34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Effect of Latency on Clustering of P300 Recordings for ADHD Discrimination

D. H. Peluffo-Ordoñez[1], J.D. Martinez-Vargas[1] and G. Castellanos-Dominguez[1]

***Abstract*— This paper is focused on testing the latency contribution as regards the quality of formed groups for discriminating between healthy and attention deficit hyperactivity disorder children. To this end, two different cases are considered: nonaligned original recordings and aligned signals according to P300 position. For latter case, a novel approach to conduct time location of P300 component is introduced, which is based on derivative of event-related potential signals. The used database holds event-related potentials registered in auditory and visual oddball paradigm. Several experiments are carried out testing both configurations of considered data matrix. For grouping input data matrices, the *k-means* clustering technique is employed. To assess the quality of formed clusters and the relevance for clustering of latency-based features, relative values of distances between centroids and data points are computed in order to apprise separability and compactness of estimated clusters. Experimental results show that time localization of P300 component is not a decisive feature in formation of compact and well-defined groups within a discrimination framework for two considered data classes under certain conditions.**

I. INTRODUCTION

One of the most common psychiatry disorders in childhood is the attention–deficit hyperactivity disorder (ADHD) [1], which is diagnosed according to frequent and developmentally-inappropriate over-activity, inattention and impulsiveness high level. Typically, its diagnosis is done by taking into consideration the clinical criteria of `DSM-IV` or `ICD-10`, supported by the conduct outlined in questionnaires applied to parents and teachers. However, until now there do not exist conclusive tests or biological markers able to properly diagnose this behavioral disorder [2].

Event-related potentials (ERPs) are brain electrical signals generated as a response to an external sensorial stimulus. They have been useful in investigations on perceptual and cognitive-processing deficits, specially in children with ADHD, since these potentials are physiologically correlated with neuro-cognitive functions. The typical features used for analyzing cognitive processes are the areas and the peaks of the ERP components, defined by the mean and peak to peak voltages, respectively, which are computed by windowing the recordings in time domain time. However, needed parameters are commonly determined by visual inspection of the averaged ERP waveforms [3], which represents a great drawback for designing an automatic classification system.

Literature has reported that certain underlying processes can be found in the sequence of characteristic peaks and troughs from ERP. Probably, P300 component is the most studied ERP component in investigations of selective attention and information processing, due partly to its relatively large amplitude and facile elicitation in experimental contexts [4]. Numerous studies have shown the existence of alterations of ERPs in children with ADHD, especially in latency and amplitude of P300 component; however there is no a final consensus determining the types of variations of these parameters regarding children with behavioral disorders. For instance, in relation to latency of P300 wave in visual tasks, [5] reported a shorter latency in children with ADHD compared with control children; on the other hand, [6], in auditory and visual tasks, suggest that there are no differences in latency of ADHD and control children, whereas [7] proved that ADHD children have a longer latency than control children.

The scope of this paper is to test the effect of latency in the conformation of clusters that are associated with both considered classes: control and ADHD. To achieve this purpose, an unsupervised technique is used to compute distances between cluster centroid generated and original feature space, when only using latency and other morphological features. Furthermore, to estimate the relevance of latency, the P300 component of all recordings is aligned according to location of that wave on a pattern signal - it is determined for each class. Then, the distances obtained with original signal clusters are compared with those generated by aligned signals clusters. Obtained results show that latency does not have a relevant effect in formation of well-defined clusters under a criteria of separability and compactness.

II. METHOD AND MATERIALS

A. Data Base

Data recordings were collected from 120 children belonging to educational institutions of the metropolitan area of the Manizales (60 labeled as healthy control and 60 as ADHD). The subjects aging between 4 and 15 years old, were medically diagnosed based on clinical criteria of `DSM-IV` and `minikid` criteria by a multidisciplinary specialist team consisting of a general physician, psychologist, neuro-psychologist and experts in children psychiatric disorders. Both groups were tested under the same lighting and noise conditions, and were defined by the following inclusion criteria: non abnormality physical examination, normal visual and hearing ability, intellectual coefficient greater than 80 and, if necessary, pharmacological management previously suspended. Subjects were verified to not be diagnosed with another neurological disorders.

