NEURAL NETWORK BASED PERSON IDENTIFICATION USING EEG FEATURES

M. Poulos,(1) *M. Rangoussi* (2) *and N. Alexandris* (1)

(1) Dept. of Informatics,
University of Piraeus,
80, Karaoli and Dimitrioy str.,
Piraeus, GR-18534, GREECE
tel: +30 1 4222803; fax: +30 1 4112463
E-mail: marios.p@usa.net

(2) Dept. of Electrical and Computer Eng.,
National Technical Univ. of Athens,
9, Iroon Polytechneioy str.,
Athens, GR-15773, GREECE
tel: +30 1 7722491, fax: +30 1 7722492
E-mail: maria@image.ece.ntua.gr

ABSTRACT

A direct connection between the ElectroEncephaloGram (EEG) and the genetic information of an individual has been suspected and investigated by neurophysiologists and psychiatrists since 1960. However, most of this early as well as more recent research focuses on the classification of pathological EEG cases, aiming to construct tests for purposes of diagnosis. On the contrary, our work focuses on healthy individuals and aims to establish an one-to-one correspondence between the genetic information of the individual and certain features of his/her EEG, as an intermediate step towards the further goal of developing a test for person identification based on features extracted from the EEG. Potential applications include, among others, information encoding and decoding and access to secure information. At the present stage the proposed method uses spectral information extracted from the EEG non-parametrically via the FFT and employs a neural network (a Learning Vector Quantizer - LVQ) to classify unknown EEGs as belonging to one of a finite number of individuals. Correct classification scores ranging from 80% to 100% in experiments conducted on real data, show evidence that the EEG indeed carries genetic information and that the proposed method can be used to construct person identification tests based on EEG features.

1. INTRODUCTION

The objective of this work is to extract genetically-specific information from a person's EEG and use this information to develop a person identification method, based on features extracted from the EEG recording. Potential applications of this new person identification method are, for example, information encoding and decoding or access to secure information. EEG recording 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 existence of genetic information in the EEG was investigated as early as in the 1930s, [2]. However, it has not been until in the 1960s that a direct connection was established between a person's EEG (especially the alpha and beta rhythms) and his/her genetic information or code, thanks to the pioneering work of Vogel with twins and family members, [13]. Results of related research over the years can be found, among others, in references [7], [5], [3], [10], [8], [1], [11].

Most of this research effort, however, has focused on the classification of genetically or pathologically induced EEG variants due, for example, to epilepsy or schizophrenia, for diagnostic purposes, [4], [12]. Recent research including linear and non-linear approaches with a neural network classification scheme has reached a 71% classification score, [4]. A key observation in these approaches is the fact that a given pathology induces a pathology-specific variation pattern on the "healthy" EEG signal. Diagnosis of the pathology is therefore based on the detection of this variation pattern, which serves as classification feature.

On the contrary, the present work focuses in principle on healthy as opposed to 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. A neural network classifier (LVQ) is employed to classify unknown EEGs as belonging to one of a set of (known) individuals. Spectral values obtained non-parametrically from the alpha rhythm spectral band of the EEG signal are used as features to form the input vectors. This is a continuation of our previous work on the same subject, using non-linear processing (computational geometry algorithms, [9]). Certain limitations of the computational geometry approach, however, as discussed in [9] - mainly complexity issues and the need for a unique target class in classification - have prompted the neural network approach taken here.

2. FEATURE EXTRACTION AND CLASSIFICATION

The basis of the signal processing employed at the present stage of our work is spectral analysis implemented via the Fast Fourier Transform (FFT). Although parametric linear or non-linear processing of the EEG signal for feature extraction is also to be investigated, the non-parametric FFT based spectral analysis was chosen first due to its obvious physical interpretation in terms of EEG rhythms. The proposed method proceeds in two steps:

Step 1 : Feature Extraction from the EEG signal. The spectral values of the EEG signal are obtained and then confined to the alpha rhythm band values only. These values are further broken into three overlapping subbands, namely [7-10Hz], [8-11Hz] and [9-12Hz], in order to investigate whether one of them is informative enough to represent the whole EEG signal, for person identification purposes.

Step 2 : Neural Network Classification. The spectral values obtained in Step 1 are used as feature vectors for classification. These vectors are fed into an LVQ classifier, [6], first for training and then for the actual classification of unknown input vectors. LVQ was 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 experimental setup. This choice was later justified by comparative experimental results with the backpropagation network.

3. EXPERIMENTAL PART

Two different experiments are conducted using real data, in order to exhibit the potential of the proposed method for person classification and, further, for person identification. The first test case focuses on identification of a given person among a variety of others. Its purpose is to show that the proposed method works and that the tools chosen are appropriate for the problem at hand. The second test case focuses on a more realistic, multiclass setup, where a group of four persons are in turn identified.

