Visual modifications on the P300 speller BCI paradigm

2009 J. Neural Eng. 6 046011

(http://iopscience.iop.org/1741-2552/6/4/046011)

IOP Publishing JOURNAL OF NEURAL ENGINEERING

J. Neural Eng. 6 (2009) 046011 (8pp) doi:10.1088/1741-2560/6/4/046011

Visual modifications on the P300 speller BCI paradigm

M Salvaris and F Sepulveda

BCI Group, School of Computer Science and Electronic Engineering, University of Essex, Wivehoe Park Colchester, Essex CO4 3SQ, UK

E-mail: mssalv@essex.ac.uk and fsepulv@essex.ac.uk

Received 5 May 2009
Accepted for publication 17 June 2009
Published 15 July 2009
Online at stacks.iop.org/JNE/6/046011

Abstract

The best known P300 speller brain-computer interface (BCI) paradigm is the Farwell and Donchin paradigm. In this paper, various changes to the visual aspects of this protocol are explored as well as their effects on classification. Changes to the dimensions of the symbols, the distance between the symbols and the colours used were tested. The purpose of the present work was not to achieve the highest possible accuracy results, but to ascertain whether these simple modifications to the visual protocol will provide classification differences between them and what these differences will be. Eight subjects were used, with each subject carrying out a total of six different experiments. In each experiment, the user spelt a total of 39 characters. Two types of classifiers were trained and tested to determine whether the results were classifier dependant. These were a support vector machine (SVM) with a radial basis function (RBF) kernel and Fisher's linear discriminant (FLD). The single-trial classification results and multiple-trial classification results were recorded and compared. Although no visual protocol was the best for all subjects, the best performances, across both classifiers, were obtained with the white background (WB) visual protocol. The worst performance was obtained with the small symbol size (SSS) visual protocol.

1. Introduction

A form of synchronous brain-computer interface (BCI) is the P300 BCI, named so due to the positive deflection of the electroencephalography (EEG) around 300 ms post stimulus. In this BCI protocol, the user is able to interact with the interface by paying attention to chosen rare target stimuli during a random sequence of target and non-target stimuli (oddball paradigm) [1–3]. The oddball paradigm relies on the fact that, on average, attended rare stimuli produce larger P300 potentials than attended frequent ones [4]. In the P300 paradigm, the subject is often required to respond mentally or physically to the target stimuli since passive stimulus processing generally produces smaller P300 amplitudes than active tasks [5]. An early example of a BCI system that utilized the unique nature of this potential was that presented by Farwell and Donchin [2], in which the user was presented with a matrix of symbols and each row and column would flash randomly. The user paid attention and mentally recorded the row or column flashes that happened to contain the symbol they desired to select. Due to the less than adequate signal-to-noise ratio of EEG, the classification of the P300 potential has to happen over a number of trials in order to achieve high-accuracy rates. Due to this fact and due to inherent physiological limitations, the transfer rate of such systems is very slow [6, 7].

The P300 BCI protocol can be split into the following three subproblems: (a) protocol design, (b) signal processing/feature selection and/or extraction, and (c) classifier training. Many attempts have been made to improve the transfer rates and accuracy of this BCI protocol by improving results in each of these subproblems [8–15]. Although modifications to matrix element dimensions, flash patterns and inter-stimulus interval (ISI) have been explored [16–19], the effects of inter-symbol distance, symbol size and differing foreground and background colours have not.

The work presented here aims to tackle subproblem (a), in other words, to ascertain whether the simple modifications to the visual protocol carried out in these experiments will provide classification differences between them and what these differences will be.

Table 1. Table of abbreviations.

Abbreviation	Full name	Description
WB	White background	Visual protocol with the white background.
BB	Black background	Visual protocol with the black background.
LISD	Large inter-symbol distance	Visual protocol with the largest distance between symbols/characters.
SISD	Small inter-symbol distance	Visual protocol with the smallest distance between symbols/characters.
SSS	Small symbol size	Visual protocol with the smallest symbol/characters.
LSS	Large symbol size	Visual protocol with the largest symbols/characters.
DT	Distance top	Distance of the top edge of first row to the top edge of the screen.
DB	Distance bottom	Distance of the bottom edge of the last row to the bottom edge of the screen.
DR	Distance right	Distance of the right edge of the right most column from the right edge of the screen.
DL	Distance left	Distance of the left edge of the left most column to the left side of the screen.
CX	Character X	Width of the character D.
CY	Character Y	Height of the character D.
ICX	Inter-character X	Horizontal distance between characters. Measured from the center of one character to the center of the next.
ICY	Inter-character Y	Vertical distance between characters. Measured from the center of one character to the center of the next.
BX	Board X	Horizontal distance from the left edge of the first column to the right edge of the last column.
BY	Board Y	Vertical distance from the top edge of the first row to the bottom edge of the last row.

