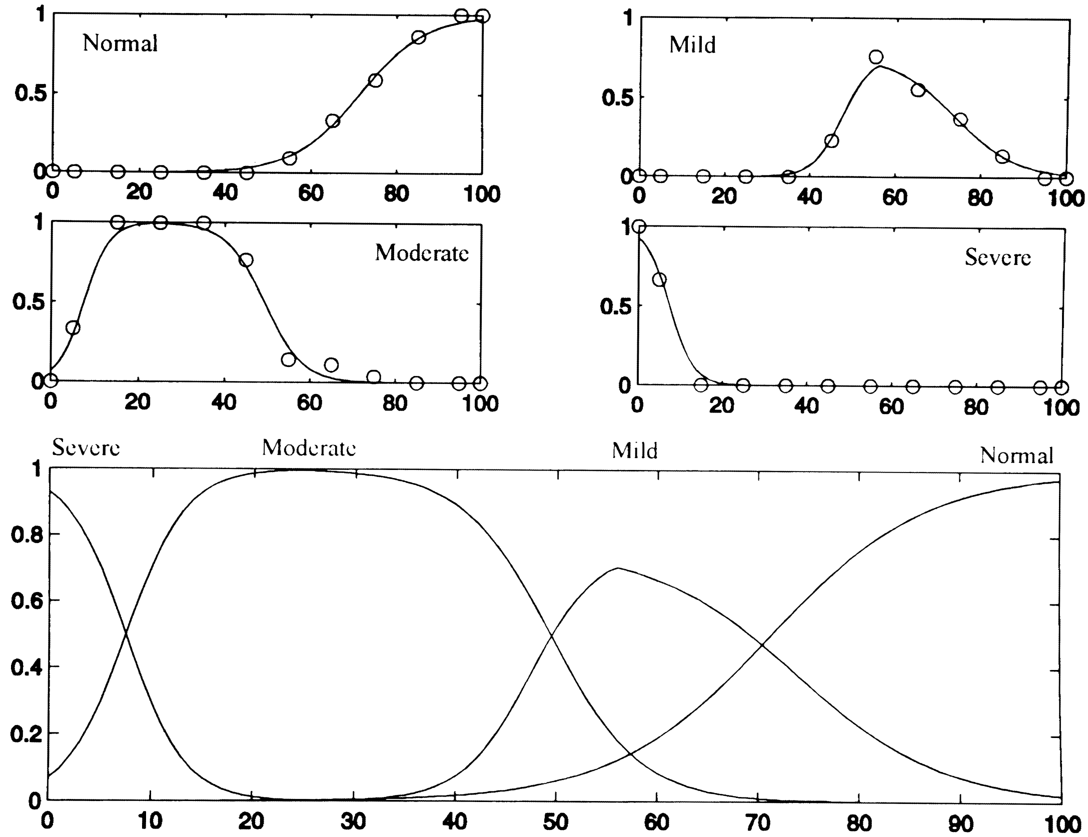

Fig. 8. Membership functions for amplitude. It is clear that the 'normal,' 'mild,' 'moderate,' and 'severe' domains have an important overlap, are not of equal width, and could not have been predicted.

5. Discussion

We have presented a relatively complex method aimed at assessing the degree of abnormality in an EEG recorded in the ICU. Our approach has been to follow a path that parallels as much as possible the human evaluation. This has the important advantage that the system could *justify* its judgment in terms that are familiar to the user: if the system finds that an EEG is moderately abnormal, it can then explain that this was caused by a mildly asymmetrical EEG combined with a moderately depressed EEG, that the asymmetry and the depression were, for instance, prominent in the left central region. The system also relies on very few thresholds and it could therefore be adapted to another type of EEG (adult for instance), or trained by other physicians, who would have their own criteria for defining mild, moderate or severe abnormality. The unbiased evaluation of a large number of EEG sections from 74 patients indicates that over 90% were classified within one half-level of the EEGer's interpretation. Only 3% were more than one level away (in fact these 5 cases out of 188 were 1.5 levels away, e.g. classified 'normal-mild' when the EEGer had called it 'moderate,' or classified 'mild' when the EEGer classified 'moderate-severe'). This

