Single trial P300 detection in children using expert knowledge and SOM

C. Morales Student Member IEEE, C.M. Held, Senior Member IEEE , P. A. Estevez, Senior Member IEEE, C. A. Perez, Senior Member IEEE, S. Reyes, P. Peirano, C. Algarin.*

***Abstract*— Preliminary results of an automatic system for single trial P300 visual evoked potential events detection are presented. For each single trial P300, several candidate events were generated, and then filtered, using 3 wave features. The surviving candidate events were fed into a SOM-based classifier. A context filter was applied before the final output. No stationary condition of the P300 is involved in the algorithms. Recordings of 27 assessment sessions, each with 120 trials, were visually inspected by experts to identify and mark the P300 events, which was accomplished in about one third of the trials. The dataset was divided in training (18) and testing (9) subsets. The system identifies the initial and end times of the P300; it obtained a sensitivity of 53.9%, a specificity of 64.0% and an accuracy of 61.2% in the testing dataset.**

I. Introduction

The event-related potential (ERP) is a bio-electric wave occurring as a response to an auditory, visual or cognitive stimulus, that can be identified in a EEG recording [1]. It is a voltage fluctuation in the electrical activity of the brain, temporally and spatially associated with a stimulus. The P300 is usually assumed as a slow wave component of the ERP with a positive peak around 300 [ms] after stimulus [1], [2]. Characteristic measures of P300 are amplitude, latency and area under the curve, estimated by measuring its initial, peak and end times.

EEG recordings of ERPs have very low signal to noise ratio (SNR) [2]. There is evidence that P300 can be identified in just 40% of the trials in adults [3]. A well-known technique to increase SNR is by averaging several trials. However, this assumes stationarity, which is somewhat controversial due to e.g. evidence of inter-trial latency and P300 amplitude variability [1], [4], or evidence of evoked alpha band activity only in some trials after stimulus [5]. Hence the interest of single trial ERP research.

* This work was supported by the Department of Electrical Engineering and INTA, Universidad de Chile, and by the Chilean Science and Technology Funding Agency, Chile, under Grants FONDECYT No. 1120319 and No. 1110513. Asterisk indicates corresponding author

C. M. Held, P. A. Estevez, C.A.Perez are with the Department of Electrical Engineering, Universidad de Chile, Santiago, Chile C. M. Held is also with Apacomint Ltda, Santiago, Chile (e-mail: cheld@apacom.cl; pestevez@cec.uchile.cl; clperez@cec.uchile.cl).

C. Morales * is with the Department of Electrical Engineering, Universidad de Chile, Santiago, Chile, (e-mail: crimoral@ug.uchile.cl).

S. Reyes, P. Peirano, and C. Algarinare with the Sleep Laboratory, Instituto de Nutrición y Tecnología de los Alimentos, Universidad de Chile, Santiago, Chile (e-mail: sreyes@inta.cl; ppeirano@inta.uchile.cl; calgarin@inta.uchile.cl).

Several studies involve automatic detection of P300 in ERPs. In the Go/Nogo paradigm, commonly used in psychophysiology, the P300 occurrence is always expected, but its amplitude, latency and topography [6] vary among different trial and subject conditions. In another application, the oddball stimulation paradigm used in Brain Computer Interface (BCI) systems, the P300 is expected to be found in association to target trials [7], [8]. In such cases, feature extraction and automatic classifiers are applied to search the P300 and identify the target trials. The accuracy of this detection in single trial is almost 40%, and it increases to almost 80% when multiple trials are averaged [7], [8].

The goal of this research work was to develop an automatic system, in which each input is the raw data of a single-channel EEG of a single trial. The output is a vector containing the initial and end times of the P300 component event if detected, otherwise, a null value vector is delivered. We used Self Organizing Maps (SOMs) [9] as part of our detection scheme because of its potential contribution as an unsupervised tool to extract hidden data structures from high dimensional feature sets.

