IJBEM
Vol. 7, No. 2, 2005


# Multivariate Autoregressive Modeling Combined with Simulated Annealing Optimization for Classifying Sources of Event Related Potentials

C.E. Vasios[1], G.K. Matsopoulos[1], E.M. Ventouras[1], C. Papageorgiou[2], V.P. Kontaxakis[2], K.S. Nikita[1] and N. Uzunoglu[1]
[1]National Technical University of Athens, School of Electrical and Computer Engineering, GREECE
[2]Department of Psychiatry, Eginition Hospital, University of Athens, GREECE

*Abstract*—**Event-Related Potentials (ERPs) provide non-invasive measurements of the electrical activity on the scalp that are linked to the presentation of stimuli and events. Brain mapping techniques are able to provide evidence for the solution of debatable issues in cognitive science. In this paper, a two-step signal classification approach is proposed, extending the use of the Low Resolution Brain Electrical Tomography (LORETA) inversion technique. The first step concerns the feature extraction module, which is based on the combination of the Multivariate Autoregressive model with the Simulated Annealing technique. The classification module, as the second step of the methodology, is implemented by means of an Artificial Neural Network (ANN) trained with the back-propagation algorithm under "leave-one-out cross-validation". The ANN is a multi-layer perceptron, the architecture of which, is selected after a detailed search. The proposed methodology has been applied for the classification of first episode schizophrenic patients and normal controls using as input signals the intracranial current sources obtained by the inversion of ERPs using the LORETA technique. Results by implementing the proposed methodology provide classification rates of up to 93.1%. Finally, the proposed methodology may be used for the design of more robust classifiers based on the head-surface measured potentials as well as on the intracranial source locations, which directly relate to cognitive mechanisms.**

*Keywords*—**Multivariate Autoregression (MVAR), Simulated Annealing (SA), Classification, Cross-validation, LORETA**

## I. Introduction

The inversion of cognitive ERPs to intracranial current sources provides a method to observe brain phenomena related to information processing mechanisms. Various methods are currently used, mainly computing discrete brain dipoles or dipolar layers, which generate potentials, on the surface of a model of the intervening volume conductor, that best fit the measured ERPs [1-4]. The use of ERPs in psychiatry could be greatly enhanced by classification systems integrated in properly designed decision-support systems (DSS).

Scalar autoregressive (AR) coefficients, extracted from biosignals and treated as feature vectors in classification methods, have been widely used in designing DSS in medicine [5][6]. The use of AR coefficients extracted from single electroencephalogram (EEG) channels may miss information existing in the relation of multiple simultaneously recorded waveforms, reflecting the underlying intracranial generators' geometry and dynamics. In such cases, the AR model can be replaced by the multivariate autoregressive (MVAR) model [7].

In contrast to the use of features extracted from scalp-recorded EEG or ERPs, for the development of classification systems, relatively fewer studies explore the use of features from intracranial quantities [8-10].

In this paper, a method for the classification of multivariate autoregressive coefficients extracted by intracranial source signals, is proposed applying cross-validation and neural network architecture selection while extending the use of LORETA inversion techniques.

## II. Subjects and ERP Recording Procedure

Fourteen (14) never medicated FES patients (8 men and 6 women) with mean age 29 (±7) years were matched for age and sex to 30 healthy controls (20 men and 10 women) with mean age 31 (±3) years. Written informed consent was obtained from both patients and controls. Patients and controls were evaluated by a computerised version of the digit span Wechsler test [11]. The parameters calculated were ERPs for each subject as well as for each of the abductions Fp1, Fp2, F3, F4, Fz, C3, C4, Cz, (C3-T5)/2, (C4-T6)/2, P3, P4, Pz, O1, O2, resulting from the twenty six (26) test repetitions of the experimental procedure. These signals were then averaged as a pre-processing de-noising step of the procedure.

## III. Brain Electrical Tomography

LORETA was used to compute the 3-dimensional intracerebral distributions of current density. The algorithm solves the inverse problem assuming related orientations and strengths of neighboring neuronal sources. Mathematically this assumption is implemented by finding the 'smoothest' of all possible activity distributions.