Table 1

Global evaluation of the entire system

Results	EEGer						
Computer	Normal	Normal-mild	Mild	Mild-moderate	Moderate	Moderate-severe	Severe
Normal	**37**	*8*	3	0	0	0	0
Normal-mild	*16*	**12**	*23*	3	3	0	0
Mild	3	*9*	**18**	*3*	2	1	0
Mild-moderate	0	1	*6*	**5**	*7*	0	1
Moderate	0	0	0	*8*	**9**	*1*	1
Moderate-severe	0	0	0	0	*5*	**2**	*0*
Severe	0	0	0	0	0	*1*	**1**

Areas in the diagonal with bold figures represent full agreement between EEGer and computer. Areas in italic represent one half-level of disagreement.

is a very low level of error and we have therefore achieved quite a reliable warning system. This system could be best conceptualized as made to warn *that a qualified person should examine the EEG*, and *not that the patient is deteriorating*. Current networking technology allows a qualified EEG reader to connect a remote review station to the station where the EEG is recorded and examine the EEG without necessarily having to physically go to the bedside.

Our method makes use of a variety of techniques, some of which have been used in EEG analysis, sometimes in a different context. Statistical comparison of features between a control group and a subject, as we do here for each feature, has been used in significance probability mapping (Duffy et al., 1979). Neural networks have been used mostly in the detection of epileptiform patterns (Gabor et al., 1996; Webber et al., 1994, 1996), but also to predict movement during anesthesia (Sharma and Roy, 1997). Expert systems have also been used in the detection of epileptiform events (Dingle et al., 1993). We are not aware of the use of fuzzy logic in EEG analysis.

It must be noticed that our computer method only used the 3 features of amplitude, symmetry and front/back differentiation, whereas the EEGers to whom it was compared interpreted the EEG as they do for a standard clinical EEG, taking into account all information at their disposal in the tracing. The relative success of the method indicates that these 3 features are probably of paramount importance in the clinical judgment.

From the membership function curves obtained by fuzzy logic for the 3 features (Fig. 8), it is clear that the 4 abnormality levels are not equally distributed in the value space. Some of these subspaces are narrower and some are wider. Moreover, the up-going and down-going slopes of most bell-shaped membership functions are not symmetrical. This indicates that the EEGer's assessment of abnormality for these features is a nonlinear process. In addition, for all 3 features, the normal subspace overlaps the moderate subspace (Fig. 8), which means that either the features are not representing exactly what the EEGer thinks (the amplitude feature does not reflect precisely how amplitude is seen by the EEGer), or the EEGer is not consistently judging an EEG with the same feature value.

Examination of the weights of the neural network gives insight into the decision-making process of the system, as well as into the decision-making process of the EEGer that the system imitates. The weights obtained for the 12-input single-layer network are listed in Table 2 for one of the 6 training sets (they are very similar for all sets). We can see that the differences of the weights between the successive nodes that represent different amplitude levels are greater than that for symmetry and front/back differentiation. It reveals that amplitude, which indicates the degree of EEG depression, is a more important feature than symmetry and front/back differentiation, and a variation in amplitude will greatly influence the assessment of abnormality. Table 2 also shows that the weight differences between the symmetrical and mildly asymmetrical nodes are much smaller than those between the moderately and severely asymmetrical nodes. This may reflect that transient or mild asymmetries are evaluated as having no pathological significance, as reported by Lombroso (1985).

When interpreting pediatric EEG, it is important to take the patients' age into account since the EEG of children varies extensively from birth up to about 10 years of age. Ideally, it would be better to define age intervals within which the characteristics of a normal EEG are not expected to vary extensively, such as less than 4 months, 4 to 12 months, 1–4 years, 4–10 years and more than 10 years,. Independent 'normal' control populations for these age intervals could then be established and the system could be trained and tested for each interval. However, in practice, the realization of this aim is difficult since many recordings for every age interval have to be collected.

