Brain–Computer Interface (BCI): Is It Strictly Necessary to Use Random Sequences in Visual Spellers?

Manson Cheuk-Man Fong, James William Minett, Thierry Blu, and William Shi-Yuan Wang
The Chinese University of Hong Kong
Hong Kong, China
cmfong@ee.cuhk.edu.hk, jminett@ee.cuhk.edu.hk, tblu@ee.cuhk.edu.hk, wsywang@ee.cuhk.edu.hk

ABSTRACT
The P300 speller is a standard paradigm for brain–computer interfacing (BCI) based on electroencephalography (EEG). It exploits the fact that the user's selective attention to a target stimulus among a random sequence of stimuli enhances the magnitude of the P300 evoked potential. The present study questions the necessity of using random sequences of stimulation. In two types of experimental runs, subjects attended to a target stimulus while the stimuli, four in total, were each intensified twelve times, in either random order or deterministic order. The 32-channel EEG data were analyzed offline using linear discriminant analysis (LDA). Similar classification accuracies of 95.3% and 93.2% were obtained for the random and deterministic runs, respectively, using the data associated with 3 sequences of stimulation. Furthermore, using a montage of 5 posterior electrodes, the two paradigms attained identical accuracy of 92.4%. These results suggest that: (a) the use of random sequences is not necessary for effective BCI performance; and (b) deterministic sequences can be used in some BCI speller applications.

Author Keywords
Brain–computer interface (BCI); P300 speller; ERP-based visual speller; electroencephalography; oddball paradigm; linear discriminant analysis (LDA).

ACM Classification Keywords
H.5.m. **[Information interfaces and presentation (e.g., HCI)]:** Miscellaneous; H.5.2. **[User Interfaces]:** User-centered design.

INTRODUCTION

A brain–computer interface (BCI) [37] is a device which translates brain signals into commands that control applications. Currently, scalp electroencephalography (EEG) is the predominant technology for realizing non-invasive BCI systems, not only because of the portability of EEG systems, but also due to the continual progress being made in eliciting different types of prominent control signals in a range of EEG-based BCI paradigms (for a review, see [5]).

The P300 speller, the focus of the present study, represents one of the most successful EEG-based BCI paradigms. Primarily, it exploits the fact that selective visual attention to a target stimulus can enhance the average electrical potential elicited, referred to as the event-related potential (ERP [14]), compared to that elicited by other stimuli that are either unattended or deliberately ignored. More precisely, it is known that when subjects respond mentally to rare, target events that are *randomly* interspersed among frequent, non-target events (the so-called *oddball paradigm*), the target events (the *oddballs*) tend to elicit a stronger ERP component, termed P300, than the non-target events [10, 33]. The component is so named as it is a positive deflection in electrical potential that peaks around 300 ms post-stimulus, and is usually more prominent at central-parietal sites. In the original formulation of the P300 speller by Farwell and Donchin in 1988 [11], a 6×6 matrix of symbols, analogous to a virtual keyboard, is displayed on-screen to the user as choices—henceforth, we shall refer to this speller as the *matrix speller* to distinguish it from other variants. To select a choice, the user is required to attend to that choice, while the rows and columns of the matrix are intensified successively in random order. Intensification of a row or column that contains the intended choice, by virtue of its rare and random occurrence, constitutes an oddball, so eliciting a stronger P300 than that elicited by intensification of other rows and columns. This difference forms the basis for the intended choice to be identified. Recent works confirmed the validity of the matrix speller as a practical BCI, through online assessment with both able-bodied [8, 13] and pathological groups [24, 30]. Numerous optimization schemes have been explored in order to increase the communication rate [21], from adjusting the various system parameters associated with the speller (e.g., matrix size [2, 29], inter-stimulus interval (ISI) [29], method of intensification [16, 34, 35], stimulus type [15, 22], etc.) to improving the methods of signal processing and classification [6, 15, 20]. This collective effort has allowed the mean input rate for the matrix speller to increase from about 12 bits/min, as estimated offline in Farwell and Donchin's original study, to about 23 bits/min, as determined online in a recent study [35]. In another line of study, other visual paradigms in which different

APCHI '12, August 28–31, 2012, Matsue-city, Shimane, Japan.

geometric configurations are used to arrange the choices have been tested. These included a four-choice paradigm in which four stimuli ('YES', 'NO', 'PASS', 'END') were successively presented centrally [28]; a 2D cursor control system, wherein four arrows were arranged on the periphery of a square [25]; and a two-level speller, referred to as Hex-o-Spell, in which six discs containing either one or multiple symbols were arranged in the corners of a hexagon [6, 36].

The present study has two primary motives, one theoretical and one practical. On the theoretical side, we examine one of the most important, but often understated, working assumptions of the P300 speller paradigm, namely, that the performance of the BCI is the best when the choices are intensified in a random order. Specifically, we compare the classification accuracies obtained for two paradigms—a random paradigm, in which the choices are successively intensified in random sequence; and a deterministic paradigm, in which the choices are successively intensified in fixed sequence (i.e., in the same order repeatedly). The use of random intensification sequences has two clear advantages: (a) rare targets in a random sequence are known to elicit a larger P300 than non-targets for almost every subject [10], providing a general basis for discriminating between the two types of stimuli; (b) although its amplitude is modulated by attention [26], the P300 can be elicited without active attention by visual oddballs, hence can be considered an automatic response [18, 31]. It is unclear, however, to what extent the utility of the P300 speller hinges on the use of random sequences for choice intensification. More precisely, is selective attention to target choices alone sufficient for the choices to be identified, regardless of whether the sequence of intensification is random or deterministic? Given that the P300 amplitude elicited by targets increases as the subjective expectancy towards the target occurrences decreases [9, 32], it is tempting to conclude that, since the target occurrences in a deterministic sequence are not only predictable, but also entirely known, the degree of subjective expectancy is maximal, which would in turn imply that the P300 amplitude elicited by the targets should be smallest under such circumstances. However, the findings regarding subjective expectancy have been obtained using random sequences, and there is no guarantee that they can be generalized to the case of deterministic sequences. In fact, a recent study reported that targets following a predictive sequence, i.e., one which allowed the subject to precisely learn of the exact times of occurrences of the targets, elicited a P300 component whose magnitude was not significantly different from that elicited by targets following a random, non-predictive sequence [12]. Although that study was not conducted within a BCI context, its findings clearly cast doubts on the assumption that the P300 amplitude elicited by the targets in deterministic sequences will be reduced. Adding to these uncertainties are the recent findings that ERP components other than P300 can also be used to discriminate targets and non-targets. In particular, for the matrix speller, these components include P2 [1, 19] and N2 [1, 20, 22, 36]. For paradigms in which the choices are arranged differently, these include a range of both earlier and later components, such as P1, N1, P2, N2 and N3, in Hex-o-spell [6, 36], and again, N2, when the stimuli are centrally presented one-by-one [3]. Given that the modulation of such a wide range of components might not all hinge on the use of random sequences, it is clearly a valid question as to whether some of these components will be modulated in the case of the deterministic paradigm.

