Classification of Event-Related Potentials Associated with Response Errors in Actors

P. A. Asvestas, E. Ventouras, I. Karanasiou, and G. K. Matsopoulos

Abstract— **Event-Related Potentials (ERPs) provide non-invasive measurements of the electrical activity on the scalp related to the processing of stimuli and preparation of responses by the brain. In this paper, an ERP-signal classification method capable of discriminating between ERPs of correct and incorrect responses of actors is proposed. A number of histogram-related features were calculated from each ERP-signal and the most significant ones were extracted using the Sequential Forward Floating Selection algorithm along with the Fuzzy C-Means clustering algorithm. The Fuzzy C-Means algorithm was also used for the classification task. The approach yielded classification accuracy 93.75% for the actors' correct and incorrect responses. The proposed ERP-signal classification method provides a promising tool to study error detection and observational-learning mechanisms in joint-action research and may foster the future development of systems capable of automatically detecting erroneous actions in human-human and human-artificial agent interactions.**

I. Introduction

Event-related potentials (ERPs) provide non-invasive measurement of electrical activity on the scalp linked to the specific stimulus or response events [1]. The study of ERPs focuses on those parts of the average waveform that contain significant local maxima and minima, called peaks or components. When subjects commit incorrect actions, a negative deflection of the ERPs is produced, peaking at around 80 msec after the initiation of the incorrect response, called error-related negativity (ERN) [2], [3]. A positivity following the ERN, the so-called error positivity has also been described (Pe) [4], showing a maximum between 200 and 500 msec after the initiation of the incorrect response. Research has shown that ERN is elicited when there is a mismatch between representations of the actual response and the correct response [5], [6]. Recently, the focus of ERN-research has been extended to include also the mechanisms related to the observation of errors committed by others, in an effort to elucidate whether the mechanisms responsible for learning 'by doing' are similar to mechanisms of observational learning [7]. In this work, an ERN was also found in a condition where subjects observed the incorrect actions of another person involved in a modified Eriksen flanker task, albeit with a lower amplitude than the ERN for self generated errors and a later occurrence of the peak.

Classification algorithms to discriminate between ERPs have been developed for various applications. In [8], ERP data obtained from both normal control subjects and chronic schizophrenic patients were classified, using a parallel principal component neural network. The proposed architecture provided overall classification accuracy up to 90%. In [9], genetic algorithm and fuzzy ARTMAP classifier were combined to identify the discriminatory subset of the feature set for classification of alcoholics and non-alcoholics using brain rhythm extracted during visual stimulus. The feature set consisted of seven spectral power ratios extracted from multi-channel visual evoked potential (VEP) recordings. The classification performance reached 95.9%. A computer-based classification system capable of distinguishing patients with depression from normal controls by ERP signals using the P600 component was presented in [10]. The proposed system used a combination of support vector machine (SVM) classifiers and a majority-vote engine. The obtained classification accuracy was up to 94%. In [11], single-trial EEGs were classified by means of a perceptron neural network. Features were extracted from multichannel EEG using an algorithm that combined common spatial subspace decomposition with Fisher discriminant analysis. The obtained classification accuracy was 84%.

The existence of differences in the ERPs of actors' correct and incorrect responses creates the challenge to develop classification systems aimed at discriminating between such actions in real-time on the basis of single-trial EEG recordings, independently of the subsequent actions that the subject committing the error might take or not. One step further, the existence of differences in the ERPs of observers, when observing correct and incorrect actions, might foster the development of classification systems capable of detecting performance errors of a human - or an artificial agent – in need of being monitored in a joint-action situation. The primary aim of the present study was the development and implementation of a classification system for discriminating correct and incorrect responses, based on scalp-recorded ERPs of actors, using histogram-related features. The present work is based on subjects' averages and is a first step towards future single-trial classification.

II. Subjects and ERP Recording Procedure

The ERP data used in the present study were collected in previous research [12]. The data were acquired from 16

Manuscript received July 5, 2008.

P. A. Asvestas and E. Ventouras are with the Department of Medical Instruments Technology, Faculty of Technological Applications, Technological Educational Institute of Athens, Ag. Spyridonos Str., Egaleo 122 10 Athens, Greece.

