Novel protocols for P300-based brain-computer interfaces

Mathew Salvaris, Caterina Cinel, Luca Citi and Riccardo Poli

***Abstract*—The oddball protocol is often used in Brain-Computer Interfaces (BCIs) to induce P300 ERPs, although, recently, some issues have been shown to detrimentally effect its performance. In this paper, we study a new periodic protocol and explore whether it can compete with the standard oddball protocol within the context of a BCI mouse. We found that the new protocol consistently and significantly outperforms the standard oddball protocol in relation to information transfer rates (33 bits/min for the former and 22 bits/min for the latter, measured at 90% accuracy) as well as P300 amplitudes. Furthermore, we performed a comparison of two periodic protocols with two less conventional oddball-like protocols that reveals the importance of the interactions between task and sequence in determining the success of a protocol.**

***Index Terms*—Brain-computer interfaces, P300, Event-related potentials, Oddball paradigm, BCI mouse**

I. Introduction

THE FIELD of brain-computer interfaces (BCIs) [1]–[3] continues to capture the imagination of many researchers. BCIs refer to systems that utilise electroencephalography (EEG) and other technologies to detect changes in the state of the brain and to make sense of such changes in order to achieve communication [3]. Because BCI systems do not require the use of the normal communication channels available to able bodied individuals, they are especially desirable for the severely disabled.

Still, the performance of BCIs is limited. This is due to numerous factors such as the non-stationary nature of brain signals, limitations of the signal processing and classification methods used and a limited understanding of the underlying cognitive processes [3]–[6]. These are also the reasons why BCI remains such an engaging field.

Amongst the numerous BCI paradigms, those based on the P300 component in EEG are among the most successful and well-studied [7], [8]. Usually, in BCI, P300 ERPs are triggered using oddball procedures [9]. Although P300-based BCIs have enjoyed remarkable success over the last few years, some considerable drawbacks have also been identified [10]–[12] (more on this in the next section).

In this paper, we propose a novel P300-based BCI paradigm that has the potential to significantly reduce some of the limitations of previous approaches (as we will discuss in Section IV). A key feature of our paradigm is that it is *not* based on the standard oddball technique involving identifying a rare and unpredictable target stimulus within a sequence of non-target stimuli. Instead, it uses *periodic protocols* where the timing of target stimuli presentation is regular and, therefore, entirely predictable and where participants have to *discriminate* between different features of the targets. We demonstrate our ideas within a specific application (a P300-based BCI Mouse) although the approach is general and can be applied to a variety of domains in BCI and possibly also in psychophysiology.

The paper is organised as follows. In Section II we provide a survey of relevant literature on the P300 component and P300-based BCIs. In Section III we review previous work in the specific application domain (a P300-based BCI Mouse) within which we test our approach. In Section IV we discuss the key ideas which led us to formulate a class of periodic protocols. In Section V we describe the two experiments we conducted to test these ideas and introduce the details of our experimental methodology. In Section VI we report the results of the experiments. In particular, we compare new protocols against the traditional oddball protocol and against two other protocols that, like the oddball, are based on random presentation sequences. These comparisons are based on classification performance and on the features of the Event-Related Potentials (ERPs) different protocols elicit. In Section VII we discuss our findings. Finally, we provide some conclusions and indications for future work in Section VIII.

II. P300 and P300-based BCIs

The P300 ERP has a centro-parietal focus and typically occurs around 300–600 ms after a task-relevant stimulus is detected amongst a group of task-irrelevant stimuli (e.g., see [13]). It is a widely explored potential not only in BCI but also, rather naturally, in cognitive neuroscience. The P300 can be elicited by different sensory modalities (auditory, tactile and visual) under certain experimental conditions [14], [15]. It is an *endogenous* component, since it is caused by internal cognitive processes rather than by some automated preattentive processing of incoming stimuli. In other words, it is 'invoked' rather than 'evoked' by the stimuli [16]. There are a number of P300-like potentials [13]: the one which is most relevant for this paper is the P3b. However, for simplicity, we will still refer to it as P300 in the rest of the paper.

Most P300-based protocols rely on the so called "oddball" paradigm [17], [18]. Typically the oddball paradigm involves the presentation of a train of standard background stimuli (non-targets) within which rare "oddball"/relevant stimuli (or targets) are included at random temporal positions. Subjects

M. Salvaris, C. Cinel and R. Poli are with the Brain-Computer Interfaces Lab, School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK. L. Citi is with the Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital — Harvard Medical School, Boston, MA, USA, and with Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA USA. emails: {mssalv,ccinel,rpoli}@essex.ac.uk, lciti@neurostat.mit.edu

are tasked with detecting the target stimuli and, typically, performing some simple task such as pressing a button every time the target has been perceived or counting the targets.[1] While the nature of the stimuli and of the task may change, there are three properties that are common to virtually all oddball paradigms: (1) targets need to be discriminated from non-targets, (2) as stimulus presentation is randomised, the target-to-target interval (TTI) is not a constant, but a random variable, and (3) targets are relatively rare.

The characteristics of the P300, such as its amplitude, scalp location and latency, are influenced by a number of factors ranging from the physiological aspects of the subject (e.g., fatigue and food intake) to properties of the protocol used to elicit P300s, as well as the motivations and imperatives imbued in the subjects and their psychological state [13], [16]. These sources of variability confound matters and make it very hard to design an optimal P300-elicitation paradigm.

Key protocol properties that affect the amplitude and latency of P300s are the target probability, the inter-stimulus interval (ISI, the time interval between the end of a stimulus and the beginning of the following one) and the stimulus onset asynchrony (SOA, the time interval between the beginning of two consecutive stimuli). The inverse relationship between P300 amplitude and target probability has been known for several decades [16] and so has the positive correlation between the amplitude of P300s with the ISI or the SOA [19]. More recently, it has been suggested that the variance observed in the P300 properties, when adjusting such parameters, might be better explained using the TTI [20].

In order to avoid confusion in the interpretation of ERPs, the SOAs chosen in cognitive psychology experiments are usually quite long (in the order of seconds). In this way, the ERPs associated with a stimulus are clearly distinguishable from those associated with the following, or previous, stimuli. In a BCI, however, this is not viable since such long SOAs would result in impractically slow communication rates. Therefore, the SOAs used in BCI are typically much shorter (around 175 ms is quite common). This triggers a whole new set of problems and effects which have recently been identified and studied [10]–[12]. These noticeably hinder classification and, therefore, reduce the information transfer rate of BCI systems.

A key effect is related to the temporal proximity of target stimuli: targets that closely follow other targets produce worse classification than targets that appear one or more stimulus presentations later. This effect has been attributed to psychophysiological phenomena such as attentional blink [21] and repetition blindness [22] as well as overlap effects [10]–[12], [23]. Some attempts have been made to either mitigate these effects [2], [10] or harness them in order to achieve better classification [11]. Results have been encouraging but the problem still remains and can only get worse as we seek to shorten the SOA to achieve higher information transfer rates.