II. Methods

A. Recordings and data set

Visual evoked potentials assessments were performed in 27 healthy 10 year old subjects, applying the visual Go/NoGo paradigm. Subjects were instructed to press a button as soon as possible with their dominant hand as response to the Go stimulus. The experiment considered 120 consecutive stimuli, and the recording included EEG, physical response and Go/NoGo trial condition. EEG activity was recorded according to the ten-twenty system [10] at a sample frequency of 200 [Hz]. A 1700 [ms] EEG window was recorded for each trial, of which 100 [ms] were pre-stimulus activity [11]. Hence our dataset consisted of a total of 3240 trials. The analysis was performed at the Pz derivation because it has been described that it shows the P300 event with larger amplitude [11],[12]. The EEG recording was re-referenced to mastoids average and low-pass filtered at a 50[Hz] cut frequency.

The training dataset consisted of 18 experiment records (2160 trials), which was used to set up and tune the detection system; the remaining 9 experiments (1080 trials) were reserved to test its performance. The P300 component was characterized as a positive voltage peak between two local minima. Experts were asked to visually identify them,

marking the initial and end time of the event (Fig. 1) which established the ground truth data. The P300 component was identified in about 33% of the trials.

B. Single trial P300 event detection

The automatic detection system was designed in 5 main stages: a) Basic P300 candidates generation, detecting all possible P300 candidates in a single trial. b) Features extraction from those candidates. c) Advanced P300 candidates selection, using the previous features. d) SOM classification into classes *P3* and *noP3*, applied upon the advanced P300 candidates. e) Expert-knowledge based context filter, processing the output of the SOM, and delivering the final output. Each step is described below.

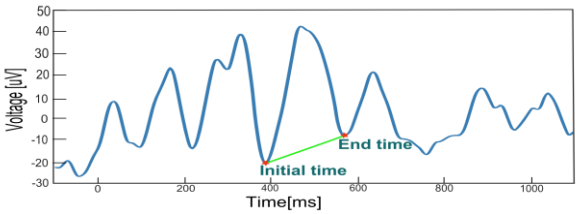


Figure 1 : Example of a one channel single trial ERP signal, with a P300 event identified by an expert .Expert markings in the training dataset were used to train the automatic system, and to test its performance in the testing dataset.

a) ***Basic P300 candidates generation:*** Single P300s present themselves in many different shapes, hence the initial identification of candidates cannot rely on form features. Initially, the system considered any pair of local minima in the time series (single trial recording) located within the [100, 900][ms] time window (w_i). If there were n local minima in a sample, the total number of pairs would be M= $\binom{n}{2} = \frac{n(n-1)}{2}$. Then, to be accepted as a basic P300 candidate, the considered pair had to have no "lower" local minimum. To reject unfeasible w_i, a straight line (r) was constructed between the initial and end times of each w_i (this baseline r was later used to calculate additional features explained below). If there existed at least one local minimum within w_i whose voltage value was equal or lower than the r voltage at the same time, w_i was rejected (Fig.2 a,d). Non rejected w_i (Fig.2 b,c) were labeled as basic P300 candidates w_c, each defined by an initial and an end time. A majority of all created w_i were rejected (70,4±4,9% rejected vs 29,6±4,9%accepted in the training dataset).

b) ***Features Extraction:*** 6 wave features were observed and measured for each wave event (w_c):

1) Initial time, corresponding to the initial local minimum of pair of points defining w_c.

2) End time, corresponding to the end local minimum of the pair of points defining w_c.

3) Frequency is the inverse of the time lapse between the initial and end times.

4) Peak to peak (*p2p*) amplitude (V_{p2p}) is the largest voltage difference within w_c; it is obtained by identifying the maximum voltage V_M within w_c, and measuring the difference between V_M and the lowest between the initial and end values of w_c .

5) Wave incision is the ratio V_{LM} / V_{p2p} : V_{p2p} is the peak to peak amplitude just explained. V_{LM} was calculated as follows: all local minima within w_c were identified (excluding the initial and end points), constituting a set $S(V_m)$. The *p2p* amplitude between each of those local minima and V_M was calculated, V_{LM} being the largest one of these differences.

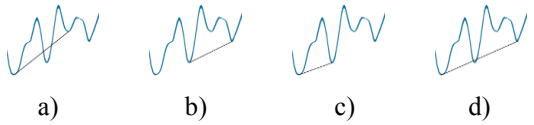


Figure 2 : Automatic basic P300 candidates generation: Examples of wave events identified as local minima pairs within the [100,900][ms] range: w. Candidates a) and d) are rejected because at least one point in the wave reaches or crosses the straight line r between the initial and end; b) and c) are feasible candidates w_c.