The LORETA version used in the present study [12] was registered to the Talairach brain atlas [13]. Based on the digitised Talairach and probability atlases of the Brain Imaging Centre (Montreal Neurologic Institute), computations were restricted to cortical gray matter and hippocampus. Tha spatial resolution of the method was 7mm and the solution space consisted of 2394 voxels. The LORETA algorithm computed at each voxel current density as the linear weighted sum of the scalp electric potentials. Thus, LORETA combines the high time resolution of the EEG/ERP with a source localization method that permits truly three-dimensional tomography of the brain electrical acitvity.

## IV. Classification System

The proposed DSS consists of two basic modules: the feature extraction and the classification modules, as shown in Fig. 1.

The input to the first module is the signals of different intracranial sources of a specific ERP data set, for all subjects. The appropriate characteristics are extracted and processed by the feature extraction module, and then fed to the classification module. The structure of the neural network classifier is selected after a detailed search. The output of the DSS is one of two classes: patients or normal subjects.

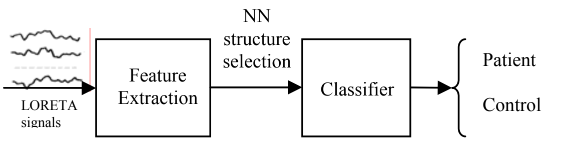


**Fig. 1:** Block Diagram of the proposed DSS for the classification of the ERPs into two classes: patients and healthy controls.

*A. Feature Extraction Module*

The feature extraction module comprises the implementation of the Multivariate Autoregressive model (MVAR model) in conjunction with the Simulated Annealing technique (SA technique), for the selection of optimum features.

The implementation of the Multivariate Autoregressive model to intracranial signals, keeps up with the idea that they are described by a linear filter fed with noise. According to this model, each value of the signal can be estimated using some previous values of it, as follows [14]:

$$x(k) = -\mathrm{A}(1)x(k-1) - \mathrm{A}(2)x(k-2) + \cdots - \mathrm{A}(p)x(k-p) + e(k)$$

where $x(k)$ is a $d-$dimensional vector of data at time $k$ and $e(k)$ is a $d-$dimensional vector of random input. The $A(i),\ i = 1,...,p$ are the $d \times d$ matrices of the autoregression coefficients to be estimated from $x(k),\ k = 1,...,N$ and $p$ is the model order. These coefficients construct the feature vector of each subject.

In this paper the Multivariate Autoregressive model is implemented in conjunction with the Simulated Annealing technique, according to the following procedure. Firstly, an optimum combination of abductions (number and kind) is found using the MVAR model in conjunction with the SA technique [15][16]. The optimum selection is based on the classification rate obtained by the Fuzzy C-Means Algorithm [17]. The optimum combination of abductions obtained by the previous step is further examined by a fine-tuning process in order to be finalized. This methodology is presented in pseudo-code as follows:

<u>Step 1:</u> Define the model order p.
<u>Step 2:</u> Search for the optimum combination of inputs using the SA technique
<u>Step 2.1:</u>Define the kind and number of inputs
Set initial Temperature.
Random selection of initial combination of inputs
For i=1 to a number of temperatures do
***Begin***
For j=1 to maximum number of combinations per temperature
***Begin***
<u>Step 2.2:</u> Selection of next combination of inputs based on the current combination of inputs and the current temperature
<u>Step 2.3:</u> Calculation of MVAR Coefficients
<u>Step 2.4:</u> Calculation of classification rate, using the Fuzzy C-Means algorithm
<u>Step 2.5:</u> Acceptance of the current combination based on the Boltzmann distribution
***End***
Reduction of Temperature
***End***

The selection of the next combination of input signals depends on the current one and the current temperature. The higher the temperature, the smaller the number of inputs that participate in each change. Given a combination of inputs, for the choice of the next combination, one of the following changes took place: a) insertion of inputs, b) abstraction of inputs, c) alteration of inputs, d) insertion and alteration of inputs, and e) abstraction and alteration of inputs. In cases of insertion or alteration of inputs, the lower the temperature, the smaller the distance between the new input and the rest.

According to the aforementioned Multivariate Autoregressive model, a feature vector is constructed for the finalized optimum combination of abductions, with dimensionality $p \times d \times d$, where $p$ is the model order and b is the number of abductions used.