2. Methods

It is important to keep in mind that the methods for this experiment were chosen in order to allow comparability between the various visual protocols with as little bias from the classifiers and the preprocessing methods used. Therefore, many state of the art methods such as discreet wavelet transforms and recursive channel elimination have been ignored in favour of simplicity and comparability [8, 11, 14].

2.1. Signal processing

The EEG data were collected with a Biosemi ActiveTwo system at a sampling rate of 512 Hz. A total of 66 channels were used, with two of the channels placed on the earlobes as references. The channels were referenced to a linked ear reference after data collection. The data were bandpass filtered at 0.1–30 Hz using an eighth-order Butterworth filter. Out of the 64 channels, eight were selected. These were Fz, Cz, Pz, Oz, P3, P4, PO7, PO8, since they have been shown to provide good classification performance [15]. Following each intensification, 800 ms segments of data were extracted from each of the chosen channels. Similarly to [15], the segments were then moving-average filtered and decimated by a factor of 19. The preprocessing technique used resulted in a feature vector of 168 (i.e., $(512 \times 0.8)/19$ samples $\times$ 8 channels) elements.

It is possible that different channels to the ones used may have provided better results. However, as mentioned above, it was not the purpose of this work to try to achieve the highest classification accuracy scores, but rather to have a set-up that was, as much as possible, comparable across subjects. Nonetheless, the channels selected have been shown to yield high classification performance in the P300 speller paradigm [20].

2.2. Abbreviations

In order to save space and make figures more readable, abbreviations have been introduced throughout this paper. They can be seen in table 1.

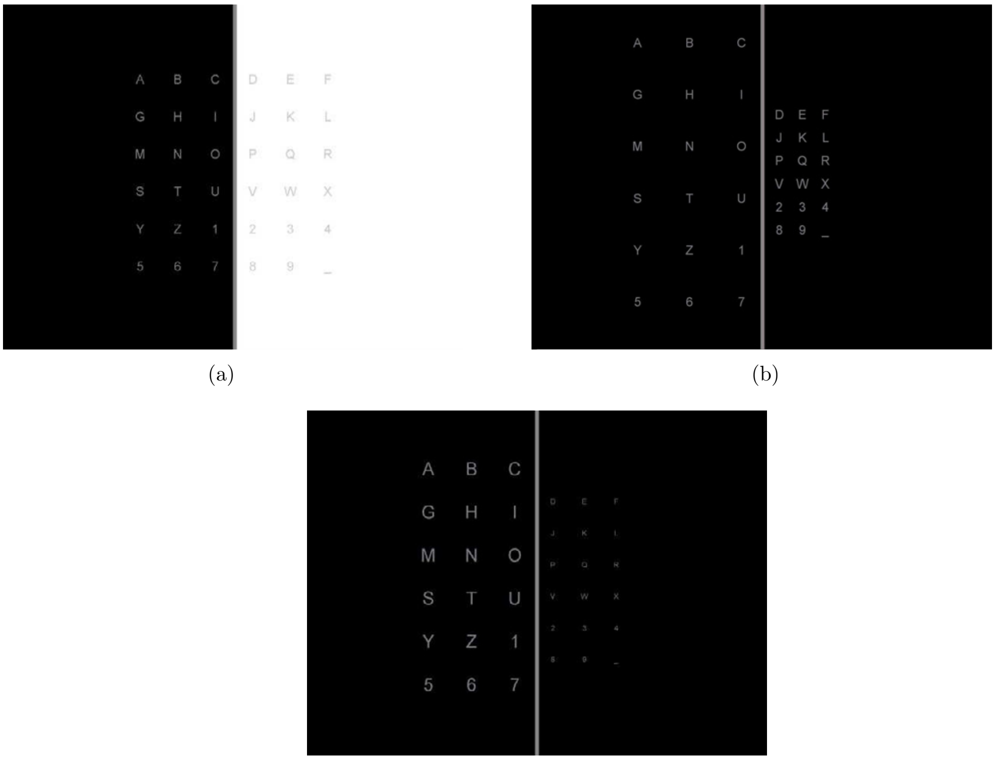
2.3. Experimental protocols

A total of eight subjects were used (six males and two females between the ages of 19 and 28, average age 22.3), all able-odied, with each subject carrying out a series of six different experiments.