[1] Universidad Nacional de Colombia, Manizales, Caldas cgcastellanosd@unal.edu.co

Recordings were acquired by means of electrodes located in the head midline (i.e., Fz, Cz, Pz) according to 10−20 international system, with a sampling frequency equals to 640. Signals acquisition procedure extracted 1 *s* before and after stimulus appeared. As evaluation protocol, the oddball paradigm in auditory and visual modalities was applied over analyzed subjects. The first procedure involves the emission of 80 *dB* tone lasting 50 *ms*, with a frequency of 1.000 *Hz* for frequent stimulus and 3.000 *Hz* for target stimulus, presented randomly every 1.5 *s*. In the visual modality of the test, the subject is asked for watching a screen located 1 *m* before showing a consistent pattern image (checkerboard with 16 squares), which is fixed as the frequent stimulus. In turn, the rare stimulus is the presentation of a target in the center of the screen with the same common pattern in the background. So, the subject had to press a button whenever the unusual stimulus had appeared. Each testing included 200 stimuli, of which 80% are non-target and 20% remaining are target stimuli.

B. Characterization

Morphological features related to time distribution of waveform are only considered in this work, consisting of parameters measured over a windowed recordings. The following 16 morphological feature set that had shown an adequate performance in other similar studies [8], [9] is used: latency (time between stimulus and P300), amplitude (signal value on P300 point), latency/amplitude ratio, absolute amplitude, positive area, negative area, total area, absolute total area, total absolute area, average absolute signal slope, peak-peak value (amplitude measured between N200 and P300 components), peak-peak value time window (time elapsed between N200 and P300 waves), peak-peak slope, zero crossings, zero crossings density and slope sign alterations.

As a result, the input data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]$ is formed, where $\boldsymbol{x}_i \in \mathbb{R}^{1 \times p}$ is a p-dimensional feature vector associated with i-th subject. Matrix $\boldsymbol{X}$ is normalized to guarantee the scale coherence in representation of data, using: $\boldsymbol{x}_i \leftarrow (\boldsymbol{x}_i - \mu(\boldsymbol{x}_i))/\sigma(\boldsymbol{x}_i)$, where $\mu(\cdot)$ and $\sigma(\cdot)$ are a mean and a standard deviation operator, respectively.

P300 Localization: Since some features directly depend on location of P300 wave, the proper detection of such wave is a decisive task. Although, some reports point out that P300 component is the trough closer to 300 *ms*, other authors refer that this wave is not necessarily present at that concrete time instant, but in contrast its latency can show variations because of the effect of neurological disorders [10]. On this account, an algorithm based on signal derivative is applied on time windows in order to automatically detect the P300 wave taking into consideration real medical criteria. To estimate the location of P300 component and calculate latency-based features, an algorithm summarized in 1 that is based on derivative of ERP is introduced.

C. Dynamic Resampling

To analyze the separability and compactness of groups formed from morphological features, all recordings are

Algorithm 1 *Localization of P300 component*

Given a signal $s(t)$
1. Set analysis window: $t \in (t_1, t_2)$
2. Localize the local minimums $\boldsymbol{p} = [p_1, \dots, p_M]$ of $s(t)$ ranged into the interval (t_1, t_2), where M is the number of peaks detected in such interval.

