PERSON IDENTIFICATION BASED ON PARAMETRIC PROCESSING OF THE EEG

M. Poulos,(1) M. Rangoussi,(2) V. Chrissikopoulos,(1) A. Evangelou,(3)

(1)Dept. of Informatics,
University of Piraeus,
80, Karaoli and Dimitrioy str.,
Piraeus, GR-18534,
GREECE
tel: +30 1 4112463
E-mail:marios.p@usa.net

(2)Dept. of Electronics,
Tech. Educ. Inst. of Piraeus,
250, Thivon str.,
Aigaleo, GR-12244,
GREECE
tel: +30 1 5381225,
E-mail:maria@em.teipir.gr

(3)Dept. of Exp. Physiology,
School of Medicine,
Univ. of Ioannina,
Ioannina, GR-45110,
GREECE
tel: +30 651 97577
fax: +30 651 32045

ABSTRACT

Person identification based on parametric spectral analysis of the EEG signal is addressed in this work - a problem that has not yet been seen in a signal-processing framework, to the best of our knowledge. AR parameters are estimated from a signal containing only the alpha rhythm activity of the EEG. These parameters are used as features in the classification step, which employs a Learning Vector Quantizer network. The proposed method was applied on a set of real EEG recordings made on healthy individuals, in an attempt to experimentally investigate the connection between a person's EEG and genetically - specific information. Correct classification scores at the range of 72% to 84% show the potential of our approach for person classification / identification and are in agreement to previous research showing evidence that the EEG carries genetic information.

1 INTRODUCTION

The target of the present study is to extract genetically-specific features from the EEG recording of a healthy individual. These features will form the basis of a person identification method, whose potential applications are centered on information security, such as information encoding and decoding or secure access to resources. The recording of the EEG is non-invasive and medically safe; it therefore constitutes a viable and under certain conditions attractive alternative to currently existing forms of person identification based on fingerprints, blood test or retinal scanning.

The influence of the genetic code of the individual on the EEG activity has already been investigated in a number of studies that have assessed EEGs of twins (see, e.g., [4], [2], [3], [8], [10], [17]). This research has produced valuable results; for example, it was shown that the morphology of the EEG of monozygotic twins is almost identical, [12], [6], [9]. The spectral analysis of the EEG of twins has shown no significant spectral differences, [16], [13], [12]. Moreover, correlation of the spectra of monozygotic and of dizygotic twins has shown that there exist many more similarities in the former case than in the latter, [9]. In the case of family members it has been shown that common genetic information was found in alpha and beta rhythm EEG activity. The degree of influence due to the family factor was determined quantitatively, [1], [17], [14]. The aim of these studies was to improve the diagnostic methods of neurologic and psychiatric disorders, [1].

On the other hand, the present work investigates healthy rather than pathological cases and aims to establish an one-to-one correspondence between the genetic information of the individual and certain appropriate features of the recorded EEG. After the extraction of these features from the EEG recording, a neural network classifier is employed to classify unknown EEGs as belonging to one of a set of (known) individuals.

Parametric spectral analysis of the alpha rhythm of the EEG is performed by fitting a linear all-pole (AR) model to the EEG spectrum. The coefficients of the fitted model are then used as features to form the input vectors for the neural network. This is a continuation of our previous work on the same subject, using non-linear processing (computational geometry algorithms, [11]). Certain limitations of the computational geometry approach, however, as discussed in [11] - mainly complexity issues and the need for an unambiguous output class result in classification - have prompted the neural network approach taken here.

0-7803-5682-9/99/$10.00©1999 IEEE.