*B. Selection of NN structure*

The selection of the topology of the feed-forward ANN is a methodological aspect that was investigated in the present work. After a preliminary investigation, it was found that satisfactory results, were achieved using two to four currents combination, with orders of 4 or 5. Therefore we chose to search for the optimum number of hidden layer neurons using source combination [881, 1266, 1496, 2226], leading to a feature vector of 64 (Table 1). First, we tested 3-layer neural networks with 1 and 2 output neurons and hidden layer neurons ranging from 4 to 40 (with steps of 4). Then we tested 4-layer neural networks with the 1st hidden layer neurons ranging from 4 to 40 (with steps of 4). For each number of neurons in the hidden layer the neurons in the second hidden layer varied from 4 to 40 (with steps of 4).

**Table 1:** Specificity and sensitivity computed for the case of source combination [881, 1266, 1496, 2226].

| | **Controls** | **FES** | **Accuracy** | |
|---|---|---|---|---|
| **Controls** (30) | 28 | 2 | 93.3 | **Specificity** |
| **FES** (14) | 3 | 11 | 78.5 | **Sensitivity** |
| **Accuracy** | 90.3 | 84.6 | 88.6 | **Overall** |
| | **Neg. predict. value** | **Pos. pred. value** | | |

The results of the tests suggested broadly similar performance for 3-layered and 4-layered networks with 1 or 2 output neurons. Furthermore, given that the input layer consisted of 64 neurons, the performance of the network was not significantly influenced by the number of neurons in the first hidden layer (which, of course, is the sole hidden layer in 3-layered ANNs), as long as that number was between 12 and 28 (approximately equal or more than 1/5 and equal or less than 2/5 of the number of input neurons). Reducing the neurons in the first layer to less than 12 led to a gradual reduction of the performance .

So, according to the above empirical results the networks used in the present study were 3-layered, with the ratio of neurons at the input and hidden layer around 5 and 1 neuron in the output layer.

### *C. Classification Module*

The second module is implemented with an ANN consisting of three layers (Fig. 2).

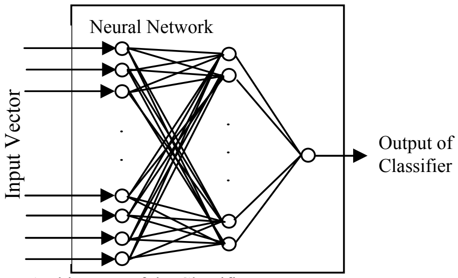


**Fig. 2:** Architecture of the Classifier.

The input layer consists of a number of neurons equal to the number of the selected features. The hidden layer contains number of neurons equal to the one fifth of the input neurons. The output layer consists of one neuron, encoding the two classes of the subjects: patient and normal (0=patient and 1=normal). The back-propagation algorithm [18] with adaptive learning rate and momentum (estimated under a trial-and-error process) has been used in order to train the ANN (Initial weights randomly selected in [–1,+1]. Log-sigmoid/tan-sigmoid used for the hidden/output layer, respectively).

In order to avoid overtraining and achieve an acceptable generalization in the classification, the "leave-one-out cross-validation" neural network implementation scheme was adopted [19]. According to this scheme, the neural network is trained using all the patients and control subjects, except from one (no mater if patient or control), which will be used for testing. The generalization ability of the specific network is tested using the single excepted subject. The above mentioned training-testing procedure is repeated using a different subject for testing, until all subjects are used once each. Under the "leave-one-out cross-validation" implementation, the applied neural networks present slight differences between each other, by inference of the slight variation of the training and validation sets and testing subject in each one. So, the aggregate sum of the correctly classified subjects can be considered as corresponding to the neural network that comes up after being trained with all patients and control subjects.

## V. RESULTS

### *A. Implementation Parameters*

The chosen time period of signals was the 500-800 msec time interval, because this time period corresponds to the P600 ERP component, which represents the completion of any synchronized brain operations concerning a decision taken after the presentation of a warning stimulus. Additionally, the order of the model used was tested for different values varying from 3 up to 15. The number of signals in a combination varies between 2 and 8.

The parameters of the SA technique were determined experimentally and set to: initial guess=random, initial temperature=5, percentage of temperature reduction at each iteration=5%, number of temperatures=50, maximum number of combinations per temperature=40.

