DALE E. OLSEN, JOHN A. CRISTION, and CHARLES W. SPAUR

AUTOMATIC DETECTION OF EPILEPTIC SEIZURES USING ELECTROENCEPHALOGRAPHIC SIGNALS

Many epileptic patients require prolonged monitoring by electroencephalography while waiting for seizures or seizure discharges to occur. Interpretation of the massive record generated by one patient would place an unreasonable demand on the neurologist; moreover, neurologists often provide different interpretations. For these and other reasons, the development of a reliable automatic seizure detection algorithm has been a goal of researchers for years. One factor that makes formulating an algorithm difficult is that data from an electroencephalogram do not form a stationary time series. By exploring a variety of techniques, such as autoregressive modeling, discriminant analysis, clustering, and artificial neural networks, promising algorithms have been developed.

INTRODUCTION

The Problem

About 1% of the U.S. population has epilepsy, and 250,000 to 500,000 of these epileptics have seizures that are not controlled by standard medications. For these patients, an electroencephalogram (EEG) recorded from up to sixty-four locations on the scalp is essential to the clinical determination of an appropriate drug therapy or surgical procedure. Many patients require inpatient EEG recordings, for it is often impossible to obtain sufficient EEG data using routine outpatient testing methods. Inpatient EEG monitoring must often be conducted until seizure discharges occur. Since a megabyte of EEG data is collected for each patient in a one-minute period and monitoring continues for days or weeks, data recognition and reduction algorithms are of great importance.

A 24-hour-per-day Epilepsy Monitoring Unit with four beds is currently operating at the Johns Hopkins Hospital (JHH). The unit brings with it the need to develop and implement automatic real-time spike and seizure detection algorithms for the facility's computers and affords an opportunity to develop and test methods for selectively recording and assessing seizure disorders.

The Objective

Our objective is to use multivariate time series (up to sixty-four correlated time series or data channels) generated by differencing EEG measurements at various locations on the head to detect seizures automatically. This has been the goal of researchers for many years.[1–5] Normally, only seizure data are desired for investigation; false detections can reasonably be screened out by human analysis, but missed seizures are lost. The interest, therefore, is in detecting seizures with high statistical power even at the expense of higher-than-usual false alarm rates, which are also known as type I errors.

The Literature and the Approach

Several authors have summarized the considerable work on the automated analysis of EEG signals.[5–8] An important part of developing a seizure detection algorithm is to translate the complex and subjective methods of EEG analysis currently used into mathematical terms. Some attempts to mimic subjective analysis methods using a computer have met with limited success.[9] A stationary series is assumed in most time series techniques (spectral analysis and autoregressive-moving-average [ARMA] modeling, for example), and some analysts have made the same supposition about EEG data.[10] Although this assumption is incorrect,[11–13] the time series techniques nevertheless suggest modeling and analysis approaches useful for reducing EEG data.

The approach used here will be to fit the parameters of an autoregressive (AR) model to a short segment of EEG data from selected scalar time series or data channels and to combine the model parameters with other signal features to obtain a feature vector (a vector representing the model parameters and signal characteristics). The elements of this vector can be used to classify time series segments. (Similar techniques have been applied to EEG data by others[6,14–17] for different purposes.) The feature vector is used as a descriptor of the electrical state of the patient's brain. The neurologist will read the EEG and will identify the patient's brain electrical state or EEG signature, which in turn will be associated with a range of feature vector values. Values of the feature vector associated with benign (normal) and seizure states are of interest.

The seizure detection algorithm is to be used for processing time series data from a single EEG channel (the potential difference between two electrodes) and identifying the patient's condition as represented by EEG signatures. States of interest include normal/awake, normal/asleep, and seizure. If multiple channels were analyzed simultaneously, far more complicated processing techniques would be required,[18,19] making real-time analysis

difficult. If each channel were analyzed independently, however, the patient state could be determined by using either a voting technique or by combining features from different channels.

An AR time series model is briefly described in the discussion that follows. The model is applied to the data to form a feature vector of the parameters used to characterize the time series, and the feature vector can thus be used to identify the patient state. Empirically derived statistical models (discriminant analysis) are used to classify the time-varying characterizations and are ultimately used to identify the state of the patient. A neural network approach is also being used as an alternative to the statistical methods. Both of the algorithms have been developed and tested using data provided by Ronald P. Lesser at JHH. The results of the testing and the treatment of the differences between patients are described.

DATA REDUCTION

Time Series Model

The EEG data are viewed as coming from an AR time series model in the form

$$\mathbf{Z}_t = \Theta_1 \mathbf{Z}_{t-1} + \Theta_2 \mathbf{Z}_{t-2} + \ \ldots \ + \Theta_{10} \mathbf{Z}_{t-10} + \epsilon_t$$

$$(t = 11, 12, \ldots, m) , \qquad (1)$$

where $\mathbf{Z}_t$ is the value of the selected channel of EEG data at time t, Θ_j denotes the parameters used to characterize the series, and ϵ_t is the model prediction error and is normally distributed with mean 0 and variance σ^2; the ϵ_t's are independent. Least-squares techniques are used to generate estimates $\hat{\Theta}_j$ and $\hat{\sigma}$ of Θ_j and σ (standard deviation). This model will be fitted to a short segment of data to create a feature vector of parameters used to characterize that segment. Ten parameters (AR[10]) were used because this number of parameters provided the smallest feature vector that could reliably detect seizures for the data sets first considered. The Θ_j parameters contain frequency information.

Other parameters can also be computed and added to the feature vector. These other parameters have been shown to be important in distinguishing between seizures and other states. They include the skewness (third moment about the mean) and kurtosis (fourth moment about the mean) of the data. These parameters were not used initially, however.