I. Karanasiou and G. K. Matsopoulos are with the School of Electrical & Computer Engineering, National Technical University of Athens, 9 Heroon Polytechniou Str., 157 80, Zografou, Athens, Greece (corresponding author I. Karanasiou: tel. +302107722289, e-mail: ikaran@esd.ece.ntua.gr).

healthy volunteers. Participants were faced in front of a table facing an experimenter, having in front of them, on the table, two joystick devices positioned to the left and right of a Led stimulus device. Sixteen participants performed a modified Eriksen flanker task in which they responded to the direction of a center arrowhead surrounded by distracting flankers pointing either in the same direction as the center arrow, or in opposite direction. EEG activity was recorded from 47 electrodes, as well as vertical and horizontal electro-oculograms (Fig. 1) with sampling rate 250 Hz. Correct and incorrect trials were averaged over a 800 ms epoch (baseline [-100 , 0] ms before response). Trials to be included in the averaging process had been selected according to an RT-matching procedure between correct and incorrect trials (described in [12]) to mitigate the differential contribution of stimulus-related activity in the ERP. A time window, starting at -6 msec and ending at 700 msec (corresponding to 176 samples) after the response, was selected for analysis. A total of 32 × 47 = 1504 ERP recordings were available for analysis. From the available recordings, 16 × 47 = 752 recordings corresponded to correct responses and the rest 16 × 47 = 752 recordings corresponded to incorrect responses.

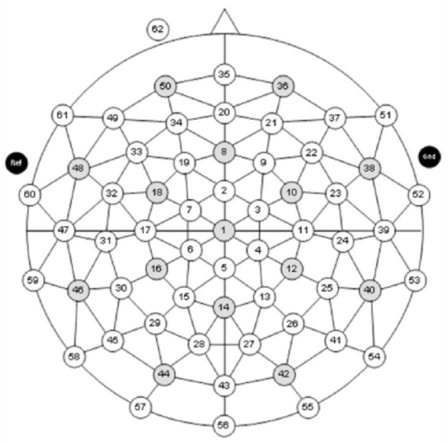

Fig. 1. Graphic representation of the electrode placement.

III. Classification Methodology

The proposed methodology consists of three stages:

- Feature calculation
- Feature selection
- Classification

Each stage is described analytically below.

A. Feature Calculation

Let $y_{\min}$ and $y_{\max}$ be the minimum and maximum value of the ensemble of the 1504 ERP recordings for the time window [-6, 700] msec, respectively. Let $\{c_k\}$, $k = 0,1,\ldots,M$ ($c_k < c_{k+1}$), be a uniform partition of the interval $[y_{\min}, y_{\max}]$ into M subintervals (or bins), namely:

$$c_k = y_{\min} + k\frac{y_{\max} - y_{\min}}{M}$$

Furthermore, each subinterval is represented by its central value:

$$\overline{c}_k = \frac{c_k + c_{k+1}}{2}, k = 0,1,\ldots,M-1$$

Given the aforementioned partition, the histogram of an ERP recording can be calculated according to the following formula:

$$H_k = \sum_{i=1}^{N}\left[s\left(y_i - c_k\right) - s\left(y_i - c_{k+1}\right)\right]$$

where $\{y_1,\ldots,y_N\}$, $N = 176$, are the samples of the ERP recording that correspond to the time interval [-6, 700] msec and $s(x) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases}$ is the Heaviside function. The value of H_k ($k = 0,1,\ldots,M-1$) provides the number of y_n ($n = 1,\ldots,N$) that lie in the subinterval $[c_k, c_{k+1})$. Therefore, the probability, p_k, that a value y_n falls in the subinterval $[c_k, c_{k+1})$ is given by the following relation:

$$p_k = \frac{H_k}{N}$$

The following features are calculated from the histogram of each ERP recording:

