THE ELECTROENCEPHALOGRAM AS A BIOMETRIC

R.B. Paranjape, J. Mahovsky*, L. Benedicenti*, Z. Koles'*

'Biomedical Engineering, University of Alberta, Edmonton, Canada
*Electronic/Software Systems Engineering, University of Regina,
3737 Wascana Parkway Regina, Canada, S4S 0A2
Raman.Paranjape@uregina.ca

ABSTRACT

This paper examines the effectiveness of the electroencephalogram (EEG) as a biometric for the identification of individual subjects in a pool of 40 normal subjects. The EEG's second order statistics are computed using autoregressive models of various order. The coefficients in these models are then evaluated for their biometric potential. Discriminant functions applied to the model coefficients are used to examine the degree to which the subjects in the data pool can be identified. The results indicate that the EEG has significant biometric potential. In this data pool, 100% of subjects are correctly classified when all data is used, and over 80% when the functions are computed from half the data and then applied to the remaining.

1. INTRODUCTION

It has long been known that brain electrical activity as recorded by the EEG has certain biometric qualities. Clinicians can see patterns and characteristics that are unique identifiers or labels specific to patients in these recordings [1]. While these biometric qualities may not be at the same level as the fingerprint, it is interesting to examine the EEG to study it is potentially as a biometric for the identification of human subjects. Currently little or no published work exists that quantitatively evaluates the EEG for its biometric potential.

1.1 The EEG

The normal functioning human brain generates both electric and magnetic fields. These fields are the result of the summation electrical signal from minute flows of ions as neurons, primarily in the cerebral cortex, respond to various stimuli. The cerebral cortex is made up of between 10^9 and 10^{10} neurons and the summated electrical signal from these cells is in fact a unique measure of human brain function [2]. The summated electrical field is easily measured using electrodes attached to the scalp and an appropriate amplification system [3]. This measured electrical signal is known as the electroencephalogram or simply the EEG. The EEG is a signal that is representative of the summated electrical activity of the functioning human brain. The recording electrodes are typically placed in standardized locations over the main anatomical structures of the brain such as the frontal, temporal or parietal lobes. As such, the electrical activation of the scalp is not uniformly or regularly sampled. Figure 1 shows some of the standardized locations where electrodes could be placed.

In this paper we focus on recordings from the P4 electrode. The signal from the P4 electrode is relatively strong and typically contains the alpha rhythm. We speculate that using multi-channel data from all EEG electrodes can enhance the results obtained from this single-channel recording.

1.2 Biometrics

A simple definition of a biometrics is any biological or physiological signal that can be used to identify a person [4]. Various types of signals or characteristics can be used as biometric measures. Systems based on fingerprint, face recognition, iris and retinal scan, or speech matching are becoming increasingly common. However, other biometric systems that utilize more diffuse physiological entities such as facial thermograms and hand geometry are being developed. A biometric system is essentially a pattern recognition system that makes a personal identification by determining the authenticity of specific physiological characteristics possessed by a person.

1.3 The EEG as a Biometric

In this paper, we examine the characteristics of the EEG as a biometric by examining Autoregressive (AR) models that are representative of the second order statistics [5] of the EEG. Autoregressive models of various orders are computed for a selected number of EEG epochs. In this work, the concept of the AR model coefficients of the EEG having some biometric potential is first graphically demonstrated with low order autoregressive models. By showing that there is a

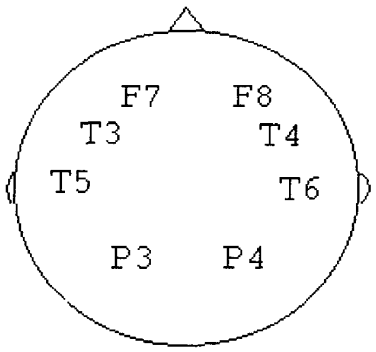

Fig. 1 Diagrammatic view of head with 8-electrode positions indicated: F-electrode overlay Frontal lobes, T-electrodes overlay Temporal lobes, and P electrodes overlay Parietal lobes.

natural clustering of AR coefficients the idea that individuals may be uniquely identified is suggested. Discriminant analysis is then applied to the AR model coefficients. It is shown that a single electrode recording of the EEG can be used to correctly classify up to 100% of subjects in our data set when all data is used to formulate the discriminant function and a sufficiently large model is used. High levels of correct classification continue to be obtained if 50% of the data is used to formulate the discriminant function and then the discriminant function is applied to the remaining data. From these results we conclude that there is clear evidence to support the use of EEG data as a biometric and that the EEG could be used in a constellation of other biometric measures that can be used to identify individuals within a group.