When the MVAR model is used, in ERP modeling, requires the definition of several parameters such as the number and kind of signals, the time interval of the examined waveforms and the order of the model used. Even though in our study we considered a fixed time interval (500-800 msec) the search space constructed by the combination of the remaining aforementioned parameters (2394 signals and model orders between 3 and 15), seems practically non-manageable. So, as a pre-processing step, we tried to reduce the initial 2394 input signals, in order to reduce the computational intensity of our classification system. We made a comparative correlation testing of numerous source neighborhoods defined by placing each source on the centre of a cube and comparing its waveform to the waveforms of the sources that belong to the vertices of this cube. Conclusions indicate that adjacent sources have similar waveforms. Such a finding seems quite reasonable first because of the LORETA's choice of the smoothest inverse solution and second because of the relatively small electrodes/sources ratio (16/2394≈1/150). So, having in mind that similar waveforms do not affect the performance of the classification system, we finally ended up with 478 sources uniformly distributed in the cortical gray matter and hippocampus, having applied a source sub-sampling by order of 5.

**Table 4:** Performance (classification rates) of the *MVAR/SA* feature extraction method implemented on LORETA source data. The first column corresponds to the model order, the second to the resulting dimension of the feature vector, while the two last columns present the classification rate achieved by each source position combination and the misclassified patients and control subjects.

| Source positions | Order | Dimensio-nality | Classifica-tion Rate (%) | Misclassified subjects |
|---|---|---|---|---|
| 766, 1521 | 4 | 16 | **93.1** | 1/30 - 2/14 |
| 1476, 1661 | 5 | 20 | **90.9** | 2/30 - 2/14 |
| 571, 1366 | 4 | 16 | **90.9** | 2/30 - 2/14 |
| 1336, 1431, 1921 | 5 | 45 | **90.9** | 2/30 - 2/14 |
| 881,1266,1496,2226 | 4 | 64 | **88.6** | 2/30 - 3/14 |

### *B. Clinical Results*

Classification results obtained with the *MVAR/SA* method, are presented in Table 4. The highest classification rate (93.1%) was achieved for the source combination [766, 1521]. For combinations [1476, 1661], [571, 366] and [1336, 1431, 1921] the classification rate was 90.9% and for combination [881, 1266, 1496, 2226] was 88.6%. The dimensions of the feature vectors produced by the feature extraction module were 16, 20, 16, 45 and 64 respectively. In Fig. 3 we present the topographical distribution of the source positions for optimum combinations.

## VI. Discussion

The requirement to use information existing in source waveforms corresponding to concurrently recorded ERP waveforms leads to the creation of an unpractical large search space for selecting the MVAR model providing the best classification rate. The combination of the MVAR model with the SA optimization technique, as proposed in the present work, provides a principled way to reduce the computational complexity of the search process.

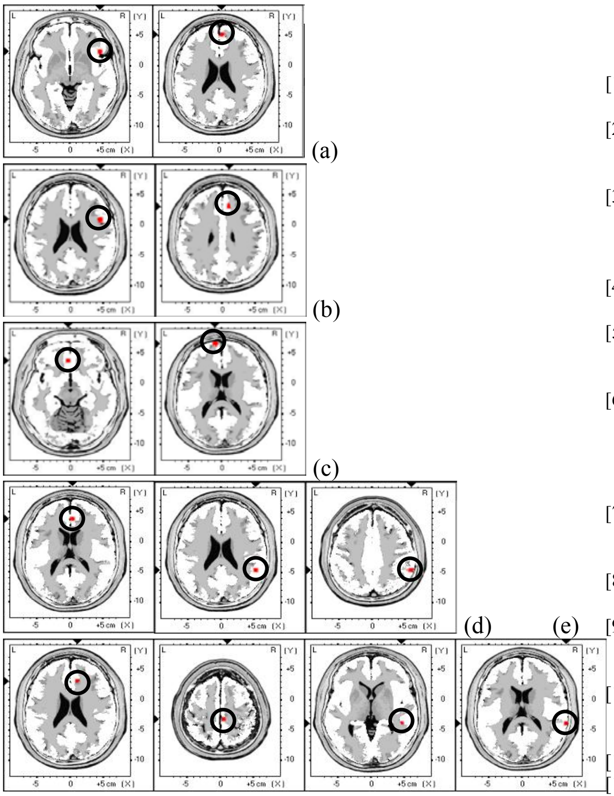


**Fig. 3:** Horizontal planes of optimum source combinations, registered to the cortical gray matter and hippocampus of the Talairach brain atlas. (a):[766, 1521], (b): [1476, 1661], (c):[571, 1366], (d):[1336, 1431, 1921], (e):[881, 1266, 1496, 2226]