For the last feature, an inner interpolated wave s is built, its initial and end points coinciding with the ones of w_c, and using the cubic spline interpolation including also all $S(V_m)$ points in between.

6)Area ratio is the ratio of the area formed between s and r and the area described by w_c and r. An example is shown in Fig. 3.

Figure 3 Example of feature 6: In a P300 candidate, the area ratio feature was calculated as the ratio between two defined areas. a) Area between baseline r and an interpolated curve including all local minima of w_c. b) Area between baseline r and w_c.

c) ***Advanced P300 candidates selection:*** The features defined above are used to further discard P300 candidates. Expert marked P300 of the training set were used to establish compatible features ranges. Histogram distribution statistics of marked P300 point to 3 individual features which could separate a large number of P300 candidates: Initial time within [190,550][ms]; end time within [340, 750][ms]; and frequency within [3,11][Hz]. Output of this stage for each trial can be zero, one or more than one P300 candidates. If no P300 candidate survived at this stage, its output would be a null vector, concluding that the single trial has no identifiable P300.

d) ***SOM-based classifier:*** A SOM Neural Network (NN) was trained and used to classify each P300 candidate into one of two possible classes, *P3* or *noP3*. All P300 markings and P300 candidates of the training set that passed the previous stages were used to train the SOM: P300 markings and corresponding P300 candidates (those with at least 90% of time coincidence with expert markings) were labeled as *P3*, all others as *noP3*. For a single trial with no marked P300, all P300 candidates were labeled as *noP3* members.

The whole training set consisted in 2160 EEG single trials. A total of 14652 candidates passed the previous steps, 1258

P3 class members, corresponding to 725 marked P300 (in some cases more than one candidate associated to the same marked P300 survived so far) and 13394 *noP3* class members. All features used in the feature extraction stage, except End time, were the applied as inputs for the SOM NN, after a standardization process. The training dataset was used to fix the parameters, and the acquired parameters were stored in order to apply the same standardization to the testing dataset.

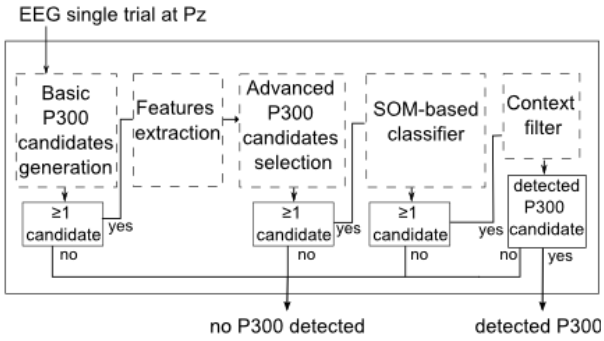


Figure 4 Block diagram of the automatic P300 detection system, showing its 5 stages: Basic P300 candidates generation, Features extraction, Advanced P300 candidates selection, SOM-based classifier and Context filter.

Unsupervised SOM training was performed using 5 inputs and 400 neurons as outputs. After training, the output neurons were associated to one of the classes as follows: Each output neuron (n_i) had an associated proportion of *P3* examples ($P300_i$) and *noP3* examples ($nP300_i$), calculated as: $P300_i = \frac{a_i}{P300}$ is the ratio between the number of *P3* examples (a_i) associated to neuron i and the total number of P300 in the training dataset (1258), and $nP300_i = \frac{b_i}{NP300}$, where b_i is the total number of *noP3* examples associated to neuron i and NP300 is the total number of class members in the training dataset (13394). These values were calculated for all output neurons. The proportions $P300_i$ and $nP300_i$ were ranked independently. Then, for each particular neuron, all other neuron proportions valued same or less than the neuron own value were added. The larger sum total defined the class membership of each neuron. Neuron ranking information is stored and used in the next step.

Output neurons distribution in the SOM obtained with the training dataset is shown in Fig.5. Dark areas in Fig 5.a) represent neurons associated to *P3* class, and the dark areas in Fig 5.b) represent neurons associated with *noP3* class. A neuron with a larger dark area means that more individuals are represented by it. After trainng, a total of 44 neurons were associated to class *P3* and 356 to class *noP3*.