3.1. Signal Acquisition and Feature Extraction

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

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. All EEG recordings lasted for three (3) continuous minutes, thus producing a 23040 samples long record each, at 128Hz sampling rate. Further processing was carried out in Matlab on a Pentium PC. Spectral values of the EEG signal were computed and alpha rhythm frequency band (7-12Hz) was retained for further processing. Alpha rhythm frequencies were next partitioned into three overlapping frequency bands of 3Hz each ([7-10Hz, 8-11Hz, 9-12Hz]), each band containing 540 spectral values at the resolution employed.

3.2. Test Case #1

Test case #1 aimed to differentiate between individual A and "non-A" individuals, the group X members serving as the "non-A" class in that case. The same experiment was subsequently carried out for individuals B, C and D, and for each one of the three alpha rhythm subbands mentioned above. In every case members of group X served as the "non-B", "non-C" or "non-D" class, respectively. Let us note here that, although this four-ply experiment was not absolutely necessary, we present here the results of all four tests, because of the indicative nature of this test. In every case, twenty five (25) feature vectors from the individual of interest (A or B or C or D, respectively) along with twenty five (25) feature vectors from group X formed the training set, which thus consisted of fifty (50) feature vectors. The remaining twenty (20) out of the total forty five (45) feature vectors of subject A, along with the remaining fifty (50) out of seventy five (75) feature vectors of group X, formed the test set for subject A. Test sets for subjects B, C and D were formed accordingly, each consisting of seventy (70) feature vectors.

The network was trained by the (A,X) training set and then classification was performed using the (A,X) test set, in each one of the three alpha rhythm subbands. Classification scores after training are shown in Table 1. Individuals B, C and D were treated analogously; results are shown in Tables 2, 3 and 4, respectively.

3.3. Test Case #2

Test case #2 addresses a more realistic, multi-target setup, where four individuals of interest, namely A, B, C and D, are to be identified. Twenty five (25) feature vectors of each one of the classes A, B, C and D were used as the training set, whose size was thus 100 vectors. The remaining twenty (20) vectors of each one of the classes A, B, C and D (80 in all) were used as the test set. Classification scores after training are shown in Table 5.

As it can be seen in Tables 1 - 5, promising classification scores are obtained in the range 80% to 100%, depending on the individual and the frequency band. No frequency band shows any clear benefit over the others as to the classification scores obtained; rather, they are all equally informative regarding the problem at hand. This conclusion limits the field for the feature extraction step; it can result, therefore, in considerable computational savings regardless of the specific algorithm used for feature extraction.

4. CONCLUSION - FURTHER RESEARCH

Person identification based on spectral information extracted from the EEG is addressed in this work - a problem that has not yet been seen in a signal processing framework, to the best of our knowledge. Neural network classification was performed on real EEG data of healthy individuals, in an attempt to experimentaly investigate the connection between a person's EEG and genetically-specific information. The proposed method has yielded correct classification scores at the range of 80% to 100%. These results are in agreement to previous research showing evidence that the EEG carries genetic information, and also show the potential of our approach regarding the problem at hand. Trimming of the network parameters and parametric spectral estimation, which will reduce the dimensionality of the feature space,

	7-10 Hz			8-11 Hz			9-12 Hz		
out: $\rightarrow$ in: $\downarrow$	A	X	Total	A	X	Total	A	X	Total
A	$\frac{19}{20}$ (95%)	$\frac{1}{20}$ (5%)	20	$\frac{19}{20}$ (95%)	$\frac{1}{20}$ (5%)	20	$\frac{20}{20}$ (100%)	$\frac{0}{20}$ (0%)	20
X	$\frac{6}{50}$ (12%)	$\frac{44}{50}$ (88%)	50	$\frac{5}{50}$ (10%)	$\frac{45}{50}$ (90%)	50	$\frac{8}{50}$ (16%)	$\frac{42}{50}$ (84%)	50
Total			70			70			70

Table 1: Subject A against group X: classification scores in the 3 alpha rhythm subbands.

	7-10 Hz			8-11 Hz			9-12 Hz		
out: $\rightarrow$ in: $\downarrow$	B	X	Total	B	X	Total	B	X	Total
B	$\frac{19}{20}$ (95%)	$\frac{1}{20}$ (5%)	20	$\frac{20}{20}$ (100%)	$\frac{0}{20}$ (0%)	20	$\frac{20}{20}$ (100%)	$\frac{0}{20}$ (0%)	20
X	$\frac{5}{50}$ (10%)	$\frac{45}{50}$ (90%)	50	$\frac{6}{50}$ (12%)	$\frac{44}{50}$ (88%)	50	$\frac{7}{50}$ (14%)	$\frac{43}{50}$ (86%)	50
Total			70			70			70

Table 2: Subject B against group X: classification scores in the 3 alpha rhythm subbands.