The experiments undertaken in this study were as follows (the first item is considered the original Farwell and Donchin protocol, although the timing parameters used were different).

(i) Black background: where a grid of white characters is displayed on a black background.
(ii) White background: where a grid of black characters is displayed on a white background.
(iii) Large symbol size: where a similar set-up to the black background is used except that the size of the symbol/characters is increased.
(iv) Small symbol size: where a similar set-up to the black background is used except that the size of the symbol/characters is decreased.
(v) Large inter-symbol distance: where a similar set-up to the black background is used except that the distance between the symbols is increased.
(vi) Small inter-symbol distance: where a similar set-up to the black background is used except that the distance between the symbols is decreased.

Samples of the visual protocols can be seen in figure 1. Each visual protocol took up the whole screen; the half screen representations depicted in figure 1 are just for brevity.



(a) (b) (c)

Figure 1. Samples of the visual protocols: (*a*) white and black background visual protocols, (*b*) large and small inter symbol distance visual protocols, and (*c*) large and small symbol size visual protocols.

For all visual protocols except WB, the background colour in the 8-bit *RGB* model was (0, 0, 0) (*black*), the default character colour was (120, 120, 120) (*dark grey*) and the highlighted row/column colour was (240, 240, 240) (*light grey*). For WB, the background colour was (255, 255, 255) (*white*), the default character colour was (200, 200, 200) (*grey*) and the highlighted row/column colour was (0, 0, 0) (*black*) (see figure 1). The metrics used for each visual protocol can be seen in table 2 and figure 2. Please refer to table 1 for an explanation of the abbreviations used.

The order in which each subject executed the experiments was randomized. In the majority of cases, the six experiments were carried out over 2–3 days (2–3 experiments each day). The number of experiments executed on each day was determined by the subjects themselves. This was to ensure that the subject would be able to provide each experiment with their full attention. The only subject for whom this was not done was subject 5, who carried out all six experiments in 1 day. This was due to the subjects time constraints. During the running of the experiments, the subjects had at least 5 min to relax before the next experiment. If the subjects required more time, this was given.

From each experiment, a total of 39 characters were spelt. These characters spelt out the sentence 'THE_QUICK_BROWN_FOX_JUMPS_OVER_LAZY_DOG'. This sentence was chosen since it includes all the letters in the alphabet. Each subject was presented with the whole sequence of letters on a piece of paper during the set-up and was allowed to retain it during the experiment. During the experiment, no feedback was given to the subjects as to whether they performed the task correctly. At the end of the experiment, the subjects reported whether they lost sequence or not, e.g., if the experiment finished earlier than they expected (subject repeated or introduced a character) or carried on for longer than expected (subject missed a character). In either case, the experiment was repeated.

The subjects were seated with their eyes at 1 m from a 19 inch TFT screen. Great care was taken to ensure that the subject was relaxed and alert throughout the whole experiment. The stimulus-onset asychrony (SOA) was 300 ms, with an inter-stimulus interval (ISI) of 150 ms. Each character epoch consisted of ten trials, with each trial consisting of 12 flashes of each row or column. The rows and columns were flashed in a random manner. During each trial, all of the rows and columns were flashed once. In each character epoch, the user focused on one character in the sentence. The sentence was spelt out in the order shown in the previous paragraph. After each character epoch, there was a pause of 2500 ms in which the subject located the next desired symbol on the board. After 500 ms into the pause, an auditory cue was presented to indicate to the user that they must move on to the next character if they have not already started to do so. This results in the user selecting a character every 38.5 sec. For a graphical representation of the protocol see figure 3.