Apart from the theoretical interests above, there are practical concerns over the use of random sequences, especially from the point of view of user-centered design. Thus far, there have been a number of unresolved weaknesses associated with the matrix speller. To attain high accuracy, it is necessary to visually intensify all symbols for multiple times, and then average the resultant responses to produce a robust signal for classification. The sustained visual stimulation of rows and columns, however, can cause discomfort [16, 17], and discourage long use. Intensifying symbols one at a time can reduce the discomfort, but this results in a reduction in the communication rate [13]. Intuitively, the random stimulation might contribute to such discomfort, since the randomness requires the user to maintain a high-level of concentration throughout the symbol-selection process, which may last up to 10 seconds. In contrast, if the sequence of stimulation is deterministic, users can vary their level of concentration accordingly during a selection, e.g., to relax between target stimulations. Thus, if the deterministic paradigm can achieve a similar accuracy as the random paradigm, it will provide the target users with an alternative system that is potentially more user-friendly.

In the present study, as a starting point for studying the differences between using random and deterministic sequences within a BCI context, a simple four-choice paradigm similar to [25] in terms of geometrical arrangement was used. The results are of direct relevance to the Hex-o-Spell paradigm, whose accuracy has been demonstrated recently to be higher than that of the matrix speller [36].

EXPERIMENTAL METHODS

Participants

Six healthy Chinese subjects (S1–S6; 3 male and 3 female), aged 22–37 (mean 26.3), took part in the experiment as volunteers. Three subjects (S2, S4 and S5) had not previously sat an EEG experiment; the other three subjects (S1, S3 & S6) all had prior experiences with the P300 speller paradigm using random sequences, but not using deterministic sequences. Informed consent was obtained from each subject.

Data Acquisition

Each subject was seated in a quiet room 60 cm in front of a 19' LCD monitor (resolution: 1280 × 1024) that displayed the experimental stimuli as images. Each image was composed of four Chinese characters, arranged geometrically according to their respective meanings (left, up, right and down). Figure 1 exemplifies an intensification of the stimulus "left". Each character is about 80 pixels wide (visual angle: 2.0°), with its center being 140 pixels (visual angle: 3.6°) from the center of the screen.

EEG data were acquired at a sampling rate of 1024 Hz using a 32-channel ActiveTwo EEG system (BioSemi B. V., Amsterdam, The Netherlands). Figure 2 shows the positionings of the 32 pin-type, Ag/AgCl active electrodes. Two flat-type electrodes were attached over the left and right mastoids for offline re-referencing. Two additional electrodes, common mode sense (CMS) and driven right leg (DRL), positioned at C1 and C2 respectively, were used to complete a feedback loop, such that the average electrical potential over all electrodes was driven to as close a voltage as possible to the amplifier reference voltage [4]. Stimulus presentation was controlled using the software E-Prime 2.0 (Psychology Software Tools, Inc.).

Experimental Procedure

Each subject completed two sets of four blocks for each of the two paradigms: *random paradigm* (**RP**) and *deterministic paradigm* (**DP**). Each block consisted of 24 experimental runs. The details of a run are as follows:

1. At the beginning of each run, subjects were shown a prompt ("Please attend to: X") on-screen for 2 seconds, where X represented one of the four possible target choices (left, up, right and down), displayed in Chinese. Each choice was selected as the target choice exactly once every 4 runs. Thus, within a block, each choice was selected as target 6 times.
2. An image showing the four choices in fully-lit color (as exemplified for "left" in Figure 1) was then presented for 2 seconds, after which all four choices were displayed in dimly-lit color for a further 2 seconds.
3. The choices were then intensified in succession for 12 sequences, each consisting of 4 intensifications. Each intensification consisted of the display of two images: the first image, in which one of the choices was fully-lit (while the others were dimly-lit), was displayed for 133 ms; the second image, in which all choices were dimly-lit, was displayed for 33 ms. The inter-stimulus interval (ISI) was therefore 166 ms. For convenience, a sequence of intensifications is referred to as a *trial* and a single intensification a *sub-trial*.
4. For the **RP** runs, the choices were intensified in pseudorandom sequences, meaning that the following criteria were satisfied: (a) each choice was intensified exactly once in a trial; and (b) across trials, none of the choices was intensified twice consecutively. For the **DP** runs, the choices were intensified in fixed sequence, i.e., in the same order for all twelve trials. The order was selected to be "left, up, right & down", i.e., clockwise starting from the Chinese character for "left".
5. Following the standard procedure for the matrix speller, subjects were asked to respond to every intensification of the target character by maintaining a mental count of the number of times the target character had been intensified within the run.
6. The duration of each run was approximately 14 seconds, and there was a 3-second rest period between runs.

Figure 1. An intensification of the choice "left", printed in Chinese.

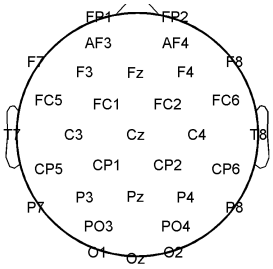

Figure 2. 32-channel ActiveTwo EEG system.

Each block lasted for approximately 7 minutes, and there was an optional pause after 12 runs. The total time for one recording session was about 1 hour.

The order of the two sets of blocks was counterbalanced across subjects, to control for any potential training effect. That is, for half of the subjects, the 4 **RP** blocks were administered first, while for the other half, the 4 **DP** blocks were administered first.