Feature Vectors

The parameter estimates $\hat{\Theta}$ and $\hat{\sigma}^2$ are used to characterize a segment of data from a scalar time series or channel. The process can be repeated for a series of segments so that the data from a channel are characterized by a series of parameter estimates. Let $\hat{\Theta}_i$ and $\hat{\sigma}_i^2$ be represented as

$$\mathbf{X}_i = \begin{bmatrix} \hat{\Theta}_i \\ \hat{\sigma}_i^2 \end{bmatrix} \qquad (i = 1, 2, \ldots, n) , \qquad (2)$$

where n is the number of time segments related to one patient state or signature category. The vectors $\mathbf{X}_i$ are the feature vectors and are used to generate the analysis results described in the next section. The feature vectors can be used in discriminant analysis[20,21] or in a neural network[22,23] to determine the patient's state.

DISCRIMINANT ANALYSIS APPROACH

After obtaining a collection of feature vectors from a patient in a given state, a statistical model for the patient is found by calculating the mean and variance of the feature vectors. Figure 1 is a plot of artificial two-dimensional feature vectors sampled from two normal distributions (one marked with circles and the other with squares) and illustrates the discrimination procedure. Although the circles and the squares cannot be discriminated using a single feature, using the joint probability distribution (both features) makes separating the two populations relatively simple. For example, if Feature 1 has a value of 1.2, it is not possible to determine the population from which the feature was sampled using this information alone. If, in addition, it is known that Feature 2 has a value of 1, then it is likely that the features were sampled from the population denoted by circles. To make the modeling easy, a normal distribution of the data is assumed in statistical discriminant analysis.

Sometimes the distribution of $\hat{\Theta}_i$ can reasonably be modeled using a normal distribution; however, the distribution of σ^2 is highly skewed. A transformation of this parameter is needed to make it approximate a normal distribution. The transformation is given by log(log σ). The empirical distribution of the transformed parameters is more nearly normal. Values of σ^2 have always been larger than 1 for the data used so that the transformation has

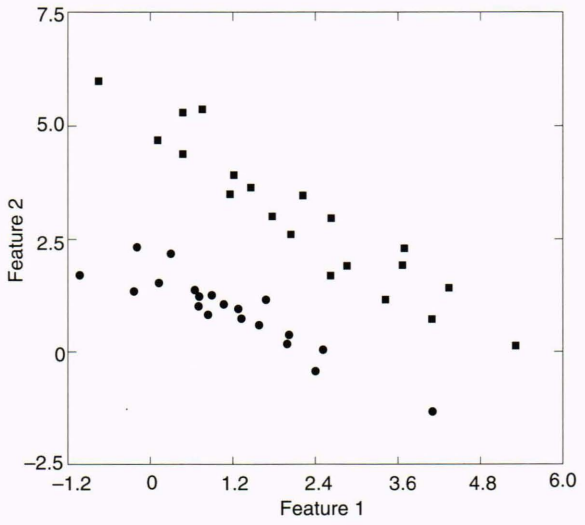


Figure 1. Sample plot of feature vectors.

always been well defined. Let $\hat{\phi}$ be the transformed value of σ^2 and let the feature vector be defined instead as

$$\mathbf{X}_i = \begin{bmatrix} \hat{\Theta}_i \\ \hat{\phi}_i \end{bmatrix} \quad (i = 1, 2, \ldots, n) . \tag{3}$$

Then let

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \tag{4a}$$

and

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} [(\mathbf{X}_i - \hat{\mu})(\mathbf{X}_i - \hat{\mu})^t] , \tag{4b}$$

where $\hat{\Sigma}$ is the variance-covariance matrix and the superscript t is the vector transpose operator.

It is assumed that when the patient is again in the same state, the new feature vectors will also be $N(\hat{\mu}, \hat{\Sigma})$-distributed (normal distribution). Correlations between data segments are ignored, since the correlations do not cause a problem in a large sample of feature vectors and because it is not apparent how to treat them reasonably.

Determining the Electrical State of the Patient's Brain

If the patient is in state j ($j = 1, 2, \ldots, J$), it is expected that a new feature vector $\mathbf{X}_i$ will be distributed $\mathbf{X}_i \sim N(\hat{\mu}_j, \hat{\Sigma}_j)$ for some j (where the subscript j has been added to distinguish different patient states). For each $\mathbf{X}_i$, the patient state is determined by finding which of the J possible states is most likely to have produced the observed feature vector. This selection is accomplished by evaluating the J values of -2 log likelihood (normality is assumed) using

$$\log n|\hat{\Sigma}_j| + (\mathbf{X}_i - \hat{\mu}_j)^t \ \hat{\Sigma}_j^{-1} \ (\mathbf{X}_i - \hat{\mu}_j) , \tag{5}$$

and then choosing the smallest (most likely) value. Since one value is always smallest, a patient state is always selected.

Clustering for Normality

The techniques just described can be used to separate seizure and normal data for time segments where the patient's normal state is stable. Since the patient's normal state varies widely, however, the feature vectors are not necessarily multivariate normal (Gaussian). Nonparametric discriminant approaches such as the one described by Rauch[24] should be investigated. The approach used here divides the feature vectors into subsets that have a more nearly normal distribution, thus enabling improved discrimination.

A series of software system cluster procedures,[25] which are procedures for grouping common data types, were used to see if they could separate the data into normal and seizure subsets. The normal and seizure feature vectors were first combined into a single data set and then clustered to see if any of the algorithms would separate the vectors into normal and seizure subsets without training. Most of the procedures combined seizure and normal data and did not produce useful clusters. This result is not surprising, since many of the elements of the feature vector provide little help in discrimination but contribute to the clustering. Ward's technique[16] separated seizure and normal data, for the most part, and plots revealed that the transformed residual variance ϕ and the first autoregressive parameter AR1 (Θ_1) are the most important in discrimination. Since this method separated seizure and normal data, it was selected as the method of clustering. Some editing was used to form subgroups of normal feature vectors.

