Size enhancement coupled with intensification of symbols improves P300 Speller accuracy

G. Gibert^{1,2,3}, V. Attina^{1,2,3}, J. Mattout^{1,2,3}, E. Maby^{1,2,3}, O. Bertrand^{1,2,3}

¹INSERM, U821, Lyon, F-69500, France
²Institut Fédératif des Neurosciences, Lyon, F-69000, France
³Université Lyon 1, Lyon, F-69000, France

Guillaume.Gibert@inserm.fr

Abstract

The P300 Speller has been proposed in 1988 by Farwell and Donchin [1]. In this Brain Computer Interface (BCI), a matrix of symbols is presented whose rows and columns are sequentially intensified. In this study, we investigated the influence of three stimulation parameters: the enhancement of the symbols while intensified, the inter-stimulus interval (ISI) and the reduction of flash duration. Results indicate that symbol enhancement increase P300 amplitude and the ensuing classification accuracy by a Fisher LDA. P300 amplitude and classification accuracy decrease with faster ISI. Finally, reduction of flash duration do not increase the P300 amplitude but yielded a better classification accuracy.

1 Introduction

Farwell and Donchin presented in 1988 [1] a BCI based on the P300 EEG evoked response. This BCI is used for communication by spelling words. The subject looks at a 6×6 matrix of grey symbols (letters, numbers...) on a black background. The subject's task is to focus attention to one of the symbols in the matrix (the target). Each row and column is flashed sequentially in a random manner and the subject has to count the number of intensifications of the same target. Each flash of the row or the column containing the target symbol produces a P300 response while non-target flashes do not. Averaging over responses to each row and each column and feeding a classifier with these average responses enables to detect the target symbol.

A large amount of work has been devoted to feature extraction [2] and classification [3] but only few studies have provided insights on the influence of stimulation parameters. Allison & Pineda [4] studied the effect of matrix size on the amplitude of the P300 response. Larger matrices evoked larger P300 amplitude but did not improve classification accuracy [5]. It was also shown that the P300 amplitude increases with slower ISI [6]. But again, it didn't improve the classification accuracy [5]. Finally, the symbol size was manipulated with no effect on classification accuracy [7].

The purpose of this study is to examine the impact of two new stimulus properties: the enhancement of the symbols coupled with intensification and the flash duration. We also manipulated the ISI. In this paper, we focus on the quantitative evaluation of experimental parameters on the ensuing features to detect the P300 response in single trial responses. Features are first compared on average. Then, an online analysis is mimicked and a linear classifier is applied to estimate the relative performance of the experimental manipulations as a function of the number of accounted observations.

2 Methods

2.1 Subjects

Four healthy volunteers (four women 21, 22, 20 and 22 years old) were paid to participate. All subjects reported normal vision. They had no previous experience with the P300 Speller paradigm,

nor with any other BCI paradigm. The protocol of this experiment was approved by the regional Ethical Committee and each subject signed an informed consent, prior to the experiment.

2.2 Stimulation device

Stimulation was handled by a C++/SDL software on a dedicated computer. While visual stimulation was sent to a CRT screen (vertical screen refresh equal to 60 Hz, resolution 1024×768), a trigger (for the flashed row or column, from 1 to 12) was sent to the EEG amplifier via parallel port. Since the trigger clock was based on the refresh screen VGA signal, the jitter between visual stimulations and triggers was less than 0.1 ms (as measured by an optoelectronic sensor).

2.3 Data acquisition and preprocessing

EEG activity was recorded continuously from 32 active electrodes (actiCap, Brain Products GmbH, Munich) at standard locations following the extended 10/10 international system, referenced to the nose and grounded to the forehead. Horizontal and vertical electro-oculograms (EOG) were recorded from the right eye. All impedances were kept below 10 k Ω throughout the experiments. EEG signals were bandpass filtered between 0.1-150 Hz (EOG signals were bandpass filtered between 0.01-150 Hz), amplified and digitized at a rate of 1 kHz using a BrainAmp amplifier (Brain Products GmbH, Munich). The EEG was collected and stored using BrainVision Recorder software from Brain Products.

2.4 Experimental setup and design

The subjects were seated in a comfortable chair at 1.2 m from the CRT screen. They were watching a 6×6 matrix of letters (a-z), numbers (1-9) and symbol (_) (see Figure 1 (a)). The experiment was divided into runs. Each run corresponded to one word (or non-word). Before each run, the entire word to be spelt was displayed at the top of the screen. The subject was instructed to focus his attention on the current target symbol (which was displayed in between brackets next to the target word) and to count the number of times this symbol was intensified. After each symbol, there was a 3 s delay. During this period, the subject indicated the outcome of his counting and then focused his attention on the next symbol. The subject could make a short break after each run. Before and after the experiment, a resting period of 3 minutes was recorded, during which the subject was instructed to remain still and to watch a black screen.



(a) Stimulation matrix.

(b) Enhancement of a row while intensified.

