World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:1, No:11, 2007

Person Identification by Using AR Model for EEG Signals

Gelareh Mohammadi, Parisa Shoushtari, Behnam Molaee Ardekani and Mohammad B. Shamsollahi

***Abstract*—** A direct connection between ElectroEncephaloGram (EEG) and the genetic information of individuals has been investigated by neurophysiologists and psychiatrists since 1960's; and it opens a new research area in the science. This paper focuses on the person identification based on feature extracted from the EEG which can show a direct connection between EEG and the genetic information of subjects. In this work the full EO EEG signal of healthy individuals are estimated by an autoregressive (AR) model and the AR parameters are extracted as features. Here for feature vector constitution, two methods have been proposed; in the first method the extracted parameters of each channel are used as a feature vector in the classification step which employs a competitive neural network and in the second method a combination of different channel parameters are used as a feature vector. Correct classification scores at the range of 80% to 100% reveal the potential of our approach for person classification/identification and are in agreement to the previous researches showing evidence that the EEG signal carries genetic information. The novelty of this work is in the combination of AR parameters and the network type (competitive network) that we have used. A comparison between the first and the second approach imply preference of the second one.

***Keywords*—** Person Identification, Autoregressive Model, EEG, Neural Network

I. Introduction

PERSON identification by EEG signals is one of the new research areas in the science which can show a connection between the genetic information and EEG of an individual. EEG recording is non-invasive and medically safe; therefore, it should be feasible to use EEG as a useful tool for person identification. The existence of genetic information in the EEG was investigated as early as in the 1930's [9].However, it has not been expanded until in the 1960's that a direct connection was established between a person's EEG and his/her genetic information [21]. Most of the previous researches have focused on the classification of genetically or pathologically induced EEG variants due, for example to epilepsy or schizophrenia for diagnostic purposes,[12]-[20].

On the contrary, the present work focuses on healthy cases and aims to establish a one-to-one correspondence between the genetic information and certain appropriate features of the recorded EEG signal of individual. A direct connection between genetic information and EEG says that EEG must be uniqueness for each person.

Although much investigation has not been done to assess the uniqueness of EEG patterns of each person in the rest, there are some proofs showing that EEG patterns are probably unique for individuals [1].

In this research, it has been tried to find out suitable EEG features as biometrics to classify individuals by employing a competitive neural network. In the sequel of this section there are brief expression about EEG data and biometrics.

A. EEG data

Brain waves (EEG) are the responses of the neural cells to various stimuli [2]; these waves, on the surface of the brain, are responses to different stimuli and what is recorded is the sum of all these responses. There are some electrodes on the scalp to record and amplify signals. These electrodes are typically placed in standardized locations over the main anatomical structures of the brain such as: Frontal, Temporal and Parietal lobs [3].

EEG signal is a time series which has a statistical properties but these properties are varies by means of time, mental state and different persons.

B. Biometrics

Any biological or physiological signal like fingerprints, retinal scans or speech matching [5] that can be used to identify a person [4] is called biometric. A biometric system uses recognizable features, possessed by a person.

In this paper we use EEG signal as an identifying signal. The features extracted are AR parameters in specific time durations and these features are given to a Competitive Neural Network to be classified. So in this paper AR parameters of EEG is biometric.

II. Materials

A. Autoregressive Model

In this model the series is estimated by a linear difference

G. Mohamadi is a M.Sc. student in the School of Electrical Engineering, Sharif University of Technology, Tehran, Iran (e-mail: gelareh_mohamadi@yahoo.com).

P. Shooshatri, is a M.Sc. student in the School of Electrical Engineering, Sharif University of Technology, Tehran, Iran (e-mail: pshooshtari@ee.sharif.edu).

B. Molaee Ardekani is with the LTSI, Université de Rennes 1 and Inserm U 642, Rennes, France, and the School of Electrical Engineering, Sharif University of Technology, Tehran, Iran. (e-mail: behnam.molaee-ardekani@univ-rennes1.fr)

M. B. Shamsollahi is an assistant professor in the School of Electrical Engineering, Sharif University of Technology, Tehran, Iran. (Corresponding author, phone: +98-21-66164356; e-mail: mbshams@sharif.edu).

equation in time domain:

$$X(t) = \sum_{i=1}^{p} a(i)X(t-i) + e(t) \quad (1)$$

Where a current sample of the time series X(t) is a linear function of P previous samples plus an independent and identically distributed (i.i.d) white noise input e(k) [6].

In this work the Yull-Walker approach has been employed to estimate the AR parameters by the use of LMS (least square method) criterion.

