Exploiting P300 Amplitude Variations Can Improve Classification Accuracy in Donchin's BCI Speller

Luca Citi, Riccardo Poli, and Caterina Cinel
School of Computer Science and Electronic Engineering
University of Essex
CO4 3SQ, Colchester, UK
Email: {lciti,rpoli,ccinel}@essex.ac.uk

Abstract—The P300 is an endogenous component of EEG event related potentials which is elicited by rare and significant stimuli. P300s are used increasingly frequently in Brain Computer Interfaces (BCI) because, being naturally elicited in response to external stimuli, users do not need special training. However, P300 waves are hard to detect and, therefore, multiple stimulus presentations are needed before an interface can make a reliable decision. While significant improvements have been made in the detection of P300s, no particular attention has been paid to the variability in shape and timing of P300 waves and its exploitation in BCI. In this paper we start filling this gap, by first documenting and then exploiting a modulation in amplitude of P300 caused by target-to-target interval (TTI) differences. We demonstrate this within the context of the Donchin's speller, which is perhaps the best known example of a BCI system relying on the detection P300 waves, where target-to-target interval variations are induced by stimuli randomisation. In particular we show that by specialising detectors to work with P300s elicited with each TTI, we can further improve the performance of the best known Donchin's speller with minimal changes.

I. INTRODUCTION

Brain-Computer Interface (BCI) systems measure specific (intentionally and unintentionally induced) signals of brain activity and translate them into device control signals (see [1] for a comprehensive review).

Many factors limit the performance of a BCI system. These include: the natural variability and noise in the brain signals measured; the limitations of the recording and signal processing methods that extract signal features and of the algorithms that translate these features into device commands; the quality of the feedback provided to the user; the lack of motivation, tiredness, limited degree of understanding, age variations, handedness, etc. in users; the natural limitations of the human perceptual system [2].

In some cases, however, amplitude and shape variations in brain waves may carry information which can be exploited to the benefit of BCI (e.g., see the discussion in [2]) even if a physiological explanation for such variations is unavailable (see for example the analogue approach used in [3] where the task of extracting information from amplitude variations was left to an evolutionary algorithm).

In this paper we will document and exploit a source of waveform variations in P300 waves (an endogenous component of EEG event related potentials which is elicited by rare and significant stimuli): namely, their modulation caused by variations in the interval between target-stimulus presentations. We illustrate our ideas using the Donchin's speller as a model, although we expect the benefits of our approach could be accrued also in other P300-based systems.

The paper is organised as follows. In Section II we provide some background on Donchin's speller, on the P300 waves and the factors that may affect their characteristics, and on what is known about how P300 characteristics are affected by the timing of stimulus presentation. In Section III we document the effects that target-to-target interval variations have on the shape of P300s. In Section IV we modify the best Donchin's speller available to date so as to take timing effects into account and show that the new system is superior to the original. We draw some conclusions in Section V.

II. BACKGROUND

A. P300 Waves and Factors Affecting their Characteristics

Among the different approaches used in BCI studies, those based on the P300 event related potential (ERP) [4] present a relatively high bit-rate and no need for user training.

ERPs are relatively well defined shape-wise variations to the ongoing EEG elicited by a stimulus and temporally linked to it. ERPs include an exogenous response, due to the primary processing of the stimulus, as well as an endogenous response, which is a reflection of higher cognitive processing induced by the stimulus [5].

The P300 wave is an endogenous component of ERPs with a latency of about 300 ms which is elicited by rare and/or significant stimuli (visual, auditory, or somatosensory). Effectively P300 potentials are ERP components whose presence depends on whether or not a user attends a rare, deviant or target stimulus. This is what makes it possible to use them in BCI systems to determine user intentions. For an overview of the cognitive theory and neurophysiological origin of the P300 see [6].

The characteristics of the P300 component (mainly its amplitude and its latency) vary depending on several factors [7]. Some factors are related to the psychophysical state of the subject [8], such as food intake, fatigue, assumption of drugs. Others factors depend on the the physical layout of the stimuli [9]–[11], such as number of symbols, their size, their relative spacing. Other important factors are related to the sequence of stimuli. For example, several studies have

reported that the P300 amplitude increases as target probability decreases (for a review, see [5]). The P300 amplitude seems also to be positively correlated with the interstimulus interval (ISI) or the stimulus onset asynchrony (SOA) (as reported, among others, in [12]–[14]). Other studies [15], [16] object that, despite the P300 being clearly affected by target stimulus probability and ISI, each of these factors also varies the average target-to-target interval (TTI) which they hypothesise to be the true factor underlying the P300 amplitude effects attributed to target probability, sequence length, and ISI.