	7-10 Hz			8-11 Hz			9-12 Hz		
out: $\rightarrow$ in: $\downarrow$	C	X	Total	C	X	Total	C	X	Total
C	$\frac{16}{20}$ (80%)	$\frac{4}{20}$ (20%)	20	$\frac{18}{20}$ (90%)	$\frac{2}{20}$ (10%)	20	$\frac{19}{20}$ (95%)	$\frac{1}{20}$ (5%)	20
X	$\frac{10}{50}$ (20%)	$\frac{40}{50}$ (80%)	50	$\frac{9}{50}$ (18%)	$\frac{41}{50}$ (82%)	50	$\frac{8}{50}$ (16%)	$\frac{42}{50}$ (84%)	50
Total			70			70			70

Table 3: Subject C against group X: classification scores in the 3 alpha rhythm subbands.

	7-10 Hz			8-11 Hz			9-12 Hz		
out: $\rightarrow$ in: $\downarrow$	D	X	Total	D	X	Total	D	X	Total
D	$\frac{17}{20}$ (85%)	$\frac{3}{20}$ (15%)	20	$\frac{18}{20}$ (90%)	$\frac{2}{20}$ (10%)	20	$\frac{18}{20}$ (90%)	$\frac{2}{20}$ (10%)	20
X	$\frac{5}{50}$ (10%)	$\frac{45}{50}$ (90%)	50	$\frac{5}{50}$ (10%)	$\frac{45}{50}$ (90%)	50	$\frac{8}{50}$ (16%)	$\frac{42}{50}$ (84%)	50
Total			70			70			70

Table 4: Subject D against group X: classification scores in the 3 alpha rhythm subbands.

are expected to further enhance the results. Certainly, extensive experiments are necessary 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.

5. REFERENCES

[1] Anoklin, A., Steinlein, O., Fisher, C., Mao, Y., Vogt, P., Schalt, E., Vogel, F. (1992), "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,* Davidson, ed. Phenomen Press, 1978.

[4] 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.

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

[6] Kohonen, T., "Self-Organization and Associative Memory," 3rd ed., *Springer-Verlag,* New York, 1988.

[7] Lennox, W., Gibbs, E., Gibbs, F., "The brain-patern, an hereditary trait," *The Journal of Heredity,* vol. 36, pp. 233-243, 1945.

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

[9] Poulos, M., Rangoussi, M., Kafetzopoulos, E., "Person identification via the EEG using computational geometry algorithms," *Proc. Intl. Conf. EUSIPCO'98,* Rhodes, Greece, Sept. 1998.

[10] 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.

[11] Sviderskaya, N. E., Korolkova, T. A., "Genetic features of the spatial organization the human cerabral cortex," *Neuroscience and Behavioral Physiology,* vol. 25, no. 5, 1995.

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

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

7-10 Hz band				
out:→ in:↓	A	B	C	D
A	$\frac{16}{20}$ (80%)	$\frac{1}{20}$ (5%)	$\frac{1}{20}$ (5%)	$\frac{2}{20}$ (10%)
B	$\frac{0}{20}$ (0%)	$\frac{17}{20}$ (85%)	$\frac{1}{20}$ (5%)	$\frac{2}{20}$ (10%)
C	$\frac{1}{20}$ (5%)	$\frac{1}{20}$ (5%)	$\frac{16}{20}$ (80%)	$\frac{2}{20}$ (10%)
D	$\frac{1}{20}$ (5%)	$\frac{1}{20}$ (5%)	$\frac{0}{20}$ (0%)	$\frac{18}{20}$ (90%)
Total: 80				

8-11 Hz band				
out:→ in:↓	A	B	C	D
A	$\frac{17}{20}$ (85%)	$\frac{0}{20}$ (0%)	$\frac{1}{20}$ (5%)	$\frac{2}{20}$ (10%)
B	$\frac{0}{20}$ (0%)	$\frac{17}{20}$ (85%)	$\frac{1}{20}$ (5%)	$\frac{2}{20}$ (10%)
C	$\frac{1}{20}$ (5%)	$\frac{1}{20}$ (5%)	$\frac{16}{20}$ (80%)	$\frac{2}{20}$ (10%)
D	$\frac{0}{20}$ (0%)	$\frac{1}{20}$ (5%)	$\frac{0}{20}$ (0%)	$\frac{19}{20}$ (95%)
Total: 80				

9-12 Hz band				
out:→ in:↓	A	B	C	D
A	$\frac{18}{20}$ (90%)	$\frac{0}{20}$ (0%)	$\frac{0}{20}$ (0%)	$\frac{2}{20}$ (10%)
B	$\frac{0}{20}$ (0%)	$\frac{16}{20}$ (80%)	$\frac{1}{20}$ (5%)	$\frac{3}{20}$ (15%)
C	$\frac{0}{20}$ (0%)	$\frac{1}{20}$ (5%)	$\frac{17}{20}$ (85%)	$\frac{2}{20}$ (10%)
D	$\frac{0}{20}$ (0%)	$\frac{2}{20}$ (10%)	$\frac{0}{20}$ (0%)	$\frac{18}{20}$ (90%)
Total: 80				

Table 5: Common pool experiment for subjects A, B, C and D. Classification scores for the 3 alpha rhythm subbands.