DATA ANALYSIS

To compare the projected performance attainable by the random and deterministic paradigms, the EEG data acquired for their corresponding runs were analyzed using the same procedure, to be detailed in this section. In brief, for both paradigms, four-fold cross-validation was performed for each subject. Specifically, for a given fold of analysis, the runs from three of the four blocks were used to train a classifier, which was subsequently applied to classify the target choices in the remaining block; the analysis was repeated four times such that each block served once as the testing data.

Preprocessing

Six preprocessing steps, all handled using EEGLAB [7], were carried out in the order stated below.

1. *Filtering*. The cutoff frequencies of the band-pass filter were set to 1.0 Hz and 40 Hz.
2. *Segmentation*. The band-passed data were segmented into sub-trials of duration 833 ms. Each sub-trial started at 333 ms pre-stimulus, and lasted until 500 ms post-stimulus. For each block, a total of 1152 sub-trials were extracted, corresponding to 288 and 864 sub-trials that were time-locked to a target and a non-target stimulus, respectively. These sub-trials are referred to as target sub-trials and non-target sub-trials.
3. *Re-referencing*. Within each sub-trial, the time-series associated with the two mastoid electrodes were averaged, and subtracted from that associated with each of the 32 main electrodes.
4. *Baseline correction*. For every sub-trial, a baseline potential between 333 ms pre-stimulus to 333 ms post-stimulus was estimated for each electrode. Such a baseline period was chosen such that it covered one sequence of intensifications, and contained, on average, one target trial and three non-target trials for both random and deterministic blocks.
5. *Artifact rejection*. EEG data are often contaminated by artifacts whose presence can be attributed to oculomotor activities, such as eye blinks and eye movement, which are reflected as large amplitude signals that are most prominent at anterior frontal sites. To reduce the effects of such outliers, sub-trials in which an absolute potential exceeding 50μV was recorded from either of the two anterior frontal sites (FP1 and FP2; see Figure 2), within a time-window spanning 333 ms pre-stimulus to 500 ms post-stimulus, were rejected. The other sub-trials belonging to the same trial were also rejected. On average, the percentage of trials that remained were 83.0 ± 16.1% and 78.0 ± 17.4% for the random and deterministic paradigms, respectively.
6. *Downsampling*. Each sub-trial was downsampled from 1024 Hz to 128 Hz.

Classification

After preprocessing, machine-learning was applied to recognize the target choice in each run. Since artifact rejection had been applied, the number of trials available for training was different for each subject and each fold of cross-validation. On average, out of a maximum of 864 trials, there were 717 and 674 trials remaining for training, for the random and deterministic paradigms, respectively.

Feature selection

The 2048 time-samples from 0–500 ms post-stimulus, corresponding to the 64 time-samples in the 32 electrodes, were first standardized, and were then used as the candidate features for constructing a feature vector. That is, given a sub-trial, the potential x measured at a particular electrode and time-point was transformed as follows:

$$x' = (x - \mu_x)/\sigma_x \quad (1)$$

where μ_x and σ_x were the pooled estimates of the mean and standard deviation of x, based on the training data and without taking classes (targets vs. non-targets) into account. The time-window covered most of the ERP components previously reported to be modulated in the random paradigm, including P1, N1, P2, N2 and P300. To optimize the classification performance and to control for effects of over-fitting, the maximum number of spatiotemporal features (K) selected for inclusion was examined at 6 levels: 50, 100, 200, 500, 1000, and 2048. Specifically, based on the training data, the two-sample t-statistics were obtained for every spatiotemporal feature, i.e., the corresponding potentials in target sub-trials and non-target sub-trials were subject to the Student's t-test. The K most-discriminative features whose t-statistics had the largest absolute values were then selected [23] for the given fold of analysis.

Linear discriminant analysis

Fisher's linear discriminant analysis (see [15], for example) was applied to the training data to obtain a binary classifier for discriminating between target and non-target sub-trials. The classifier corresponds to a decision hyperplane defined by:

$$\boldsymbol{w} \cdot \boldsymbol{x} - b = 0 \quad (2)$$

where $\boldsymbol{x}$ is a vector in the space of feature vectors, $\boldsymbol{w}$ a normalized vector of feature weights, and b a bias term. The sum $\boldsymbol{w} \cdot \boldsymbol{x} - b$ is referred to as the standardized discriminant function [27], which tends to be positive for targets and negative for non-targets. In this paper, the elements of $\boldsymbol{w}$ will be referred to as the standardized discriminant function coefficients.

Calculation of classification performances

To determine the minimum number of intensification sequences necessary to achieve accurate performance, runs were classified using the data associated with N trials, for $N = 1,\ldots,12$. Specifically, the intended choice in each run was determined using the classifier obtained, and compared to the actual choice that was specified to the subject. For a given run, N scores were calculated for each choice c as follows:

$$Score(N,c) = \sum_{k=1}^{N} s_{k,c} = \sum_{k=1}^{N} \boldsymbol{w} \cdot \boldsymbol{x}_{k,c} \ , N = 1,\ldots,12 \quad (3)$$

where $s_{k,c}$ represents the sub-score for the sub-trial that was timelocked to the intensification of the choice c in the k^{th} trial within the run, and $\boldsymbol{x}_{k,c}$ the feature vector associated with that sub-trial. The choice with the highest score was determined to be the target for the given run. Note that: (a)

the constant, bias term b in (2) was not included in calculation of the sub-scores, since this term would have no effect on the relative scores across choices; (b) the sub-scores associated with a rejected sequence were all replaced by zeros. Overall, the above calculations were performed for both training and testing runs, to evaluate whether or not there was serious problem of over-fitting.