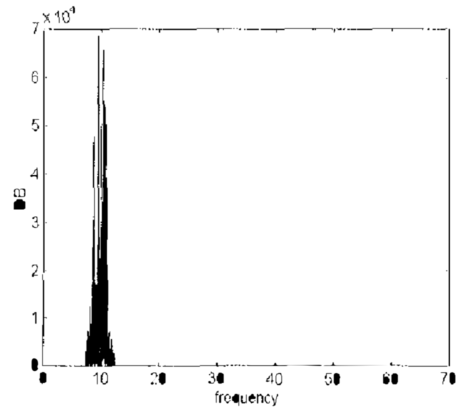

Figure 1: Spectrum of the signal $x(n)$ containing only the alpha rhythm EEG activity.

2 THE PROPOSED METHOD

2.1 Data Acquisition

A set of EEG measurements were carried out and the real EEG data obtained formed the basis for the application of the proposed method. All recordings were taken using a digital electroencephalograph with the PHY-100 Stellate software. Subjects were at rest, with closed eyes. Voltage difference (in mVolts) was recorded between leads O2 and CZ (one channel). All EEG recordings lasted for three (3) continuous minutes, thus producing a 23040 samples long record each, at 128 Hz sampling rate. A 1-30 Hz bandpass filter was applied on the data to retain useful EEG activity. Further processing was carried out off-line, in Matlab, on a Pentium PC.

2.2 Feature extraction-classification

The processing of the EEG recordings proceeds in the following steps:

Step 1. *Preprocessing of the EEG signal* :

- Isolation of the α rhythm signal.
  The recorded signal is spectrally analyzed using Fourier Transform and the band between 7.5 Hz and 12.5 Hz (alpha rhythm) is retained for further processing. The corresponding discrete signal $x(n)$ is regenerated via an inverse Fourier Transform. The spectrum of this signal is shown in Figure 1.
- Model fitting.
  A linear rational model of the ARMA type is fitted to the alpha rhythm EEG signal $x(n)$, which is treated as a superposition of a signal component (deterministic) plus additive noise (random), due to imperfections of the recording process. Standard parameter estimation procedures can be used for the estimation of the model order and model paremeters, using second or third order correlations of $x(n)$ (see, e.g., [15]).
  At the present stage of our work we considered purely AR models. The reason for that was to keep the complexity of the proposed method low, as it is known that for MA parameter estimation one should either resort to non-linear minimization or employ third order correlations (cumulants) of the data. EEG signals can be adequately handled via second order correlations, since there is no evidence of non-causality present. Therefore, the following overdetermined set of equations, based on autocorrelation alone, was formed and solved for the AR parameters $\mathbf{a}_p = [1\ a(1)\ a(2)\ \dots\ a(p)]^T$,

$$c_{2,x}(\tau) + \sum_{i=1}^{p} a(i) c_{2,x}(\tau - i) = 0, \quad (1)$$

  where $q+1 \leq \tau \leq M+p+q$, or, in matrix form,

$$\mathbf{C}_{2,x} \cdot \mathbf{a}_p = \mathbf{0}_p, \quad (2)$$

  where p, q are the corresponding AR and MA part moder orders, $M > 0$ accounts for the overdeterminancy present in the linear equation system and matrix $\mathbf{C}_{2,x}$ consists of the corresponding autocorrelation lags and is of dimensions $(M+p) \times (p+1)$.
  In our case, model orders were set to $q = 0$ (AR model) and $p = 8$. The later choice was made experimentally; it should be noted that it is in accordance with existing studies ([5]), where AR orders ranging from $p = 8$ to $p = 12$ are proposed for EEG recordings longer that 1 min.
  The estimated AR parameters $\mathbf{a}_p$ are used as features describing the EEG signal $x(n)$.

Step 2. *Neural Network Classification* :

The AR parameters extracted from the preprocessing step are fed into an artificial neural network that performs the classification step. The neural network employed is Kohonen's Linear Vector Quantizer (LVQ), [7]. LVQ is based on a different philosophy than neural networks / algorithms of the multilayer perceptron / back-propagation type. In brief, a subset of codebook vectors is assigned to each class S_k and then the codebook vector m_i which has the smallest Euclidean distance from sample vector x is sought. Sample x is assigned to the same group as the closest m_i. LVQ was therefore chosen because of its ability to classify incoming vectors into classes that are not linearly separable in the feature space - a clearly desirable property given the nature of our feature space. This choice was later justified by comparative experimental results with the multilayer perceptron.