Besides the amplitude, symmetry and front/back differentiation, some other EEG features could eventually be incorporated into this system, such as lack of spontaneous variation over a long period of time, which Bricolo et al. (1973) have associated with a poor prognosis, or generalized slowing. These are clearly features that are of importance in assessing the overall trend of EEG activity. A feature representing them could be developed and be integrated to the expert system in a similar way to the features used here.

In addition to background abnormalities, seizures provide important prognostic information. Recent long-term monitoring studies reveal that seizures among patients with acute brain injury are more common than previously recognized. Jordan (1993) reported a significant incidence, up to 35%, of seizures in an analysis of 124 monitored patients. Lowenstein and Aminoff (1992) described 47 comatose patients, in 80% of whom EEGs suggested nonconvulsive seizures. Therefore, the combined use of the background EEG monitoring system presented here and computer-based automatic seizure detection systems (Gotman, 1990; Gotman et al., 1997a,b) will provide more complete diagnostic information.

The present monitoring system was built to receive a 6 h

Table 2

Weights obtained after training the 12-input single-layer network with the first data set

Feature	Amplitude				Symmetry				Front/Back Differentiation			
Class	Nor	Mild	Mod	Sev	Nor	Mild	Mod	Sev	Nor	Mild	Mod	Sev
Weight	0.30	0.14	–0.05	–0.18	0.15	0.12	0.04	–0.15	0.11	0.06	0.09	–0.01

The separation between the weights of a particular feature reflects the importance of the feature.

section of unprocessed EEG as input and determine its abnormality level. The reason for using a 6 h section as the analysis unit is that it provides information that is not affected by short-term changes or transient situations. This information will have to be updated regularly in order to provide the critical aspect of trend, which is of such clinical importance. This could for instance be achieved by providing a new statement every 30 or 60 min, relating to the last 6 h. We could also make this analysis for the last 6 h and, separately, another analysis of a similar type for the last 30 min, thus providing information on *shorter-term abnormalities* in the context of a longer-term record. Nuwer (1994) has presented an extensive review of the importance of EEG monitoring in the neurosurgical ICU, emphasizing the importance of detecting fluctuations of background abnormalities, but also its difficulty because of the need to have interpretation by qualified personnel. The assessment of EEG trend is also used for high-risk patients to determine the adequate depth of anesthesia (Halimi et al., 1990). More often, the EEG trend is monitored for changes in the CNS function and condition during surgery. In a study of 92 patients monitored by an EEG trend-analyzer during carotid surgery, the EEG trend-analyzer showed a sensitivity of 100% for ischemic events (Loeprecht et al., 1985). The question of EEG monitoring during carotid endarterectomy was also extensively studied by Chiappa et al. (1979). In this case, however, the situation is different from the system we propose because changes may be rapid, but a qualified interpreter is always present.

In clinical use, non-pathological factors such as medication or patient's anesthetic level need also be taken into account when interpreting results of the monitoring system. Drug-induced amplitude attenuation, burst suppression, or seizures have to be excluded. High-frequency components introduced by sedatives should also be considered carefully.

Our system was not designed for a rapid detection of abrupt changes in the EEG. It is based on the premise that a qualified person must interpret the EEG before a conclusion is reached; in practice, this person is not likely to be near the ICU and may not always be available on short notice. That person's response time may therefore be of the order of several minutes. We view this system as a replacement for *no recording at all*, or short recordings at several hours interval, a situation in which the response time is very long. It is obvious that if we could also detect rapidly and reliably short-term changes, the system would be even better. It is also obvious that EEG monitoring would be best combined with the monitoring of other variables, such as heart rate, blood pressure, intra-cranial pressure (Pronk, 1986), as well as with the measurement of somatosensory evoked potentials, a sensitive indicator of the integrity of cortical and subcortical function (Chiappa, 1990).