B. Competitive Neural Network

In a competitive neural network only one of the output neurons, having the highest level, will win the competition.

In this paper, it has been used a reinforced learning algorithm. In this approach there is a supervisor determining the winner neuron for the training set but it doesn't determine the real quantity of expectative output so this network is considered as an unsupervised one. Here is the error signal binary (zero or one) for example if the expectative neuron became the winner, error signal is zero otherwise, error signal is one and the network will try to adjust its weight till the expectative neuron win.

In this network there are N neurons in the output layer and each neuron has its own weights (W_n).(fig.1)

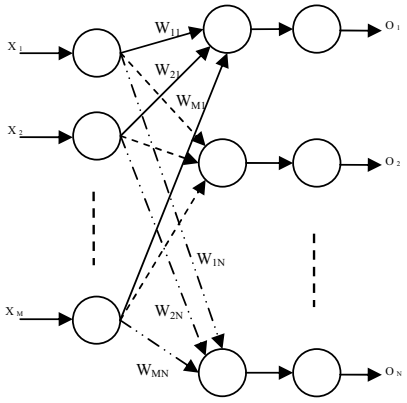

Fig.1 Competitive Neural Network

If X vector is given as an input data, the output for each neuron is calculated by:

$$O_n = \sum_j W_{jn} X_j \quad (2)$$

The neuron which has the maximum quantity is the winner. When the correct neuron does not win the competition the signal error is existed and the weight matrix of correct neuron will be reinforced by the following rule but the others don't have any change:

$$W_n(k) = W_n(k-1) + \alpha(k)(\underline{X} - W_n) \quad (3)$$

Where $\alpha(k)$ is usually a small positive number, called training coefficient, this parameter can change during training time or it can be constant.

C. . Data Set

The used data in this paper have been recorded from ten healthy volunteer subjects. (Six men and four women)

For each person the data is taken from 100 channels in 24 sec duration and the rate of sampling is 170 Hz.

So for each channel there are 4080 samples which are divided to eight epochs of 3 sec duration for each one.

III. Approaches

In this paper two methods have been proposed for feature vector constitution, single channel and multichannel method. There is a little difference between these two approaches but a perfectly clear difference can be seen in the results. In the sequel, both of the methods and the results are presented.

A. Single channel Method

We want to demonstrate that the EEG of a person is probably uniqueness which means that, as explained before, there is a connection between the EEG and the genetic information of individuals. In this research we have 10 subjects and there are eight epochs of 3 sec duration for each person on 100 channels; so there is a data set of 80 epochs for each channel.

The goal of this research is to extract some features which have a discriminant property in individuals. Here the AR parameters of each epoch are considered as features. Although EEG is not a stationary signal in its nature but we assume it stationary in each epoch and calculate the parameters for each epoch. Depends on the order of model, the number of parameters is different. Due to a specific order, the parameters are calculated.

In the single channel method, the obtained parameters of each epoch are arranged in a vector and this vector is considered as the feature vector. A portion of these vectors is given to a supervised neural network as the training set and after training period, to validate the classification and similarity of EEG parameters in a same individual, total vectors are given to the network as testing set. The applicate network is a competitive neural network with a reinforced learning algorithm. In the current approaches, 50 vectors of 80 for each channel are used as the training set. This processing was carried out in Msoftware ATLAB on a Pentium four PC. In the presented experiments, one of the 100 channels was selected and the processing was performed on the each person epochs of the selected channel, then these epochs divided to two parts as the training and testing set, as pointed out before.

Here we didn't exanimate the effect of channel place on the results accurately but a visual inspection yield that there is a higher correctness score in the back of the scalp over the parietal channels than is in the other channels.(e.g. here channel number 004 and 080). Classification scores according to the different orders of AR model for a number of random selected channels are showed in Table 1 and Table 2. So briefly, at first the AR parameters with a particular order is calculated for each epoch and the feature vector is formed by these parameters; then by the use of a classifier , which is a Competitive Neural Network in this research, the feature vectors are classified. After training the network with all the training set the network will be tested and the percentage of correct identification is calculated.