In fact, in addition to being modulated by target probability, the P300 is also sensitive to the order of nontarget and target stimuli since this temporally modifies target stimulus probabilities. There is a positive correlation between P300 amplitude and the number of nontarget stimuli preceding a target (e.g. [17], [18]).

The influence of the sequence preceding a target on the P300, could be partially explained as the result of "recovery cycle" limitations in the mechanisms responsible for component generation [15]. Smaller potentials could be produced after a short TTI, because the system has not yet reacquired the necessary resources to produce large ERPs.

Others ascribe the lower amplitudes associated with shorter TTIs (i.e., few nontargets preceding a target stimulus) to an inability to consistently generate a P300, rather than producing a P300 with small amplitude [19]. In other words, they attribute the lower amplitude of the averaged ERP, to an increase in the percentage of responses to target stimuli that do not show a P300 component, whereas the amplitude of the P300 component (for the responses that do show a P300) would be unaffected.

B. Donchin's P300-based speller

Back in 1988, Donchin and his student Farwell designed a speller based on the P300 component. The user was presented with a 6 by 6 matrix of characters (see Fig. 1) whose rows and columns were randomly highlighted. The user's task was to focus attention on the chosen character. Every 12 flashes (one per each row and column), the 2 containing the desired character represented a rare target stimulus, therefore able to elicit a P300-like response. By averaging the ERP related to each row and column and looking for the largest P300 response, it was possible to infer the target character with sufficient accuracy.

Fig. 1. The matrix of characters used in the Donchin speller.

III. TTI EFFECTS IN DONCHIN'S SPELLER

A. Methods

To test whether TTI-modulated P300 variability also occurs with Donchin's speller, we studied the training set of the two subjects of dataset II from the BCI competition III [20].

In the competition data were collected with the Donchin speller protocol described above, using a SOA of 175 ms. For each subject, the training set consisted of 85 characters, each one containing 15 sequences of 12 intensifications. Further details on the data can be found in [20].

The signals were further bandpass-filtered in the band 0.15 - 5 Hz (HPF: 1600-tap FIR; LPF: 960-tap FIR) to reduce exogenous components at 5.7 Hz and multiples. The 1-second epochs following each flash were extracted. Therefore, for each subject, a total of 15300 ($85 \times 15 \times 12$) epochs were available, of which 2550 ($85 \times 15 \times 2$) were targets.

The set of epochs was partitioned according to the number of nontargets presented between the previous target and the current epoch. For example, for the sequence ...TNNNT..., the second target (T) is assigned to partition 3 because it is preceded by three nontargets (N). Then, for each partition, average target and nontarget responses were determined as follows.

First, outlier epochs were removed. For each set of epochs, the first, $q_1(i)$, and third, $q_3(i)$, quartiles at each sample, i, were found. Then an acceptance "strip" was defined as the time-varying interval $[q_1(i) - 1.5 \Delta q(i), q_3(t) + 1.5 \Delta q(i)]$ where $\Delta(i) = q_3(i) - q_1(i)$ is the interquartile range. Responses falling outside the acceptance strip for more than one tenth of the epoch were rejected. The remaining responses were averaged. The mean, $m(i)$ and the standard error $ste(i)$ of the responses for each class were finally evaluated using the remaining epochs.

B. Results

Fig. 2 shows the average responses obtained from the epoch-partitioning and artifact-rejection procedure described above. The results confirm that despite Donchin's speller being characterised by fast SOA, significant modulations of the P300 amplitude due to TTI variations are present.

The most significant effect is visible around the peak of "t", the average P300, that is about 450 ms for subject A and 350 ms for B. It is not surprising that in the average response of the partition "t00" (two targets in a row), the P300 is almost completely absent, as the previous P300 (approximately 175 ms before) has not yet faded away. Similar considerations apply to "t01".

The other averages show an increase in the P300 amplitude for increasing number of nontargets separating the flash from the previous target (proportional to the TTI).