1. *Mean value*, which quantifies the central value of a distribution:
$$\mu = \sum_{k=0}^{M-1} p_k \overline{c}_k$$
2. *Standard deviation*, which is a measure of variability around the mean value:
$$\sigma = \sqrt{\sum_{k=0}^{M-1}\left(\overline{c}_k - \mu\right)^2 p_k}$$
3. *Skewness*, which characterizes the degree of asymmetry of a distribution around its mean:
$$skew = \frac{\sum_{k=0}^{M-1}\left(c_k - \mu\right)^3 p_k}{\sigma^3}$$
4. *Kurtosis*, which measures the relative peakedness or flatness of a distribution:
$$kurt = \frac{\sum_{k=0}^{M-1}\left(c_k - \mu\right)^4 p_k}{\sigma^4}$$
5. *Entropy*, which is a measure of uniformity of the histogram:
$$Entr = -\sum_{k=0}^{M-1} p_k \log_2 p_k$$
6. *Energy*: $Ener = \sum_{k=0}^{M-1} p_k^2$
7. *Median*, which is the number separating the higher half of a distribution, from the lower half :

$med = \bar{c}_{k_m}$, where k_m such that $\sum_{k=0}^{k_m} p_k \geq 0.5$ and $\sum_{k=k_m}^{M-1} p_k \geq 0.5$.

Furthermore, the following features are calculated for each ERP recording:

8. Maximum value of samples: $\max_n \{y_n\}$
9. Minimum value of samples: $\min_n \{y_n\}$
10. Index of maximum value of samples: $\arg\max_n \{y_n\}$
11. Index of minimum value of samples: $\arg\min_n \{y_n\}$

In total, from each participant's ERPs $47 \times 11 = 517$ features are calculated.

B. Feature Selection

Due to the high number of calculated features, it is necessary to eliminate features that are linearly correlated or carry no diagnostic information. Therefore, a process of feature selection is applied prior to classification, with the purpose of discovering a subset of features that optimize the classification process, in terms of accuracy. The sequential float forward search (SFFS) technique has been employed as a feature selection process, which is formulated as follows [13]: let $Y = \{1, 2, \ldots, P\}$ be the available features ($P = 517$ in our case), $P_{\max}$ is the maximum number of features to be extracted ($P_{\max} \leq P$), $X_k \subseteq Y$ is a subset of features containing k features and $J : \Omega \to R$ is an evaluation function (e.g. the accuracy of a classifier), where Ω denotes the set of all possible subsets of Y. Then,

- $X_0 = \varnothing$ and $k = 0$
- *while* $k < 2$
  - *Find the most significant feature*: $y = \arg\max_{a \in (Y - X_k)} J(X_k \cup \{a\})$
  - $X_{k+1} = X_k \cup y$
  - $k = k + 1$
- *while* $k \leq P_{\max}$
  - *Find the most significant feature:* $y = \arg\max_{a \in (Y - X_k)} J(X_k \cup \{a\})$
  - $X_{k+1} = X_k \cup y$
  - $k = k + 1$
  - *while* $k > 2$
    - *Find the least significant feature*: $x = \arg\max_{a \in X_k} J(X_k - \{a\})$
      - *if* $J(X_k - \{x\}) > J(X_{k-1})$
        - $X_{k-1} = X_k - \{x\}$
        - $k = k - 1$
      - *else break*
  - *end*
- *end*

The best feature set, X^*, is the one with the minimum cardinality for which the maximum value of the evaluation function is obtained, namely:

$$X^* = X_{k^*}$$
$$k^* = \arg\min_k \{J(X_k) = J_{\max}\}$$
$$J_{\max} = \max_k \{J(X_k)\}$$

In the present study, the selected evaluation function was the clustering accuracy of the fuzzy c-means (FCM) algorithm. In particular, FCM is an unsupervised clustering algorithm which allows one feature vector to belong to two or more clusters. It is based on minimization of the following objective function:

$$G_m = \sum_{i=1}^{Q} \sum_{j=1}^{C} u_{ij}^m \left\| \mathbf{x}_i - \mathbf{c}_j \right\|^2$$

where $\mathbf{x}_i$ is a *d*-dimensional feature vector, *m* is any real number greater than 1 (in our case, $m = 2$), u_{ij} is the degree of membership of $\mathbf{x}_i$ in the cluster *j* ($\sum_{j=1}^{c} u_{ij} = 1$), $\mathbf{c}_j$ is the d-dimensional center of the cluster, *Q* and *C* are the number of data vectors and clusters, respectively and $\| \ \|$ is any norm expressing the similarity between any measured data and the center.