[1]While focusing attention on the target stimulus is a requirement, the amplitude of the P300 (P3b) is significantly enhanced if, additionally, a task is assigned to the occurrence of a such stimulus [16] as long as task demands are still within a subject's capabilities (an excessive task difficulty, however, can have a negative effect on P300 amplitude).

III. BCI MICE BASED ON THE P300

Given the point-and-click nature of most modern user interfaces, an important application of BCI is controlling 2–D pointer movements. There have been some attempts to develop BCI systems for this purpose, the most successful of which, to date, being those based on the detection of μ or β rhythms [24], and those using invasive cortical interfaces (e.g., [25]). The former, however, require lengthy training periods before users can control them, while the latter are not very practical, being very invasive.

These problems can be overcome by systems based on the use of P300s. Some success with this approach has been reported in [26] where an ISI of 2.5 seconds was used, leading to the pointer moving at the rate of one movement every 10 seconds. Shorter ISIs were used in [27] resulting in a speed of one cursor movement every 4 seconds.

A more responsive P300-based mouse (producing one movement per second) was presented in [28]. Initial offline results were reasonably good, however, preliminary online tests and further offline analysis showed that the mouse pointer was hard to control.

We explored a variety of alternatives to this approach in [29]. Results indicated that the best protocols are those where stimuli are randomly flashed. The protocols used in [29] are an improvement over those used in [28], [30]. For example, at the faster SOA and using a vectorial output-integration strategy, our mouse could perform an analogue movement every 100 ms. However, the information transfer rate achieved were still insufficient to produce a responsive and accurate control over the mouse pointer.

A key feature of these and other P300-based BCIs is their use of the traditional oddball paradigm. As we have seen in Section II, this paradigm suffers from some drawbacks, particularly at the fast presentation rates used in BCI. It is then natural to ask whether it would be possible to define protocols which move away from the classical oddball, while still producing robust P300s.

In the following section we discuss some ideas on how this could be achieved.

IV. KEY IDEAS

As we mentioned in Section II, there are three features that are common to virtually all oddball paradigms: (1) targets need to be discriminated from non-targets, (2) TTIs are random, and (3) targets are relatively rare. If we take these properties as our working definition of "oddball protocol", we should note that feature (1) constrains the *task* that subjects may be required to carry out in an oddball protocol, while features (2) and (3) constrain the *sequence* of stimuli used in an oddball protocol. It stands to reason that departing from the oddball paradigm requires protocols that differ from it in sequence, task or both.

Let us consider the stimulus sequence first. If a key problem with the oddball paradigm is the variability of target presentation, perhaps we could design a paradigm that had a constant temporal distance between successive targets. In this case, would the resulting periodic target presentation generate robust P300s given that the activation of the target stimulus would be predictable? The evidence from the literature suggests that in

order to obtain P300s of appreciable amplitude, it is necessary to create uncertainty in the subject [16], [31]. Given this, the answer to the previous question would seem to be in the negative. However, if the predictability in the timing of targets could be compensated for with more engaging stimuli and tasks, perhaps targets would still produce clear P300s.

This leads us to the second issue: the task. Typically tasks involve two components: a *perceptual component* (i.e., what analysis of the stimuli a subject needs to do upon their presentation) and an *action component* (i.e., what the subject is asked to do after the analysis of the stimuli). In the traditional oddball paradigm, the former consists of distinguishing targets from non-targets, while the latter is either counting or pressing a button. In principle we could change either or both components of the task while exploring alternative protocols. For example, one could change the action component of the task into one which required higher levels of cognitive processing. More radical changes to the protocol would become possible if one modified the perceptual component of the task first. Naturally, this would require changes to the *stimuli* themselves since, with fixed target and non-target stimuli, there can be very little freedom as to the perceptual task. Changes to the perceptual component of the task would then give us further freedom in choosing the action component. For example, one could add extra perceptual features to the stimuli thereby making it possible to define perceptual tasks based on these features. This would make it possible to create sequences of stimulus presentations where one feature is varied periodically while others are varied randomly.

We carried out a preliminary test of these ideas in [32], where we defined three protocols that use a periodic stimulation sequence in combination with various tasks. Only one of them proved capable of producing robust P300s. In this paper we will study this periodic protocol and explore whether it can compete with the standard oddball protocol within the context of a BCI mouse. We will also perform a comparison of variants of the protocol with corresponding variants of the oddball protocol to better understand the interactions between task and sequence.

V. Methodology

In the present study we performed two experiments where we tested a total of *five distinct protocols*.

Experiment 1 was carried out to perform a pairwise comparison between the best periodic protocol from [32] (which below we will call Periodic Random-Colour Name) against the state of the art: a traditional oddball paradigm (later called Random Single-Colour Count).

Experiment 2 was conducted to replicate Experiment 1 (bringing the total number of subjects that did the experiment to 16) as well as to better understand the combined effects of the changes in sequence (random vs periodic) and changes in task (counting vs colour naming) on the P300 component and more generally on ERPs, for a fixed choice of stimuli.

In all the protocols, participants were presented with a display containing eight circles (with each circle representing a direction of movement for the mouse cursor) that formed an imaginary circle at the centre of the display (see Fig. 1). During each trial, the circles in turn flashed (by changing rapidly from a baseline colour/intensity to a different colour/intensity and back), either randomly or periodically, depending on the protocol. At the beginning of the trial, participants were assigned a target circle and asked to perform a task every time it flashed.

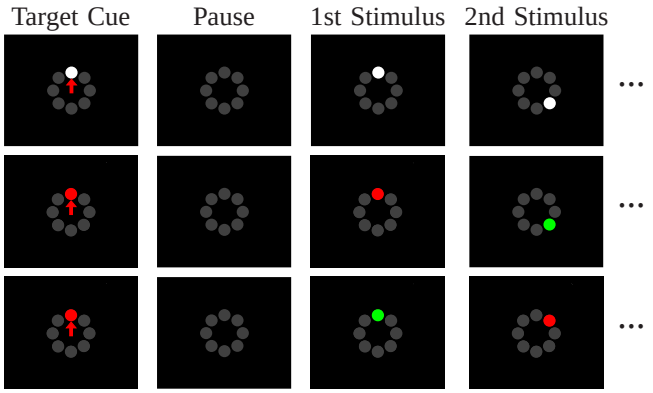

Fig. 1: The start of a sequence of stimulus presentations for the Random Single-Colour Count protocol (top), the Random Random-Colour Count and Random Random-Colour Name protocols (middle), and the Periodic Random-Colour Name and Periodic Random-Colour Count protocols (bottom).

Below we provide further details on the methodology.

A. Participants

In Experiment 1, data were collected from a total of 8 participants with an average age of 33. They all had normal or corrected to normal vision except for subject 5 who had strabismus with exotopia in the left eye.

In Experiment 2, data were collected from 8 subjects with an average age of 26.5. They all had normal or corrected to normal vision. None of the subjects of Experiment 2 had participated in the previous experiment.