2.0 APPROACH

2.1 The Data Set

A data set of 8-channel EEG recordings from 40 normal volunteer subjects was used in this study. The subject's EEG was recorded while performing the simple activity of resting with eyes open (EO) and resting with eyes closed (EC). Electrodes were placed over the traditional areas of the brain, the frontal (F7, F8), temporal (T3, T4, T5, T6) and parietal (P3, P4) lobes of the brain in accordance with Figure 1. Recordings were carried out over an extended period of time with data stored in epochs of 8.533 sec duration. A trained neurologist evaluated the epochs and those epochs containing appreciable muscle (EMG), cardiac (ECG), or other noise signals were removed from the data set. Thus, while each epoch contained contiguous data, epochs were not necessarily contiguous in time. For each subject typically there were about 8 epochs available. Each epoch was composed of 1024 digital samples of EEG data acquired at a sampling rate of 120 samples/second for each of 8 electrodes.

2.2 Autoregressive Data Modeling

In this work we develop Autoregressive models for single EEG traces using the well known Lattice Equivalent Model and Levinson Recursion [4]. Model orders from 3 through 21 were generated rapidly and efficiently using this method. As the order of the model increases the accuracy of the model, as a predictor of the next value in the EEG time series is increased.

2.3 Discriminant Function Analysis

In this work we investigate the extent to which these EEG data sets can be used to identify the individual that generated the EEG by formulating discriminant analysis functions which minimize differences in EEG epochs generated by the same individual while maximizing the differences between EEG epochs generated by different individuals. Discriminant analysis, an exploratory method of data evaluation, is performed as a two-stage process. First the total variance/covariance matrix for all variables is computed, and then the within-groups variance/ covariance matrix is computed. The two matrices are then inverted and a function is computed that minimizes the variance within group while maximizing the variance between groups. The discriminant functions were computed using a commercial statistical package called Statistica [6].

3.0 RESULTS

3.1 Graphical View of Low Order AR Model Coefficients

An initial test of the concept examines the degree of clustering of the AR coefficients for low order models. A 3rd order AR models is computed using the P4 electrode for five (5) EC subjects drawn at random from the data set all available subjects. We plot (See Figure 2) these model coefficients for the five subjects identifying each subject using a unique symbol. For each subject there were at least 6 epochs of data.

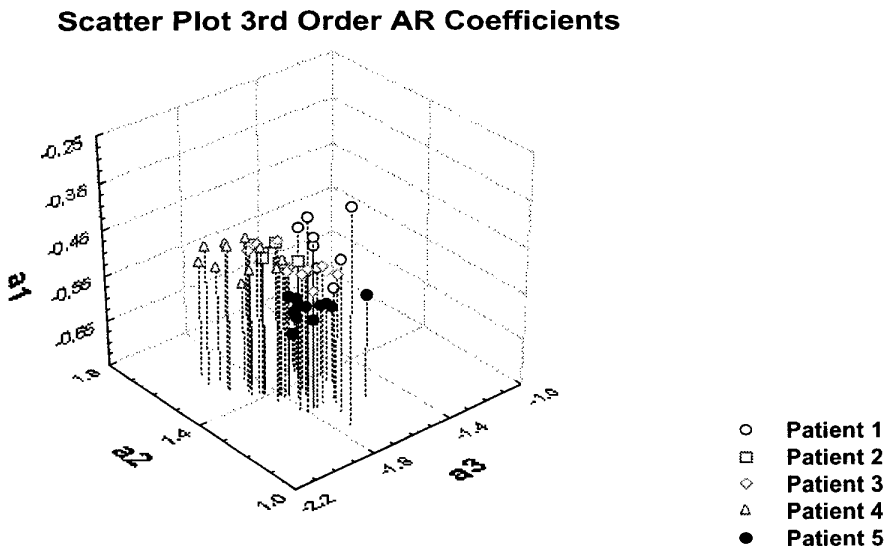

Fig. 2 Scatter plot shows distribution of 3rd order AR coefficients for five subjects. Each subject is shown with a different symbol. The clustering of subject coefficients is see in this figure.

The axes in Figure 2 indicate magnitude of the AR first, second and third coefficients. An examination of Figure 2 shows the distinct clustering of each subject. This scatter plot clearly suggests that the AR model parameters have some biometric value and that each subject occupies a unique cluster in the AR coefficient domain.

3.2 Analysis of AR parameters using Discriminant Functions

In order to evaluate further the biometric characteristics of the AR model parameters, discriminant function analysis was applied to the full EO EEG data set. This was followed by the computation of a discriminant function to try to distinguish individual subjects in the data set.