The discrimination procedures then were asked to identify which of several populations most likely produced the feature vector. For data sets with a sufficiently large collection of seizure data, the seizure feature vectors were divided into two clusters; the normal sets were divided into three or four clusters. The discrimination procedures were usually able to identify the cluster group as well as the state of the patient correctly.

Figure 2 is a plot of the first autoregressive parameter AR1 (Θ_1) versus the transformed residual variance ϕ in which the parameters have been clustered into three normal data groups and three seizure groups. The clustering is based on more than the two parameters shown. A similar plot with additional time points would show seizure and normal points scattered over much more area. Discrimination would require more parameters.

Agreement Probabilities

A particular feature vector could be from any of several models ($\hat{\mu}_j, \hat{\Sigma}_j$), or it could be an unlikely outcome of any of the models. To quantify how well the classification procedure "likes" the choice of state, a χ^2 probability provides the likelihood that the observation is farther away from the mean given that the sample is from the jth model. The probability is computed using

$$P_{ij} = \Pr[\ \chi^2_{(11)} > (\mathbf{X}_i - \hat{\mu}_j)^t \ \hat{\Sigma}_j^{-1} \ (\mathbf{X}_i - \hat{\mu}_j)] , \tag{6}$$

where $\chi^2_{(11)}$ is a random variable with 11 degrees of freedom ($\mathbf{X}_i$ has eleven elements), Pr is the probability operator, and the superscript t is the vector transpose operator. If the probability P_{ij} is more than 0.05, then the vector resembles the model; values farther from zero indicate even more consistency between the model

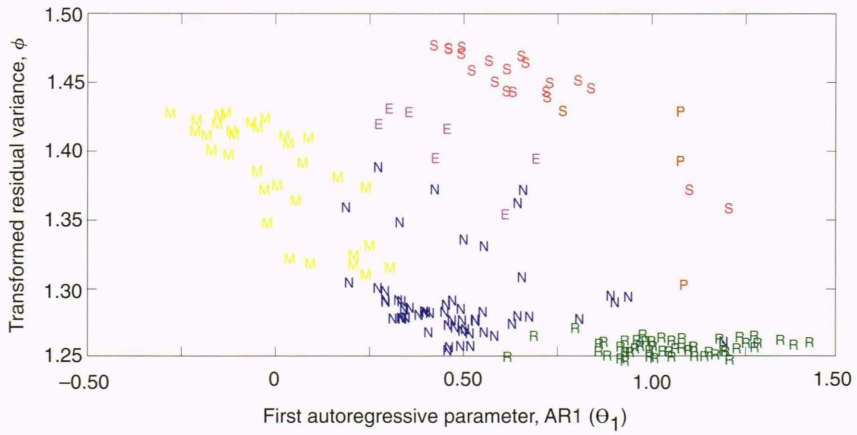


Figure 2. Plot of the first autoregressive parameter AR1 (Θ_1) versus the transformed residual variance ϕ. The parameters have been clustered into three normal data groups (N = normal, M = muscle artifacts, R = relaxed state) and three seizure groups (E = early seizure, S = seizure, P = postseizure).

prediction and the observed feature vector. If the probability for the jth model is less than 0.001, then it is concluded that the feature vector is different for the jth model. The χ^2 probabilities will be referred to here as agreement probabilities.

Patient Differences

Because of individual variations, the statistical characteristics of EEG's were expected to vary greatly between patients. The models developed for one patient were used to classify the data from different patients partly to verify the expected differences. Although the agreement probabilities indicated that the data were different, the algorithms often correctly classified the data. In most cases, seizures were correctly indicated. Apparently, the patients chosen had disorders that manifest themselves in similar ways.

Calibration Using Discriminant Analysis

Models can be generated for two patients in similar states, and the relationship between these models can be determined. If a large library of models exists for the first patient, the relationship can be used to calibrate the library for use in analyzing new data from the second patient. This procedure could be used to identify the first seizure recorded in a new patient.

Suppose $\hat{\boldsymbol{\mu}}_{1j}$, $\hat{\Sigma}_{1j}$ represents the model for the first patient in state j and $\hat{\boldsymbol{\mu}}_{2j}$, $\hat{\Sigma}_{2j}$ represents the model of the second patient in a similar state. State j could be defined as a relaxed/awake state in the Epilepsy Monitoring Unit shortly after the patient enters the unit. A feature vector from the first patient having a distribution

$$\mathbf{X}_{1i} \sim \mathrm{N}(\hat{\boldsymbol{\mu}}_{1j}, \hat{\Sigma}_{1j}) \tag{7}$$

can be transformed into $\mathbf{X}_{2i}$ so that

$$\mathbf{X}_{2i} \sim \mathrm{N}(\hat{\boldsymbol{\mu}}_{2j}, \hat{\Sigma}_{2j}) \tag{8}$$

using

$$\mathbf{X}_{2i} = \hat{\Sigma}_{2j}^{1/2}\, \hat{\Sigma}_{1j}^{-1/2}\, (\mathbf{X}_{1i} - \hat{\boldsymbol{\mu}}_{1j}) + \hat{\boldsymbol{\mu}}_{2j}\,. \tag{9}$$

In a similar way the entire library for the first patient can be calibrated to create a library for the second patient. The kth modeling elements would be transformed using

$$\mathbf{X}_{2k} = \hat{\Sigma}_{2j}^{1/2}\, \hat{\Sigma}_{1j}^{-1/2}\, (\hat{\boldsymbol{\mu}}_{1k} - \hat{\boldsymbol{\mu}}_{1j}) + \hat{\boldsymbol{\mu}}_{2j} \tag{10a}$$

and

$$\hat{\Sigma}_{2k} = \hat{\Sigma}_{2j}^{1/2}\, \hat{\Sigma}_{1j}^{-1/2}\, \hat{\Sigma}_{1k}\, \hat{\Sigma}_{1j}^{-1/2}\, \hat{\Sigma}_{2j}^{1/2}\,. \tag{10b}$$

The library of mean vectors $\hat{\boldsymbol{\mu}}_{1k}$ is transformed in the same way feature vectors would be transformed.