TABLE 1
SCORES OF SINGLE CHANNEL METHOD

Order / Channel	001	002	003	004
3	92.5	50	62.5	55
5	87.5	57.5	67.5	92.5
7	90	75	65	82.5
9	92.5	95	87.5	95
11	90	90	85	97.5
13	92.5	90	85	97.5
15	92.5	87.5	85	97.5
17	92	95	85	97.5
19	92.5	92.5	85	97.5

Percentage of correct identification according to different order for channel 001, 002, 003, 004

TABLE 2
SCORES OF SINGLE CHANNEL METHOD

Order / Channel	014	030	050	080
3	87.5	57.5	50	85
5	90	67.5	80	87.5
7	82.5	62.5	90	90
9	95	87.5	95	97.5
11	95	92.5	92.5	97.5
13	95	92.5	92.5	97.5
15	95	95	97.5	97.5
17	95	95	95	100
19	95	95	95	97.5

Percentage of correct identification according to different order for channel 014, 030, 050, 080

B. Multichannel Method

In this approach all the data are like previous, in the single channel method, but we tried to introduce a feature vector with better score in classification. Here we use AR parameters as the features again, but the formed vector is a combination of AR parameters of different channels. It means, after acquisition of AR parameters for each channel according to a specific order, the parameters of two, three or more channels are assigned in one vector as the feature vector. And as aforesaid, the network is trained by a portion of vectors set and tested by the total vectors. In the presented experiments like the single channel method the network use 50 vectors for training. The scores of this approach according to different orders of model for a number of channels are showed in Table 3 and Table 4 for a combination of two random selected channels parameters and in the Table5 and Table 6 for a combination of three random selected channels. These results obviously imply higher scores than do in the single channel method results. So, we can say that a combination of parametric model of EEG in different channel shows a higher relation between EEG and genetic information of a person than using one channel parameters. With a little attention to the order of AR model and the scores it is clear that if the order of model increases to a certain value (here, of order 11), the scores becomes better, but increasing more than this value approximately has no effect on the scores. It seems that the value of appropriate order depends on the number of subject; and by increasing the number of subject, the algorithm needs to a higher order to have satisfactive results. Although, calculating high order AR model parameters for a much number of subjects, in order to identifying them, may not be reasonable or practical, but the only purpose in this paper is to show the potential of this parameters to classifying persons and demonstrating the probability of EEG uniqueness for individuals which reveal a direct relation between genetic information and EEG.

IV. CONCLUSION

Person identification based on AR parameters extracted from EEG is addressed in this work. A neural network classification was performed on real EEG data of healthy individuals in an attempt to experimentally investigate the relation between a person's EEG and genetically-specific information. In this paper two methods have been proposed; first a single channel method which uses the AR parameters of one channel as a feature vector and second a multichannel method which uses a combination of the AR parameters of different channel as a feature vector. These approaches have yielded correct classification scores at the range of 80% to 95% for the first method and at the range of 85% to 100% in the second one. Obviously it can be seen that combination of the AR parameters from different channels improve the score and if the number of channel, combined, increases there is a visible amendment in the percentage of correct classification. These results are in agreement with the previous researches showing evidence that the EEG carries genetic information, and also show the potential of our approach to classify known EEGs. 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.

The total result shows that the results in the back of the scalp over the parietal channels have a better identification than do in the other locations of the scalp. It can also be seen that by increasing order of model more than 11 in this results it isn't a specific change in percentage of correctness in the current experiments with 10 subjects; but for more subjects, it seems that probably the least suitable order of AR model is higher.

Although, calculating high order AR model parameters for a much number of subjects in order to identifying may not be reasonable or practical, but the only purpose in this paper is to show the potential of these parameters to classifying persons and demonstrating the probability of EEG uniqueness for individuals which reveal a direct relation between genetic information and EEG.

This team in its recent works tries to recognize EEG signals of an individual, recorded in distinct time, among others, and they reach to rather desirable results which will be spread soon.

TABLE 3
SCORES MULTICHANNEL METHOD

Order / Channel	001,002	002,003	003,004	004,005
3	90	57.5	70	95
5	95	77.5	95	95
7	92.5	82.5	92.5	97.5
9	100	90	95	100
11	100	90	95	100
13	100	92.5	95	100
15	100	90	95	100
17	100	87.5	97.5	100
19	100	90	97.5	100

Percentage of correct identification according to different order. (Combination of two channels)

TABLE 4
SCORES MULTICHANNEL METHOD

Order / Channel	004,014	014,030	030,050	050,080
3	95	87.5	77.5	75
5	97.5	97.5	95	92.5
7	85	92.5	92.5	95
9	97.5	97.5	97.5	100
11	97.5	97.5	95	100
13	97.5	97.5	100	100
15	95	97.5	97.5	100
17	97.5	97.5	95	100
19	97.5	97.5	95	100

Percentage of correct identification according to different order. (Combination of two channels)

