An Open-Access P300 Speller Database

Claudia Ledesma-Ramirez, Erik Bojorges-Valdez, Oscar Yáñez-Suarez, Carolina Saavedra, Laurent Bougrain, Gerardo Gabriel Gentiletti

4th International BCI Meeting

http://akimpech.izt.uam.mx/p300db

An Open-Access P300 Speller Database

Claudia Ledesma-Ramirez[1], Erik Bojorges-Valdez[1], Oscar Yáñez-Suarez[1], Carolina Saavedra[2], Laurent Bougrain[2], Gerardo Gabriel Gentiletti[1,3]

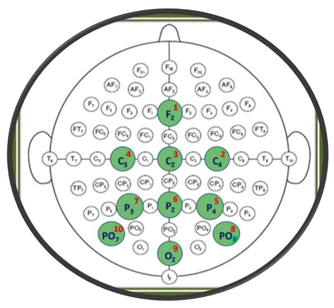
[1] Laboratorio de Neuroimagenología, Universidad Autónoma Metropolitana (UAM), Mexico
[2] Cortex team-project, Nancy University/INRIA Nancy Grand Est, France
[3] Laboratorio de Ingeniería en Rehabilitación e Investigaciones Neuromusculares y Sensoriales, Universidad Nacional de Entre Ríos (UNER), Argentina

GOAL:

The P300 Speller is probably the best known application in BCI [1]. Over the years, many improvements over the pioneering systems have been made and some performance comparisons exist [2]. To contribute to the improvement process, we propose an open access to a large database obtained from first-time users of the P300 speller application implemented within the BCI2000 platform [3] (Figure 1). The database is documented with associated classifier designs and objective performance measures, readily available for comparison and reference. We also propose a set of Matlab functions that help in the preparation of data for alternative classifier design and testing.

Figure 1. The database website.

DATABASE:



Notch	Band-Pass
Chebyshev	Chebyshev
4th order	8th order
58 - 62 Hz	0.1- 60 Hz

Figure 2. Recorded EEG channels and filter parameters

The database includes recordings from 30 healthy subjects (18 Males/ 12 Females, age 21-25) controlling various conditions (sleep duration, drugs, etc).

Each subject participated to 4 sessions with 15 sequences:

1) Three copy-spelling runs.
2) One copy-spelling run with feedback using a classifier trained on data from session one.
3) Three free-spelling runs (user-selected words, around 15 characters per subject).
4) Variable free-spelling runs with reduced number of sequences as indicated by bit-rate analysis.

10 channels (Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz) have been recorded at 256 sps using the g.tec gUSBamp with acquisition characteristics shown in Figure 2. The stimulus is highlighted for 62.5 ms with an inter-stimuli interval of 125 ms.

We also propose a set of Matlab functions to extract and average target and non-target responses specifying for example the number of sequences to average and the duration of the response and to save it in Matlab or ASCII format.

The database, a complete description of the parameters used for the speller and the code are available at: http://akimpech.izt.uam.mx/p300db.

REFERENCE CLASSIFIERS AND PERFORMANCE

SWLDA (step-wise linear discriminant analysis) classifiers have been trained for each subject. In order to provide the users of the database with an objective, comparable measure of performance -that takes into account the choice of features and is independent of the training/testing set- the relative (receiver) operating characteristic or ROC curve has been selected. Summarized by the area under de curve (Az), the ROC reflects intrinsic class separability: higher values of Az correspond to better classifier designs.

As a reference, results for each subject are available on the web site. Accuracy using SWLDA with 15 training sequences can be established in terms of an 86.7% of the participants having 100% correct spelling, while the lowest percentage of correctly detected characters reached by the rest of the database population was 85%. ROC areas above 0.95 were reached by 76.7% of the population in about 10 sequences. Thus, for 15 sequences the general performance is very good. Classifier features were selected mainly from P08, Oz, PO7 and Pz electrodes and within the 100-290 ms window. This shows that EP related to visual stimulation and its recognition play an important role in the high accuracy of the classifier (See Table 1 and Figure 3).

ne	accuracy 100%	95% 100%	90% 95%	85% 90%	≤ 85%
15	26	1	2	0	1
14	22	4	3	1	0
13	25	2	2	1	0
11	19	3	5	0	3
10	20	3	3	4	0
8	18	3	2	3	4
5	8	2	7	6	7
3	3	0	3	6	18

Table 1. Distribution of database cases as a function of classifier accuracy and number of averaged epochs (ne).

