

Design, Implementation and Evaluation of a Real-time P300-based Brain-Computer Interface System

Armagan Amcalar

Faculty of Engineering and Natural Sciences
Sabanci University
Istanbul, Turkey
aamcalar@sabanciuniv.edu

Mujdat Cetin

Faculty of Engineering and Natural Sciences
Sabanci University
Istanbul Turkey
mcetin@sabanciuniv.edu

Abstract— We present a new end-to-end brain-computer interface system based on electroencephalography (EEG). Our system exploits the P300 signal in the brain, a positive deflection in event-related potentials, caused by rare events. P300 can be used for various tasks, perhaps the most well-known being a spelling device. We have designed a flexible visual stimulus mechanism that can be adapted to user preferences and developed and implemented EEG signal processing, learning and classification algorithms. Our classifier is based on Bayes linear discriminant analysis, in which we have explored various choices and improvements. We have designed data collection experiments for offline and online decision-making and have proposed modifications in the stimulus and decision-making procedure to increase online efficiency. We have evaluated the performance of our system on 8 healthy subjects on a spelling task and have observed that our system achieves higher average speed than state-of-the-art systems reported in the literature for a given classification accuracy.

Keywords—Brain-Computer Interface, P300

I. INTRODUCTION

A brain-computer interface (BCI) is intended to help disabled subjects gain control over their environment with the use of their brain activity. A computer maps this electrical activity to functions the subject is in need of. It collects the data from an EEG amplifier and by using signal processing techniques, analyzes and makes a decision of what to do with the data.

A BCI is most useful at helping disabled subjects make choices about things such as their needs (medication, nurse, pain etc), channels in a TV remote, answers to basic YES / NO questions, or maybe letters of the alphabet.

In this paper, the performance of a BCI application, a P300 speller is tackled. P300 is an event related potential (ERP) that occurs in brain signals when the subject is exposed to visual or auditory stimulation. The P300 speller paradigm we use was first introduced by Farwell and Donchin in [1]. They reported their results on 4 healthy subjects, with a rate of 2.3 letters per minute with 95% accuracy.

Since then, various aspects of this paradigm are tackled to increase performance; electrode selection, stimulus shape,

timings, and presentation, data sampling, feature extraction, filtering, classifier algorithm and other processing procedures. Donchin et al. increased this rate up to 4.3 letters/min with 95% accuracy [2]. Meinicke increased this rate up to 5.5 letters/min with above 90% accuracy [3]. Kaper reported a rate of 47.26 bits/min [4] and Serby [5] reported 5.45 letters/min and 23.77 bits/min. These results are summarized in Table II, along with a re-interpretation with respect to this work.

We have developed a new end-to-end P300-based real-time BCI system to explore possibilities to increase this speed. Our system offers flexible stimulus mechanisms and evaluation possibilities assessable via user defined preferences. We have explored various choices in the classifier, in search of better classification results. The speed – letter/min. – we achieved in this study improves the published work so far. The average speed we achieve in offline experiments for 100% accuracy is 9.363 letters/min and 48.4073 bits/min, where in online experiments, this result is 11.14 letters/min, making sure that the subject types what he/she wants.

II. METHODOLOGY

A. Hardware Setup

All the experiments in this study are conducted with the same hardware setup. The data are recorded and digitized with a 64-channel BioSemi ActiveTwo EEG amplifier in a Faraday Cage, void of electromagnetic interference. Active electrodes are utilized, attached to the electrode cap with conductive gel. The recorded data are digitized at 2048 Hz and sent to a laptop with a dual-core processor, which records the incoming data to a hard disk. The laptop is also used for stimulation and is responsible for sending trigger signals to the amplifier during the experiment.

B. Software Setup

For offline analysis, the data are recorded in BioSemi ActiView software, and for online analysis in MATLAB, via a modified version of a MEX interface developed by Hoffmann. The classification and other analyses are also done in MATLAB and the visual stimulus is developed in C#.