Limited testing of the calibration procedure shows that it has promise, and further refinements are expected. Perhaps the calibrating transformations should be built on the basis of more than one set of matching models. Methods of matching patient states with library states must be tested. An extensive library of both normal states and seizure states (seizure states vary with patients) will be needed.

NEURAL NETWORK APPROACH

An artificial neural network also uses feature vectors but provides a very different approach for separating seizure and normal data. Neural networks resemble their biological counterpart only slightly, but they are similar in that information is stored and processed in a distributed manner.

The neural network we used is referred to in the literature as a feed-forward network with backward propagation of errors. Such networks are many times called "back-prop" nets.[22] In this type of network, information (the feature vectors) is processed as it propagates through the network to produce a number between 0 and

1. During training, errors are determined by comparing the network output with a scalar target value used to characterize the patient state (0 = normal, 1 = seizure [see Fig. 3]). The accumulated error over all of the output vectors is then propagated backward through the network, and adjustments in the connection strengths between the network nodes are made. By iterative presentation of input and output vector pairs followed by adjustments of the associated connection weights, a system of connections is derived that models the input/output vector relationships. According to Lippman,[23] the model derived from this process is nonparametric, and it makes weak assumptions about the shapes of underlying parametric distributions.

Feed-Forward/Back-Propagation Network Description

The artificial neural network nodes used have predetermined connections as shown in Figure 3. The strengths of these connections were changed as the network was trained using a gradient search. The input feature vector is presented to the network by feeding one component of the vector into each input node in parallel. The input nodes in this network implementation pass the input value upward unchanged. The data values are then amplified or inhibited by the connection weights associated with each input-layer to hidden-layer connection. The input to a hidden node j is thus given by

$$I_j = \sum_{i=1}^{12} W_{ji} O_i , \tag{11}$$

where W_{ji} is the weight associated with the connection from input i to hidden node j, O_i is the output value of input node i, and, by definition, $O_{12} = 1$ (learnable bias). Each hidden node then computes an activation function or output between 0 and 1 given by

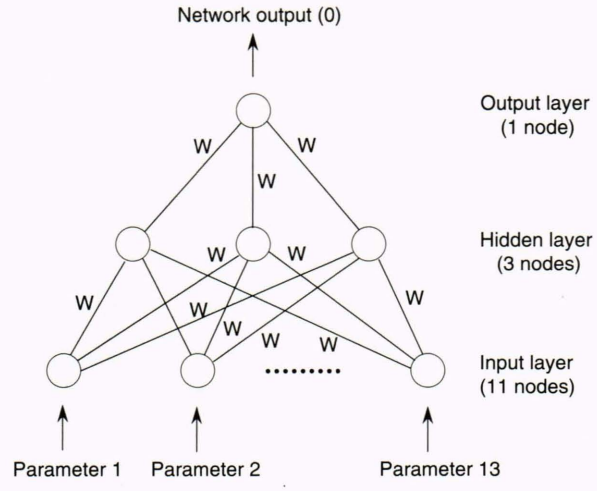


Figure 3. Neural network structure (W = connection weight).

$$f(I_j) = \frac{1}{1 + e^{-I_j}} . \tag{12}$$

These activation values are then broadcast to the output layer nodes. (In this application only one output node was used.) The value received by the output node is given by

$$I = \sum_{j=1}^{4} W_j O_j , \tag{13}$$

where W_j is the weight associated with the connection from hidden node j to the output node, O_j is the output value of hidden node j, and, by definition, $O_4 = 1$ (learnable bias). The output node then computes its activation function, which is the output of the network,

$$f(I) = \frac{1}{1 + e^{-I}} . \tag{14}$$

During training, the output value is compared with the desired output or target vector. For this application, a seizure feature vector has a scalar target value of 1, and a nonseizure feature vector has a scalar target value of 0. Errors are corrected by adjusting the connection weights so that each weight will be proportional to the gradient of the error with respect to that weight. This technique is known as gradient descent error minimization.

TESTING THE ALGORITHMS

Discriminant Analysis Performance Evaluation

The initial testing of the discriminant analysis method was conducted in two phases. Both phases used only the autoregressive (AR) parameters and the residual standard deviation as feature vector elements.

Phase I—Classifying Patient Status. A series of tests used data from four different patients who were awake and active. The data contained normal/awake brain activity, muscle motion, and other artifacts, each of which made seizure detection more difficult. For each patient, the data contained a preseizure period of about 400 seconds, a seizure lasting thirty or more seconds, and a postseizure period (see Fig. 4). For each patient, a first model ($\hat{\mu}_1$, $\hat{\Sigma}_1$) was developed using about 300 seconds of normal/awake data, and a second model ($\hat{\mu}_2$, $\hat{\Sigma}_2$) was developed using most of the available seizure data. This procedure is called training the algorithm. Only these two models were then used to classify all of the data for a given patient. Neither clustering nor postseizure data were used.