B. Stimuli and procedure

As indicated above, in all our experiments we used displays showing eight circles (with each circle having a diameter of 1.5 cm and subtending the centre of the screen by $1.61°$) arranged around an imaginary circle at the centre of the display (see Fig. 1). The background was black and the stimuli were normally grey. However, during each trial, the circles in turn flashed. A flash of a circle consisted of a momentary change in its colour. As we will explain below, in some protocols the colour of the flashing stimuli was white, while in others the flashing colour could be either red or green.

The experiments included different combinations of five protocols: three based on random sequences (like the traditional oddball) and two periodic. Below we will describe the protocols in more detail. Before we do that, however, we would like to explain our naming scheme for the protocols. Protocols names have three parts: (1) the type of sequence (random or periodic), (2) the features of the target stimuli (single-colour or random-colour), and (3) the task to be carried out by subjects. For example, "Periodic Random-Colour Count" refers to a protocol with a periodic sequence of target presentations, where the colour of the targets is randomised, and where the task is to mentally count the number of targets.

With this notation in hand, we are now in a position to describe the protocols. These were as follows:

Random Single-Colour Count (RSCC): This protocol (effectively a standard oddball protocol) randomly highlighted one of the circles for the predefined SOA, then another randomly-chosen one, and so on, as illustrated in the top row of Fig. 1. Subjects were tasked with mentally counting the number of times the target stimulus was highlighted. This protocol can be considered to be the reference design since most P300-based BCI protocols utilise this type of stimulation and task combination, including the first P300-based BCI created by Farwell and Donchin [9].

Random Random-Colour Count (RRCC): This protocol was identical to the previous protocol except that the stimuli were highlighted randomly in red or in green (see Fig. 1, second row). Again, the subjects were tasked with counting the number of times the target stimulus was highlighted (irrespective of whether it was highlighted in red or in green).

Random Random-Colour Name (RRCN): This protocol is identical to the Random Random-Colour Count except for the task: here the subject had to mentally name the colour of the highlighted target.

Periodic Random-Colour Name (PRCN): This protocol was identical to the Random Random-Colour Name protocol, except that stimuli flashed in a periodic clockwise order (Fig. 1, bottom row). Each subject was asked to mentally name the colour of the target stimulus when this was highlighted.

Periodic Random-Colour Count (PRCC): This protocol differed from the Periodic Random Colour Name protocol in task only. In this protocol the subject had to count the number of times the target was highlighted.

Note that in all these protocols *the targets are defined by their position* in the ring of 8 circles displayed on the screen and not by any other feature.

The order in which the subjects carried out the protocols was randomised in all experiments. Protocols were blocked according to the task, so, for example, a subject would perform the protocols with a counting task first and then the protocols with a colour naming task or *vice versa*.

All protocols used an SOA of approximately 100 ms (more precisely 6×60^{-1} s, as permitted by the 60 Hz refresh rate of the LCD monitor used) with an ISI of 0 ms. This meant that all eight stimuli would be flashed within 800 ms. In the protocols where stimuli were flashed in a random order, the same stimulus could not flash twice in succession and also two adjacent stimuli were not allowed to flash within one SOA.

Each protocol was divided into runs, which we will call *direction epochs*, because in the context of our BCI mouse each circle represents a direction of motion for the mouse pointer. During a direction epoch, the participant was first greeted by a blank screen and after a predefined period the stimuli appeared near the centre of the screen. A red arrow then appeared for 1 second pointing to the target (see Fig. 1, leftmost column). Participants were instructed to focus their attention on the target stimulus and perform the task associated with the protocol. After 2 seconds the flashing of the stimuli started. This stopped after 20 to 24 trials, with a trial consisting of a single flashing of each of the 8 stimuli. After the direction epoch had been completed, and depending on the task that was assigned, the subject was requested to verbally communicate either the number of times the target stimulus flashed or the colour of the target during its final flash.

In the experiments with the Random Single-Colour Count and Periodic Random-Colour Name protocols, each participant carried out 16 direction epochs. Thus, each circle was a target in two different direction epochs. In the experiment involving the Periodic Random-Colour Count, Random Random-Colour Count and Random Random-Colour Name protocols, the subjects carried out 8 direction epochs and, so, the 8 possible directions were covered once.

Participants were seated comfortably at about 80 cm from an LCD screen. Data were collected from 64 electrode sites using a BioSemi ActiveTwo EEG system. The EEG channels were referenced to the mean of the electrodes placed on either earlobe. The data were initially sampled at 2048 Hz.

C. Classification

Classification of epochs into targets and non-targets was performed on the data collected with the Random Single-Colour Count and Periodic Random-Colour Name protocols (where we had 16 direction epochs per subject). Classification was carried out using a linear Support-Vector Machine (SVM) [33]. All 64 channels were used since it has been shown that channel selection, when used in conjunction with SVMs, only mildly improves results [34]. Furthermore, in order to obtain this marginal improvement the channel selection would have to be done on a subject by subject basis making it computationally expensive. The data were band-pass filtered between 0.15 and 30 Hz and initially downsampled to 128 Hz. Then, from each channel an 800 ms epoch was extracted which was further decimated to 32 Hz.

The classification results for each protocol were estimated using 8 fold cross-validations, where each fold was created by leaving out two direction epochs (selected randomly without replacement). The penalty parameter for the SVM was determined by an inner cross-validation loop for each of the outer cross-validation loops.

After training, the output of the SVM can be interpreted as a measure of how closely the feature vector associated with the stimulus matches the target. By applying a threshold to this measure, one can transform it into a binary decision regarding the presence of a target. Naturally, the higher the threshold is, the less likely a false positive error will be. However, unavoidably a higher specificity brings a lower sensitivity (i.e., an increased number of false negatives) with it.

D. Performance measures

The behaviour of our classifiers in relation to changes in their thresholds can be well represented using Receiver Operator Characteristics (ROC) curves. These are plots of the true-positive *vs* the false-positive rate for a binary classifier as its discrimination threshold is varied.

To assess the performance associated with the different protocols we used the Area Under the Curve (or AUC) of the ROC of the SVM output. The AUC is a well known summary for ROC curves that has been used widely in machine learning.

In addition, we used the information transfer rate and the classification accuracy to characterise our classifiers.

We compute the information transfer rate (ITR) as in [35]. Accuracy was computed based on the number of matches between the target direction for each direction epoch and the predicted directions given by the formula $\arg\max_D \left(\mathrm{SVM}_J(D)\right)$, where D is one of the 8 possible directions and $\mathrm{SVM}_J(D)$ is the sum of the raw outputs produced by the SVM in the J most recent flashes of the stimulus corresponding to direction D.

To compare the separation between targets and non-targets we also used the signed squared correlation coefficient, r^2, which represents the proportion of the total variance in a signal feature that is accounted for by the users intent [1], [36].