Figure 1: On the left, Stimulation matrix for the word SIX with current target S. On the right, an example of the *enhanced* condition, symbols on the second row are larger while intensified.

Five conditions were tested in blocks, randomly presented across subjects. Conditions are detailled in Table 1. In each condition, the user had to spell 5 words (or non-words) made of 2 to 4 letters (or numbers) such that each condition involved the same following 15 symbols [A E I L M O P R S T U V X 9 5] to be spelt in. These 15 selected symbols were evenly distributed all over the screen matrix. There was at least 3 symbols in each quarter of the screen. For each symbol, each row and each column was flashed 15 times in a random sequence.

Condition	Flash	ISI	Font	Accuracy	Bitrate
	duration		size	(%)	(bits/min)
enhanced slow	100 ms	500 ms	+5	91.25	2.86
enhanced medium	100 ms	$350 \mathrm{ms}$	+5	91.11	4.08
enhanced fast	100 ms	200 ms	+5	85.83	6.42
classic	100 ms	200 ms	+0	81.66	5.90
short flash	$50 \mathrm{ms}$	200 ms	+0	85.13	6.34

Table 1: Description and mean results (accuracy and bitrate for 15 repetitions using a LDA classifier and the entire ERP from 0 to 500 ms) of the different conditions of stimulation.

dimension of the matrix was $8.67^{\circ}H \times 11.33^{\circ}W$. Each symbol subtended $0.48^{\circ}H \times 0.48^{\circ}W$ of visual angle and the distance between each character was $1.24^{\circ}H \times 1.62^{\circ}W$. In the *enhanced* conditions, symbols were larger (font size +5) while intensified (see Figure 1 (b)). In that case, they subtended $0.62^{\circ}H \times 0.62^{\circ}W$ of visual angle. Three inter-stimulus intervals (ISI) and 2 flash durations were tested in different blocks (see Table 1).

2.5 Evoked Potential Analysis

Averaging was done separately for each of the 12 rows and columns in a time window ranging from 300 ms before to 1000 ms after stimulus onset. A baseline correction was applied to each event-related potential (ERP) using random epochs of 300 ms from the resting period recorded at the beginning of the experiment. Then, another average was performed across the 15 symbols in each condition, leading to one ERP by condition and subject computed over 15 symbols \times 15 flashs \times 2 (1 target row + 1 target column), i.e., 450 trials. Finally, these evoked potentials were digitally filtered with a band-pass filter (butterworth filter, 0.2-30 Hz, slope 24 dB/octave).

2.6 Fisher Linear Discriminant Analysis

Single trials were epoched between 0 and 500 ms after stimulus onset. For each condition, averaging within symbol and row/column was performed using 1 up to all 15 trials When averaging less than all 15 trials, the trials were randomly selected. Evoked potentials were digitally band-pass filtered (butterworth filter, 0.2-30 Hz, slope 24 dB/octave) and downsampled at 120 Hz. Two kinds of features were used: the vector of 60 samples (corresponding to the whole ERP) and the maximum amplitude of the ERP between 300 ms and 500 ms. This yielded 15 symbols $\times 2$ (1 target row and 1 target column) = 30 target samples and 15 symbols $\times 10 = 150$ non-target samples for each subject and condition. A leave-one-out strategy for cross-validation was adopted. This means that the data were split into a training set of size N - 1 (where N is the total number of samples) and a test set of size 1. Then the average of the squared error on the left-out pattern over the N possible ways of obtaining such a partition was used as test and the rest as training of a Fisher LDA classifier.

3 Results

3.1 Behavioral measure

The subjects were asked to report the number of flashes they counted for each target. Of course, no information about the actual number of flashes per symbol was delivered to the subject prior the experiment. Even though there were always 30 flashes per target, subjects made errors. The accuracy of this counting was used as a behavioral measure of task difficulty as well as a measure of attention. The performance rate were: enhanced slow (98.2%±0.7) > enhanced medium (98.1%±1.0) > enhanced fast (97.8%±1.2) > classic (96.9%±1.7) > short flash (96.3%±1.9). Note that, as expected, the worst result was obtained in the most difficult task: short flash (short ISI and

short flash duration) whereas the best result was obtained in the easiest condition: *enhanced slow* (longer ISI).



3.2 Evoked Potential Comparison

(c) ERP for each *enhanced* conditions.

Figure 2: Grand-average ERP (over subjects) for target stimulus recorded at Pz electrode and global topographies in the different conditions.

ERP scalp topographies at the maximum of the P300 (389 ms) were computed using spherical spline interpolation [8] for each condition (see Figure 2 (a)). The maximum of the P300 response is located on the Pz electrode for all the conditions. Therefore, the study focuses on this particular electrode.

The averaged waveforms over subjects for the *classic*, the *enhanced fast* and the *short flash* conditions are shown in Figure 2 (b). We observe that the *enhanced fast* condition elicits larger positivity than the *classic* condition. Across the 4 subjects, the mean rank of condition *enhanced fast* and *classic* are significantly different for the maximum amplitude (Kruskall-Wallis test, p=0.0264) and the mean amplitude (Kruskall-Wallis test, p=0.0264) between 300 ms and 500 ms.