2.4. Classification

The classifiers used in this work were a support vector machine (SVM) with a radial basis function (RBF) kernel

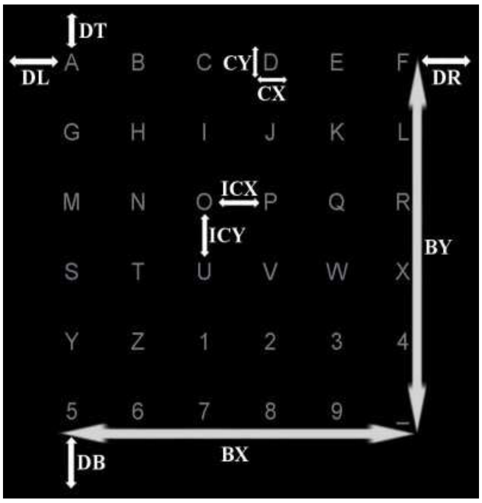


Figure 2. Size metrics for visual protocols.

and a Fisher's linear discriminant (FLD). The process and data distributions used to train both classifiers were identical except that parameter tuning was carried out for the SVM, which was unnecessary for the FLD. The presence or absence of the target P300 from the chosen EEG features is a binary classification problem, and the decision hyper-plane of the discriminant function is defined by

$$w \cdot f(x) + b = 0, \tag{1}$$

where $f(\cdot)$ is the transformation function, x is the feature vector, w is the weights vector and b is the bias or threshold. For the FLD $f(x) = x$, but for the RBF SVM $f(\cdot)$ is the kernel transformation shown in equation (4). From this binary classification, the desired row and column are selected through equations (3) and (2), where J is the number of trials. The selection of the row/column is equivalent to selecting the response that strongly represents the characteristic of the target P300, as defined by the training data, by selecting the feature vector with the largest positive distance from the trained hyper-plane. This makes the bias term in equation (1) irrelevant for the row/column selection. The character that appears at the intersection of the predicted row and column in the matrix is the one chosen:

$$\text{predicted column} = \underset{\text{col}}{\operatorname{argmax}} \left[\sum_{j_{\text{col}}=1}^{J} w \cdot f(x_{j_{\text{col}}}) \right], \tag{2}$$

$$\text{predicted row} = \underset{\text{row}}{\operatorname{argmax}} \left[\sum_{j_{\text{row}}=1}^{J} w \cdot f(x_{j_{\text{row}}}) \right]. \tag{3}$$

The reasons for choosing the FLD and RBF SVM were so that a comparison could be made between classifiers with and without regularization, linear and nonlinear. Furthermore, their popularity in BCI makes their use here even more pertinent [9, 10].

2.4.1. Support vector machine. SVMs use a discriminant hyper-plane to classify data into one of two classes. They maximize the margin from the nearest training points in order to find the optimal hyper-plane with the best generalization capabilities. SVMs use a regularization parameter C that enables them to cope with outliers and therefore allows for errors on the training set. This leads to better generalization and avoids over-fitting. An SVM such as the one described is known as a linear SVM since it classifies data by using linear boundaries. SVMs are also able to create nonlinear

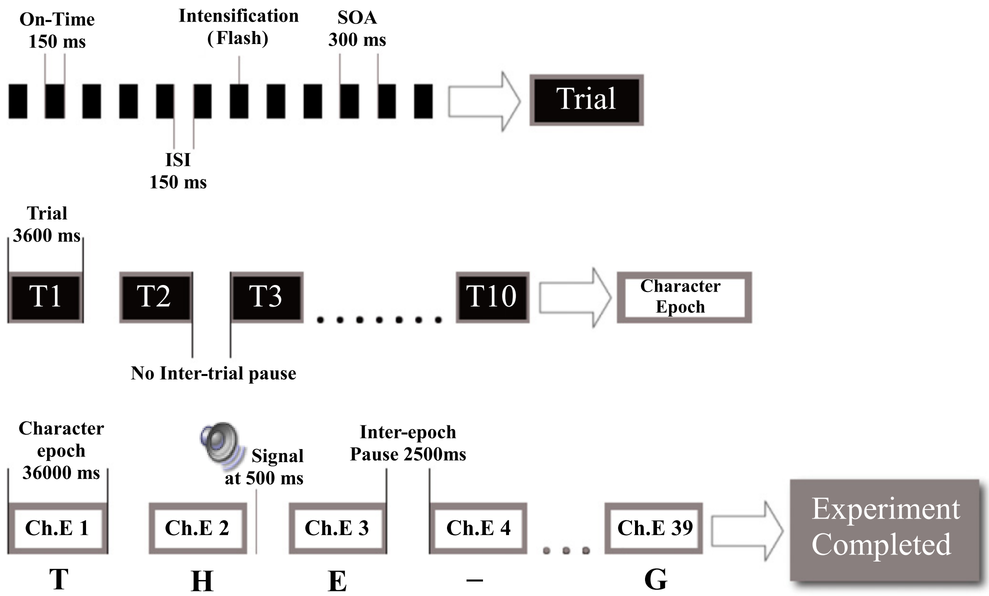


Figure 3. Experiment protocol—a trial was made up of 12 subtrials, with SOA of 300 ms and ISI 150 ms. Each epoch was made up of a series of ten trials which last 3.6 sec, with no inter-trial pause. Thus, the whole experiment was made up of 39 character epochs. There was a 2.5 sec pause between each character epoch.