C. Stimulus

Our flexible stimulus system allows any matrix size, cell content customization (letters or shapes), different coloring and stimulation schemes as displayed in Figure 1. Also, each flash duration and ISI (inter-stimulus interval) can be specified. Overall, these options can be saved as presets to be used again later on. We present our results in the most well-known stimulus type, a 6x6 matrix of characters, so that they can be compared to existing results based on this stimulus. This stimulus is a 6x6 matrix originally proposed by Farwell and Donchin in [1] that incorporates letters and numbers in each cell. The rows and columns of the matrix are highlighted in a block-randomized fashion; i.e. in 12 flashes, each row and column is flashed exactly once. Each flash lasts for 125 ms, and after each one is a period of 175 ms where none of the cells are highlighted. Therefore, each stimulus lasts 300 ms. Note that in order to define a letter, there should be at least two flashes, one row and one column, where the cell at the intersection holds the target letter. Offline analyses are done in the standard grey/white matrix. Online analyses are done in a random-colored matrix where each highlight is in a different color.

D. Terminology

In the context of this paper, offline analysis means the experimenter has prior knowledge on the letters for both training and test sets, and the analysis is done after raw data are recorded. Online analysis means the experimenter dictates only the letters for the training set and has no prior knowledge of letters in the test set and the system produces estimated letter and displays it to the subject in real time. Each flash of a row or column is called a trial. With block-randomization in mind, 12 flashes that include all the rows and columns flashing constitute a trial group. According to timings reported in the previous section, a trial group lasts for 3.6 s. A determined number of trial groups make up a run. In this study, this number is 10 for offline experiments. Online experiments have variable numbers of trial groups since they depend on classifier output. Recording and stimulation goes on without any interruptions in each run, and the target is the same letter. After a run ends, there is a brief period where the user is informed about the next target letter, and then the next run begins. A determined number of runs constitute a session. In this study, there are 8 runs (8 letters) in a session. A session group is a dataset that includes more than one session (e.g. one training session and one test session). There are breaks between recordings of sessions in a session group, to let the subject rest and prepare for the next session. A trigger signal is an indication of the highlighted row/column and is sent over to the acquisition device. Trigger data are recorded alongside with regular EEG data. An epoch is a determined period of recorded data that includes a trial.



Figure 1. Different stimuli

E. Data Acquisition

The electrodes used are Fp1, Fp2, P3, P4, PO7, PO8, Fz, Cz, Pz and Oz. Two reference electrodes are attached to each mastoid channel. Although Fp1 and Fp2 are generally ignored due to eye-blink artifacts, we have included them in our analysis to explore their effect on classification.

F. Preliminaries

For offline analyses in this study, there are two sessions in a session group, one being the training session and the other, the test session. Other than a few minor exceptions, the training session of each subject featured 8 runs that had “D E D E D E D E” as targets. The test sessions also featured 8 runs and included random letters, chosen either by the subject or the experimenter beforehand. Each epoch lasts for 1 second. The classifier is trained on the first session and tested on the second.

G. Data pre-processing

Proper pre-processing is an important factor in classification performance. We have conducted several different pre-processing schemes and observed that no scheme is best for all subjects. The definitive scheme used in all offline analyses is similar to [6] and is as follows:

To get rid of irrelevant frequency components, the data are filtered with a 6th order Butterworth band-pass filter with a pass-band of 1 – 12 Hz. ActiView saves the data with respect to the common-mode sense (CMS) electrode. To obtain a greater SNR, the data are re-referenced to the average of two mastoid channels.

For better performance of the classifier, the data should be normalized. But data with peaks lose resolution when normalized; therefore the data are first winsorized in a 10% frame, and zero-mean normalization follows next. Lastly, the data are decimated by 64. After decimation, each epoch is represented with 32 samples.

The feature vector for each epoch is then the concatenation of filtered data from each electrode, i.e. a vector of 320 samples for 10 electrodes.

We found out that in general, subjects blink rarely during each run and Fp1 and Fp2 contribute positively to the classification performance, especially when eye-blink artifacts are removed by winsorization.

We have observed that half of the subjects performed better with normalization and winsorization, and the other

half performed better without them. In offline analysis, the results are generated according to the scheme the subject was best at. In online analysis, normalization and winsorization are applied to all subjects.

H. Classification

For the classification algorithm, we used Bayesian Linear Discriminant Analysis (BLDA), mentioned in [7], A derivative of Fisher’s LDA, BLDA gives probabilistic output of test data, incorporates feature selection based on discriminative power and learns regularization parameters automatically from the training set.

Averaging of multiple trials is frequently used to increase the SNR of P300 waves. Rather than using averaging, in our work, we incorporate information from multiple trials by probabilistic updates as new trial data are received. In particular, BLDA calculates a score for each epoch of test data, reflecting its similarity to the underlying classes. Scores are added up in consecutive trial groups until a firm separation between scores is present.