Results for all four patients were similar. When the training data were used to classify the state of the patient,

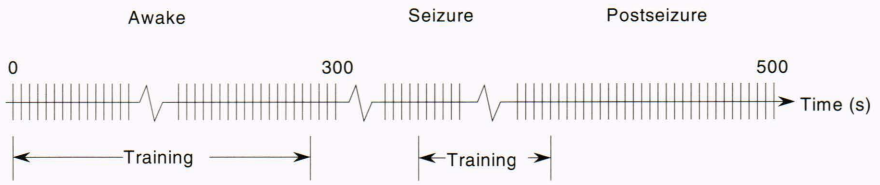


Figure 4. Data format for algorithm evaluation.

correct decisions were always made. (Correct classification, even for the training data, does not occur with some algorithms.) The data from the patients in normal states were classified as normal data, and the data from the patients in seizure states were classified as seizure data. The agreement probabilities were usually more than 0.05. Although all of the data were correctly classified, a few epochs had small agreement probabilities (see the earlier discussion of agreement probabilities). Different channels for each of the patients were also analyzed, and the results were consistently similar. This finding suggests that when a seizure is detected, it should be detected in several channels.

Typically, a 100-second data interval existed between the normal/awake state and the seizure state. These data were not used in training or in estimating $\hat{\boldsymbol{\mu}}_j$, $\hat{\boldsymbol{\Sigma}}_j$. For these data, also, the patient states were correctly classified. Although the neurologist often classified more data at the beginning of the seizure as seizure data, this judgment was based on having seen the data that follow and then backtracking. The sequential algorithm developed here detected the seizures as soon as they were clearly apparent. The agreement probabilities were normally more than 0.05. On occasion, however, the probabilities were smaller.

The first part of the postseizure data was classified as seizure data; the next part was followed by a period of mixed classification; the final part was classified as normal data. The agreement probabilities were usually less than 0.001, indicating that the postseizure period had different statistical characteristics than the two training periods. This result suggests a need for more training classes.

For one of the patients, five data sets were analyzed. These sets included one taken when the patient was in light sleep, another when he was in deep sleep, a third when he was awake while making scalp movements (to introduce artifacts into the data), and two with seizures. The sleep states were clearly distinguished from each other and from other states. The artifactual data resembled other normal/awake data and were not identified as seizures.

These results show that the discriminant approach employed with AR parameters is valuable and can be used to characterize short periods of data well enough to identify patient states properly. The results do not, however, demonstrate the algorithm's ability to characterize the data the next day or from different patients.

Phase 2—Testing Using Independent Data Sets. To validate the algorithm, two ten-minute data sets from each of three patients were analyzed. Each set contained at least one seizure lasting from fifty to ninety seconds and a range of other patient states. The first data set from each patient was used for training and the second to determine if discriminant analysis can be used to classify an independent data set. The paired data sets were generated at different times of the day or on different days.

The clustering methods described previously used the first three autoregressive parameters AR1, AR2, and AR3 (Θ_1, Θ_2, and Θ_3) and log(log σ) to separate the feature vectors into three normal groups and one seizure group. When the resulting discriminants were used to classify the training data sets, classification of the seizure and normal data for all three patients was 100% reliable, and the particular cluster could be identified with 99% reliability.

Neural Network Performance Evaluation

As stated earlier, the neural network, which used only the AR parameters and the residual standard deviation as feature vector elements, attempted to classify each feature vector as seizure (1) or nonseizure (0). The actual values from the network ranged continuously from 0 to 1.0. For our purposes, the following classifications of network outputs were made: an output value greater than or equal to 0.5 implies seizure; an output value less than 0.5 implies nonseizure. Most network values were greater than 0.9 or less than 0.1, indicating little ambiguity in the classification. During the transitions in and out of seizure states, however, the network output took on several intermediate values.

Intercomparing Seizure Identification Method Results

An intercomparison of seizure identification method results is presented in Table 1. The numbers in the table under each heading indicate which of the 204 epochs were classified as seizures by discriminant analysis, by the neural network, and by an interpretation of the EEG tracing.

The classification of actual seizure patterns is subject to interpretation. Some seizures seem to fade away, whereas others have clear ends. Some of the clear ends fall in the middle of the epoch used to estimate a feature vector, so the feature vector could be classified correctly either way. The early part of a seizure EEG tracing most often looks very different than it does ten or fifteen seconds later; thus, an algorithm trained on an entire seizure record may not classify the early part of a seizure as being typical. (Sufficient data do not exist to form an early seizure cluster.)

For the first patient, the training seizure provided limited examples of seizure feature vectors. As a result, not all of the test seizure vectors were properly identified, although the seizure was clearly detected. Surprisingly, when the discriminant from the second patient was

Table 1. Intercomparison of seizure identification method results.

Patient	Epochs of discriminant-identified seizure patterns	Epochs of neural-network–identified seizure patterns	Epochs of actual seizure patterns
1	19–21	18–22	17–19 early seizure
	25		20–32 $\frac{1}{2}$ seizure
	31–33		
1[a]	18–31		17–19 early seizure
	160		20–32 $\frac{1}{2}$ seizure
2			50–53 early seizure
	54–64	58, 61, 62	54–64 seizure
		66, 67, 69	65 → seizure fades
	88–122	92–94, 101–122	88–118 seizure
		126, 127, 129–131, 136	119 → seizure fades
3	17 preseizure alarms		
			117–119 early seizure
	119–137	120–140	120–139 seizure
			140–142 seizure fades
	3 trailing alarms		

[a]Patient 1 classified using patient 2 discriminants.

used to classify data from the first, classification was reliable (see Table 1).

The test data set for the second patient contained two seizures, and both were accurately identified. The training data apparently provided a richer variety of feature vectors than did the training data from the first patient.