Once the SOM was trained, it was applied as a classifier to the training dataset. In the testing phase the SOM is only applied as classifier, using the parameters established with the training set. The 5 features were fed to the SOM, and the candidate was classified according to the winner-output neuron (best matching unit). If it activated a *P3* neuron, the candidate was still considered a P300 candidate, otherwise it was discarded. Output of this stage can deliver zero, one or more than one P300 candidates.

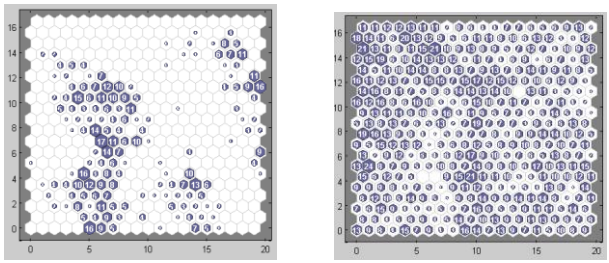
e) ***Context filter:*** If no P300 candidate survived the previous stage, the classification output is a null vector, and it concluded that the single trial had no identifiable P300. If one candidate was classified in the *P3* class, it was established as the P300 of its single trial. If more than one candidate were classified into *P3* class, context criteria were applied. The P300 candidate that classified to the highest ranked neuron (max $P300_i$) among candidates in the SOM was selected as output**.** If more than one candidate shared the same highest rank, and their time coincidence was less than 10%, then all candidates were rejected and the single trial was labeled as without P300. If their time coincidence was equal or above 10%, the system chose the P300 candidate that started first; if more than one candidate shared that starting point, the one that ended first was chosen.

The results were obtained with the testing dataset. The system confidence was analyzed with a confusion matrix. Correct performance for each single trial was established as when:

a) There was no P300 marked by the expert, and non was detected by the system (true negative: TN)

b) The P300 was marked by the expert, and was also detected by the system, with a relevant coincidence (90%) in the time span for both (true positive: TP).

A false negative (FN) was a trial where the expert marked a P300 event and system did not detect any. False positive (FP) meant either that the expert did not find a P300 but the system found one (FP1), or that both the expert and the system found a P300, but the time coincidence did not meet the minimum threshold criteria (FP2).



a) b)

Figure 5 SOM obtained with the training dataset. For visualization, the neurons outputs distribution is shown in separate grids. Dark areas correspond to positive associations to each class. a) neurons associated to expert marked P300 events; b) neurons associated to non-marked P300 candidates. Neurons are then associated to *P3* or *noP3* classes according to an accumulated proportion criterion.

III. RESULTS

The classification results for both the training and testing data set are shown in the confusion matrices in table I.

The sensitivity of the results was low, reaching 62.4% on the training and 53.9% on the testing dataset. To gain insight about this performance, we evaluated intermediate FN outputs at two stages using the training dataset. 240 marked P300 (33.1% of all marked P300) were not identified by the system and labeled as FN; 47 of the marked P300 (19.6% of

the FN) did not meet the features conditions at the advanced P300 candidate selection, the other 193 cases (80.4% of the FN) did not classify into *P3* class neurons in the SOM.

Several single trials without expert markings got assigned as P300 (FP) by the system. We analyzed the 526 cases showing up with the training set data, and described two kinds of FP (FP1 and FP2). It turned out that most cases (439) corresponded to detections in events without P300 markings (FP1 make to 83.5% of all FP). The other 87 cases (16.5% of all FP) corresponded to FP2, where 17.2% (15 of 87) had no temporal intersection with P300 markings, 11.5% (10 of 87) had less than 50% time coincidence, and 71.3% (62 of 87) had 50% or more, but less than 90%, time coincidence with P300 markings.

TABLE I. Confusion Matrices obtained for the training and testing datasets

	18 subject Training set		9 subject Test set	
Proposed System Classification	Ground Truth		Ground Truth	
	P300	noP300	P300	noP300
P300	398	526	161	281
noP300	240	996	138	500
	Sensitivity: 62.4% Specificity: 65.4% Accuracy: 64.5%		Sensitivity: 53.9% Specificity: 64.0% Accuracy: 61.2%	

IV. DISCUSSION

The output results at the different stages of the system for the training dataset hinted at the difficulty of this classification task. The output of the basic P300 candidates generation (first stage) still included all the marked P300. Only 91,3% (662 of 725) of them fulfilled the features within ranges and therefore continued as candidates after the advanced P300 candidates selection (third stage). At the output of the SOM-based classifier, *P3* class neurons included a mere 64.2% of the total P300 examples of the training dataset. The negative classification did better: SOM neurons corresponding to the *noP3* class covered 93.3% of P300 candidates without corresponding markings.