3. Compute $\tilde{s}(t)$ as the derivative of $s(t)$
4. Localize all peaks (maximums and minimums) $\tilde{\boldsymbol{p}} = [\tilde{p}_1, \dots, \tilde{p}_N]$ of $\tilde{s}(t)$, where N is the total number of peaks of $\tilde{s}(t)$
5. Determine the time location of all points of $\boldsymbol{p}$ on $\tilde{s}(t)$
6. For each time location determined in step 5, compute the euclidian distance between its immediately previous and posterior peak as follows
$$\delta = \sqrt{(\tilde{s}(\tilde{p}_j) - \tilde{s}(\tilde{p}_{j-1}))^2 + (\tilde{p}_j - \tilde{p}_{j-1})^2}$$
7. Form the distance vector $\boldsymbol{\Delta} = \{\delta_i, i = 1, \dots, M\}$
8. Choose P300 component at $\max_i\{\boldsymbol{\Delta}\}$

aligned and re-sampled to locate the P300 component at the same time point according to a pattern-signal previously chosen from the given observation dataset. Pattern-signal $\boldsymbol{s}_p$ is defined as the signal with higher correlation score stored in an averaged correlation vector $\boldsymbol{\rho}$, which is calculated from the upper triangular correlation matrix $\boldsymbol{R}$ given by:

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & 0 \\ \vdots & \vdots & & \\ \vdots & \vdots & \dots & 0 \\ r_{n1} & 0 & 0 & 0 \end{bmatrix} = [\boldsymbol{r}_1 | \cdots | \boldsymbol{r}_n] \tag{1}$$

where $r_{ij} = \text{corr}(\boldsymbol{s}_i, \boldsymbol{s}_j)$, $\boldsymbol{s}_i$ represents the signal associated with i-th subject, and $\text{corr}(\cdot,\cdot)$ is a correlation operator standard. The averaged correlation vector $\boldsymbol{\rho} \in \mathbb{R}^n$ can be defined as: $\boldsymbol{\rho} = [\mu(\boldsymbol{r}_1), \dots, \mu(\boldsymbol{r}_n)]$, where $\boldsymbol{r}_k$ is the k-th column vector. Then, $\boldsymbol{s}_p$ is chosen as the signal that corresponds to $\max\{\boldsymbol{\rho}\}$.

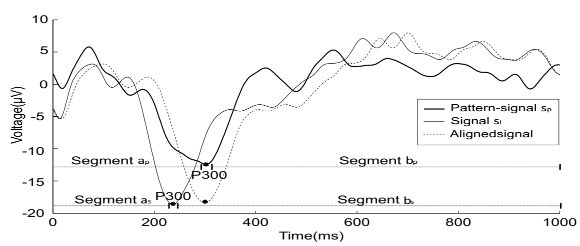

Fig. 1. Alignment of ERP signal according to pattern-signal

After determining pattern signal, P300 component is located on such waveform through locating algorithm described in section II-B (henceforth termed *P300-pattern*), and then, remaining signals are aligned doing to coincide their P300 components with P300-pattern. For this end, all signals are divided into two segments: segments $\boldsymbol{a}_s$ and $\boldsymbol{b}_s$ as shown in Figure 1. These two segments are re-sampled to equal the length of their corresponding segments from $\boldsymbol{s}_p$, i.e., $\boldsymbol{a}_p$ and $\boldsymbol{b}_p$ at sampling frequencies f_a and f_b, defined as: $f_a = \ell(\boldsymbol{a}_p)/\ell(\boldsymbol{a}_s)$; $f_b = \ell(\boldsymbol{b}_p)/\ell(\boldsymbol{b}_s)$, where $\ell(\cdot)$ denotes the number of samples of its argument.

D. Unsupervised Grouping

To assess the compactness and separability of data set $\boldsymbol{X}$ divided into two homogeneous clusters (each associated to one class), unsupervised techniques for grouping are employed. Since the main interest of this work is to show the discriminant capability of feature set in terms of separability and compactness between formed groups, a basic grouping technique is employed. For this purpose, *k-means* algorithm is implemented as described in [11]. Particularly, from the centroids obtained with *k-means* algorithm, a distance matrix $\boldsymbol{D} = [d_{ij}] \in \mathbb{R}^{n\times k}$ is formed, where each entry ij is calculated as $d_{ij} = \mathrm{d}(\boldsymbol{x}_i, \boldsymbol{q}_j)$, $\boldsymbol{q}_j$ denotes j-th centroid, $i = 1,\ldots,n$ and $j = 1,\ldots k$. In this case $k = 2$.