IV. IMPROVING PERFORMANCE TAKING TTI EFFECTS INTO ACCOUNT

A. Methods

Our next step was to test whether knowledge of the effects of the target-to-target-interval on the P300 can be exploited to
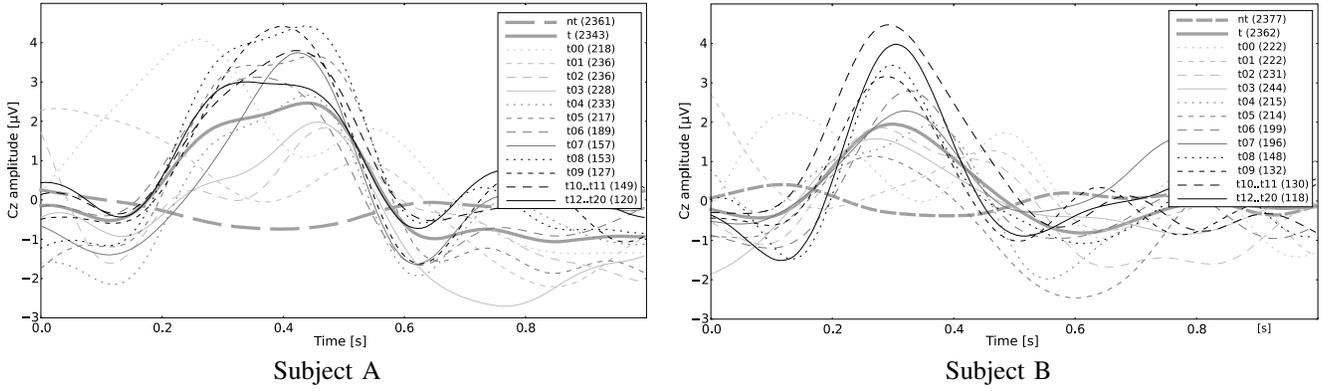
Subject A

Subject B

Fig. 2. The thin lines are averages of Cz for the epochs partitioned according to the number of nontargets preceding the current flash (lines "t00" to "t12..t20"). The thick line labelled "t" is the average of all targets while "nt" is the average of an equally-sized random sample of nontargets. The numbers in parentheses represent the number of epochs used to form each average, after the artifact rejection procedure described in the text. The standard error is approximately 0.15 μV for "t" and "nt" while it is approximately 0.5 μV for the others. The effect on the P300 amplitude of the number of nontargets preceding the flash is evident because the colour and texture of the thin lines fade from light (short TTI) to dark (long TTI) for increasing values of the amplitude they take at the time when the "t" line peaks (see also Fig. 3).
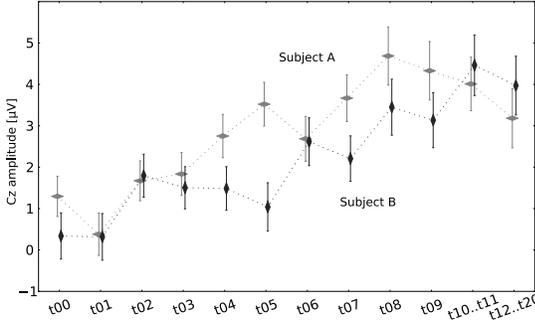

Fig. 3. Slice view of Fig. 2 at the time when the average target ERP "t" peaks. The two lines, for subject A and B, are slightly horizontally displaced to ease reading. The whiskers represent the standard error.

build better classifiers. Instead of building a new full-blown ad-hoc classifier, we decided to test first whether a thin layer built on top of a high-performing existing approach could improve the performance.

We borrowed from the work of Rakotomamonjy and Guigue [21] which resulted winner of the III BCI Competition for the Donchin speller. Namely, we used the approach they called "Ensemble SVM without channel selection" because it is easier to implement and outperforms other alternatives when using only 5 sequences to classify a character.

They used an ensemble of classifiers approach, where the datasets were split in several subsets and a linear support vector machine (SVM) classifier was trained on each of them. The outputs of all classifiers were summed up to build the final decision. When using J sequences, the character identified by row r and column c was scored

$$S_{r,c} = \sum_{k=1}^{K} \sum_{j=1}^{J} f_k(x_{r(r,j)}) + f_k(x_{c(c,j)}) \qquad (1)$$

where $f_k(x)$ is the output of the kth classifier, x_i is a vector with features from the ith epoch, $r(r,j)$ is a mapping

returning the ordinal position of the flash where the row r was target during the jth sequence, and similarly for $c(c,j)$ [1]. The character with higher score was returned.