E. Artifact Removal

For the purposes of classification, for averaging and for the computation of r^2 values, the EEG signals for each subject were divided up into 800 ms epochs starting with each stimulus presentation. We applied to these epochs an artifact rejection procedure which involved computing the first ($Q_1(t)$) and third ($Q_3(t)$) quartiles of the voltages at each time step across all the epochs. The procedure then removed epochs where the signal was outside the range $\left[Q_1(t) - 1.5\left(Q_3(t) - Q_1(t)\right),\ Q_3(t) + 1.5\left(Q_3(t) - Q_1(t)\right)\right]$ for more than 10% of the samples in an epoch.

To maximise the use of valid information, whenever possible (e.g., for averaging) this procedure was applied on a channel by channel basis. So, for some channels (e.g., centro-parietal ones) we had slightly more epochs than for others (e.g., frontal ones) for averaging purposes. For classification and related statistics, however, where all channels of an epoch are required to form a feature vector, we applied the artifact rejection procedure described above to entire epochs, by requiring that the average (across all channel) proportion of samples outside our acceptance envelope did not exceed 10%.

VI. RESULTS

A. Comparing the Periodic Random-Colour Name protocol with the Random Single-Colour Count protocol

The objective of Experiment 1 (later partially replicated in Experiment 2) was a direct comparison between our best periodic protocol found in [32] (Periodic Random-Colour Name) with a traditional oddball protocol (Random Single-Colour Count), such as those used in most P300-based BCI, including our own previous work on BCI mice. We based this comparison on both classification performance and ERP waveforms.

1) Classification: Let us start by considering the AUCs obtained with the two protocols.

Table I provides an AUC-based performance comparison between the Random Single-Colour Count and Periodic Random-Colour Name protocols. Because these two protocols were included in the design of both Experiment 1 and Experiment 2, the table shows the data from both experiments (similarly, other analyses for the two protocols were performed on the whole set of data). As one can see, the oddball protocol gives very good results. However, the Periodic Random-Colour Name protocol provides not only better average results but also, and perhaps more importantly, superior AUCs for 15 subjects out of 16. Importantly, the *improvements* in AUC are highly negatively correlated with the AUCs of the Random Single-Colour Count protocol (Pearson's correlation coefficient -0.903), the greatest improvements occurring for the subjects (e.g., S05, S16 and S07) that performed worst with the oddball protocol.

TABLE I: Area under the curve (AUC) values for the Periodic Random-Colour Name and the Random Single-Colour Count protocols used in Experiments 1 and 2. Subjects S01 to S08 took part in Experiment 1, while subjects S09 to S16 took part in Experiment 2.

Subject	Random Single-Colour Count	Periodic Random-Colour Name
S01	0.918	0.941
S02	0.927	0.928
S03	0.852	0.907
S04	0.877	0.904
S05	0.674	0.878
S06	0.813	0.862
S07	0.797	0.912
S08	0.883	0.944
S09	0.933	0.935
S10	0.889	0.942
S11	0.806	0.883
S12	0.884	0.934
S13	0.963	0.929
S14	0.919	0.972
S15	0.824	0.847
S16	0.760	0.885
Mean±StdDev	0.857±0.073	0.913±0.033

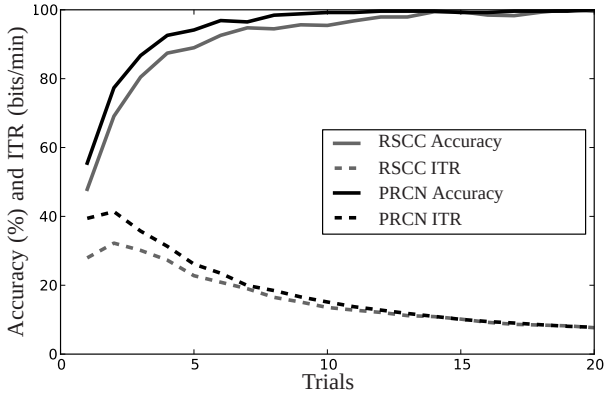

Fig. 2: Accuracy and information transfer rate (ITR) of the Periodic Random-Colour Name and the Random Single-Colour Count protocol obtained in Experiments 1 and 2.

Under the assumption of normality for both distributions and equal variance, a paired two tailed *t*-test reveals that the AUC results obtained with the two protocols are statistically significantly different (*p*-value = 0.0014). Unfortunately, the two distributions do not have equal variance and with AUC values being limited to the interval $[0, 1]$ the distributions should not, as per default, be considered normal. Therefore, we ran a more appropriate test — the paired Wilcoxon signed rank test — which also found that the results are highly statistically significant (*p*-value = 0.00043).

Let us now turn our attention to the accuracy of the protocols and their information transfer rate. These are reported in Fig. 2 as a function of the number of repetitions ('Trials') of a

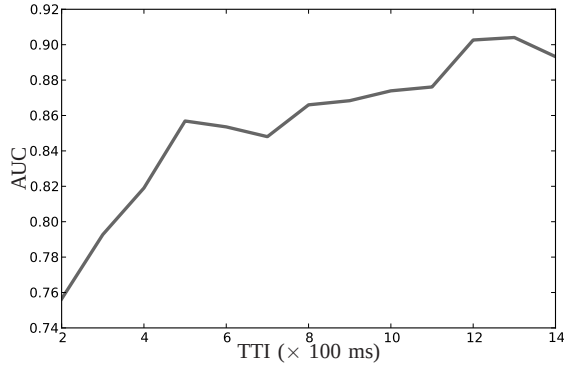

Fig. 3: AUC as a function of TTI in the Random Single-Colour Count protocol for the 16 subjects in Experiments 1 and 2.

sequence of stimuli before a decision is made by the classifier. We see that the Periodic Random-Colour Name protocol provides consistent improvements for both accuracy and ITR with respect to the Random Single-Colour Count protocol. For example, the highest average ITR for the Periodic Random-Colour Name protocol is 41 bits/min with an accuracy of 77%, while the highest average ITR for the Random Single-Colour Count is 32 bits/min at an accuracy of 69%. More realistically, the periodic protocol achieves an average ITR of 33 bits/min at an accuracy of 90% (effectively confirming the corresponding result in [32]), while the oddball protocol achieves an average ITR of 22 bits/min at the same accuracy. To test the significance of this ITR difference, we compared the results in a pairwise fashion. Both the paired t-test and the Wilcoxon rank sum test were significant (p-value = 0.0041 for the former and p-value = 0.0098 for the latter).

As mentioned in Section II, some detrimental effects of temporal proximity of targets on the characteristics of P300s have been reported in the literature. In an effort to understand the reasons for the relative performance differences shown by the two protocols studied in this section, we sought to determine whether a similar effect is present in our data with the Random Single-Colour Count protocol. To this end we divided up the positive (target) trials in the oddball data-set on the basis of the time elapsed since the previous target presentation (the TTI). We then computed ROC curves and AUCs for each of the TTIs. Fig. 3 shows a plot of the AUC as a function of the TTI. Results do indeed show a positive trend with the AUC improving as the TTI increases, reaching a level which is close to that of the Periodic Random-Colour Name protocol. Also, we see that relatively low AUC values are associated with the shortest TTIs. These contribute negatively to the overall performance of the Random Single-Colour Count protocol. Naturally, the periodic protocol's stimulus sequences have a constant TTI, and, so, no such effect can be observed in them.