An accumulated distance matrix $\widetilde{\boldsymbol{D}} \in \mathbb{R}^{k\times k}$ is obtained from matrix $\boldsymbol{D}$, whose main diagonal is constituted by the sum of distances between centroids $\boldsymbol{C}_k$ and data points of its respective cluster k, and off-diagonal elements are the sum of distances between centroid of cluster k and the data points belonging to remaining clusters, thus:

$$\widetilde{\boldsymbol{D}} = \begin{bmatrix} \sum\limits_{i\in\boldsymbol{C}_1} \mathrm{d}(\boldsymbol{x}_i,\boldsymbol{q}_1) & \sum\limits_{i\in\boldsymbol{C}_1} \mathrm{d}(\boldsymbol{x}_i,\boldsymbol{q}_2) \\ \sum\limits_{i\in\boldsymbol{C}_2} \mathrm{d}(\boldsymbol{x}_i,\boldsymbol{q}_1) & \sum\limits_{i\in\boldsymbol{C}_2} \mathrm{d}(\boldsymbol{x}_i,\boldsymbol{q}_2) \end{bmatrix} = \begin{bmatrix} \tilde{d}_{11} & \tilde{d}_{12} \\ \tilde{d}_{21} & \tilde{d}_{22} \end{bmatrix} \quad (2)$$

where $\boldsymbol{C}_k \in \mathbb{R}^{n_k\times k}$ is the k-th cluster and n_k is its corresponding number of data points. Then, the relative value vector associated with $\widetilde{\boldsymbol{D}}$ is calculated as follows:

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} |\tilde{d}_{11} - \tilde{d}_{12}|/\tilde{d}_{11} \\ |\tilde{d}_{21} - \tilde{d}_{22}|/\tilde{d}_{22} \end{bmatrix} \quad (3)$$

As can can be seen vector $\boldsymbol{v}$ is an indicator of grouping quality, since it takes into account the difference of intra–and between–classes distances. In addition, to avoid sensibility to the magnitude of values, difference between elements $\tilde{d}_{k1}$ and $\tilde{d}_{k2}$ is normalized with respect to element $\tilde{d}_{kk}$. Fisher's ratio is a typical measure commonly used for measuring the classification performance. By employing values from matrix $\widetilde{\mathbf{D}}$ (See Eq. (2)), Fisher's ratio can be estimated as:

$$J = \frac{\tilde{d}_{12} + \tilde{d}_{21}}{\tilde{d}_{11} + \tilde{d}_{22}} \quad (4)$$

III. RESULTS AND DISCUSSION

In order to test the influence of latency in separability of clusters, four different experiments were carried out:

1) Firstly, clusters are formed using only latency as feature vector.
2) For the second case, 16 characteristics are used, including latency.
3) In the third experiment, the whole morphological feature set is used excepting latency.
4) Lastly, the second case is carried out but using a feature matrix obtained from aligned ERP signals through technique describe in section II-C.

Table I shows relative values of distances calculated in each performed experiment. It can be observed that the greatest values are obtained for the fourth case, when clusters are formed using the data matrix $\boldsymbol{X}$ estimated from aligned ERP signals. The fact of aligning P300 wave (on the same time point for all ERP signals belonging to same class) implies that temporal location of such component will be irrelevant to characterize with other morphological feature. Then, it is possible to say that latency is not a determinant feature in formation of well–separated and compact clusters, for considered dataset.

Relative value	Test 1	Test 2	Test 3	Test 4
v_1	0,0431	0,3731	0,3883	0,5124
v_2	0,0525	0,3711	0,6171	1,3546
J	0.9519	0.9828	0.9830	1,0390

TABLE I

ESTIMATED NON-SUPERVISED MEASURES OF CLUSTERING

Fisher's ratio, J, indicates a good clustering if its value is maximum. In these experiments the relative values have shown to be more sensitive to the conformation of the groups. This can be attributed to the Fisher's ratio is obtained from a ratio while the relative values are obtained from a sum, which in this case showed more sensitivity to changes in the grouping.