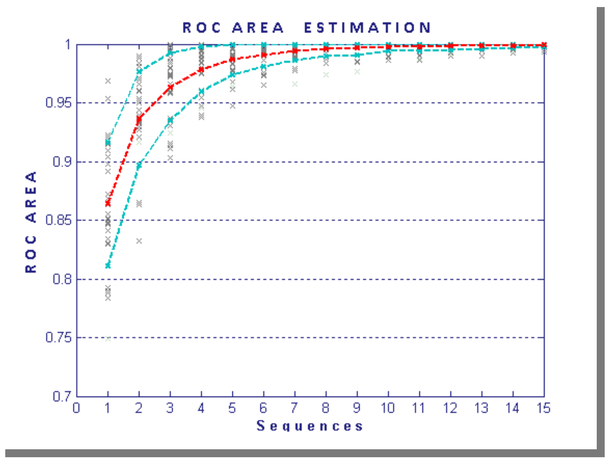


Figure 3. Mean ROC area for all cases using SWLDA analysis. Each x represents an individual case, blue lines are standard deviation.

DATABASE APPLICATION:

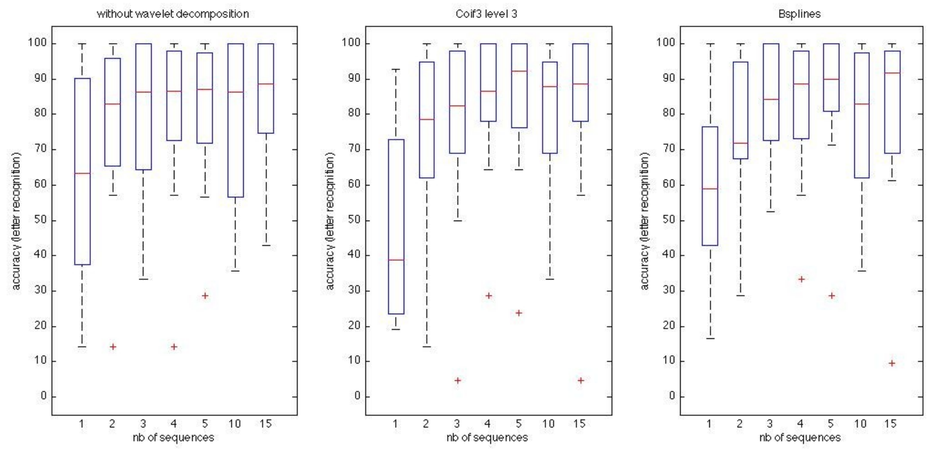


Figure 4. Impact of different preprocessing schemes (none,coiflet decomposition, b-splines decomposition) on SVM classifier accuracies.

DISCUSSION:

This open-access P300 database includes recordings from 30 healthy subjects. Data is available in BCI2000 and Matlab formats. A set of Matlab functions for the extraction of the information that might be needed for a given application is also included.

The database website provides, together with the data, a description about conditions of each subject that has been recorded. Individual results, accuracy, ROC area, and performace for every sequence count are also reported. Given the individual accuracies and ROC areas for the reference SWLDA classifier, it could be argued that overall data quality is high.

We hope the work will contribute to better compare classifier techniques as related to the P300 detection problem and applications, by providing fair comparison grounds and reference data.

REFERENCES:

[1] Farwell L. A. and Donchin E. “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials.” Electroenceph. Clin. Neurophysiol. Vol. 70, pp.510-23 (1988).

[2] Krusienski, D. J., Sellers E. W., Cabestaing F. “A comparison of classification techniques for the P300 Speller.” Journal of Neural Engineering. Vol. 3, pp. 299-305 (2006).

[3] Schalk G., Mc Farland D., Hinterberger T., Birbaumer N., Wolpaw J. “BCI2000: A General-Porpose Brain-computer Interface (BCI) System.” IEEE Trans. Biomed. Eng. Vol. 51, pp. 1034-1043 (2004).