RESULTS

Classification Performances

Figure 3 summarizes the mean classification accuracies obtained with different number of trials (N = 1, ..., 12) for both the random paradigm (**RP**) and the deterministic paradigm (**DP**). Each curve represents the accuracies obtained for a particular number of features (K = 50, 100, 200, 500, 1000 and 2048). Individual accuracies were averaged across the four folds of validations. As expected, the average accuracy generally increased with increasing N for both paradigms. Also, since it was evident that the average accuracy increased with K, only the cases for K = 2048 are reported in detail from this point onward. Most importantly, accurate performance was obtained for both paradigms, with the accuracies for N = 3 and N = 12 being 95.3 ± 1.7% and 99.0 ± 1.3% for **RP**, and the corresponding accuracies were 93.2 ± 4.9% and 98.3 ± 1.4% for **DP**. The training accuracies were also obtained for both paradigms. Although the testing accuracy for each N was lower than the corresponding training accuracy for both paradigms, such effects of over-fitting were not serious. Specifically, the differences between training accuracy and testing accuracy (training – testing) at N = 3 and N = 12 were 2.7 ± 1.3 % and 0.2 ± 0.6% for **RP**, and 3.1 ± 2.2% and 0.4 ± 0.5% for **DP**, respectively.

The Spatiotemporal Features Underlying Classification

In this sub-section, the basis for discriminating target and non-target sub-trials in the present experiment is demonstrated and characterized using three methods: (a) event-related potentials; (b) standardized discriminant function analysis; and (c) supplementary classification analyses, by montages (i.e., subsets of electrodes) and by time-intervals. For practical considerations, it is essential to determine whether accurate performance is attainable with only a small set of electrodes. Thus, our main focus is on locating the electrodes that are the most informative. The temporal aspects were also inspected systematically to determine which time-interval was the most informative.

(a) Event-related potentials

Figure 4 shows the grand-averaged ERP waveforms (i.e., average ERP waveforms across subjects) elicited by both targets and non-targets for both paradigms, as recorded over various positions on the scalp, taking all 4 experimental blocks into account for all subjects. In both cases, a prominent positivity peaking between 150–300 ms post-stimulus was observed for the waveform elicited by targets but not by non-targets, over a broad range of electrodes, as highlighted in Figure 4 for the electrodes Fz, Cz and Pz. However, the grand-averaged ERP waveforms do not capture the individual differences across subjects. Thus, this positivity between 150–300 ms does not necessarily comprise the most informative features on a per-subject basis. To demonstrate the individual differences, the two-sample t-statistics comparing the potential associated with every spatiotemporal feature in the target and non-target sub-trials were obtained for all six subjects. These t-statistics are shown as feature maps in Figure 5 for both paradigms. A highly positive (or negative) value would imply that the mean potential for target sub-trials was significantly greater (or less) than that for non-target sub-trials. Consistent with the grand-averaged ERP shown in Figure 4, prominent positivities within the 150–300 ms interval post-stimulus were observed for most subjects at many frontal, central and parietal locations, for both paradigms. However, for all subjects except S4, a prominent peak that emerged within the 100–200 ms interval post-stimulus was also observed at some occipital locations. For S1–S3, the peak was positive, while for S5–S6, the peak was negative.

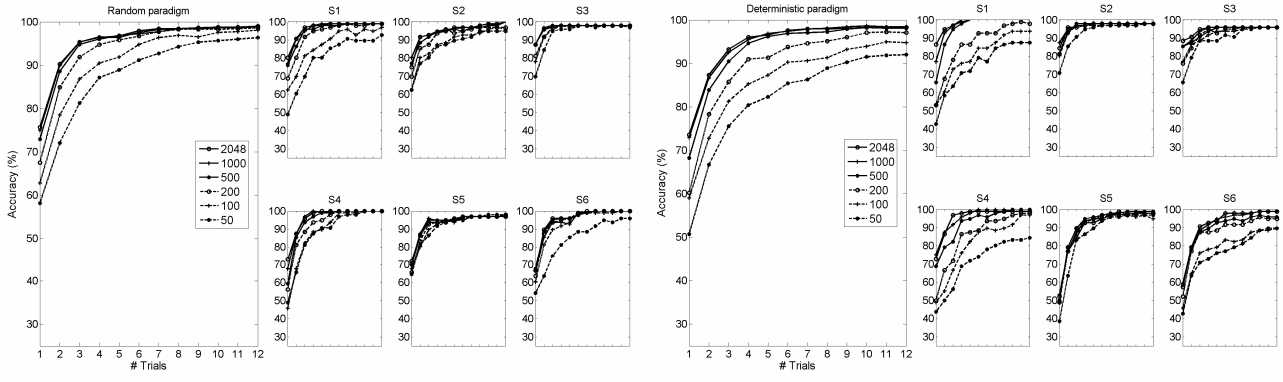

Figure 3. Mean classification accuracies versus the number of trials employed for classification, for the random paradigm (left) and deterministic paradigm (right). Each curve was obtained using a different number of features (K = 50, 100, 200, 500, 1000, and 2048) for classification.

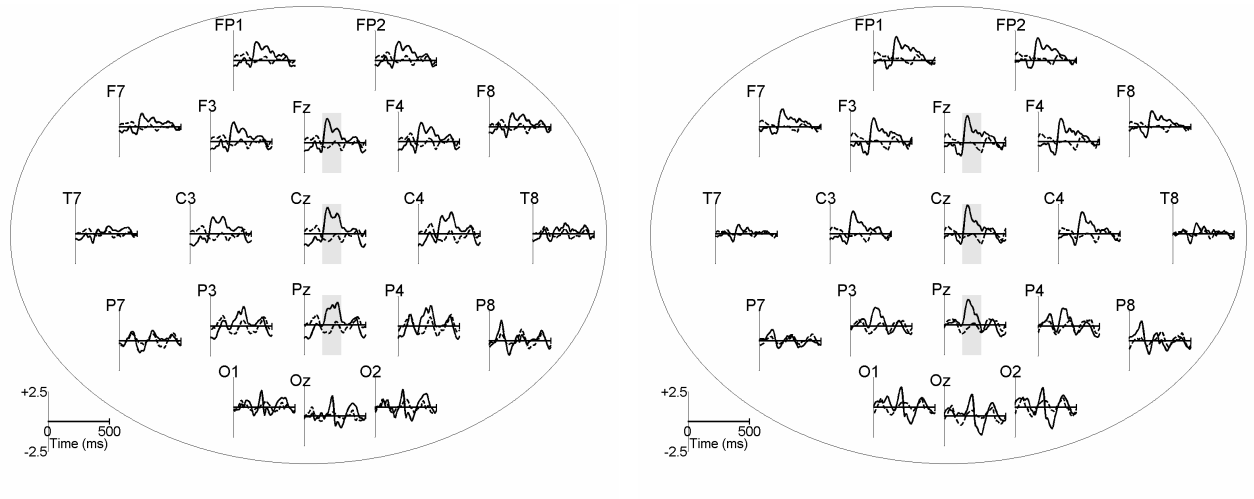

Figure 4. Grand-averaged event-related potential (ERP) obtained at 20 selected electrodes, for both the random (left) and deterministic (right) paradigms. The interval between 150–300 post-stimulus is highlighted in gray for Fz, Cz & Pz, to illustrate the positivity observed.