2) ERP properties: Naturally, the analysis of ERP components elicited by the Random Single-Colour Count and the Periodic Random-Colour Name protocols was performed, as usual, by looking at averages over a large number of trials. Due to the large number of overlapping components in the oddball protocol and to the blurring effect of variable-latency ERPs, the ERP averages produced in the case of the oddball protocol are expected to be less similar to the "canonical" ERPs one finds in psychophysiology research reports. On the contrary, thanks to the constant TTI of the periodic protocol, the ERPs produced by it should be less affected by blurring and should be less atypical.

Average ERPs for the two protocols for three different channels and for targets and non-targets are reported in the first two columns of Fig. 4. From the averages it is easy to discern that the traditional oddball protocol produces on average smaller P300s than the periodic protocol. As expected, the ERP averages of the periodic protocol are considerably clearer when compared to the oddball ERP averages.

The TTIs in the oddball protocol we used are very short compared with the TTIs used in previous BCI work and much shorter than those typically used in psychophysiology. It has frequently been reported that short TTIs lead to reduced P300 amplitudes. Also, phenomena such as attentional blink and repetition blindness, which have been shown to detrimentally affect the P300 amplitude, are much more likely to occur at such short TTIs. We believe these are the reasons for the small P300 obtained with the Random Single-Colour Count protocol.

These results suggest that part of the performance advantage of the Periodic Random-Colour Name protocol is due to sequence effects. As we will see in Section VI-B, the chosen task is another feature contributing to its success.

3) Target vs non-target discrimination: In [36], [37] the importance of the early components in the classification of ERP data was highlighted. While such dependency is to be expected in visual protocols, the finding may be important for the clinical application of BCIs, because users may have limited visual acuity on which early potentials rely. With this in mind we sought to determine which components are particularly important for distinguishing targets from non-targets in the Random Single-Colour Count and Periodic Random-Colour Name protocols.

We started from an analysis of the signed r^2 function defined in [1] for the two protocols. The last column of Fig. 4 shows how, in the two protocols tested, the signed squared correlation coefficient varies across time, channel and protocol. It is apparent that, in the channels chosen, the periodic protocol offers a much greater discrimination between targets and non-targets than the classic oddball protocol. It is also clear that the most discriminative information between targets and non-targets is in the first part of an epoch for the oddball protocol, whereas for the periodic protocol the most discriminative information is in the middle of the epoch.

We also checked which of these two protocols' classification performance relies the most on the early components. To this end we carried out classification using epochs which discarded ever increasing time intervals after the onset of the stimuli. When the start time was 0 ms after event onset, the classification epoch lasted 800 ms; when the start was at 50 ms after event onset, the classification epoch lasted 750 ms; and so on. The results of this procedure can be seen in Fig. 5. It is easy to discern that not only does the Periodic Random-Colour Name protocol provide better results than the Random Single-Colour Count protocol but it is also least affected by the removal of the early components. For example, when the

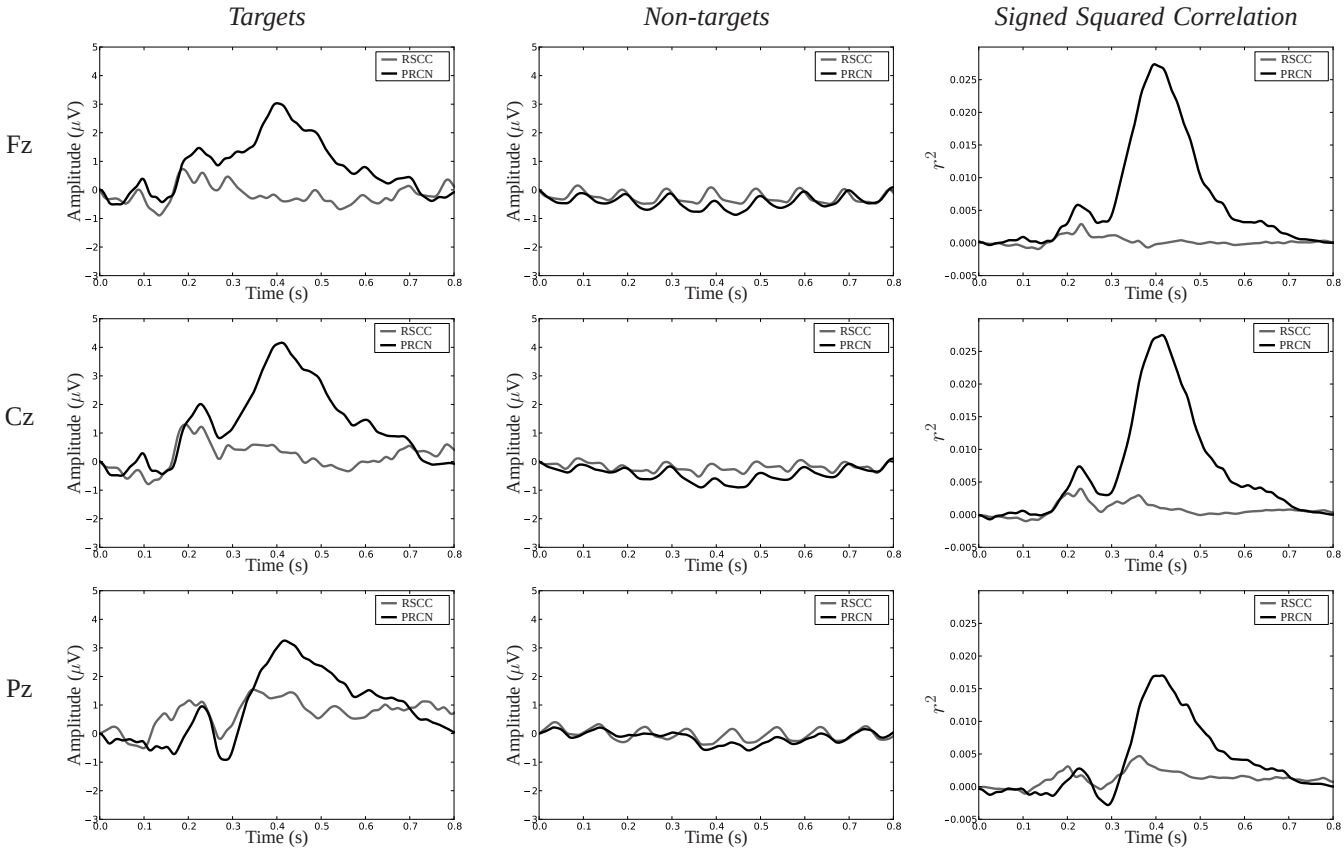

Fig. 4: ERP averages of the targets (first column) and non-targets (second column), and signed squared correlation coefficients (third column) for of the Periodic Random-Colour Name (PRCN) and Random Single-Colour Count (RSCC) protocols obtained in Experiments 1 and 2 for channels Fz, Cz and Pz. The thinner lines in panels represent the standard error of the mean.