Acknowledgements

We are grateful to the technologists of the EEG laboratory of the Montreal Children's Hospital, Carol Leitner, Elyse Chevrier and Sophie Lalonde, for their active collaboration in this project. This work was supported in part by a grant from the Medical Research Council of Canada and Stellate Systems through the University-Industry program of the Medical Research Council (Grant MRC-UI-11514).

Appendix A Fuzzy logic

We describe here the details of the use of fuzzy logic to allow the computer system to imitate the human decision regarding normality of a particular feature. These different normality levels can be formalized as fuzzy sets.

Fuzzy memberships are interpreted as the notion of frequency probability. A frequency probability $P(A)$ is a number around which the frequency of the occurrence of the event A $V_N(A) = K_N(A)/N$ oscillates, where $K_N(A)$ is the number of occurrence of A in N trials. By definition, $P(A) = \lim_{N\to\infty} V_N(A)$. The membership function of the elements u in the fuzzy set S is identified with the probability density $\mu_S(u) = P(\chi_S(u) = 1) = P_S(u)$, where

$$\chi_S(u) = \begin{cases} 1, \ldots & u \in S \\ 0, \ldots & u \notin S \end{cases} \tag{9}$$

A practical algorithm was developed to evaluate the frequency probabilities $V_N(A)$. Consider that there are 150 6 h EEG sections used as training examples, each of which has a data structure shown in Fig. 6. To build the membership curve for 'normal amplitude,' the data used here were the 150 quantitative values for EEGs'amplitude and the corresponding EEGer's classifications. The detailed algorithm

Table 3

Evaluation of the frequency probabilities of 'normality' from the training data

B_i	≤0	1–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–99	≥100
AN_i	1	6	1	3	9	13	21	18	27	22	22	14
ANN_i	0	0	0	0	0	0	2	6	16	19	22	14
Fp_i	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.33	0.59	0.86	1.00	1.00

The top row shows the intervals into which the amplitude values are divided. The second row shows the total number of observations for each amplitude interval. The third row shows the observations that were called 'normal' by the EEGer. The last row shows, for each amplitude interval, the proportion of observations that were called 'normal.'

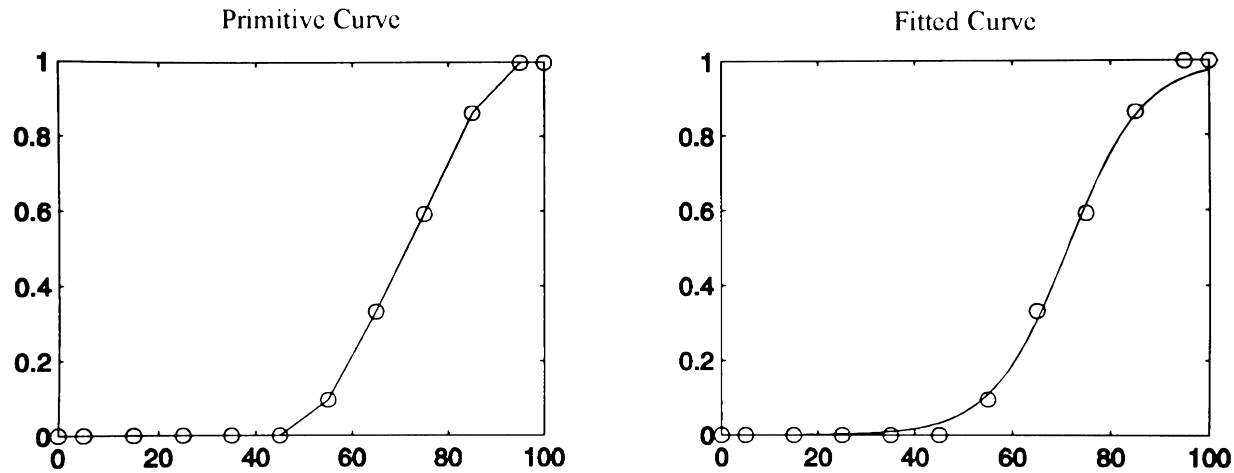

Fig. 9. (A) Primitive membership curve for 'Normal amplitude'. For each 'o,' the abscissa represents one of the 12 possible values of the amplitude feature (= 0, 1–10, 11–20...), and the ordinate represents the proportion of EEGs called normal by the EEGer. For example, the 8th point indicates that 33% of EEGs with an amplitude feature between 60 and 70 were judged to have a normal amplitude (see). (B) Same curve as in A, but fitted with an exponential S-shaped curve to obtain a smooth function.