Table 2. Metric values for each visual protocol (cm) (see table 1 for the explanation of abbreviations).

Board metric	WB	BB	LISD	SISD	SSS	LSS
Distance top	4.5	4.5	0.8	8.3	6.4	2.7
Distance bottom	5.1	5.1	1.6	9	6.8	3.8
Distance right	8.7	8.7	5	12.5	10.3	7.2
Distance left	8.7	8.7	5	12.5	10.3	7.2
Character X	0.7	0.7	0.7	0.7	0.4	1
Character Y	0.8	0.8	0.8	0.8	0.5	1.3
Inter-character X	3.8	3.8	5.3	2.3	3.2	4.3
Inter-character Y	3.8	3.8	5.3	2.3	3.2	4.3
Board X	20.1	20.1	27.3	12.6	16.5	23.4
Board Y	20.1	20.1	27.3	12.6	16.5	23.4

decision boundaries through the use of the 'kernel trick', which allows the SVM to map the data onto another space of higher dimensionality by using a kernel function. A number of different kernels exist, but the most popular in BCI literature is the Gaussian radial basis function kernel [9, 21, 22]. The use of the RBF kernel adds another parameter that needs to be tuned through cross-validation, i.e., γ (kernel bandwidth—see equation (4)):

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \qquad \gamma > 0. \tag{4}$$

SVMs are known to be insensitive to overtraining and to the curse of dimensionality, allowing them to have good generalization properties [9, 21]. All of this of course comes at the expense of execution speed.

The data were normalized to mean zero and unit variance. The parameters for the SVM were found by five-fold cross-validation for each sample distribution. The performance of each validation set was determined by equation (5), taken from [8], where *fp*, *tp* and *fn* are the numbers of false positives, true positives and false negatives, respectively:

$$C_{cs} = \frac{tp}{tp + fp + fn}. \tag{5}$$

The reason for using this performance criterion and not simply the classification performance is the fact that the latter does not take into account the number of true negative examples, which is important for unbalanced data sets, since its omission helps the parameter selection procedure focus on parameters that give positive scores to target examples. The classification performance was then determined by three-fold cross-validation, with each split containing 26 character epochs for training and 13 character epochs for testing. The toolbox used to train and test the SVM was LIBSVM [23].

2.4.2. Fisher's linear discriminant. The FLD, or Fisher's LDA, is a linear classifier which, much like the SVM, aims to use hyper-planes to separate the data representing the two classes [9]. The FLD's main advantages are its conceptual and computational simplicity. Its linearity is often cited as its main disadvantage [9, 10]. Regularization of the FLD is possible [24], although it is not often used in BCI [9, 10]. Despite this the FLD has been successfully applied to a number of BCI problems [9]. It is important to note that the FLD is equivalent to the LDA when the assumptions made by the LDA are satisfied. These are: the classes are normally distributed and the class covariances are equal [25]. The FLD was trained using a similar process to that carried out for the SVM, except that there was no cross-validation to tune parameters since the FLD was used without regularization. The toolbox used to train and test the FLD was the statistical pattern recognition toolbox [26].

3. Results

3.1. Subjective choice of visual protocol

At the end of the experiments, each subject was asked to declare whether they had a preference for a particular protocol or whether they all seemed similar. The answers were collated into the list shown below:

- Subjects 2, 3, 5, 6, 7 and 8 declared that no visual protocol was preferable.
- Subject 1 declared a preference for the white background visual protocol.
- Subject 4 declared a preference for the small inter-symbol distance visual protocol.