The LVQ neural network used has a competitive (hidden) layer with 4 nodes, which segments the feature space into 4 subclasses, which are then grouped into 2 final classes by the second, linear layer.

3 EXPERIMENTAL PART

Although the ultimate goal of this work is person identification, the tests presented here aim at the more tractable goal of classifying an individual as one of a finite set of (known) persons. Note that even this result has practical meaning in applications involving secure access to resources or information. The tests reported here are of a limited scope: far from being exhaustive, they are intended to show that the proposed method works and that the tools chosen are appropriate for the problem at hand.

For this purpose, a set of forty five (45) EEG recordings were taken for each of four (4) individuals, referred to as A, B, C and D. In addition, one EEG recording was taken from each one of seventy five (75) different individuals, to form a group named X. The final pool of EEG recordings thus contained (4 x 45 + 75 x 1=) 255 EEG recordings.

The first test aimes to differentiate between individual A and "non-A" individuals, the group X members serving as the "non-A" class. The same experiment was subsequently carried out for individual B, C and D. In every case, members of group X served as the "non-B", "non-C" or "non-D" class, respectively. In the test for A, twenty (20) feature vectors formed from the EEGs of individual A, along with thirty (30) feature vectors from group X, formed the training set, which thus consisted of fifty (50) feature vectors. The remaining forty five (45) out of the total seventy five (75) feature vectors of group X, along with twenty five (25) feature vectors from subject A, formed the test set (see Table 1). Training and test sets for experiments on subjects B, C and D were formed accordingly, each consisting of (20+30=) 50 and (25+45=) 70 feature vectors, respectively.

Classification scores after training are shown in Tables 2, 3, 4 and 5, for individuals A, B, C, and D, respectively. Correct classification scores in the range of 72% to 84% are encouraging, in the sense that they show the potential of the proposed method to achieve person classification / identification. Note that group X in the tests has no homogeneity whatsoever in itself; consequently, the network can not be trained to recognize members of X as a class. Therefore, this is not a typical two-class classification problem. This accounts for the non-zero off-diagonal entries of Tables 2, 3, 4, 5.

4 CONCLUSIONS

Person identification based on parametric spectral estimation of the EEG signal was the subject of this work. Experiments were carried out with four different individuals, whose EEG recordings were taken, analyzed according to the proposed method and compared to those of a group of different individuals. An AR model was fitted to the alpha rhythm activity of the EEG recording. AR parameters were next fed into a Linear Vector Quantizer neural network, for classification. Correct classification scores at the range of 72% to 84% show the potential of the proposed approach. These results are in agreement to previous research showing evidence that the EEG carries genetic information. Certainly, extensive experimentation is required in order to obtain statistically significant results and thus "prove" the conjecture of the neurophysiologists about the one-to-one correspondence between the EEG and the genetic code.

References

[1] Anoklin, A., Steinlein, O., Fisher, C., Mao, Y., Vogt, P., Schalt, E., Vogel, F., "A genetic study of the human low-voltage electroencephalogram," *Human Genetics,* vol. 90, pp. 99-112, 1992.

[2] Berger, H., "Das Elektrenkephalogramm des Menschen," *Nova Acta Leopoldina,* Bd. 6. Nr. 38, 1938.

[3] Buchbaum, M. S., Gershon, E.S., "Genetic factors in EEG, Sleep and evoked potentials," In: *Human consiousness and its transformations,* Ed. Davidson, Phenomen Press, 1978.

[4] Davis, H., Davis, P.A., "Action potentials of the brain in normal persons and in normal states of cerebral activity," *Archives of Neurological Psychiatry,* vol. 36, pp. 1214-1224, 1936.