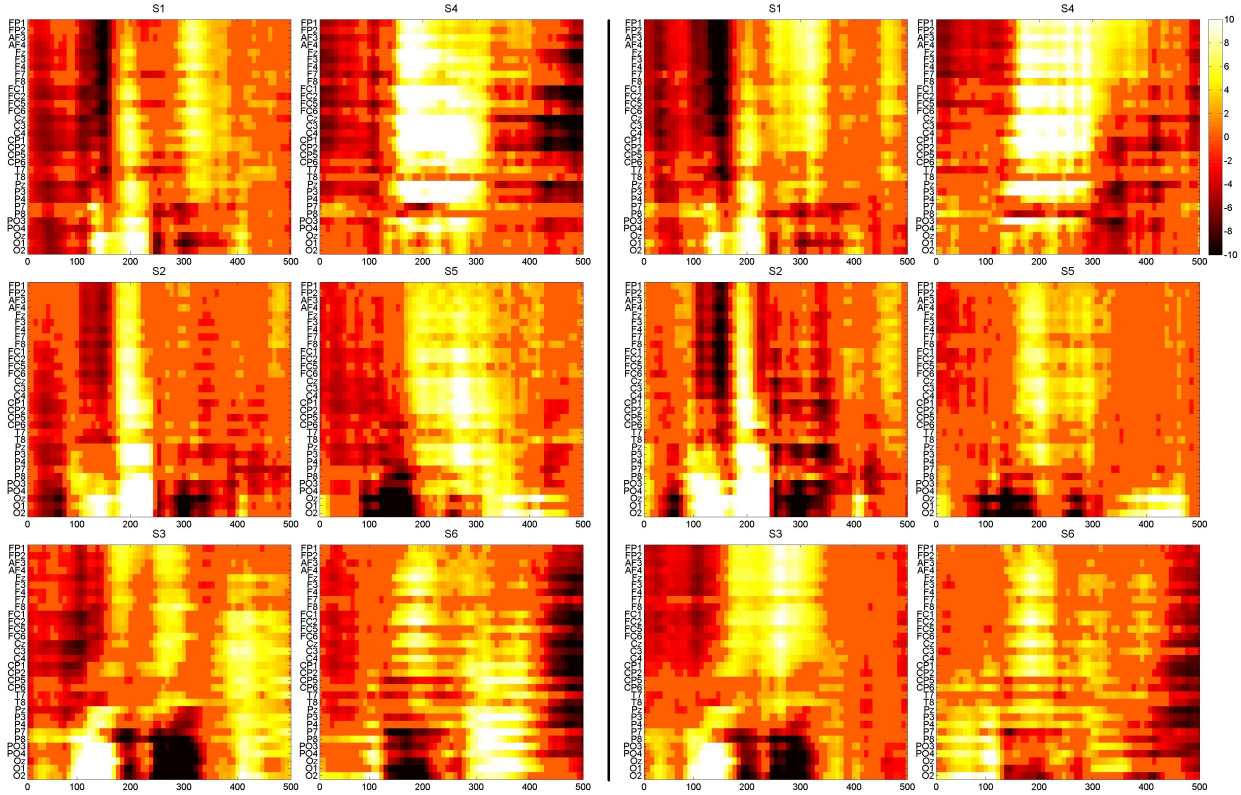

Figure 5. T-value maps obtained for individual subjects, for both the random (left) and deterministic (right) paradigms. The t-statistics associated with the features for which the difference across the target and non-target conditions was non-significant (p>0.05) were displayed as zero-values.

(b) Standardized discriminant function analysis

The contribution made by each spatiotemporal feature for group separation can be ranked according to its associating coefficient in the standardized discriminant function [27]. To shed light on the nature of the primary features, the square of the coefficients were summed over time, to compare the contributions of different electrodes, and over space (i.e., electrodes), to measure the contributions of different time-points. Figure 6 shows, for both **RP** and **DP**, the variation in contributions (by percentage) along the

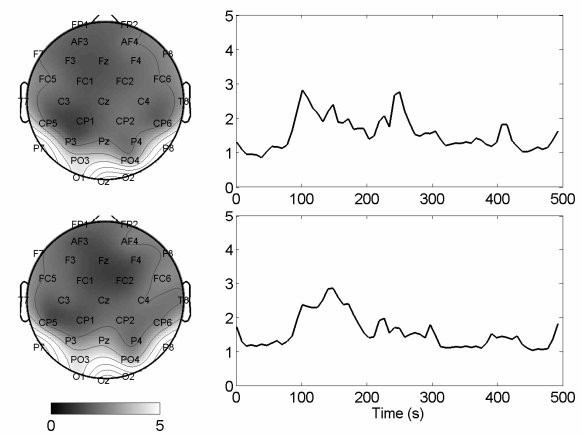

Figure 6. The contributions (by %) made by different electrodes, shown as a topomap (left), and different time-points, shown as a time-series (right), as indicated by the sum of squares of the standardized discriminant function coefficients along the appropriate dimension. Top: random paradigm; bottom: deterministic paradigm.

spatial and temporal dimensions, averaged across the six subjects and for the four folds of analyses. The results of the spatial analysis are shown in detail in Table 1 for the seven most important electrodes in each paradigm. Had the relative contributions been the same for the 32 electrodes, the total contributions made by each electrode would have been 3.125 %. Consistent with the similarities shown in Figure 6, the first six electrodes were the same across paradigms, namely Oz, O1, O2, P7, P8, and PO3—all posterior electrodes. Also, their order of importance was almost identical, except that O2, by proportion, was consistently found to contribute to the discrimination more for **RP** than for **DP**, for all six subjects.