3.2. Support vector machine classification performance

A box-plot of P300 versus No-P300 classifier performance for each visual protocol across all subjects can be seen in figure 4(*a*). The corresponding character classification performance can be seen in figure 4(*b*). An explorative study was carried out on the character classification performance to assess the normality and homogeneity of variance of the results. The Kolmogorov–Smirnov test revealed a significant ($p < 0.05$) departure from a normal distribution for SSS. The Shapiro–Wilk test also showed ($p < 0.05$) a departure from normal distribution for SSS as well as WB and LSS. Levene's test was carried out to test the homogeneity of variance. The test result was non-significant indicating that the variances could be considered to be homogeneous. Friedman's test was used to assess the statistical significance of the character classification performance. The classification performance of the SVM did significantly change over the six visual protocols tested ($\chi^2(5) = 13.925$, $p < 0.05$). Wilcoxon tests were used to follow up this finding. A Bonferroni correction was applied and so all effects are reported at a 0.0033 level of significance. Using the SVM, the only two visual protocols that showed statistically significant difference in performance between them were WB and SSS, $T = 0$, $r = -0.63$). All statistical tests were carried out on the mean character classification performance across all trials.

3.3. Fisher's linear discriminant classification performance

The classification performance of the FLD can be seen in figures 5(*a*) and (*b*). Similarly to the SVM results, the FLD results were subjected to an explorative study to assess the normality and homogeneity of variance of the data. Both the Shapiro–Wilk and Kolmogorov–Smirnov tests revealed a significant ($p < 0.05$) departure from normal distribution for SSS, whereas the Kolmogorov–Smirnov test showed a

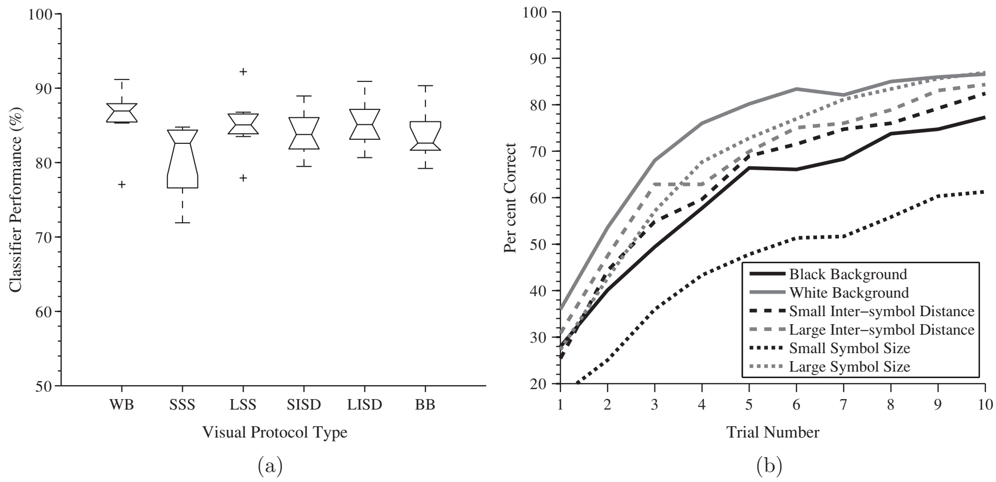


Figure 4. SVM accuracy results for all subjects: (*a*) mean SVM performance for binary classification of P300 versus No-P300 epochs and (*b*) mean character classification performance of SVM.

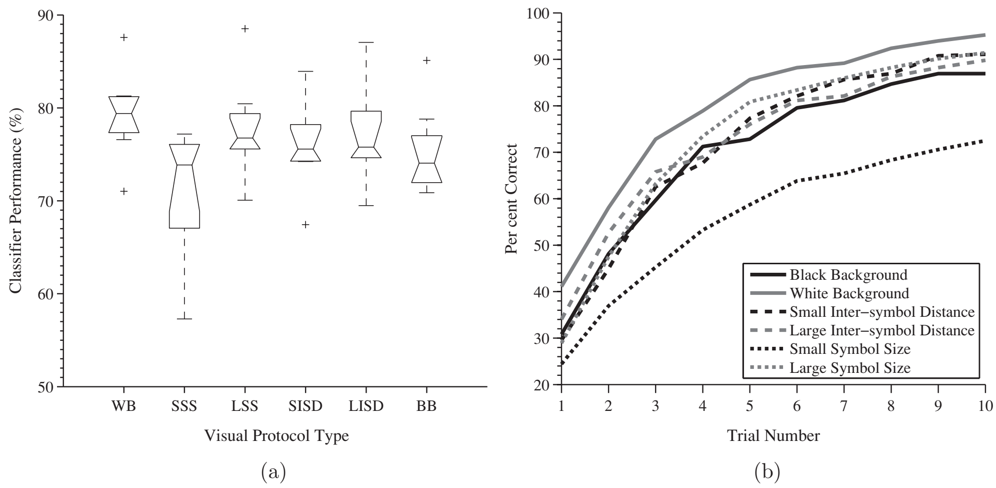


Figure 5. FLD accuracy results for all subjects: (*a*) mean FLD performance for binary classification of P300 versus No-P300 epochs and (*b*) mean character classification performance of FLD.