We decided to start from their approach and just change the scoring function in order to account for the effect of the number of nontargets preceding each epoch. The following hypotheses were made: the ERP response was considered to be a Gaussian random process whose mean is shifted up around 300 ms poststimulus when the flash is target; the amount by which the mean increases depends, among many factors, on the number of nontarget stimuli preceding the current one (Fig. 2 can be interpreted as the timecourse of the mean in the different conditions). The discriminability, or equivalently the reliability of the classification, of each epoch as target or nontarget depends on the distance between the corresponding target and nontarget class. As a result, the output of the classifier will be more unreliable when few nontargets separate the flash from the previous target.

Using these considerations, the scoring function was changed to

$$S_{r,c} = \sum_{k=1}^{K} \sum_{j=1}^{J} w_k(h(r(r,j),r,c)) f_k(x_{r(r,j)}) \qquad (2)$$
$$+ w_k(h(c(c,j),r,c)) f_k(x_{c(c,j)})$$

where $w_k(h)$ is a mapping which associates a weight to the number of nontargets preceding a flash, while $h(i,r,c)$ is a mapping returning the number of nontargets preceding the ith flash in the hypothesis that the row r and the column c identified the target character.

The function $w_k(h)$ was found for each classifier using the part of the training set which was not used to build that particular SVM classifier. A stochastic hill climber was

---

[1]The definition of the scoring given here is slightly different from the one given in equation (2) of [21] but equivalent. Our definition is functional to equation (2) in the present article.
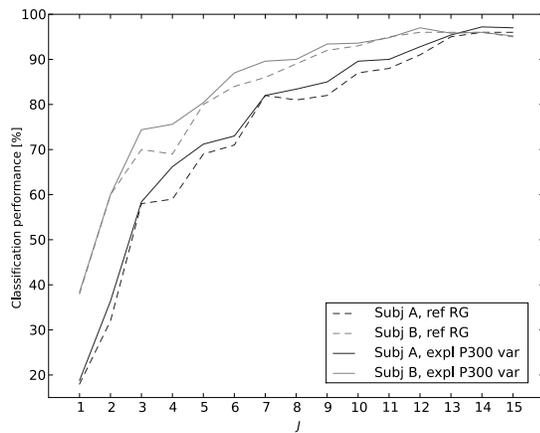
Fig. 4. Classification performance in terms of percentage of correctly recognised characters as a function of the number of sequences used (J in (1) and (2)). For the two subjects, the performance of the approach introduced in this paper ("expl P300 var") is compared to the reference algorithm ("ref RG") by Rakotomamonjy and Guigue [21].

trained to maximise the number of correct characters using 5 sequences. The starting points were in the range $[0.8, 1]$ and each $w_k(h)$ was normalised to have unit sum.

Finally, the algorithm was tested on the test set and the results compared with those of the original method in [21].

B. Results

As reported in Fig. 4, the proposed approach outperforms the original approach in [21]. The improvement, albeit moderate, is consistent. On average (for all the values of J considered) there is a 2.1% advance for subject A and 1.5% for B. Using a small number of sequences, $J = 1$–5, the average improvement is 3% for A and 2.4% for B; using $J = 6$–10 sequences, 2% for A and for B; using $J = 11$–15 sequences, 1.2% for A and 0.2% for B. The maximum boost happens for both subjects for $J = 4$ and is 7% for A and 6% for B.

Significant improvements arise when 3 to 7 sequences are used, which is a range characterised by a reasonable speed vs accuracy compromise.

V. CONCLUSION

In this paper we first document and then exploit, within the context of the Donchin's speller, a modulation in amplitude of P300 caused by target-to-target interval differences. In particular, we show that by specialising detectors to work with P300s elicited with each TTI, we can consistently improve performance of the best known classification algorithm for Donchin's speller with minimal changes. In the future we intend to explore the possibility of obtaining similar improvements within other BCI paradigms based on P300s.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting." *IEEE Trans Rehabil Eng*, vol. 8, no. 2, pp. 164–173, Jun 2000.

[2] C. Cinel, R. Poli, and L. Citi, "Possible sources of perceptual errors in P300-based speller paradigm," *Biomedizinische Technik*, vol. 49, pp. 39–40, 2004, Proceedings of 2nd International BCI workshop and Training Course.