Table 2 shows the total contributions made by four time-intervals (T1: 0–125 ms, T2: 125–250 ms, T3: 250–375 ms & T4: 375–500 ms), for each subject and for both paradigms. It is apparent that the feature weights tend to be concentrated in time-interval T2 (125–250 ms), exceptions being S3 and S5 for **RP**, wherein the feature weights were more concentrated in T3 and T1, respectively. These temporal characteristics suggest that the time-interval from 125–250 ms was more responsible for the discrimination than the other time-intervals.

(c) Supplementary classification analyses

Two types of classification analyses were performed to characterize the spatiotemporal characteristics present in the two paradigms. In the first analysis (the spatial analysis), the data associated with 8 montages (see Figure 7), each comprising 5 electrodes, were used for classification. In half of the montages, a midline electrode (Fz, Cz, Pz or Oz) was included together with four neighboring electrodes; in the other half, a midline electrode was included together with four laterally arranged electrodes.

	Random							**Deterministic**						
Sub.	Oz	O2	P7	P8	O1	PO3	T7	Oz	P7	P8	O2	O1	PO3	PO4
S1	10.5	7.9	9.9	4.8	6.1	3.1	4.1	9.9	9.4	7.7	7.1	5.2	3.1	2.7
S2	11.9	7.1	8.1	7.3	5.8	4.7	2.7	7.7	8.4	7.8	5.9	6.0	2.3	4.6
S3	10.5	8.5	5.6	8.7	4.0	3.5	3.3	12.6	7.5	9.8	6.5	5.6	2.1	3.2
S4	2.1	3.9	11.8	9.5	3.0	5.4	4.0	2.0	11.6	8.4	3.2	2.3	11.8	3.0
S5	13.1	9.1	3.5	8.8	4.3	5.7	1.8	12.7	5.3	7.0	7.5	4.3	5.0	3.3
S6	7.6	12.2	8.4	4.0	5.5	3.2	3.4	12.3	6.7	2.8	11.4	5.5	1.8	3.7
Avg.	9.3	8.1	7.9	7.2	4.8	4.3	3.2	9.5	8.2	7.3	6.9	4.8	4.3	3.4

Table 1. The relative contributions made by the 7 most informative channels, for both the random and deterministic paradigms. Note that the two sets of informative electrodes were largely the same.

	Random				**Deterministic**			
Sub.	T1	T2	T3	T4	T1	T2	T3	T4
S1	19.9	**36.0**	22.8	21.3	26.7	**32.4**	17.8	23.0
S2	25.3	**30.3**	22.8	21.6	24.8	**31.5**	23.5	20.3
S3	18.8	29.6	**30.9**	20.7	21.7	**36.8**	23.2	18.2
S4	20.1	**35.8**	24.7	19.3	19.9	**40.4**	22.1	17.6
S5	**32.9**	23.9	21.9	21.2	27.8	**28.6**	21.1	22.5
S6	23.8	**28.7**	26.7	20.7	25.8	**29.9**	21.2	23.1
Avg.	23.5	**30.7**	25.0	20.8	24.5	**33.3**	21.5	20.8

Table 2. The relative contributions made by different time intervals (T1: 0–125 ms; T2: 125–250 ms; T3: 250–375 ms; T4: 375–500 ms), for both the random and deterministic paradigms. The value corresponding to the most informative interval per-subject was shown in bold. Note that on average, T2 was the most informative for both paradigms.

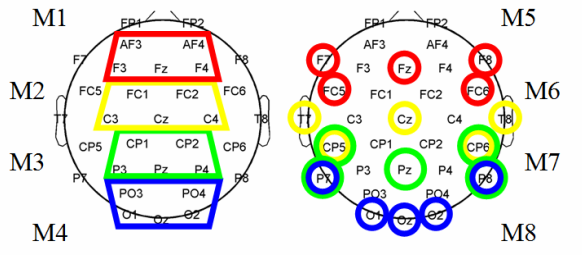

Figure 7. Eight candidate montages (M1–M8), each consisting of 5 electrodes, for which classification accuracy was obtained.

Given that the standardized discriminant function analysis in part (b) indicated an advantage for posterior electrodes, it was expected that the two montages M4 (Oz, O1, O2, PO3 & PO4) and M8 (Oz, O1, O2, P7 & P8) would give rise to the most accurate performances. However, the standardized discriminant function is multivariate in nature, i.e., it indicates the importance of a set of variables in the presence of other variables. Thus, it is not necessary that the set of features carrying the most weights in the analysis in section (b) give rise to the highest classification accuracies when employed separately. In the second analysis (the temporal analysis), consistent with the analyses in part (b), the four time intervals (T1–T4) were examined by using the data from all 32 channels for classification.

Figure 8 shows the results of the two analyses. For the spatial analysis, it is clear that the classification accuracies for both paradigms increased along the anterior-posterior axis, reaching a maximum at M4 and M8 (Figure 8, top panel). In particular, using 3 trials for classification, the accuracies for **RP** and **DP** were respectively 92.4 ± 8.4% and 92.4 ± 6.2% for M4, and 89.4 ± 10.0% and 88.4 ± 14.5% for M8. For the temporal analysis, the classification accuracies peak at T2 (125–250 ms) for both paradigms (Figure 8, bottom panel). Thus, both results are consistent with the conclusions made in part (b).

DISCUSSIONS

In the random and deterministic paradigms, by using the 32-channel data associated with 3 trials, target choices could be recognized at a mean accuracy of 95.3% and 93.2%, respectively. Moreover, for both paradigms, the accuracy reached at least 90% for all six subjects by the end of the fourth trial. In addition, consistent with the indications from the standardized discriminant function analysis, the highest accuracy was achieved by a montage consisting of 5 posterior electrodes (Oz, O1, O2, PO3 & PO4), being 92.4% for both paradigms. Evidently, as far as the classification performance is concerned, neither paradigm has an apparent advantage over the other. Consistent results regarding the temporal aspects were also obtained for the standardized discriminant function analysis and the supplementary classification analyses; in both analyses, the time-interval between 125–250 ms was found to be the most informative for both paradigms. Such spatiotemporal distributions of primary features are not typically associated with P300 responses, which have a central-parietal distribution that peak around 300 ms post-stimulus. They are, nonetheless, consistent with the growing evidence accumulated that the P300 responses do not necessarily comprise the primary features for classification in this type of visual paradigms. For this reason, some authors have advocated the more generic term of ERP-based visual speller [36]. In the present experiment, the early responses recorded at the posterior sites were found to be the primary features for both the random and deterministic paradigms, suggesting that these responses, unlike the P300 responses, might be qualitatively similar across paradigms, and might not be modulated by randomness in the intensification sequences. These conclusions have to be confirmed with a larger data set.