In the overall picture, the high FP rate seem to be caused by misclassification at the SOM output, not because of the high threshold time coincidence imposed on the detected P300 with markings.

V. CONCLUSION

A single trial P300 event automatic detection system was designed, and preliminary results are presented. We made no assumption about the stationarity of the single trial events. A labeled database was generated applying expert visual detection of P300 events, marking the initial and end times. For each single trial, the system generated basic P300 candidates, and then evaluated certain features of them. The system showed a high trial rejection rate, just as the expert, who marked P300 in about a third of all single trials. A sensitivity of 53.9%, specificity of 64.0% and accuracy of 61.2% were reached in testing dataset.

In future work, the low sensitivity of the system must be addressed, e.g. find out if there was an under-representation of examples in the evaluated feature space, or if some P300 had a particularly low SNR, or other reasons. An enhanced context filter (fifth stage), and including topographic context variables, are other options to explore. Also, the low sensitivity of the SOM points to trying a different classifier in order to enhance P300 identification.

ACKNOWLEDGMENT

This work is supported by the Department of Electrical Engineering and INTA, Universidad de Chile, and Conicyt Fondecyt grants Nº 1120319 and Nº1110513.

REFERENCES

[1] *S.J. Luck, An Introduction to Event Related Potential Technique. Cambridge, Massachusetts: MIT Press, 2005.*

[2] *J.T. Cacioppo, L.G. Tassinary and G.G. Berntson, Handbook of psychophysiology. New York: Cambridge Univ Press, 2007.*

[3] *A R. Haig, E. Gordon, G. Rogers, and J. Anderson, "Classification of single-trial ERP sub-types: application of globally optimal vector quantization using simulated annealing.," Electroencephalogr Clin. Neurophysiol., vol. 94, no. 4, pp. 288–97, Apr. 1995.*

[4] *C. Michel, T. Koenig, D. Brandeis, L.R. Gianotti, J. Wackermann, Electrical Neuroimaging. New York: Cambridge Univ Press., 2009*

[5] *S. Hanslmayr, W. Klimesch, P. Sauseng, W. Gruber, M. Doppelmayr, R. Freunberger, T. Pecherstorfer, and N. Birbaumer, "Alpha phase reset contributes to the generation of ERPs," Cereb. Cortex, vol. 17, no. 1, pp. 1–8, Jan. 2007.*

[6] *J. Polich, "Updating P300: An integrative theory of P3a and P3b". Clinical Neurophysiology. vol. 118. pp. 2128-2148. 2007.*

[7] *Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R.. "A comparison of classification techniques for the P300 Speller". Journal of Neural Engineering, 3(4), pp. 299–305. 2006*

[8] *Manyakov, N. V., Chumerin, N., Combaz, A., & Van Hulle, M. M. (2011). Comparison of classification methods for P300 brain-computer interface on disabled subjects. Computational Intelligence and Neuroscience, 2011..*

[9] *H.A. Jasper," The ten-twenty electrode system of the International Federation". Electroencephal. and Clinical Neurophys.. vol.10, no. 2, pp. 371–375. 1958.*

[10] *T. Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, vol. 30, Springer, 2001, 3rd.*

[11] *C. Algarín, C. A Nelson, P. Peirano, A. Westerlund, S. Reyes, and B. Lozoff, "Iron-deficiency anemia in infancy and poorer cognitive inhibitory control at age 10 years.," Dev. Med. Child Neurol., vol. 55, no. 5, pp. 453–8, May 2013.*

[12] *T.W. Picton, S. Bentin, P. Berg, E. Donchin, S.A. Hillyard, Jr. R. Johnson, G.A. Miller, W. Ritter, D.S. Ruchkin, M.D. Rugg, and others, "Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria". Wiley Online Library. Psychophysiology. vol.37, no.2, pp. 127-152. 2000.*