significant ($p < 0.05$) result for LSS as well. Levene's test did not provide a significant result indicating that the variances could be considered to be homogeneous. Friedman's test revealed that the character classification performance of the FLD did significantly change over the six visual protocols tested ($\chi^2(5) = 13.07, p < 0.05$). Multiple pairwise comparisons were carried out using Wilcoxon tests with Bonferroni correction, so all effects are reported at a 0.0033 level of significance. Similarly to the results obtained for the SVM, the two visual protocols that had significantly different performance from each other were WB and SSS, $T = 0, r = -0.63$.

In any P300-based experiment fatigue can confound results. This is also true for this experiment. To determine whether performance of the visual protocols was significantly affected by the sequence order of the experiments, statistical analysis was carried on the character classification results of the SVM and the FLD. As mentioned earlier, the subjects completed the whole set of six experiments over 1, 2 or 3 days. One subject completed the experiment in 1 day, five in 2 days and two in 3 days. Hence the data were split into groups according to the number of experiments carried out each day. Friedman's test revealed no significant variation in performance across the three experiments carried out each day. For the experiments carried out in 3 days (two experiments per day), a Wilcoxon signed rank test revealed no statistically significant variation in performance. No statistical tests were carried out for the subject who completed all experiments in 1 day. The variation of the performance was simply observed, and although the last two experiments did perform poorly, they were not the worst performing. Also the best performance for that subject was achieved by the fourth experiment in the sequence.

4. Discussion

The statistical test results indicate that the only visual protocols that substantially differ in performance are WB and SSS. This could be interpreted as the P300 classification being unaffected by the majority of the visual modifications made here. We believe this to be incorrect since the results presented show that certain protocols (WB, LSS) consistently outperform other visual protocols. The results presented in figures 4(*a*) and 5(*a*) show box-plots of the classifier performance for the RBF SVM and the FLD across the differing visual protocols. The highest median value in both situations was recorded with the WB visual protocol. What is of note is the large variability in the range of the data in the visual protocols and the number of outliers. The character classification results presented in figures 4(*b*) and 5(*b*), similarly to the box-plots, show the superior performance of the WB visual protocol. A more accurate conclusion would be that the differences in performance obtained by the various visual protocols are confounded by the effects of individual variation, habituation, fatigue and memory load on P300 classification performance. The negative effects of fatigue may have been accentuated by the memory load of the experimental protocol. The increase in memory load will have arisen from the constant requirement of the subject to remember their current position in the character sequence. The statistical analysis of whether the order of the experiments affected performance proved that the effects were not significant. However, the statistical analysis of the sequence effects may have not been powerful enough, since the number of subjects in each group tested is quite small. Thus, the results neither prove nor disprove conclusively the effects of fatigue. They simply indicate that fatigue was not the only influence on classification performance. Although the presentation of the target symbol before each character epoch was dismissed in favour of the chosen method, in order to limit possible perceptual confounds, the results indicate that the method used may be less preferable.

As mentioned in an earlier section, if a subject perceived that they had lost sequence by running out of characters before the experiment completed or finishing the experiment before completing the spelling of all the characters, then the subject was asked to repeat the sequence. Throughout the study, only three experiments out of 48 had to be repeated due to loss of sequence. This does present one problem, however, if the subject misses a character and then introduces one, then there will be no perceived discontinuity. The phrase used to spell is quite common and easily committed to memory. Hence loss of sequence should rarely occur. The possibility of missing a character and then further introducing or repeating a character is even less likely.

The P300 versus No-P300 classification performance of the RBF SVM is better than the FLD's, but the character classification performance of the FLD is better than the RBF SVM's. The most likely reason for this is that the RBF SVM trained to classify the majority of the negative samples (i.e., do not contain a target P300) correctly, which make up 83.3% of the data. This would lead to higher classification rates than the FLD's in simple classifier performance but ultimately lead to worse character classification performance. Furthermore the nonlinear nature of the RBF SVM did not seem to provide any classification advantage over the FLD, which is consistent with the findings by Krusienski *et al* [15].