The results from the third patient were not as good. Although the seizures were detected, some of the normal feature vectors in the test data set were incorrectly classified. As a further check, the test and training data sets were interchanged. For this second test, the normal data were correctly classified by the algorithm, but some of the seizure epochs were classified as normal. An examination of some feature vector element plots revealed that similar values were classified in one data set as signifying a normal state and in the other as indicating a seizure. This result calls into question both the algorithm and the classification of the training data. Since patients being treated in the Epilepsy Monitoring Unit often experience a series of seizures or seizure discharges, some of the data thought to indicate a normal state may in fact be seizure data. Another possibility is that additional features are required, or perhaps a different discriminant analysis method is needed. It is clear, however, from looking at the EEG records that both seizure and normal data vary in appearance. Adequate training therefore requires more data than were used in these tests. The neural network approach yielded better results with this third patient.

The network classified all of the test patterns correctly for the first patient (see Table 1), performing better than the discriminant analysis technique using the same data. The second patient presented the greatest difficulty, for the network failed to detect the early part of the seizure, first announcing the seizure about twenty seconds (seven epochs) late. For the second seizure, the network identified the seizure state about twelve seconds late. The network then continued to identify the seizure state even though it was not clearly evident from the EEG tracings. Thus, the discriminant analysis method performed better than the network for the second patient.

The network performed quite well for the third patient, although it identified the seizure state about ten seconds late. Nevertheless, neural network analysis performed much better than the discriminant approach. Plots of the feature vector elements (see Fig. 2) suggest that the parameters log(log σ) and AR1 (Θ_1) are most important for statistical discrimination. A look at the weights found by the neural network gradient search indicates that AR6 (Θ_6) is particularly important for network discrimination. This result may help explain why the neural network performed better than discriminant analysis for the third patient and less well for the second patient.

IMPROVING THE ALGORITHMS

Testing has established the validity of our basic feature vector approach but has also revealed that distinguishing a seizure from a wider variety of brain states is more difficult than if only one or two states are consid-

ered. This finding suggests that more extensive algorithm training is required. Three fundamental problems remain: (1) collecting enough data to train and test the algorithms, (2) determining if the autoregressive (AR) parameters contain sufficient information for effective discrimination or if other parameters would be useful, and (3) choosing the best discrimination method (discriminant analysis, neural networks, or logistic regression analysis).

Collecting Data

To solve the first problem partially, over thirty data sets were collected both day and night over four days from a single patient. Each data set represented 336 seconds of data. Four nonseizure data sets were selected that contained data taken while the patient was watching television, going to the bathroom, falling asleep, and sleeping. Four additional data sets, each of which contained a seizure, were stripped of all nonseizure data. The data were then divided into three-second epochs. Randomly selected epochs were used for training, and the remaining epochs were used for testing.

Each three-second epoch of data from one channel provides 600 digitized EEG measurements. A ten-parameter AR model was fitted to each epoch, providing the ten AR parameters and a string of 600 residuals. To see if other feature vector elements could provide useful information, the second, third, and fourth moments (variance, skewness, and kurtosis) of the residuals were computed to yield three more parameters, and the skewness and kurtosis of the original EEG measurements were computed to complete the set of fifteen parameters composing the feature vectors. The 600 EEG measurements from each epoch were thus reduced to a fifteen-element feature vector. The new parameters can now be evaluated as discriminants in a more challenging and realistic setting.

Using New Features for Discrimination

Because many brain states are represented by the set of normal feature vectors, clustering was used to divide the 672 nonseizure vectors into six different clusters automatically. The same procedure was used to divide the 150 seizure feature vectors into three different clusters. The means of the cluster elements are plotted in Figures 5A, 5B, and 5C.

Figure 5A illustrates how the residual standard deviation (σ), skewness, and kurtosis of the residuals can be used to separate the four normal clusters, as represented by the green pyramids and needles, from the others. The values for all three of the parameters in Figure 5A are smaller for these (green) normal clusters than for the remaining clusters. (For the purpose of illustration, the standard deviations of the clusters are ignored.) In addition, the pink seizure cluster can clearly be separated from the others because all of the plotted parameters are bigger for this seizure class. The problem remains to separate the gold and red seizure clusters from the blue and black nonseizure clusters.

Figure 5B provides a different view of the same cluster means with the residual skewness and kurtosis replaced by the skewness and kurtosis of the raw series. For the parameters shown here, the clusters are better separated, except for the red seizure group and the blue normal group. Using only these data, the algorithm could not reliably separate the clusters represented by red and blue pyramids, and some epochs of data would be misclassified.

If we look at Figure 5C, however, the red and blue clusters are clearly separated. The autoregressive parameters AR1, AR3, and AR6 (Θ_1, Θ_3, and Θ_6) can be used to separate these clusters. The use of fundamentally different parameters makes the distinction possible. The additional feature vector elements will make discrimination easier.

An algorithm based on these figures would first look at the feature vector and determine if the moments belong to the pink or one of the green clusters. (Each of the clusters has a small spread as compared with the spread of all of the seizure or all of the nonseizure data.) If so, the algorithm can accurately identify the patient state. If,