The algorithm is composed of the following steps:

1. *Initialize matrix U randomly,* $U^{(0)} = \left[u_{ij}^{(0)}\right]$ *and set* $G_m^{(0)} = 0$
2. *For* $k = 1, 2, \ldots$

$$\mathbf{c}_j^{(k)} = \frac{\sum_{i=1}^{N} \left(u_{ij}^{(k-1)}\right)^m \mathbf{x}_i}{\sum_{i=1}^{N} \left(u_{ij}^{(k-1)}\right)^m}$$

$$u_{ij}^{(k)} = \frac{1}{\sum_{p=1}^{C} \left(\frac{\left\| \mathbf{x}_i - \mathbf{c}_j^{(k)} \right\|}{\left\| \mathbf{x}_i - \mathbf{c}_p^{(k)} \right\|} \right)^{\frac{2}{m-1}}} \quad (1)$$

3. *If* $\left| G_m^{(k)} - G_m^{(k-1)} \right| < \varepsilon$ *then STOP; otherwise return to step 2.*

As already mentioned, data are bound to each cluster by means of a membership function, which represents the fuzzy behavior of this algorithm. The data vector $\mathbf{x}_i$ is assigned to the cluster k using the following rule:

$$k = \arg\max_{j} \left\{ u_{ij} \right\} \quad (2)$$

Then, a confusion matrix can be obtained by counting the number of data vectors that are assigned to each cluster. For example, for two clusters, C = 2, the confusion matrix is as follows:

$$CM = \begin{bmatrix} cm_{11} & cm_{12} \\ cm_{21} & cm_{22} \end{bmatrix}$$

where cm_{pj} ($p, j = 1,2$) denotes the number of data vectors from class p that are assigned to cluster j. Consequently, the clustering accuracy, CA, can be quantified as follows:

$$CA = \frac{\max\left\{ cm_{11} + cm_{22}, cm_{21} + cm_{22} \right\}}{Q}$$

C. *Classification*

The FCM algorithm was also used for the classification in conjunction with the leave-one-out cross-validation procedure [14]. The leave-one-out procedure was adopted in order to test the performance of the FCM classifier in a reliable manner, taking into account the limited number of cases available in the classes, and in the same time achieving an acceptable generalization in the classification. The procedure that was used to discriminate between correct and incorrect responses of actors is described below. According to this procedure, the two clusters were formed by means of the FCM algorithm using feature vectors from both response types (correct and incorrect), except from one (no matter whether it was a correct or incorrect response), that was used for testing afterwards. Each cluster was assigned a label according to the class label of the majority of its member feature vectors: if the majority of the features vectors that formed the cluster corresponded to correct (incorrect) responses then the cluster was assigned the label “Correct” (“Incorrect”). The generalization ability of the classifier was then tested using the feature vector that was singled out. This feature vector was assigned to a cluster according to (1) and (2) and it was assumed to be classified correctly if it had the same label with the cluster. The above procedure was repeated using different subject feature vectors for testing, until all subject feature vectors were used once. The classification parameters were computed by the aggregate sums of correctly classified or misclassified correct and incorrect responses.

IV. Results

As was mentioned in Section III.A 517 features were calculated from each participant’s ERPs. The SFFS algorithm was applied using the 32 available feature vectors. The maximum number of features to be extracted was set to $P_{\max} = 10$ and 111 subintervals (bins) were used for the histogram calculation. The obtained value of the evaluation function, $J_{\max}$, was 0.969. In Table I, the features and the corresponding electrodes that were finally selected are shown. Furthermore, the center of each cluster, when using the selected features with the FCM algorithm, is included in the Table.