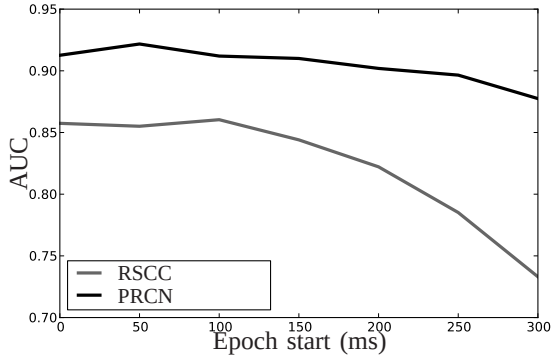

Fig. 5: AUCs of the Random Single-Colour Count (RSCC) and Periodic Random-Colour Name (PRCN) protocols as the start time (after event onset) of classification epochs is varied.

classification epoch starts 300 ms after event onset, the drop in AUC is 0.124 (14.5%) for the Random Single-Colour Count protocol but only 0.0347 (3.8%) for the Periodic Random-Colour Name protocol.

B. Effects of sequence and task on P300s

In Section VI-A we have seen that the Periodic Random-Colour Name protocol produces clearer P300s in the presence of targets than the traditional Random Single-Colour Count protocol. However, since these two protocols differ by sequence, stimuli and task, it is reasonable to ask how much each difference contributed to this result. Experiment 2 was designed to answer this very question, in addition to providing further support for the results of Experiment 1. In particular, we compared the differences in the P300 components elicited by four protocols: Periodic Random-Colour Name, Periodic Random-Colour Count, Random Random-Colour Name and Random Random-Colour Count. These protocol differ in the task (counting vs naming) and sequence (random vs periodic) employed, while they all share the same stimuli.

The average ERPs for the four protocols for three midline channels and for targets and non-targets are reported in the first two columns of Fig. 6. The target averages unequivocally show a greater P300 amplitude for the Periodic Random-Colour Name protocol in all the channels shown. However, we also see a much improved P300 amplitude in the Random Random-Colour Name protocol when compared to the standard oddball (Random Random-Colour Count). Overall the Periodic Random-Colour Name protocol tends to produce ERPs that have a more central focus (note how the P300 component for such a protocol has the largest amplitude in Cz) whereas both oddball protocols tend to have a parietal focus. Note that the Periodic Random-Colour Count produces no discernible P300 component whereas in the Periodic Random-Colour Name (where the task is changed to naming of the highlighted colour) the amplitude is the largest of all.

In summary, from a qualitative observation of average ERPs one can deduce that the task carried out by subjects is a very

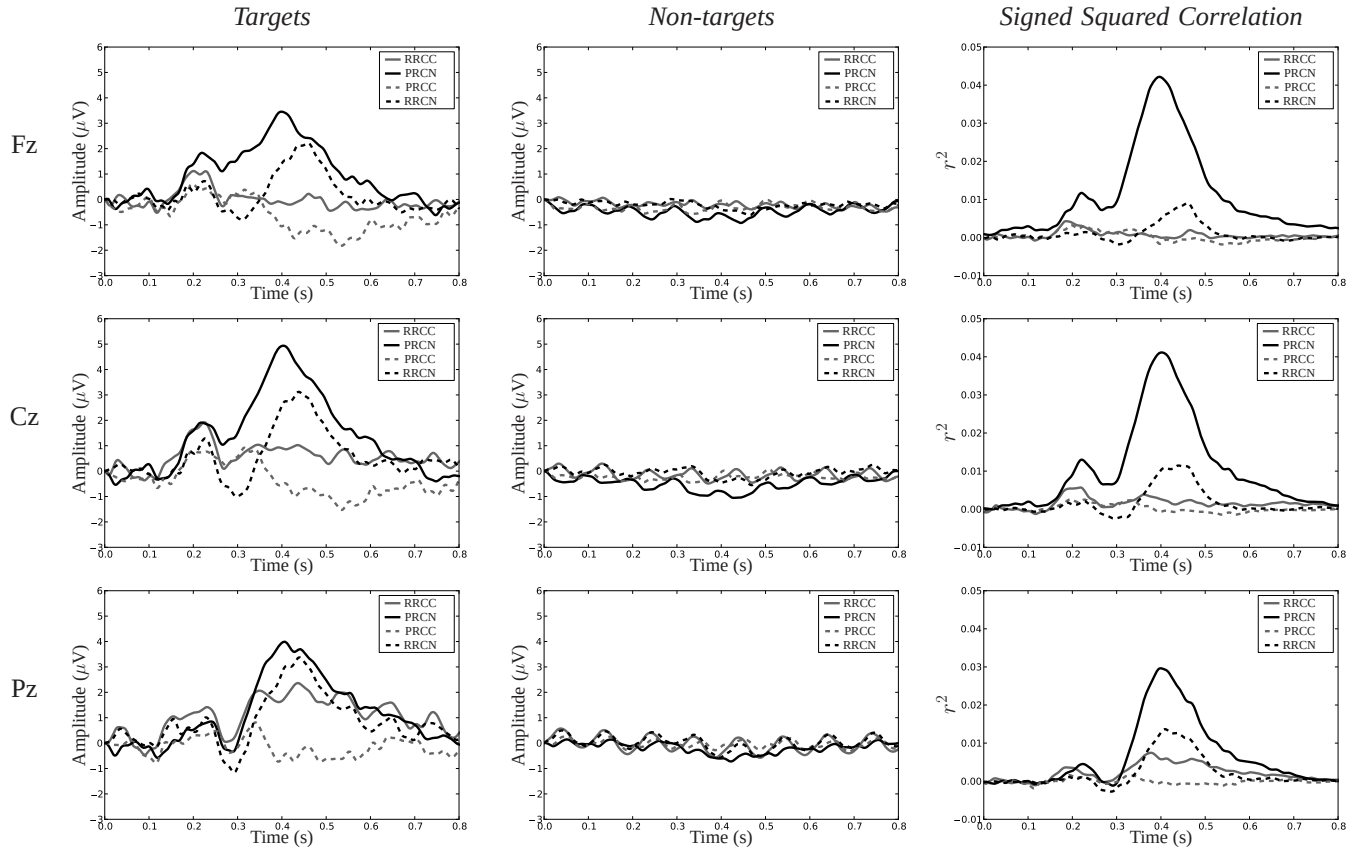

Fig. 6: ERP averages for targets (first column) and non-targets (second column), and signed squared correlation coefficients (third column) of the Periodic Random-Colour Count (PRCC), Periodic Random-Colour Name (PRCN), Random Random-Colour Count (RRCC) and Random Random-Colour Name (RRCN) protocols tested in Experiment 2, for channels Fz, Cz and Pz. The thinner lines in panels represent the standard error of the mean.

important factor for the elicitation of robust P300 components, but also that periodic sequences offer considerable benefits when coupled with an appropriate task.