The averaged waveforms over subjects are shown in Figure 2 (c) for the three *enhanced* conditions. We observe that the longer the ISI, the larger the positivity between 300 and 500 ms. The maximum amplitude of the P300 response is also a function of the ISI. However, across the 4 subjects, the difference between the three *enhanced* conditions appeared not to be significant, neither in terms of maximum amplitude (Kruskall-Wallis test, p=0.1462), nor in terms of mean amplitude between 300 and 500 ms (Kruskall-Wallis test, p=0.2106).

3.3 Classification performance



Figure 3: Mean accuracy over the four subjects of the Fisher LDA classifier for each condition and for two kinds of feature (on the left, the maximum of amplitude between 300 and 500 ms and on the right, the entire ERP from 0 to 500 ms).

The averaged accuracy over subjects of the Fisher LDA output is represented in Figure 3 for both the maximum amplitude and the entire ERP features. Consistent with [9], the accuracy grows with the number of flashing repetitions (i.e. with the quantity of information) used for averaging whatever the condition or feature type. Furthermore, whatever the type of feature and the number of accounted repetitions, the best results were obtained with the *enhanced slow* condition closely followed by the *enhanced medium* condition. Finally, the *enhanced fast* condition elicited better accuracies than the *classic* and the *short flash* conditions for most of the number of accounted repetitions. These results are consistent with our observations derived from the averaged waveforms, either based on amplitudes or areas under the curves.

4 Discussion

This study demonstrates several points. First, the size enhancement of the symbol during intensification both affects the amplitude of the P300 response and increases the accuracy of a LDA classifier. As the symbols increase in size when intensified, the stimulus intensity is higher. One possible explanation would be that the amplitude of the P300 response increases with stimulus intensity [10]. A way to confirm this hypothesis would be to vary the intensity difference between target and non target symbols by manipulating the color or the contrast. Second, short flash stimulation (50 ms) renders task more difficult than the classical paradigm. Incorrect number of counts is higher than for the classical paradigm but there is no statistical difference in terms of P300 amplitude. LDA accuracy is higher maybe because this stimulation yields shorter peak latency. To confirm this hypothesis, different time windows of analysis should be compared. Third, increasing ISI leads to increased amplitude of the P300 response as well as increased accuracy. However, increasing the ISI yields a slower BCI. The optimal compromise in terms of bitrate [11] is the *enhanced fast* condition (see Table 1). We are currently acquiring more subjects in order to confirm and generalize these findings. Finally, applying other classifiers should enable us to assess which are the most relevant parameters and to ensure that the results do not depend on the type of classifier.

5 Conclusion

This study has proposed three modifications of the classical P300 paradigm: the size enhancement of the symbols during intensification, the use of longer ISI and the reduction of the flash duration. Results have shown that variations of the flash duration has no effect on the amplitude of P300 response but that the accuracy of a Fisher LDA is higher than in the classical condition. Moreover, the symbol size enhancement during intensification yields bigger amplitude of the P300 and very high classification rates.

6 Acknowledgements

We would like to thank Gaëlle Charles for her help during the experiments. This work was supported by grant ANR05RNTL01601 of the French National Research Agency and the National Network for Software Technologies within the Open-ViBE project.

References

- L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6):510–523, Dec 1988.
- [2] B. Rivet and A. Souloumiac. Subspace estimation approach to P300 detection and application to brain-computer interface. Conf Proc IEEE Eng Med Biol Soc, 2007:5071–5074, 2007.
- [3] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A comparison of classification techniques for the P300 speller. *J Neural Eng*, 3(4):299–305, Dec 2006.
- [4] B. Z. Allison and J. A. Pineda. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):110–113, Jun 2003.
- [5] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. *Biol Psychol*, 73(3):242–252, Oct 2006.
- [6] B. Z. Allison and J. A.Pineda. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. Int J Psychophysiol, 59(2):127– 140, Feb 2006.
- [7] M. S. Salvaris and F. Sepulveda. Robustness of the Farwell & Donchin bci protocol to visual stimulus parameter changes. Conf Proc IEEE Eng Med Biol Soc, 2007:2528–2531, 2007.
- [8] F. Perrin, J. Pernier, O. Bertrand, M.-H. Giard, and J.-F. Echallier. Mapping of scalp potentials by surface spline interpolation. *Electroencephalogr Clin Neurophysiol*, 66(1):75–81, Jan 1987.
- [9] J. Cohen and J. Polich. On the number of trials needed for P300. Int J Psychophysiol, 25(3):249–255, Apr 1997.
- [10] J. W. Covington and J. Polich. P300, stimulus intensity, and modality. *Electroencephalogr Clin Neurophysiol*, 100(6):579–584, Nov 1996.
- [11] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Braincomputer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767–791, Jun 2002.