TABLE I
EXTRACTED FEATURES, CORRESPONDING ELECTRODES AND CLUSTER CENTERS

Feature	Electrode Number	Cluster Center	
		Cluster 1 (“Correct”)	Cluster 2 (“Incorrect”)
Skewness	34	0.0207	-0.0414
Skewness	44	0.2108	0.0796
Kurtosis	2	3.0145	3.3946
Kurtosis	47	2.7621	2.9504
Entropy	24	3.3424	3.3517
Mean	37	-1.1160	-1.1666
Mean	49	-1.6298	-1.5994

The placement of the selected electrodes is shown in Fig. 2.

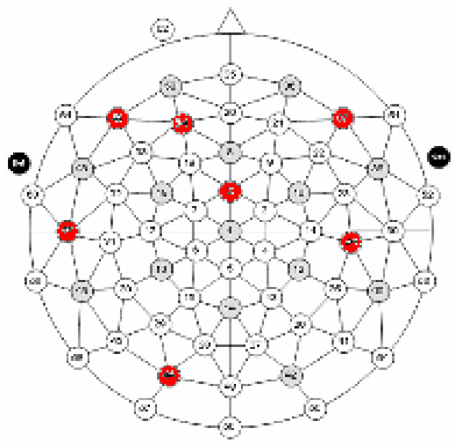

Fig. 2. Graphic representation of the electrodes that were finally selected.

Considering the results that are listed in Table I, the following conclusions can be drawn:

- Cluster 1 has a positive value of skewness for electrode 34 and Cluster 2 has a negative one. This observation signifies that the histograms for ERPs from correct (incorrect) responses tend to have an asymmetric tail extending out towards more positive (negative) values.
- Cluster 1 has a larger positive value of skewness for

electrode 44 than Cluster 2, which in turn means that the histograms of ERPs from correct responses have a larger asymmetric tail towards positive values.

- The kurtosis for electrodes 2 and 47 are positive for both Cluster 1 and Cluster 2 with larger values occurring in Cluster 2. This means that the corresponding histograms are leptokurtic for both clusters and the most leptokurtic ones are from Cluster 2.
- The entropy of electrode 24 is slightly larger in Cluster 2 than in Cluster 1, which in turns means that the corresponding histograms of Cluster 1 are slightly more uniform than these of Cluster 2.
- The mean value of electrode 37 is slightly larger for members of Cluster 1 than Cluster 2. The opposite is observed for the mean value of electrode 49.

The classification results, using the obtained features and the procedure described in Section III.C, are shown in Table II. The first column of the Table refers to the numbering of each feature vector. The second column contains the actual class of the feature vector, where a "1" ("2") indicates a feature vector obtained from a participant's ERPs for correct (incorrect) response. The next two columns list the membership of the feature vector to each cluster. Finally, the last column shows if the feature vector was classified correctly (√ mark) or not (x mark).

TABLE II
CLASSIFICATION RESULTS FOR EACH FEATURE VECTOR USING THE FCM ALGORITHM AND THE LEAVE ONE OUT PROCEDURE

Testing Feature Vector	Actual Class	Membership Value Cluster 1 ("Correct")	Cluster 2 ("Incorrect")	Correctly Classified
1	1	0.5497	0.4503	√
2	1	0.5673	0.4327	√
3	1	0.5119	0.4881	√
4	1	0.5909	0.4091	√
5	1	0.5545	0.4455	√
6	1	0.5885	0.4115	√
7	1	0.5322	0.4678	√
8	1	0.5887	0.4113	√
9	1	0.5774	0.4226	√
10	1	0.5448	0.4552	√
11	1	0.5787	0.4213	√
12	1	0.5234	0.4766	√
13	1	0.5728	0.4272	√
14	1	0.5550	0.4450	√
15	1	0.5954	0.4046	√
16	1	0.5352	0.4648	√
17	2	0.4825	0.5175	√
18	2	0.4947	0.5053	√
19	2	0.4615	0.5385	√
20	2	0.4733	0.5267	√
21	2	0.4388	0.5612	√
22	2	0.4975	0.5025	√
23	2	0.5024	0.4976	x
24	2	0.4523	0.5477	√
25	2	0.4556	0.5444	√
26	2	0.4731	0.5269	√
27	2	0.4767	0.5233	√
28	2	0.4798	0.5202	√
29	2	0.5863	0.4137	x
30	2	0.4989	0.5011	√
31	2	0.4607	0.5393	√
32	2	0.4853	0.5147	√