These observations are confirmed by an analysis of the signed squared correlation coefficients [1]. The signed r^2 values for three midline channels for the four protocols can be seen in the last column of Fig. 6. From these it is evident that the Periodic Random-Colour Name protocol produces the clearest separation between targets and non-targets and that this is focused around the time frame most commonly associated with the P300 component. The second best in terms of separability of targets and non-targets is the Random Random-Colour Name protocol, which certainly deserves further future investigation. Third best is the Random Random-Colour Count protocol, which shows an r^2 profile not much different from that of the traditional Random Single-Colour Count (oddball) protocol reported in the last columns of Fig. 4. Finally, the Periodic Random-Colour Count protocol is confirmed to be unreliable at discriminating targets from non-targets.

In order to quantitatively assess the P300 component differences, the peak amplitudes of the voltages recorded in each epoch in the 250–500 ms window were extracted. Peak voltages were averaged on a protocol-by-protocol and subject-by-subject basis and the resulting averages were statistically analysed. A test for normality was carried out and it was found that, for a number of protocols and midline electrodes, amplitudes failed to conform to normality according to the Shapiro-Wilks normality test ($p < 0.05$). We, therefore, chose to carry out a Friedman test (a non-parametric statistical test similar to ANOVA) with amplitude being the dependent variable and the protocol being the independent variable. For Fz and Cz, the protocol was found to affect the P300 amplitude in the 250 to 500 ms window (χ^2=12.9, df=3, p-value < 0.0049 and χ^2=16.8, df=3, p-value < 0.0008, respectively). For Pz the results approached significance (p-value = 0.066).

In order to determine which protocol produced the largest P300 peak, a multiple comparisons Wilcoxon rank sum test was carried out on these amplitude datasets, using the Holm method to adjust for the multiple comparisons. The Periodic Random-Colour Name protocol was found to be significantly different from all other protocols at Cz and significantly different from Periodic Random-Colour Count and Random Random-Colour Count at Fz (at the standard 5% significance level).

VII. Discussion

The purpose of this study was to find an alternative to the oddball paradigm typically used in BCI, in order to overcome some of the drawbacks that are associated with it. In particular, the experiments described above investigated the effects of presenting targets at regular intervals (instead of using the random intervals typical of the oddball protocols). We found that, when combined with a more engaging task (compared

with the typical counting task used in BCI), periodic target presentations are a valid alternative to oddball protocols.

Below we discuss the main findings of our two experiments. Given the practical objectives of this paper, in doing so we will mainly concentrate on classification performance and P300 amplitudes, occasionally offering possible psychophysiological explanations for the observed differences.

A. Periodic Random-Colour Name vs *Random Single-Colour Count*

In Experiments 1 and 2 we studied the Periodic Random-Colour Name protocol by directly comparing it against a standard oddball protocol (our Random Single-Colour Count). We found that the former statistically significantly outperforms the latter in terms of AUC scores, providing also a marked improvement in accuracies and information transfer rates. Furthermore, we found that in the periodic protocol the desired ERP components were clearer and of greater amplitude. An analysis of the signed r^2 correlation coefficients revealed that maximum discriminability between targets and non-targets is achieved at different times in the two protocols: early in the Random Single-Colour Count protocol, much later in the Periodic Random-Colour Name protocol. Indeed, while we found that there are significant detrimental effects of removing the early components from the classification epochs (as firstly reported in [36]), these effects were very modest in the periodic protocol.

How does the periodic protocol achieve all this? This is partly due to the sequence of stimulation used: by providing a constant, relatively long interval between successive target flashes, the protocol prevents cognitive phenomena such as attentional blink and repetition blindness, and minimises target-ERP overlap, thus producing clearer ERPs. Some of these phenomena are responsible for the drop in classification accuracy at short TTIs observed in matrix spellers [11] which we showed to affect also the Random Single-Colour Count protocol (see Fig. 3). This cannot happen in periodic protocols by design. They thus avoid many of the identified pitfalls associated with fast P300-based oddball protocols.

A second component of the success of the Periodic Random-Colour Name protocol is the task. This has been changed from counting to colour naming. Subjects reported the naming of target colours being a harder task than the counting required in the oddball paradigm. It is difficult to speculate as to precisely which mental processes have been affected by the task change. However, based on the P300 literature, the increased amplitude in the P300s elicited by the Periodic Random-Colour Name protocol (see Section VI-A2) suggests that the colour discrimination task may require more focused attention.

B. Importance of Task and Sequence

In Experiment 2 the importance of the task and sequence employed was studied by comparing four protocols, all using identical stimuli.

Comparing the Periodic Random-Colour Name protocol with the Periodic Random-Colour Count protocol, which are identical except for the task employed, we see very clear (and statistically significant) differences in the P300 amplitude. A similar (although less extreme) pattern of differences is observed when comparing the Random Random-Colour Count and the Random Random-Colour Name protocols, in channels Fz and Cz and to a lesser extent Pz. It is thus reasonable to infer that the colour naming task (enabled by changes in the features of stimuli) can enhance the amplitude of P300 components, perhaps to an even greater extent than changes to the nature (random vs periodic) of the stimulation sequence. However, task and sequence do interact. For example, the Periodic Random-Colour Count protocol produces no discernible P300 peak whereas the Periodic Random-Colour Name protocol produces the largest P300s of all protocols tested.

Finally, if we compare the ERP averages and the r^2 plots produced by the Random Random-Colour Count protocol in Experiment 2 (Fig. 6) with the same plots for the Random Single-Colour Count protocol in Experiment 1 (Fig. 4), we see that they are quite similar. The task and sequence of these protocols are identical, the only difference is the stimulus flashes (random-colour flashes vs single-colour flashes). This seems to suggest that the differences in the colour of flashes, or the fact that in one case the colour is always the same while in the other it changes randomly, have little effect on the P300 component in these two protocols.

VIII. Conclusion and future work

We have described a novel class of protocols for the elicitation of P300s based on discrimination of periodic targets. In testing the protocols within the context of a BCI mouse, we have shown that the best of our periodic protocols can significantly increase the average information transfer rate achieved over the traditional oddball protocol. This is accomplished by a change in the task (from mentally counting targets to mentally naming their colour) as well as ensuring a constant time interval between successive target intensifications. This avoids psychophysiological phenomena such as attentional blink and repetition blindness, eliminates variations in P300 amplitude and latency due to TTI effects and also mitigates overlap effects. Furthermore, the increase in performance is mainly due to the increase in amplitude of the later components, limiting this P300-protocol's reliance on early components.

Naturally, many questions still remain to be answered in future research. For example, would increasing the number of target colours in the Periodic Random-Colour Name protocol bring further benefits? Similarly, it is unclear at this stage if our choice of equiprobable colours is optimal. Also, we don't know if the particular feature chosen to be randomly varied is optimal. For example, the identification of colours could be changed for identification of motion. Finally, while we used the same TTI and number of colours for all subjects, performance might be further boosted by a per-subject choice of such parameters.