For offline analysis, the sum of scores are checked at the end of each trial group and the row and column with the maximum scores are selected as answers of classification, and are compared with actual targets to generate the accuracy plots in Figure 2.

Since actual targets are unknown to the experimenter in online analyses, the classifier has to decide by itself when to end each run. This is done by using margins in scores. A safe margin is determined and when the column and the row with the highest scores have that margin between themselves and the next best ones, the character at the intersection of these two is presented as the decisive answer of the classifier.

III. RESULTS

A. Offline Analysis

The offline analysis results in Figure 2 are presented in a format compatible with [6]. Figure 2(c) shows average offline classification performance of our system, which improves upon the results in [6]. The x-axis shows the number of trial groups, the left y-axis and solid lines indicate the percentage of correct results and the right y-axis and dashed lines indicate the bit rate.

7 healthy subjects took part in this study, whose ages ranged between 19 and 26. No subject had any prior BCI experience. The computation of the bit rate is performed as in [8]:

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 [(1 - P)/(N - 1)] \quad (1)$$

where P is the accuracy of classification and N is the number of elements in the matrix. Since one trial group lasts for 3.6 s and a pause of 1.4 seconds for displaying the next letter is assumed, there can be 12 trial groups in a minute, i.e., 12 letters can be written at maximum. The maximum possible bit rate of our system for offline classification is then 62.0391 bits/min.

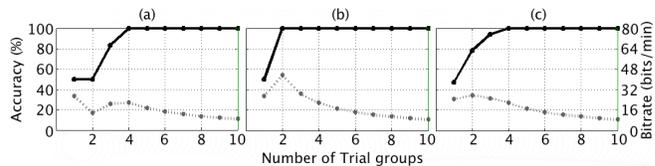


Figure 2. Offline performance. (a) Worst performing subject (S5), (b) Best performing subject (S6), (c) Average of 7 subjects

Evaluation of performance in Figure 2 is done at intervals of 3.6 s. For calculating accuracy, every run is classified separately and accuracy is the total number of correct classifications in a session over total number of classifications. Table I presents the average offline performance in letters/min.

B. Online Analysis

We have developed a greedier version of our algorithm that relies on the fact that the classifier produces probabilistic scores. In the beginning of each run, each row and column receives a score of 0. If their epoch includes a P300 wave, they get a positive score, with its magnitude reflecting the resemblance to the training set, and irrelevant epochs get a negative score. With this approach, there is usually no need to evaluate all the 12 epochs for a decision. If the score of an epoch already satisfies the margin, the decision can already be made. We have conducted online analyses with 6 subjects, 5 of whom also participated in offline analyses.

Table I presents average online performance in detail, listing Right and Wrong classification results and two kinds of accuracy vs. rate values. The first one allows errors in results, so typing rate is calculated omitting the error in classification; therefore it is faster. The latter makes sure the subject types the exact letter he/she wants, so time spent on wrong classification results are taken into calculation as lost time.

Considering this information, on average, 100% correct classification of a given set of 100 runs will last for 178 trial groups. This shows that each run is on average classified in 1.78 trial groups. Assuming a period of 3.6 s for a trial group, a letter can be classified in 6.408 s.

TABLE I. PERFORMANCE VALUES

	Avg. Offline Perf. (l/m)	Average Online Performance (Color matrix)				
		R	W	Avg. Rate		Avg Rate (l/m exc. W)
				Acc (%)	Rate (l/m)	
S1	8.3	43	8	84	9.5	8
S2	9.5	16	2	89	15.5	13.8
S3	7.8	9	0	100	10.4	10.4
S4	12.1	x	X	x	x	x
S5	7.7	9	1	90	14.5	13
S6	11.1	30	5	85	15.3	12.7
S7	10.6	x	x	x	x	x
S8	x	46	4	92	9.7	8.9
+	9.4	153	20	88	12.5	11.1

TABLE II. PERFORMANCE VALUES IN LITERATURE

Reference	Reported				Reinterpreted	
	Letters/min	Bits/min	ISI	Acc.	Letters/min	Acc.
[1]	2.3	10.67	500ms	95%	unk.	unk.
[2]	4.3	19.83	125ms	95%	9.367	100%
[3]	5.5	24	300ms	90%	11.037	95%
[4]	-	47.26	140ms	44%	19.7	92%
[5]	5.45	23.77	125ms	92%	15.209	100%
This work	11.111	49.39	300ms	93%	Offline tests	
	9.363	48.41	300ms	100%	Offline tests	
	20.48	105.88	125ms	w/o err	Online tests	
	20.44	91.58	125ms	94%	Online tests	
	24.42	126.25	125ms	100%	Offline tests	

C. Discussion

Figure 2 tells us that 48% of the time the classifier predicts the right answer in the first trial group. The classifier has correct answers in 2 trial groups 81% of the time and so on.