Table 1 presents summary results from the discriminant analysis of 5 to 40 EO subjects for various model orders using the P4 electrode data.

The table reports the percentage of epochs that were correctly classified by the discriminant function for model order versus the number of patients in the data set. The final column in the table shows the results of discriminant analysis when the error power is included in the analysis. This increases in the model order by one in each case.

	% Correct Classifications					
	# Patients					
	5	10	20	30	40	40 + Error
	# of Epochs					
Order	44	95	191	281	349	349
3	80%	83%	72%	69%	61%	77%
6	93%	95%	87%	87%	86%	92%
9	93%	97%	92%	93%	92%	96%
12	98%	99%	93%	94%	94%	96%
15	100%	99%	98%	98%	97%	98%
18	100%	100%	99%	99%	99%	99%
21	100%	100%	99%	99%	99%	99%

Table 1 shows the percentage of correct classification of EEG Epochs for various AR model orders and sizes of data sets.

Table 1 shows that it is possible to obtain high rates of correct classification of subjects using EEG data. Table 1 also indicates that as the number of subjects and epochs considered increases the order of the AR model must also increase in order to obtain the same high rate of correct classification obtained with less subjects.

3.3 Discriminant Analysis with 50% Training and 50% Test

In order to determine if the discriminant analysis in fact shows a true clustering of AR coefficients that is unique to the individual, the data was divided into two equal sets and the discriminant functions were computed using the training data set and tested using the test data set. Note that both the training set and the test data set have data from 40 subjects. There are only fewer epochs for each subjects in each data set.

Table 2 shows the percentage of correct classifications for the training and test sets for this analysis. We clearly see that the classification rate increases as the model order increases until about 12 and then is high and relatively stable.

	% Correct Classifications	
	Train	Test
	# of Epochs	
Order	175	174
3	68%	49%
6	88%	69%
9	93%	69%
12	99%	76%
15	99%	85%
18	99%	79%
21	99%	82%

Table 2 Shows the percentage of correct classification for the training data set and the test data set.

4.0 DISCUSSION

The results in Figure 2 clearly show that there is a natural clustering of EEG AR coefficients. However, from Table 1 we see that the ability of the discriminant function to classify EEG epochs using only a 3rd order AR model's coefficients is high only for small numbers of subjects and epochs. From a maximum of 83% correct classification with 10 subjects and 95 epochs, we see that only about 61% of subjects are correctly classified when 40 subjects and 349 epochs are used.

On the other hand, as the model order is increased from 3 to 21 the level of correct classification increases and remains high across an increasing number of subjects. This suggests that the AR model coefficients are useful parametric measures from a biometric perspective and that they can be used to identify subjects.

Examining the classification of subjects using 50% of the data for developing the discriminant function and the remaining 50% to test the function further supports this observation. Table 2 indicates that while the levels of correct classification are reduced they are still about 80% correct classification when the discriminant functions face novel data. It is further observed that the classification rate is generally flat after AR model orders of either 12 or 15. Clearly with 40 patients and 349 epochs the model order can not be allowed to increase too high beyond this level in order for the discriminant analysis to remain meaningful.

5.0 CONCLUSIONS

This study examines the potential of using EEG data as a biometric to identify human subjects. The results indicate that in this data set, the EEG does have some biometric potential and that subjects can be identified by their EEG. This report focuses only on a single electrode and further work on other electrodes and a multi-channel analysis is forth coming. In addition this work focuses on normal volunteers in the mental states of eyes open and eyes closed. Other types of data must be considered. Overall however, this work does indicate that there is some evidence to support the basic hypothesis that the EEG as a physiological signal has biometric potential and can be used to help identify subjects.

6.0 REFERENCE

1. Personal communication - Dr. A. Kasmia (Neurologist), Oct. 2000.
2. Integrative functions of the Nervous System, R. Schmidt, in Human Physiology, editors Schmidt R., Thews G., Springer Verlag, Berlin 1983.
3. EEG Informatics. A didactic review of methods and applications of EEG data processing, A. Remond (editor), Elsevier Scientific Publishing Inc., New York, 1997.
4. Biometrics Homepage Michigan State University (East Lansing Campus), Supervisor: Professor A.K. Jain, http://biometrics.cse.msu.edu/info.html
5. Nonlinear Methods in Spectral Analysis, S. Haykin (editor), Springer Verlag, Berlin, 1983.
6. Statistica'99, A Comprehensive System for Statistics, StatSoft Inc., 1999.