4.1. Conclusion

From the results presented, it is apparent that the choices made in visual protocols are reflected in the classification obtained. Evidently, the results presented here are limited and the variation between visual protocols is small, but the indication is that greater accuracies can be achieved by simple subject-dependent choice of visual protocols. It is also important to note that the subjective choice of protocol did not correlate with the best performance. Furthermore, most changes in accuracy did not seem to be classifier dependent. The classification difference between the visual protocols only reached statistical significance between the white background and small symbol size visual protocols. However, the results showed that the white background consistently outperformed all other visual protocols.

References

[1] Wolpaw J, Birbaumer N, McFarland D, Pfurtscheller G and Vaughan T 2002 Brain–computer interfaces for communication and control *Clin. Neurophys.* **113** 767–91

[2] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23

[3] Vaughan T, McFarland D, Schalk G, Sarnacki W, Krusienski D, Sellers E and Wolpaw J 2006 The wadsworth BCI research and development program: at home with BCI *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 229–33

[4] Polich J 2004 Clinical application of the P300 event-related brain potential *Phys. Med. Rehabil. Clinics North Am.* **15** 133–61

[5] Polich J 2007 Updating P300: An integrative theory of P3a and P3b *Clin. Neurophys.* **118** 2128–48

[6] Blankertz B *et al* 2004 The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials *IEEE Trans. Neural Syst. Rehabil. Eng.* **51** 1044–51

[7] Blankertz B, Müller K R, Krusienski D J, Schalk G, Wolpaw J R, Schlogl A, Pfurtscheller G, Millan J R, Schroder M and Birbaumer N 2006 The BCI competition III: validating alternative approaches to actual BCI problems *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 153–59

[8] Rakotomamonjy A and Guigue V 2008 BCI Competition III: Dataset II- Ensemble of SVMs for BCI P300 Speller *IEEE Trans. Biomed. Eng.* **55** 1147–54

[9] Lotte F, Congedo M, L'ecuyer A, Lamarche F and Arnaldi B 2007 A review of classification algorithms for EEG-based brain–computer interfaces *J. Neural Eng.* **4** R1–R13

[10] Bashashati A, Fetourechi M, Ward R K and Birch G E 2007 A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals *J. Neural Eng.* **4** 32–57

[11] Donchin E, Spencer K M and Wijesinghe R 2000 The mental prosthesis: assessing the speed of a P300-based brain–computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **8** 174–9

[12] Serby H, Tov E Y and Inbar G F 2005 An improved P300-based brain–computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 89–97

[13] Hoffmann U, Garcia G N, Diserens K, Vesin J M and Ebrahimi T 2005 A boosting approach to P300 detection

with application to brain–computer interfaces *Proc. IEEE EMBS Neural Eng. Conf. (Arlington, VA)*

[14] Glassman E L 2005 A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs *IEEE Trans. Biomed. Eng.* **52** 1851–62

[15] Krusienski D J, Sellers E W, Cabestaing F, Bayoudh S, McFarland D J, Vaughan T M and Wolpaw J R 2006 A comparison of classification techniques for the P300 Speller *J. Neural Eng.* **3** 299–305

[16] Allison B Z and Pineda J A 2006 Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system *Int. J. Psychophysiol.* **59** 127–40

[17] Allison B Z and Pineda J A 2003 ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 110–3

[18] Sellers E W, Krusienski D J, McFarland D J, Vaughan T M and Wolpaw J R 2006 A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance *Biol. Psychol.* **73** 242–52

[19] Meinicke P, Kaper M, Hoppe F, Heumann M and Ritter H 2003 Improving transfer rates in brain computer interfacing: a case study *Adv. Neural Inf. Process. Syst.* **15** 1131–8

[20] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21

[21] Kaper M, Meinicke P, Grossekathöfer U, Linger T and Ritter H 2004 BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm *IEEE Trans. Biomed. Eng.* **51** 1073–6

[22] Thulasidas M, Guan C and Wu J 2006 Robust classification of EEG signal for brain–computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 24–9

[23] Chang C and Lin C 2001 *LIBSVM: A Library For Support Vector Machines* software available at http://www.csie.ntu.edu.tw/cjlin/libsvm

[24] Blankertz B, Curio G and Müller K R 2002 Classifying single trial EEG: towards brain computer interfacing *Adv. Neural Inf. Process. Syst.* **14** 157–64

[25] Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* 2nd edn (New York: Wiley)

[26] Franc V and Hlaváč V 2004 Research report Center for Machine Perception *Statistical Pattern Recognition Toolbox for Matlab* (Czech Republic: Czech Technical University Prague)