The above statement can also be evidenced in Figures 2 and 3, in which bi-dimensional scatter plot of data is shown. In order to observe the effect of latency regarding separability and compactness of formed groups, Fig. 2 depicts the interaction between latency and morphological features related to area of ERP signals, with which latency showed least overlapped and more compact groups. On the whole, it can be seen the overlapping of clusters is relatively large; beside, circumference size that contain the clusters indicates that variance of data points belonging each group is much greater than those shown in Figures 3 and 4. For all figures Cluster 1 ($\circ$) denotes normal subjects and Cluster 2 ($+$) refers to AHD ones.

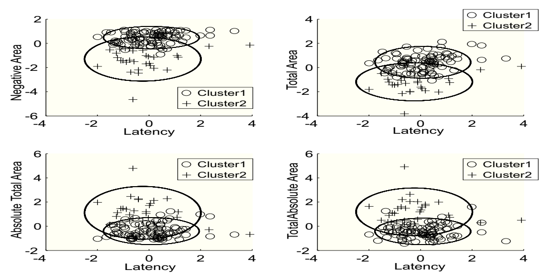

Fig. 2. Scatter plot of latency extracted of original signals

For sake of good visualization of clustering method applied on data points, pairs of features whose scatter plot shows well- defined clusters have been selected. These groups were defined under a criterion of maximum distance between the two centroids from each group. Along possible combinations of pair-features, scatter plots displayed in figure 3 present more separated and compact clusters than the ones shown in Fig. 2. However, it is evident the occurrence of some overlapping, further data points are more scattered comparing with clusters of Fig. 4. Overlapping could be avoided or reduced by applying another clustering algorithms that enhance the decision boundary, for instance soft methods or those based on densities. But, in this study, the most distance-based traditional technique was employed because the study is concerned more in characterization than clustering results.

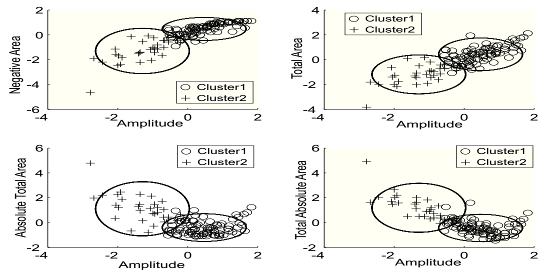

Fig. 3. Scatter plot of features extracted of original signals

Scatter plots of figure 4 were obtained from a data matrix calculated with aligned ERP recordings. In this figure, couple of features that shows the best defined clusters in experiment N° 4 are displayed. In accordance with fifth column of table I that contains the greatest relative values, in figure 4 it can be seen clusters with a greater separability (without any overlapping) and groups are well-defined and compact.

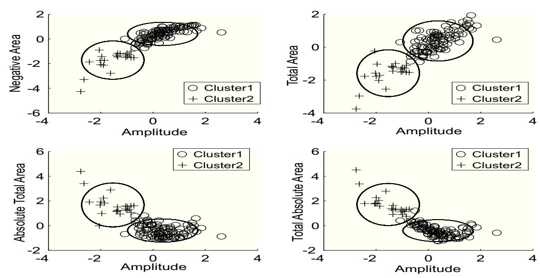

Fig. 4. Scatter plot of features extracted of aligned signals

Moreover, it can be seen that figure 2 to 4 correspond to relative values of table I, thus showing a greater separability of clusters when their relative values are also greater. Literature refers to increase and decrease in latency of ADHD children with regard to control children [5][6][7], showing the latency as a feature discriminant between studied classes. However, in this paper it has been demonstrated that latency is not a determinant parameter for a clustering process oriented to group control and pathological classes.