This study concentrated on off-line testing. Since we used an SOA of 100 ms, using the same vectorial-integration strategy as in our previous work, our periodic protocol would produce one analogue mouse-pointer movement every 100 ms if used online. This, in conjunction with the significantly increased

ITR offered by the protocol, is hoped to provide a good starting point for an robust online system. On-line tests with the protocol will be another topic for future research.

The methodology used here to generate distinct ERPs for a BCI mouse could also be applied to a number of other ERP-based BCI protocols such as the Farwell and Donchin matrix speller [9] or the Hex-O-Spell [37] within the domain of visual evoked potentials, as well as within other modalities.

Finally, we should note that in our exploration of alternative protocols, we also tested a novel protocol (Random Random-Colour Name) which is not periodic but, thanks to a task change, still elicits stronger P300s than a traditional oddball protocol. While inferior to our Periodic Random-Colour Name protocol, we believe this protocol has shown enough promise to deserve further future investigation.

ACKNOWLEDGEMENTS

We thank the UK Engineering and Physical Sciences Research Council (grant EP/F033818/1) for financial support. We would also like to thank the reviewers for their help in improving the manuscript.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–91, Jun. 2002.

[2] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. E. Frye, C. K. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns," *Clinical Neurophysiology*, vol. 121, no. 7, pp. 1109–1120, Jul. 2010.

[3] G. Schalk, "Brain-computer symbiosis," *Journal of Neural Engineering*, vol. 5, no. 1, pp. P1–P15, Mar. 2008.

[4] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface systems: progress and prospects." *Expert review of medical devices*, vol. 4, no. 4, pp. 463–474, Jul 2007.

[5] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 479–483, Mar. 2006.

[6] J. del R Millán, R. Rupp, G. R. Müller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kübler, R. Leeb, C. Neuper, K. Mller, and D. Mattia, "Combining Brain-Computer interfaces and assistive technologies: State-of-the-Art and challenges," *Frontiers in neuroscience*, vol. 4, no. 161, Sep. 2010.

[7] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain-computer interface (BCI)?" *Neuroscience letters*, vol. 462, no. 1, pp. 94–98, Oct 2009.

[8] F. Nijboer, N. Birbaumer, and A. Kübler, "The influence of psychological state and motivation on BrainComputer interface performance in patients with amyotrophic lateral sclerosis a longitudinal study," *Frontiers in neuroscience*, vol. 4, no. 55, Jul. 2010.

[9] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials." *Electroencephalography and clinical neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec 1988.

[10] S. M. M. Martens, N. J. Hill, J. Farquhar, and B. Schölkopf, "Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential." *Journal of neural engineering*, vol. 6, no. 2, p. 026003, Apr 2009.

[11] L. Citi, R. Poli, and C. Cinel, "Documenting, modelling and exploiting P300 amplitude changes due to variable target delays in donchin's speller," *Journal of Neural Engineering*, vol. 7, no. 5, p. 056006, 2010.

[12] M. Salvaris and F. Sepulveda, "Perceptual errors in the Farwell and Donchin matrix speller," in *4th International IEEE EMBS Conference on Neural Engineering*, Antalya, Turkey, 2009, pp. 275 – 278.

[13] J. Polich, "Updating P300: an integrative theory of P3a and P3b." *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, Oct 2007.

[14] A.-M. Brouwer and J. B. F. van Erp, "A tactile P300 Brain-Computer Interface." *Frontiers in neuroscience*, vol. 4, p. 19, 2010.

[15] A. Furdea, S. Halder, D. J. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces." *Psychophysiology*, vol. 46, no. 3, pp. 617–625, May 2009.

[16] W. S. Pritchard, "Psychophysiology of P300," *Psychological Bulletin*, vol. 89, no. 3, pp. 506–540, May 1981.

[17] M. Fabiani, G. Gratton, D. Karis, and E. Donchin, "Definition, identification, and reliability of the P300 component of the event-related brain potential." in *Advances in Psychophysiology*, P. K. Ackles, J. R. Jennings, and M. G. H. Coles, Eds. New York: JAI Press, 1987, vol. 2.

[18] 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.

[19] D. L. Woods and E. Courchesne, "The recovery functions of auditory event-related potentials during split-second discriminations," *Electroencephalography and Clinical Neurophysiology*, vol. 65, no. 4, pp. 304–15, Jul. 1986.

[20] 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.

[21] J. E. Raymond, K. L. Shapiro, and K. M. Arnell, "Temporary suppression of visual processing in a RSVP task: An attentional blink?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 18, no. 3, pp. 849–860, 1992.

[22] N. G. Kanwisher, "Repetition blindness: type recognition without token individuation," *Cognition*, vol. 27, no. 2, pp. 117–43, Nov. 1987.

[23] 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.

[24] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proceedings of the National Academy of Sciences*, vol. 101, no. 51, pp. 17 849–17 854, 2004.

[25] J. Donoghue, "Connecting cortex to machines: recent advances in brain interfaces," *Nature Neuroscience*, vol. 5, pp. 1085–1088, 2002.

[26] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, and S. Giove, "User adaptive BCIs: SSVEP and P300 based interfaces," *PsychNology Journal*, vol. 1, no. 4, pp. 331–354, 2003.

[27] J. B. Polikoff, H. T. Bunnell, and W. J. Borkowski Jr, "Toward a P300-based computer interface," in *Rehabilitation Engineering and Assistive Technology Society of North America (RESNA'95)*. Arlington, Va: Resna Press, 1995, pp. 178–180.

[28] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, "P300-based BCI mouse with genetically-optimized analogue control," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 16, no. 1, pp. 51–61, Feb. 2008.

[29] M. Salvaris, C. Cinel, R. Poli, L. Citi, and F. Sepulveda, "Exploring multiple protocols for a brain-computer interface mouse," in *Proceedings of 32nd IEEE EMBS Conference*, Buenos Aires, September 2010, pp. 4189–4192.

[30] R. Poli, L. Citi, F. Sepulveda, and C. Cinel, "Analogue evolutionary brain computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 4, no. 4, pp. 27–31, November 2009.

[31] C. C. Duncan-Johnson and E. Donchin, "On quantifying surprise: the variation of event-related potentials with subjective probability," *Psychophysiology*, vol. 14, no. 5, pp. 456–67, Sep. 1977.

[32] M. Salvaris, C. Cinel, and R. Poli, "Novel sequential protocols for a ERP based BCI mouse," in *5th International IEEE EMBS Neural Engineering Conference*. IEEE Press, 2011, pp. 352–355.

[33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[34] A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–1154, Mar 2008.

[35] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, and T. Vaughan, "Brain-computer interface technology: a review of the first international meeting," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 2, pp. 164–173, 2000.

[36] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk, "Does the P300 speller depend on eye gaze?" *Journal of Neural Engineering*, vol. 7, no. 5, p. 056013, 2010.

[37] M. S. Treder and B. Blankertz, "(C)overt attention and visual speller design in an ERP-based Brain-Computer Interface," *Behavioral and brain functions*, vol. 6, p. 28, 2010.