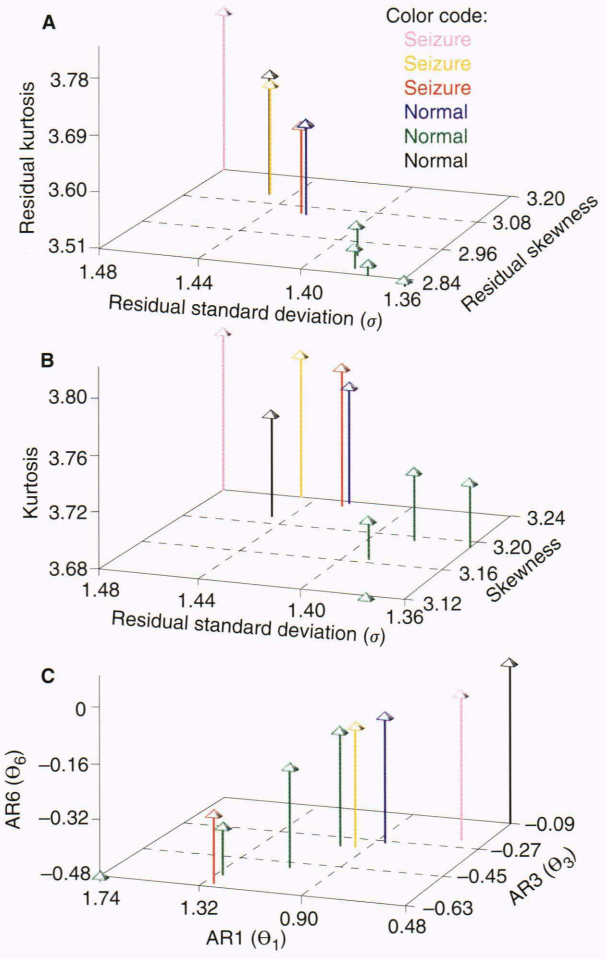


Figure 5. The process of separating seizure from nonseizure data using additional parameters. **A.** Residual kurtosis versus residual standard deviation (σ) versus residual skewness. **B.** Kurtosis versus residual standard deviation (σ) versus skewness. **C.** Autoregressive parameter AR6 (Θ_6) versus AR1 (Θ_1) versus AR3 (Θ_3).

however, the moments do not provide for a clear judgment, then the AR parameters would be used to reach a final decision.

Although the AR parameters capture much of the information necessary to separate seizure and normal epochs (using the data from only one channel), it is clear that the addition of other parameters can enhance the discrimination capabilities of both the discriminant analysis and neural network approaches. Thus, parameters of a significantly different nature can be used to aid in discrimination.

Choosing the Best Discrimination Method

One of the promising algorithms under investigation uses an approach similar to the one just illustrated. Classical discriminant analysis is used to identify the most likely cluster. If the choice is clear, a decision is made. If not, a choice is made between the two most likely populations using principal component analysis to choose the linear combination of feature vectors that best distinguishes these two clusters.

In addition to neural networks and statistical discrimination, logistic regression analysis[26,27] provides a promising method. Although this technique is primarily aimed at separating two groups, such as seizure and normal, it can be extended to work with more groups, such as the clusters. Logistic analysis will be used to determine which parameters are particularly useful in separating groups.

Remarks on Algorithm Improvement

The larger data sets collected for off-line algorithm development will be useful in an on-line algorithm. Much on-line testing will be required, however, before a reliable system is operational. The on-line algorithm will likely use both AR and other parameters to detect epochs containing seizures. The careful selection and evaluation of parameters are critical to the success of the algorithm. Limited experience with new data indicates that data from just two channels and two epochs and the parameters described in this article could be used to discriminate data epochs reliably for the patients considered.

CONCLUSION

Work is in progress at APL to develop algorithms for detecting seizures seconds after they begin using only a single channel of scalp data. At the Epilepsy Monitoring Unit, up to sixty-four channels of EEG data are being recorded and temporarily stored. If seizures are detected sometime during their first minute, the recently collected data can be stored for analysis. The efforts at APL are aimed at developing a very reliable system that has potential for other seizure detection applications. The techniques developed may also be applied to detect other mental states, including, perhaps, those appropriate for crew readiness.

REFERENCES

1 Frost, J. D., Jr., "Microprocessor-based EEG Spike Detection and Quantification," *Int. J. Bio-Med. Comput.* **10**, 357–373 (1979).

2 Gevins, A. S., Yeager, C. L., Diamond, S. L., Spire, S. P., and Zeitlin, G. M., "Automated Analysis of the Electrical Activity of the Human Brain (EEG): A Progress Report," *Proc. IEEE* **63**, 1382–1399 (1975).

3 Gevins, A. S., Blackburn, J., and Dedon, M., "Very Accurate Computer Recognition of Three-per-Second Generalized Spike-and-Wave Discharge," in *Advances in Epileptology; the Xth Epilepsy International Symposium*, Wada, J. A., and Dentry, S. K. (eds.), Raven Press, N.Y., pp. 121–128 (1980).

4 Gotman, J., and Gloor, P., "Automatic Recognition and Quantification of Epileptic Activity in the Human Scalp (EEG)," *Electroencephalogr. Clin. Neurophysiol.* **48**, 551–557 (1980).

5 Gotman, J., "Practical Use of Computer-Assisted EEG Interpretation in Epilepsy," *J. Clin. Neurophysiol.* **2**, 251–265 (1985).

6 Barlow, J. S., "Methods of Analysis of Nonstationary EEGs, with Emphasis on Segmentation Techniques; A Comparative Review," *J. Clin. Neurophysiol.* **2**, 267–304 (1985).

7 Frost, J. D., Jr., "Automatic Recognition and Characterization of Epileptiform Discharge in the Human EEG," *J. Clin. Neurophysiol.* **2,** 231–249 (1985).

8 Jansen, B. H., "Analysis of Biomedical Signals by Means of Linear Modeling," *CRC Crit. Rev. Biomed. Eng.* **12**(4), 343–392 (1985).

9 Gotman, J., "Quantitative Measures of Epileptic Spike Morphology in the Human EEG," *Electroencephalogr. Clin. Neurophysiol.* **48**, 551–557 (1980).

10 Steinberg, H.-W., Gasser, G. T., and Franke, J., "Fitting Autoregressive Models to EEG Time Series: An Empirical Comparison of Estimates of the Order," *IEEE Trans. Acoust. Speech Signal Process.* **33**, 143–150 (1985).

11 Isaksson, A., and Wennberg, A., "Spectral Properties of Non-Stationary EEG Signals, Evaluated by Means of Kalman Filtering: Application Examples from a Vigilance Test," in *Quantitative Analytic Studies in Epilepsy*, Kellaway, P., and Peterson, I. (eds.), Raven Press, N.Y., pp. 389–402 (1976).