TABLE 5
SCORES MULTICHANNEL METHOD

Order / Channel	001,002 ,003	003,004 ,005	004,014 ,030	014,030 ,050
3	87.5	87.5	90	87.5
5	85	95	97.5	97.5
7	90	97.5	92.5	100
9	100	100	97.5	100
11	100	100	97.5	100
13	100	100	97.5	100
15	100	100	97.5	100
17	100	100	97.5	100
19	100	100	97.5	100

Percentage of correct identification according to different order. (Combination of three channels)

TABLE 6
SCORES MULTICHANNEL METHOD

Order / Channel	030,040 ,050	050,060 ,070	040,060 ,070	080,090 ,100
3	95	92.5	95	95
5	97.5	92.5	97.5	95
7	97.5	92.5	100	97.5
9	100	95	100	100
11	100	95	100	100
13	100	95	100	100
15	100	95	100	100
17	100	95	100	100
19	100	95	100	100

Percentage of correct identification according to different order. (Combination of three channels)

REFERENCES

[1] A. Kasmia (Neurologist), "Personal Comunication," Oct.2000. Internet
[2] R.Schmidt , G.Thews " Integrative function of Nervous System," Spriger Verlag,Birlin1983.
[3] A.Remond(editor) "EEG informatics. A didactic review of methods and applications of EEG data processing," Elsevier Sientific Publishing Inc.,New York,1997.
[4] A.K.Jain (supervisor), "Biometrics Homepage Michigan State University,".http://biometrics.cse.msu.edu/info.html
[5] R.Paranjape,J.Mahovsky,L.Benedicenti , "On AR model and other metrics of the EEG for Subject identification," Internet.
[6] S.Jain,G.Deshpande, "Parametric Modeling of Brain Signals," Internet.
[7] S.Y.Kung, "digital Neural Netwok," (PTR Prentic Hall,Englewood Cliffs,New Jersey 1993)
[8] M.Poulos, M.Rangoussi, N.Alexandries, "Neural network based person identification using EEG features," Proceeding IEEE of the international conference on Acoustic , Speech,and Signal Processing, ICASSP'99,Arizona,USA,March 1999,PP. 1117-1120
[9] A. Anoklin, O. Fisher, Y. Mao, P. Vogt, E. Schalt, F. Vogel,(1999), "A genetic study of the human low-voltage electroencephalogram," Human Genetic ,vol. 90,PP. 99-112,1992.
[10] M. Poulos, M. Rangoussi, V. Chrissicopoulos, A. Evagelou, " person identification based on parametric processing on the EEG," Processings IEE of the Sixth International Conference on Electronics, Circuits and Systems, ICECS'99, Pafos, Cyprous. September 1999, PP. 283-286.
[11] M. Poulos, M. Rangoussi, V. Chrissicopoulos, A. Evagelou, "parametric person identification from the EEG using computational geometry," Processings IEE of the Sixth International Conference on Electronics, Circuits and Systems, ICECS'99, Pafos, Cyprous. September 1999, PP. 1005-1012.

[12] N. Hazarika, A. Tsoi, A. Sergejew, "Nonlinear Consideration in EEG signal classification," IEEE Transaction on Signal Processing,vol. 45, No. 4,PP. 829-836,1997.
[13] N. Nielsen, B. Harvand, "The electroencephalogram in univular twins brought up apart", Acta Genetica,vol. 8,PP. 57-64,1958.
[14] T. Kohonen, "Self-organization and associative memory," 3rd ed., Springer-Verlag, New York, 1988
[15] R. Plomin , "The role of inheritance in behavior," Science, vol. 248, PP. 183-188,1990.
[16] W. Lennox, E. Gibbs, F. Gibbs, "The brain-patern, an hereditary trail," The Journal of Heredity, vol. 36, PP. 233-243, 1945.
[17] M. Poulos, M. Rangoussi, E. Kafetzopoulos, "Person identification via the EEG using computational geometry algorithms," Proc. Intl. Conf. EUSIPCO'98, Rhodes, Greece, Sept. 1998.
[18] H. H. Stassen, G. Bomben, P. Propping, "Genetic aspects of the EEG: an investigation into the withinpair similarity of monozygotic and dizygotic twins with a new method of analysis," Electroencephalography and Clinical Neurophysiology,vol. 66, PP. 489-581, 1987.
[19] N.E. Sviderskaya, T.A. Korolkova, "Genetic Features of spatial organization the human cerebral cortex," Neuroscience and Behavioral Physiology, vol. 25, no. 5, 1995.
[20] J. Varner, R. Potter, J. Rohrbaugh, "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.
[21] F. Vogel, "The genetic basis of the normal EEG," Human Genetic, vol. 10, PP. 91-114, 1970.