The results of Table II are summarized in Table III, in the form of a confusion matrix

TABLE III
CONFUSION MATRIX

Predicted Class	Actual Class "Correct"	"Incorrect"
"Correct"	16	0
"Incorrect"	2	14

The confusion matrix indicates that:

- 100% of the feature vectors from ERPs of correct responses were classified correctly
- 87.5% of the feature vectors from ERPs of incorrect responses were classified correctly
- 93.75% of the total number of feature vectors was classified correctly.

V. DISCUSSION

Several issues should be discussed about the proposed methodology. The first issue regards the feature selection approach. The proposed approach is a so-called wrapper-based feature selection, which means that the performance of the classifier is used as an evaluation function. On the other hand, there are the filter-based approaches where an independent criterion, such as the mutual information [15], is used for the evaluation function. The rationale for adopting a wrapper-based approach is that it is more probable to obtain accurate classification results using features that maximize the performance of the classifier itself, than using features that optimize a criterion that does not depend on the classifier. The feature selection is performed by means of the SFFS method. Although the SFFS is a sub-optimal method that cannot guarantee to provide the best subset of features, its performance has been found to be very good compared with other search methods and, it is computationally much more efficient than the branch and bound method.

A second issue that should be considered is the classifier, which was also used in the feature selection process. As already mentioned, the FCM algorithm is an unsupervised clustering algorithm, which means that it does not require the splitting of the available feature vectors in training and testing sets, as happens with other classification algorithms (for example, k-Nearest Neighbor, Neural Networks, etc). This is an advantage in our case, since the available data are rather limited. Furthermore, FCM is easy to implement and is characterized by very good performance. If a larger set of data were available, then the k-Nearest Neighbor or the Support Vector Machines (SVM) [16] algorithm could be used.

A very significant issue is the number of bins that were used for calculating the histograms. There is no "best" number of bins, and different bin sizes can reveal different aspects of the data. Several formulas [17], [18] have been proposed in order to obtain an optimal number of bins, but these formulas generally make strong assumptions about the shape of the distribution. An interesting alternative would be to use kernel density estimation (Parzen window) [16], where the underlying distribution of the data is modeled by the mixture of known probability density functions (usually Gaussian). As was mentioned before, the number of subintervals (bins) was 111. This number was determined after performing the feature selection process for different numbers of bins and recording the obtained value of the evaluation function, $J_{\max}$. In Fig, 3, it is shown a plot of the $J_{\max}$ with respect to the number of bins. As is evident, the maximum value of $J_{\max}$ occurs when 111 bins are used for the histogram calculation

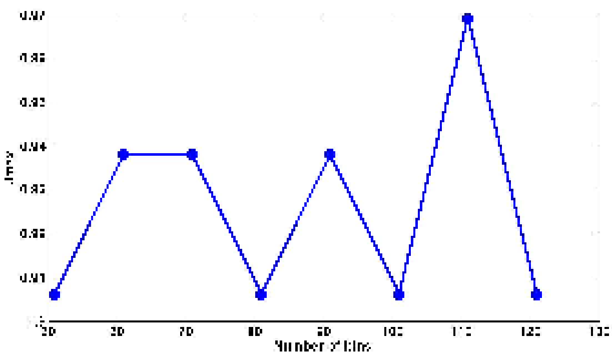

Fig. 3. Plot of the maximum value of the evaluation function for feature selection versus the number of histogram bins.

VI. Conclusion

In this paper, a methodology capable of discriminating between an actor's brain potentials that accompany correct and incorrect responses was presented. The methodology consisted of two steps: the feature selection, which was a combination of a sub-optimal search method and the FCM algorithm, and the classification which was based on the FCM algorithm using a leave one out procedure. The proposed methodology reduced significantly the initial large number of features, providing highly accurate results.

Acknowledgment

This research project is co-financed by E.U.-European Social Fund (80%) and the Greek Ministry of Development-GSRT (20%).