12 Lopes Da Silva, H. A., Disk, A., and Smits, H., "Detection of Nonstationarities in EEGs Using the Autoregressive Model—An Application of EEGs of Epileptics," in *CLEAN: Computerized EEG Analysis*, Dolce, G., and Kunkel, H. (eds.) , Gustav Fischer Verlag, Stuttgart, pp. 180–199 (1975).

13 Sharma, K. L. S., and Mahalanobis, A. K., "Harmonic Analysis via Kalman Filtering Technique," *Proc. IEEE* **61**, 143–150 (1973).

14 Jansen, B. H., Hesman, A., and Lenten, R., "Piecewise Analysis of EEGs Using AR-Modeling and Clustering," *Comput. Biomed. Res.* **14**, 168–178 (1981).

15 Jansen, B. H., Bourne, S. R., and Ward, S. W., "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis," *IEEE Trans. Biomed. Eng.* **26**, 630–637 (1981).

16 Ward, J. H., "Hierarchical Groupings to Optimize an Objective Function," *J. Am. Stat. Assoc.* **58**, 236–244 (1963).

17 Sato, K., Ono, K., Chiba, G., and Fukata, K., "On Some Methods for EEG Pattern Discrimination," *Int. J. Neurosci.* **7**, 201–206 (1987).

18 Morf, M., Vierira, A., Lee, D. T., and Kailath, T., "Recursive Multichannel Maximum Entropy Spectral Estimation," *IEEE Trans. Geosci. Electron.* **16**, 85–94 (1978).

19 Penczek, O., Grochulski, W., Greya, J., and Kowalczyk, M., "The Use of Multi-Channel Kalman Filter Algorithm in Structural Analysis of the Epileptic EEG," *Int. J. Bio-Med. Comput.* **20**, 135–151 (1987).

20 Hand, D. J., *Discrimination and Classification*, John Wiley and Sons, New York (1981).

21 Klecka, W. R., *Discriminant Analysis*, Sage Publications, Beverly Hills, Calif. (1980).

22 Rummelhart, D. E., and McClellan, J. L., *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, MIT Press, Cambridge, Mass. (1986).

23 Lippmann, R. P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Mag.* **4**(2), 4–22 (1987).

24 Rauch, A., "A Nonmetric Approach to Linear Discriminant Analysis," *J. Am. Stat. Assoc.* **84**, 173–183 (1989).

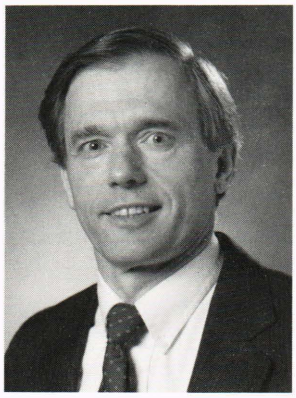
25 *SAS/STAT User's Guide*, SAS Institute, Inc., Cary, N.C. (1988).

26 Hosmer, D. W., and Lemeshow, S., *Applied Logistic Regression*, John Wiley and Sons, New York (1989).

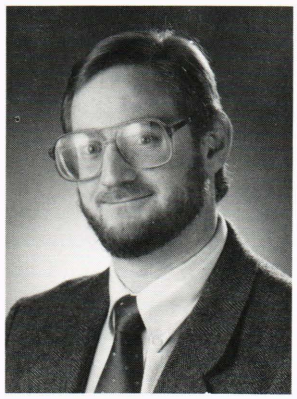
27 Cox, D. R., and Snell, E. J., *Analysis of Binary Data*, Chapman & Hall, London (1989).

ACKNOWLEDGMENT: We would like to thank Reggie Adkins for his software support.

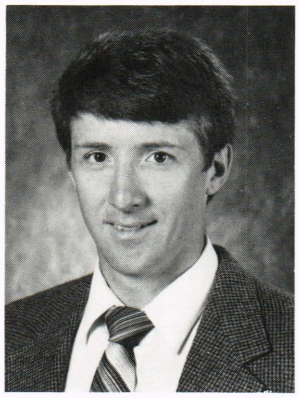
THE AUTHORS



DALE E. OLSEN received his A.B. degree in mathematics from California State University, Chico, in 1965. After working as a research engineer for the Boeing Company for three years, he returned to school and earned an M.S. and Ph.D in statistics from Oregon State University. Dr. Olsen joined APL in 1973 and began developing and applying methods used in analyzing the reliability and accuracy of weapon systems. In 1988, he received Laboratory funding to support work in the analysis of electroencephalogram data. Government interest in this work resulted in funding for the polygraph scoring algorithm. Currently, Dr. Olsen is an Assistant Group Supervisor and a project manager at APL. He is also a member of the Principal Professional Staff.



JOHN A. CRISTION received a B.S. degree in electrical engineering from Drexel University in 1986 and an M.S. degree in electrical engineering from The Johns Hopkins University G.W.C. Whiting School of Engineering in 1991. He is a member of the Associate Professional Staff at APL and works in the Signal Processing Group of the Strategic Systems Department. Mr. Cristion is currently working on a system for automatic polygraph scoring, an Independent Research and Development (IRAD) effort to develop an automated seizure detection algorithm, and an IRAD effort to study the use of neural networks in nonlinear adaptive controllers.



CHARLES W. SPAUR was born in Chariton, Iowa, in 1958. He received his B.S. degree in electrical engineering and computer science from the University of Colorado in 1986 and an M.S. in computer science from The Johns Hopkins University G.W.C. Whiting School of Engineering in 1988. In 1986, Mr. Spaur joined the APL Space Department as an engineer. Since then, he has been involved in the design and development of computer algorithms and software for spacecraft command processing, onboard signal processing, and spacecraft maneuver analysis.