[5] Hazarika N., Tsoi, A., Sergejew, A.,. "Nonlinear Considerations in EEG signal Classification," *IEEE Transactions on Signal Processing,* vol. 45, no. 4, pp. 829- 836, 1997.

[6] Juel-Nielsen, N., Harvand, B., "The electroenceplalogram in univular twins brought up apart," *Acta genetica,* vol. 8, pp. 57-64, Basel, 1958.

[7] Kohonen, T., "Self - Organization and Associative Memory," 3rd ed., Spinger-Verlang, New York, 1988.

[8] Lennox, W., Gibbs, E., Gibbs, F., "The brain-pattern, an hereditary trait," *The Journal of Heredity,* vol. 36, pp. 233-243, 1990.

[9] Lykken, D.T., "Research with twins: the concepts of emergenesis," *Phychophysiology,* vol. 19, pp. 361-373, 1982.

[10] Plomin, R., "The role of inheritance in behavior," *Science,* vol. 248, pp. 183-188, 1990.

[11] Poulos, M., Rangousi, M., Kafetzopoulos, V., "Person identification via the EEG using computational geometry algorithms," *Proc. Intl. Conf. EUSIPCO'98,* Greece, Sept. 1998.

[12] Stassen, H.H., Bomben, G., Propping, P., "Genetic aspects of the EEG: an investigation into the within-pair similarity of monozygotic and dizygotic twins with a new method of analysis," *Electroencephalography and Clinical Neurophysiology,* vol. 66, pp. 489-501, 1987.

[13] Eischen, S.E., Luckritz, J.Y., Polish, J., "Spectral analysis of EEG from families," *Biological Psychology,* vol. 41, pp. 61-68, 1995.

[14] Sviderskaya, N.E., Korol'kova, T.A., "Genetic features of the spatial organization the human cerebral cortex," *Neuroscience and Behavioral Physiology,* vol. 25, no. 5, 1995.

[15] Swami, A., Mendel, J.M., "AR identification using cumulants," *Proc. Workshop on Higher-Order Spectral Analysis,* Vail, CO, USA, pp. 1-6, 1998.

[16] Varner, J., Potter, R., Rohrbaugh, J., "A procedure for automatic classification of EEG genetic variants," *Processing of biological signals,* 30.9-3, Annual nternational Conference of the IEEE Engineering in Medicine and Biology Society, vol. 13, no. 1, 1991.

[17] Vogel, F., "The genetic basis of the normal EEG," *Human Genetics,* vol. 10, pp. 91-114, 1970.

Class	Training set	Test set	Total
A	20	25	45
X	30	45	75
Total	50	70	120

Table 1: Training and test set sizes for experiment on subject A. Sets for the experiments on subjects B, C and D are of the same size.

classified as: → belongs to class: ↓	A	X	Total
A	$\frac{20}{25}$ (80%)	$\frac{5}{25}$ (20%)	25
X	$\frac{8}{45}$ (18%)	$\frac{37}{45}$ (82%)	45
Total			70

Table 2: Classification scores of the experiment on subject A.

classified as: → belongs to class: ↓	B	X	Total
B	$\frac{19}{25}$ (76%)	$\frac{6}{25}$ (24%)	25
X	$\frac{10}{45}$ (22%)	$\frac{35}{45}$ (78%)	45
Total			70

Table 3: Classification scores of the experiment on subject B.

classified as: → belongs to class: ↓	C	X	Total
C	$\frac{21}{25}$ (84%)	$\frac{4}{25}$ (16%)	25
X	$\frac{8}{45}$ (18%)	$\frac{37}{45}$ (82%)	45
Total			70

Table 4: Classification scores of the experiment on subject C.

classified as: → belongs to class: ↓	D	X	Total
D	$\frac{18}{25}$ (72%)	$\frac{7}{25}$ (28%)	25
X	$\frac{10}{45}$ (22%)	$\frac{35}{45}$ (78%)	45
Total			70

Table 5: Classification scores of the experiment on subject D.