The authors would like also to thank Hein van Schie and Ellen de Bruijn from the Nijmegen Institute for Cognition and Information (NICI), The Netherlands, for kindly providing the data of their experiments and for their contribution to initial stages of the research.

References

[1] R. Jr. Johnson, "Event-Related Brain Potentials and Cognition," in *Handbook of Neuropsychology*, vol. 10, Elsevier, Amsterdam, The Netherlands, 1995.

[2] M. Falkenstein, J. Hohnsbein, J. Hoormann, L. Blanke, "Effects of errors in choice reaction tasks on the ERP under focused and divided attention" in Psychophysiological Brain Research, C.H.M. Brunia, A.W.K. Gaillard, and A. Kok, (Eds.) Tilburg Univ. Press, Tilburg, The Netherlands, pp. 192–195.

[3] W.J. Gehring, B. Goss, M.G.H. Coles, D.E. Meyer, E. Donchin, "A neural system for error detection and compensation," *Psychol. Sci.*, vol. 4, pp. 385–390, 1993.

[4] M. Falkenstein, J. Hohnsbein, J. Hoormann, L. Blanke, "Effects of cross-modal divided attention on late ERP components: II. Error processing in choice reaction tasks," *Electroencephalogr. Clin. Neurophysiol.,* vol. 78, pp. 447– 455, 1991.

[5] M.G. Coles, M.K. Scheffers, C.B. Holroyd, "Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components and the theory of error-processing," *Biol. Psychol.*, vol. 56, pp. 173-189, 2001.

[6] M. Falkenstein, J. Hoormann, S. Christ, J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biol. Psychol.,* vol. 51, pp. 87–107, 2000.

[7] H. van Schie, R. B. Mars, M. G. H. Coles, H. Bekkering, " Modulation of activity in medial frontal and motor cortices during error observation," *Nature Neuroscience*, vol. 7, pp. 549-554, 2004.

[8] J. R. Sveinsson, J. A. Benediktsson, S. B. Stefansson, K. Davidsson, "Parallel principal component neural networks for classification of event-related potential waveforms," *Med. Eng. Phys*., vol. 19, pp. 15-20, 1997.

[9] R. Palaniappan, R. Paramesran, "Using genetic algorithm to identify the discriminatory subset of multi-channel spectral bands for visual response," *Appl. Soft. Comput*., vol. 2, pp. 48-60, 2002.

[10] I. Kalatzis, N. Piliouras, E. Ventouras, C. C. Papageorgiou, A. D. Rabavilas, D. Cavouras, "Design and Implementation of an SVM-based Computer Classification System for Discriminating Depressive Patients from Healthy Controls using the P600 Component of ERP Signals", *Comput. Meth. Prog. Biol*., vol. 75, pp. 11-22, 2004.

[11] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao, F. Yang, "BCI Competition 2003— Data Set IV: An Algorithm Based on CSSD and FDA for Classifying Single-Trial EEG," *IEEE Trans. Biomed. Eng*., vol. 51, pp. 1081-1086, 2004.

[12] H. van Schie, R.B. Mars, M. G .H. Coles, H. Bekkering, "Modulation of activity in medial frontal and motor cortices during error observation," *Nature Neuroscience*, vol. 7, pp. 549-554, 2004.

[13] P. Pudil, J. Nonovocova, J. Kittler, "Floating search methods in feature selection," *Pat. Recog. Let*., vol. 15, pp. 1119-1125, 1994.

[14] B. Schenker, M. Agarwal, "Cross-validated structure selection for neural networks," *Comput. Chem. Eng*., vol. 20, pp. 175-186, 1996.

[15] T. W. S. Chow and D. Huang, "Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information," *IEEE Trans. Neural Networks*, vol. 16, pp. 213 – 224, 2005.

[16] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, New York : Wiley, 2001.

[17] D. W. Scott, "On optimal and data-based histograms", *Biometrika*, vol. 66pp. 605–610, 1979.

[18] D. Freedman, P. Diaconis, "On the histogram as a density estimator: L2 theory", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 57, pp. 453–476, 1981.