[3] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, "P300-based BCI mouse with genetically-optimized analogue control," *IEEE Trans Neural Syst Rehabil Eng*, vol. 16, no. 1, pp. 51–61, Feb. 2008.

[4] 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, no. 6, pp. 510–523, Dec 1988.

[5] E. Donchin and M. G. H. Coles, "Is the P300 a manifestation of context updating?" *Behavioral and Brain Sciences*, vol. 11, pp. 355–372, 1988.

[6] J. Polich, "Neuropsychology of P3a and P3b: A theoretical overview," in *Brainwaves and mind: Recent developments*, N. C. Moore and K. Arikan, Eds. Kjellberg Inc., 2004, pp. 15–29.

[7] ——, "Updating P300: an integrative theory of P3a and P3b." *Clin Neurophysiol*, vol. 118, no. 10, pp. 2128–2148, Oct 2007. [Online]. Available: http://dx.doi.org/10.1016/j.clinph.2007.04.019

[8] J. Polich and A. Kok, "Cognitive and biological determinants of P300: an integrative review." *Biol Psychol*, vol. 41, no. 2, pp. 103–146, Oct 1995.

[9] G. F. Hagen, J. R. Gatherwright, B. A. Lopez, and J. Polich, "P3a from visual stimuli: task difficulty effects." *Int J Psychophysiol*, vol. 59, no. 1, pp. 8–14, Jan 2006. [Online]. Available: http://dx.doi.org/10.1016/j.ijpsycho.2005.08.003

[10] 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*, vol. 73, no. 3, pp. 242–252, Oct 2006. [Online]. Available: http://dx.doi.org/10.1016/j.biopsycho.2006.04.007

[11] 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*, vol. 11, no. 2, pp. 110–113, Jun 2003. [Online]. Available: http://dx.doi.org/10.1109/TNSRE.2003.814448

[12] J. Polich, "Probability and inter-stimulus interval effects on the P300 from auditory stimuli." *Int J Psychophysiol*, vol. 10, no. 2, pp. 163–170, Dec 1990.

[13] P. G. Fitzgerald and T. W. Picton, "Temporal and sequential probability in evoked potential studies." *Can J Psychol*, vol. 35, no. 2, pp. 188–200, Jun 1981.

[14] 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*, vol. 59, no. 2, pp. 127–140, Feb 2006. [Online]. Available: http://dx.doi.org/10.1016/j.ijpsycho.2005.02.007

[15] C. L. Gonsalvez and J. Polich, "P300 amplitude is determined by target-to-target interval." *Psychophysiology*, vol. 39, no. 3, pp. 388–396, May 2002.

[16] R. J. Croft, C. J. Gonsalvez, C. Gabriel, and R. J. Barry, "Target-to-target interval versus probability effects on P300 in one- and two-tone tasks." *Psychophysiology*, vol. 40, no. 3, pp. 322–328, May 2003.

[17] K. C. Squires, C. Wickens, N. K. Squires, and E. Donchin, "The effect of stimulus sequence on the waveform of the cortical event-related potential." *Science*, vol. 193, no. 4258, pp. 1142–1146, Sep 1976.

[18] C. J. Gonsalvez, E. Gordon, J. Anderson, G. Pettigrew, R. J. Barry, C. Rennie, and R. Meares, "Numbers of preceding nontargets differentially affect responses to targets in normal volunteers and patients with schizophrenia: a study of event-related potentials." *Psychiatry Res*, vol. 58, no. 1, pp. 69–75, Sep 1995.

[19] B. Bonala, N. N. Boutros, and B. H. Jansen, "Target probability affects the likelihood that a P300 will be generated in response to a target stimulus, but not its amplitude." *Psychophysiology*, vol. 45, no. 1, pp. 93–99, Jan 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1469-8986.2007.00613.x

[20] B. Blankertz, K.-R. Mller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlgl, G. Pfurtscheller, J. del R Milln, M. Schrder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems." *IEEE Trans Neural Syst Rehabil Eng*, vol. 14, no. 2, pp. 153–159, Jun 2006. [Online]. Available: http://dx.doi.org/10.1109/TNSRE.2006.875642

[21] A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II- ensemble of SVMs for BCI P300 speller," *IEEE Trans Biomed Eng*, vol. 55, no. 3, pp. 1147–1154, Mar 2008. [Online]. Available: http://dx.doi.org/10.1109/TBME.2008.915728