for evaluating the frequency probability $P(\chi_S(u) = 1)$ is:

1. Divide the range of the computer's value, from 0 to 100, into 12 bins: = 0, 1–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, 90–99, = 100.
2. For every bin B_i ($i \in$ ([0,11]), calculate the number of training examples (AN_i) having a computer's value of amplitude in bin B_i. For example, if among 150 training examples, there are 22 for which the computer's calculations of amplitude is in the range [80,90], then $AN_9 = 22$.
3. For every bin B_i ($i \in$ ([0,11]), in the training examples which account for AN_i, calculate the number of events (ANN_i) for which the EEGer's evaluation of amplitude is normal. If in the 22 examples above, 19 are classified by the EEGer as having normal amplitude, then $ANN_9 = 19$.
4. For every bin B_i ($i \in$ ([0,11]), calculate its frequency probability (V_i) for 'normal amplitude': $V_i = ANN_i/AN_i$. For the example above, V_9 of normal amplitude = 19/22 ≈ 0.86.

Table 3 shows the AN_i, ANN_i and V_i calculated for 'normal amplitude.' By plotting V_i against the twelve bins, we acquire the primitive membership curve (Fig. 9A) for 'normal amplitude.'

Since $V_N(A)$ equals $P(A)$ only when N approaches infinity, an estimation error is inevitably introduced into this primitive membership curve. Therefore a better membership function can be achieved by using a well-defined function to fit the primitive curve. Here we choose the S-shaped reference function due to the fact that it resembles the normal probability distribution. Fig. 9B shows the original membership curve and its fitted exponential function by using Levenberg-Marquardt nonlinear optimization method (Press et al., 1992).

Here we constructed the fuzzy membership functions only for the 4 main classes, not the 3 intermediate classes. The reason is that the available training examples were too small to construct membership functions for 7 classes, and the embedded noise could have made the construction impossible. Therefore, any attribute evaluated by the EEGer as an intermediate class was set to the nearest lower class. For instance, if the amplitude of an example was evaluated as mild-normal, it was transformed to mild to implement this strategy.

In this step, the 3 values (from 0 to 1) calculated for the amplitude, asymmetry and front/back differentiation were thus converted to the system's best estimate of how the EEGer would have classified the corresponding aspects of the EEG.

References

Bickford, R.G., Billinger, T.W., Fleming, N.I. and Stewart, F. The compressed spectral array (CSA). A pictorial EEG. Proc. San Diego Biomed. Symp., 1972, 11: 365–370.

Bricolo, A., Turella, G., Ore, G.D. and Terzian, H. A proposal for the EEG evaluation of acute traumatic coma in neurosurgical practice. Electroenceph. clin. Neurophysiol., 1973, 34: 789.

Chiappa, K.H., Burke, S.R. and Young, R.R. Results of electroencephalographic monitoring during 367 carotid endarterectomies. Use of a dedicated minicomputer. Stroke, 1979, 10: 381–388.

Chiappa, K.H. (Ed.), 1990. Evoked Potentials in Clinical Medicine, 2nd ed. Raven Press, New York.

Devijver, P.A., 1982. Pattern Recognition: A Statistical Approach. Prentice Hall, Englewood Cliffs, NJ.

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

Duffy, F.H., Burchfiel, J.L. and Lombroso, C.T. Brain electrical activity mapping (BEAM): a method for extending the clinical utility of EEG and evoked potential data. Ann. Neurol., 1979, 5: 309–321.