Nonetheless, it is important to quote that the latency was not a discriminative feature for this particular case, in which, oddball paradigm is used and signals were previously averaged and filtered by the acquisition system. Then, one can declare that latency is not a general feature to classify ERPs into normal and AHD. There exist particular cases - e.g. this approach - in which aligned signals shows greater or equal performance than non-aligned, discarding the latency as a relevant feature. But, this fact could be attributed to the conditions under which experiments were carried out such as: feature set chosen, paradigm employed, nature of initial signals, aligning procedure, among others.

IV. Conclusions

In terms of pattern recognition, automatic diagnosis is related to grouping of homogeneous patterns in such way classes of interest can be identified. Therefore, features to be analyzed must represent properly considered signals as well as generate a good separability. Diagnostic features, such as the latency and other latency-based features, are often taken into account in the design of automatic systems for pathology detection. In particular, for ERP analysis oriented to ADHD diagnosis latency can intuitively be an important feature for manual inspection because it may change according to such pathology. Nonetheless, there is still no a standard to determine and analyze latency changes. Besides in this work, it is showed that for designing a computer-aided system, latency in comparison with other morphological features, could not be a relevant feature to achieve a high quality clustering measured via classes separability and compactness. Obviously, regarding the conditions for experiments latency can be useful, then each case must be separately analyzed with both aligned and non-aligned signals.

Acknowledgement

This work is supported by COLCIENCIAS project *Identificación Automática del Trastorno por Déficit de Atención y /o Hiperactividad sobre registros de Potenciales Evocados Cognitivos*.

References

[1] M. A. Idiazábal, A. Palencia-Taboada, J. Sangorrín, and J. Espadaler-Gamissans, "Potenciales evocados cognitivos en el trastorno por déficit de atención con hiperactividad," *Rev Neurol*, vol. 34, pp. 301–305, 2002.

[2] R. A. Barkley, *Attention deficit hyperactivity disorder: a handbook for diagnosis and treatment*, 3rd ed., Guilford Press, New York, 2005.

[3] V. Bostanov, "Data sets Ib and IIb: Feature extraction from eventrelated brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, p. 1057–1061, 2004.

[4] S. H. Patel and P. N. Azzam, "Characterization of n200 and p300: Selected studies of the event-related potentials," *International Journal of Medical Sciences*, vol. 2, no. 4, pp. 147–154, 2005.

[5] P. Robaey, F. Breton, M. Dugas, and B. Renault, "An event-related potential study of controlled and automatic processes in 6-8 year old boys with attention deficit hyperactivity disorder," *Electroencephalography and Clinical Neurophysiology*, vol. 82, pp. 330–340, 1992.

[6] J. Satterfield, A. Schell, and T. Nicholas, "Preferential neural processing of attended stimuli in attention deficit hyperctivity disorder and normal boys," *Psychophysiology*, vol. 31, pp. 1–10, 1994.

[7] R. J. Strandburg, J. T. Marsh, W. S. Brown, R. F. Asarnow, J. Higa, R. Harper, and D. Guthrie, "Continuous-processing-related event-related potentials in children with attention deficit hyperactivity disorder," *Biol Psychiatry*, vol. 40, pp. 964–980, 1996.

[8] I. Kalatzis, N. Piliouras, E. Ventouras, C. Papageorgiou, A. Rabavilas, and 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," *Computer Methods and Programs in Biomedicine*, vol. 75, pp. 11–22, 2004.

[9] V. Abootalebi, M. H. Moradi, and M. A. Khalilzadeh, "A new approach for EEG feature extraction in P300-based lie detection," *Computer methods and programs in biomedicine*, vol. 94, p. 48–57, 2009.

[10] C. C. Duncan, R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen, J. Polich, I. Reinvang, and C. V. Petten, "Event-related potentials in clinical research: Guidelines for eliciting, recording,and quantifying mismatch negativity, P300, and N400," *Clinical Neurophysiology*, vol. 120, p. 1883–1908, 2009.

[11] P. Hansen, N. Mladenovic, and E. Commerciales, "J-means: A new local search heuristic for minimum sum-of-squares clustering," *Pattern Recognition*, vol. 34, no. 2, pp. 405–413, 2001.