In terms of classification performance, our results compare favorably with a previous study in which a four-choice paradigm was employed [28]. In that study, the best-performing subject managed an accuracy of 75% at 4 trials, and 92% at 19 trials. A possible explanation for our much higher accuracies is our use of a much shorter ISI (166 ms, vs. 1,400 ms in [28]). In the original matrix speller study

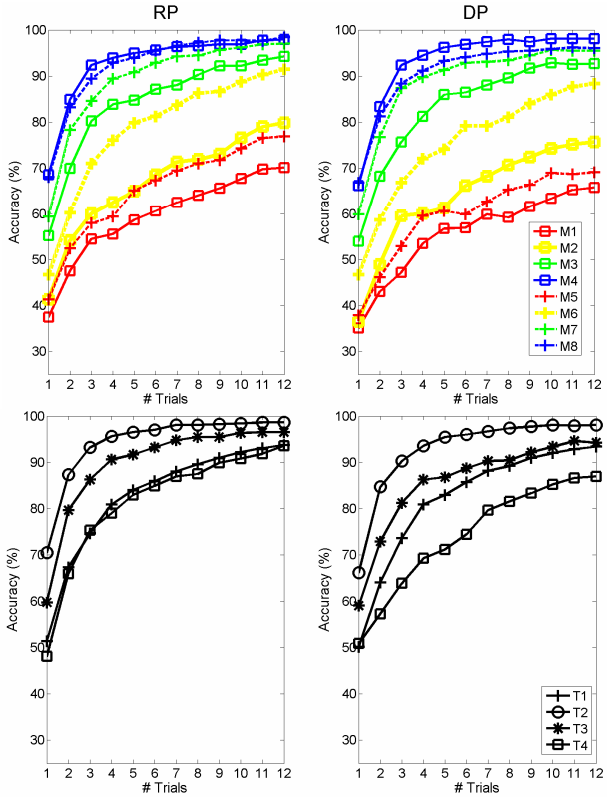

Figure 8. Mean classification accuracies versus the number of trials employed for classification, for the random (left) and deterministic (right) paradigms. Top panel: each curve was obtained using different montages (M1–M8; see Figure 7) for classification. Bottom panel: each curve was obtained using different time-intervals (T1–T4) for classification.

[11], a higher accuracy was observed at an ISI of 500 ms than at 125 ms, suggesting that a longer ISI could be helpful for good performance. On the other hand, as pointed out in [15], this conclusion might hold only within certain limits of ISIs. Our result that the classification performance is higher at an ISI of 166 ms (than at 1,400 ms) is consistent with this view. A systematic investigation of the ISI variable, after controlling for other confounding factors (such as geometrical configuration), is necessary to determine if any general conclusion could be drawn regarding the dependence of accuracy on ISI.

Finally, of the two criteria (randomness and low target probability) necessary for an oddball sequence, the present study has investigated only the former criterion. For practical considerations, it is especially important to determine if the present results, obtained with a four-choice paradigm, can be extended to other paradigms that allow more choices to be selected. If so, the use of deterministic sequences may offer a more user-friendly alternative than random sequences for ERP-based visual spellers. Work is under way to test if the deterministic paradigm works as well as the random paradigm as the number of choices is increased, e.g., from 4 to 8.

CONCLUSION

The primary aim of the present study was to determine whether the use of random sequences for stimulus intensification is crucial for accurate performance in ERP-based visual spellers by comparing the accuracies obtained when the choices were intensified either in random or deterministic sequences. It was found that the intended choices could be recognized at a mean accuracy of 95.3% and 93.2%, respectively, using the 32-channel data associated with 3 intensification sequences. The time-interval from 125–250 ms post-stimulus was found to be the most informative. To assess the projected performance of a BCI system in which the number of electrodes available is expected to be limited, further analyses were performed using the data associated with 8 candidate montages, each comprising 5 electrodes. A comparable mean accuracy of 92.4% was achieved in both paradigms for the best montage, consisting of 5 posterior electrodes: Oz, O1, O2, PO3 & PO4. These results suggest that: (a) the use of random sequences is not necessary for effective BCI performance; and (b) deterministic sequences can be used in some BCI speller applications.

ACKNOWLEDGMENTS

This work is supported in part by grants made to the Chinese University of Hong Kong by the Research Grants Council of the Hong Kong SAR (T. Blu: CUHK410110), the Office of the Government Chief Information Officer (W.S-Y. Wang: af-006), and the Patent Committee of The Chinese University of Hong Kong (TBF/11/ENG/001).

REFERENCES

1. Allison, B.Z. and Pineda, J.A. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system. *International Journal of Psychophysiology 59*, 2 (2006), 127-140.
2. Allison, B.Z. and Pineda, J.A. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering 11*, 2 (2003), 110-113.
3. Bandt, C., Weymar, M., Samaga, D. and Hamm, A.O. A simple classification tool for single-trial analysis of ERP components. *Psychophysiology 46*, 4 (2009), 747-757.
4. BioSemi. *Active Two User Manual (Version 3.2, July 3, 2007).* Amsterdam.
5. Birbaumer, N. Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control. *Psychophysiology 43*, 6 (2006), 517-532.
6. Blankertz, B., Lemm, S., Treder, M., Haufe, S. and Müller, K.-R. Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage 56*, 2 (2011), 814-825.
7. Delorme, A. and Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods 134*, 1 (2004), 9-21.
8. Donchin, E., Spencer, K.M. and Wijesinghe, R. The mental prosthesis: assessing the speed of a P300-based brain–computer interface. *IEEE Transactions on Rehabilitation Engineering 8*, 2 (2000), 174-179.
9. Duncan-Johnson, C.C. and Donchin, E. On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology 14*, 5 (1977), 456-467.
10. Fabiani, M., Gratton, G., Karis, D. and Donchin, E. The definition, identification and reliability of measurement of the P300 component of the event-related brain potential. In Ackles, P., Jennings, J. and Coles, M.G.H. eds. *Advances in psychophysiology*, JAI Press, Greenwich, CT, 1987, 1-78.
11. Farwell, L.A. and Donchin, E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology 70*, 6 (1988), 510-523.
12. Fogelson, N., Wang, X., Lewis, J.B., Kishiyama, M.M., Ding, M. and Knight, R.T. Multimodal Effects of Local Context on Target Detection: Evidence from P3b. *Journal of Cognitive Neuroscience 21*, 9 (2009), 1680-1692.

13. Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., Gramatica, F. and Edlinger, G. How many people are able to control a P300-based brain–computer interface (BCI)? *Neuroscience Letters 462*, 1 (2009), 94-98.

14. Handy, T. *Event-Related Potentials: A Methods Handbook*. The MIT Press, Cambridge, MA, USA, 2004.

15. Hoffmann, U., Vesin, J.-M., Ebrahimi, T. and Diserens, K. An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience Methods 167*, 1 (2008), 115-125.

16. Hong, B., Guo, F., Liu, T., Gao, X. and Gao, S. N200-speller using motion-onset visual response. *Clinical Neurophysiology 120*, 9 (2009), 1658-1666.

17. Ikegami, S., Takano, K., Saeki, N. and Kansaku, K. Operation of a P300-based brain–computer interface by individuals with cervical spinal cord injury. *Clinical Neurophysiology 122*, 5 (2011), 991-996.

18. Jeon, Y.-W. and Polich, J. P3a from a passive visual stimulus task. *Clinical Neurophysiology 112*, 12 (2001), 2202-2208.

19. Jin, J., Allison, B.Z., Brunner, C., Wang, B., Wang, X., Zhang, J., Neuper, C. and Pfurtscheller, G. P300 Chinese input system based on Bayesian LDA. *Biomedizinische Technik Biomedical engineering 55* (2010), 5-18.

20. Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M. and Wolpaw, J.R. Toward enhanced P300 speller performance. *Journal of Neuroscience Methods 167*, 1 (2008), 15-21.

21. Mak, J.N., Arbel, Y., Minett, J.W., McCane, L.M., Yuksel, B., Ryan, D., Thompson, D., Bianchi, L. and Erdogmus, D. Optimizing the P300-based brain–computer interface: current status, limitations and future directions. *Journal of Neural Engineering 8*, 2 (2011), 025003.

22. Minett, J.W., Zheng, H.-Y., Fong, M.C.-M., Zhou, L., Peng, G. and Wang, W.S.-Y. A Chinese text input brain–computer interface based on the P300 speller. *International Journal of Human-Computer Interaction 28*, 7 (2012), 472-483.

23. Müller, K.-R., Krauledat, M., Dornhege, G., Curio, G. and Blankertz, B. Machine learning techniques for brain–computer interfaces. *Biomedizinische Technik 49*, Suppl 1 (2004), 11-22.

24. Nijboer, F., Sellers, E.W., Mellinger, J., Jordan, M.A., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D.J., Vaughan, T.M., Wolpaw, J.R., Birbaumer, N. and Kübler, A. A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology 119*, 8 (2008), 1909-1916.

25. Piccione, F., Giorgi, F., Tonin, P., Priftis, K., Giove, S., Silvoni, S., Palmas, G. and Beverina, F. P300-based brain computer interface: Reliability and performance in healthy and paralysed participants. *Clinical Neurophysiology 117*, 3 (2006), 531-537.

26. Polich, J. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology 118*, 10 (2007), 2128-2148.

27. Rencher, A.C. *Methods of Multivariate Analysis*. John Wiley & Sons, NY, USA, 2002.

28. Sellers, E.W. and Donchin, E. A P300-based brain–computer interface: Initial tests by ALS patients. *Clinical Neurophysiology 117*, 3 (2006), 538-548.

29. Sellers, E.W., Krusienski, D.J., McFarland, D.J., Vaughan, T.M. and Wolpaw, J.R. A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology 73*, 3 (2006), 242-252.

30. Silvoni, S., Volpato, C., Cavinato, M., Marchetti, M., Priftis, K., Merico, A., Tonin, P., Koutsikos, K., Beverina, F. and Piccione, F. P300-based brain–computer interface communication: evaluation and follow-up in amyotrophic lateral sclerosis. *Frontiers in Neuroscience 3*, 60 (2009).

31. Sommer, W., Leuthold, H. and Matt, J. The expectancies that govern the P300 amplitude are mostly automatic and unconscious. *Behavioral and Brain Sciences 21*, 01 (1998), 149-150.

32. Squires, K., Wickens, C., Squires, N. and Donchin, E. The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science 193*, 4258 (1976), 1142-1146.

33. Sutton, S., Braren, M., Zubin, J. and John, E.R. Evoked-potential correlates of stimulus uncertainty. *Science 150*, 3700 (1965), 1187-1188.

34. Takano, K., Komatsu, T., Hata, N., Nakajima, Y. and Kansaku, K. Visual stimuli for the P300 brain–computer interface: A comparison of white/gray and green/blue flicker matrices. *Clinical Neurophysiology 120*, 8 (2009), 1562-1566.

35. Townsend, G., LaPallo, B.K., Boulay, C.B., Krusienski, D.J., Frye, G.E., Hauser, C.K., Schwartz, N.E., Vaughan, T.M., Wolpaw, J.R. and Sellers, E.W. A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clinical Neurophysiology 121*, 7 (2010), 1109-1120.

36. Treder, M. and Blankertz, B. (C)overt attention and visual speller design in an ERP-based brain–computer interface. *Behavioral and Brain Functions 6*, 1 (2010), 28.

37. Vidal, J.J. Toward direct brain–computer communication. *Annual review of biophysics and bioengineering 2* (1973), 157-180.