Emmerson, R.G., Chiappa, K.H., 1988. Electrophysiologic monitoring. In: Ropper, A.H., Kennedy, S.K., Zervas, N.T. (Eds.), Neurologic and Neurosurgical Intensive Care, 2nd ed. Aspen, Rockville, MD, p. 13.

Gabbouj, M., Coyle, E.J. and Gallagher, N.C. Jr. An overview of median and stack filtering. Circuits Systems Signal Process., 1992, 11: 7–45.

Gabor, A.J., Leach, R.R. and Dowla, F.U. Automated seizure detection using a self-organizing neural network (ANN). Electroenceph. clin. Neurophysiol., 1996, 99: 257–266.

Gasser, T., Bacher, P. and Mocks, J. Transformation towards the normal distribution of broad band spectral parameters of the EEG. Electroenceph. clin. Neurophysiol., 1982, 53: 119–124.

Gotman, J. Automatic seizure detection: improvements in evaluation. Electroenceph. clin. Neurophysiol., 1990, 76: 317–324.

Gotman, J., Flanagan, D., Zhang, J. and Rosenblatt, B. Automatic seizure detection in the newborn: Methods and initial evaluation. Electroenceph. clin. Neurophysiol., 1997, 103: 356–362.

Gotman, J., Flanagan, D., Rosenblatt, B., Bye, A. and Mizrahi, E.M. Evaluation of an automatic seizure detection method for the newborn EEG. Electroenceph. clin. Neurophysiol., 1997, 103: 363–369.

Halimi, P., Gozal, Y., Cphen, M. and Gozal, D. Computerized electroencephalographic monitoring in anesthesia. Cahiers d'Anesthesiologie, 1990, 38: 309–317.

Jordan, K.G. Continuous EEG and evoked potential monitoring in the neuroscience intensive care unit. J. Clin. Neurophysiol., 1993, 10: 445–475.

Loeprecht, H., Wolfle, K., Heudorfer, J. and Reich, H. Can EEG monitoring with the Trend Analyzer replace stump pressure measurement in carotid surgery?. Langenbeck's Arch. Chir., 1985, 366: 333–338.

Lombroso, C.T. Neonatal polygraphy in full-term and premature infants: a review of normal and abnormal findings. J. Clin. Neurophysiol., 1985, 2: 105–155.

Lowenstein, D.H. and Aminoff, M.J. Clinical and EEG features of status epilepticus in comatose patients. Neurology, 1992, 42: 100–104.

Nuwer, M.R. Electroencephalograms and evoked potentials: Monitoring cerebral function in the neurosurgical intensive care unit. Neurosurg. Clin. North Am., 1994, 5: 647–659.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (Eds.), 1992. Numerical Recipes in C, 2nd ed. Cambridge University Press, Cambridge, pp. 683–688.

Pronk, R.A.F., 1986. Peri-operative monitoring. In: Lopes da Silva, F.H., Storm van Leeuwen, W., R :émond, A. (Eds.), Handbook Of Electroencephalography and Clinical Neurophysiology, Vol. 2, Clinical Applications Of Computer Analysis Of EEG And Other Neurophysiological Signals. Elsevier, Amsterdam, pp. 93–130.

Sharma, A. and Roy, R.J. Design of a recognition system to predict movement during anesthesia. IEEE Trans. Biomed. Eng., 1997, 44: 505–511.

Slater, G.E. and Torres, F. Frequency-Amplitude Gradient: A new parameter for interpreting pediatric sleep EEGs. Arch. Neurol., 1979, 36: 465–470.

Watanabe, K., Miyazaki, S., Hara, K. and Hakamada, S. Behavioural state cycles, background EEGs and prognosis of newborns with perinatal hypoxia. Electroenceph. clin. Neurophysiol., 1980, 49: 618–625.

Webber, W.R., Litt, B., Wilson, K. and Lesser, R.P. 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.

Webber, W.R., Lesser, R.P., Richard, R.T. and Wilson, K. An approach to seizure detection using an artificial neural network (ANN). Electroenceph. clin. Neurophysiol., 1996, 33: 106–112.