In applications, where measured data are expensive or difficult to find and therefore limited, like FES patients, the design of neural-network based models is even more difficult, because the learning procedure gets worse with less data. Cross-validation uses a dynamic split of data managing to use all of the available data. In our study we applied cross-validation under the "leave-one-out" procedure, which consists of a cyclic allocation of the available data to the training and testing set, and allows use of a large part of the data in each cycle as training set and uses the remaining as testing set.

The rationale for investigating the capacity of source waveforms in providing satisfactory classification performance is that they are expected to be more robust to overlearning, since they might be more related to the actual pathophysiological processes [1].
The existence of frontal and right temporal source positions in combinations providing satisfactory classification results, may indicate the significance of these brain regions, in differentiating between normal and pathological mechanisms in schizophrenia, in agreement to research results concerning the involvement of those regions in schizophrenia [20][21].

Further research is currently carried out concerning the evaluation of other ERP inversion techniques that have been proposed in the literature, in order to compare the classification performance provided by the various parameters modeling brain electrical phenomena.

## References

[1] J.Z.Koles, Trends in EEG source localization, *Electroenceph Clin Neurophysiol*, 106 (1998) 127-137.
[2] M.Scherg, T.Bast and P.Berg, Multiple source analysis of interictal spikes: goals, requirements and clinical value, *J Clin Neurophysiol*, 16 (1999) 214-222.
[3] Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E: Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26: 55-67, 2000.
[4] C.Michel et al., Electric source imaging of human brain functions, *Brain Research Reviews*, 36 (2001) 108-118.
[5] R.Palaniappan and R.Paramesran, Using genetic algorithm to identify the discriminatory subset of multi-channel spectral bands for visual response, *Applied Soft Computing*, 2 (2002) 48-60.
[6] A.C.Tsoi, D.S.C.So and A.Sergejew. Classification of electroenchephalogram using artificial neural networks. In Jack D.Cowan, Gerald Tesauro and Joshua Alspector, editors, Advances in *Neural Information Processing Systems*, 6:1151-1158, Morgan Kaufmann, 1994.
[7] C.Vasios et al., A Decision Support System of Evoked Potentials for the Classification of Patients with First-Episode Schizophrenia, *German Journal of Psychiatry*, 5 (2002) 78-84.
[8] T.Dierks et al., Electrical brain activity in schizophrenia described by equivalent dipoles of FFT-data, *Schiz. Res.*, 14(1995) 145-154.
[9] D.Pizzagalli et al., Affective attitudes to face images associated with intracerebral EEG source location before face viewing, *Cognitive Brain Research*, 7 (1999) 371–377.
[10] C.Huang et al., Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study, *Clin Neurophysiol*, 111 (2000) 1961-1967.
[11] Jasper H. *Electroenc. Clin. Neurophysiol*. 10, 371-375 (1958)
[12] R.D. Pascual-Marqui, "Review of Methods for Solving the EEG Inverse Problem", *Intern. Jour. of Bioelectromag*., 1:75-86, 1999.
[13] J. Talairach and P. Tournoux, "Co-planar stereotaxic atlas of the human brain", Thieme, New York, 1988.
[14] Inger L. Simulated Annealing: practice versus theory. *J. Math. Comput. Modeling*,18:29-57, 1993
[15] W.H.Press et al., *Numerical Recipes in C*. Camb. Univ. Press, 1994.
[16] M.R.Rezaee et.al., Segmentation of MR images by a fuzzy c-mean algorithm. *Computers in Cardiology*, 21 – 24, 1995.
[17] Bishop, C.M., *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995.
[18] Haykin, S. *Neural Networks*, 2nd ed. New York: Macmillan, 1999.
[19] B. Schenker and M. Agarwal, Cross-validated structure selection for neural networks, *Computers Chem. Engng,* 12(2) 175-186, 1996
[20] Tekin S. et al.,. Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. J. Psychosom. Res. 2002;53(2):647-54.
[21] Grossberg S. The imbalanced brain: from normal behavior to schizophrenia. Biol Psychiatry. 2000;48(2):81-98.