If we assume no delay between each run, then on the average our system achieves a rate of 9.363 letters/min.

In practice, we also spend 1.4 sec. between each run to display the next letter to be typed on the screen. When we take that extra time into account, the average offline rate becomes 7.6844 letters/min.

If one assumes 93% accuracy, 100 runs will be classified in 150 trial groups with wrong results in 7 runs, which yields a result of 11.111 letters/min without interruption and 8.823 letters/min with pauses.

In online analysis, the average speed is 12.48 letters/min for error ignorant results with accuracy of 88%, and 11.14 letters/min when errors are taken into consideration.

D. Reinterpretation of Table II

The reported results in Table II are re-interpreted according to our method of performance calculation, which takes into consideration that these letters are presented in a sequence, therefore the accuracy results have to be chained, as told in II.C. The reinterpreted results are the results of rough estimations done over plots in related references, and are directly compatible with our results.

E. Decreasing the ISI

As the beginnings of a new set of experiments, we have lowered the ISI to 125 ms, where in the first 50 ms a target is highlighted, and for the next 75ms the matrix is dim.

We have conducted both online and offline experiments with this setting on two subjects, whose average results are also listed in Table II. For offline experiments, our subjects achieved an average rate of 24.42 letters/min and 126.25 bits/min with 100% accuracy and for online experiments, the average rate was 20.44 letters/min and 91.58 bits/min for 94% accuracy and 20.48 letters and 105.88 bits/min for the case that the subjects typed all the letters they wanted (i.e. re-typed erroneous letters).

IV. CONCLUSION

In this paper, we have demonstrated the flexibility and performance of our end-to-end BCI system with experiments done with 8 able-bodied subjects. The highest rate achieved by a subject using our system with 300ms ISI is 12.1 letters/min and 62.55 bits/min for 100% accuracy in offline analysis, and 15.5 letters/min and 63.71 bits/min in online analysis. On the other hand, for 125ms ISI, the highest rate for offline analysis was 32 letters/min and 165.44 bits/min for 100% accuracy and the highest rate for online analysis was 23.19 letters/min and 98.88 bits/min for 90.91% accuracy.

We have demonstrated that our system can achieve higher rates (for a given classification accuracy) than current state-of-the-art systems both for offline and for online experiments.

ACKNOWLEDGMENT

This work was supported by the Scientific and Technological Research Council of Turkey under Grant 107E135, and by a Turkish Academy of Sciences Distinguished Young Scientist Award.

REFERENCES

- [1] 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*, vol. 70, pp. 510-23, Dec 1988.
- [2] E. Donchin, *et al.*, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Trans Rehabil Eng*, vol. 8, pp. 174-9, Jun 2000.
- [3] P. Meinicke, M. Kaper, F. Hoppe, M. Huemann, H. Ritter, "Improving transfer rates in brain computer interface: A case study". *NIPS*, pp. 1107-1114, 2002.
- [4] M. Kaper and H. Ritter "Generalizing to new subjects in brain-computer interfacing," *Conf Proc IEEE Eng Med Biol Soc*, vol. 6, pp. 4363-6, 2004.
- [5] H. Serby, *et al.*, "An improved P300-based brain-computer interface," *IEEE Trans Neural Syst Rehabil Eng*, vol. 13, pp. 89-98, Mar 2005.
- [6] U. Hoffmann, *et al.*, "An efficient P300-based brain-computer interface for disabled subjects," *J Neurosci Methods*, vol. 167, pp. 115-25, Jan 15 2008.
- [7] U. Hoffmann, *et al.*, "Application of the evidence framework to brain-computer interfaces," *Conf Proc IEEE Eng Med Biol Soc*, vol. 1, pp. 446-9, 2004.
- [8] J. R. Wolpaw, *et al.*, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, pp. 767-91, Jun 2002.