2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Mitigating the Impact of Psychophysical Effects During Adaptive Stimulus Selection in the P300 Speller Brain-Computer Interface*

Xinlin J. Chen, Leslie M. Collins, and Boyla O. Mainsah

Abstract— **Stimulus-driven brain-computer interfaces (BCIs), such as the P300 speller, rely on using sensory stimuli to elicit specific neural signal components called event-related potentials (ERPs) to control external devices. However, psychophysical factors, such as refractory effects and adjacency distractions, may negatively impact ERP elicitation and BCI performance. Although conventional BCI stimulus presentation paradigms usually design stimulus presentation schedules in a pseudo-random manner, recent studies have shown that controlling the stimulus selection process can enhance ERP elicitation. In prior work, we developed an algorithm to adaptively select BCI stimuli using an objective criterion that maximizes the amount of information about the user's intent that can be elicited with the presented stimuli given current data conditions. Here, we enhance this adaptive BCI stimulus selection algorithm to mitigate adjacency distractions and refractory effects by modeling temporal dependencies of ERP elicitation in the objective function and imposing spatial restrictions in the stimulus search space. Results from simulations using synthetic data and human data from a BCI study show that the enhanced adaptive stimulus selection algorithm can improve spelling speeds relative to conventional BCI stimulus presentation paradigms.**

Clinical relevance—**Increased communication rates with our enhanced adaptive stimulus selection algorithm can potentially facilitate the translation of BCIs as viable communication alternatives for individuals with severe neuromuscular limitations.**

I. INTRODUCTION

Brain-computer interfaces (BCIs) have broad applications for restoring or replacing neural output that has been lost due to injury or disease, including nerve stimulation, limb prosthetic control, and communication. BCIs measure and analyze brain signals to convert a user's intent into commands to control external devices [1].

The P300 speller is a widely researched BCI for individuals with severe neuromuscular diseases, such as individuals with late-stage amyotrophic lateral sclerosis (ALS) [2]. Noninvasive P300 spellers rely on eliciting and detecting ERPs embedded in electroencephalography (EEG) data via sensory stimuli. Infrequent target stimuli are presented within a sequence of nontarget stimuli; the presentation of the target stimulus elicits an ERP. Users of the visual P300 speller focus on a target character within a grid layout while visual stimuli are presented as sequentially highlighted groups of characters, termed flash groups. For example, in the row-column (RC) stimulus presentation paradigm, groups of rows or columns of characters are randomly presented [2]. Fig. 1 shows a visual P300 speller interface with a highlighted row flash group. After each stimulus presentation, EEG data from a time window immediately after stimulus onset are processed. A decoding algorithm determines the user's target character by using a classifier to identify ERPs embedded in EEG data and selecting the character whose presentation pattern most closely matches the classifier's predictions.

Current BCI communication rates are relatively slow and impractical for everyday use due to their reliance on inherently noisy and nonstationary EEG data. Factors that may lower the signal-to-noise ratios (SNRs) of ERPs include fatigue and psychophysical effects arising from properties of the presented stimuli, such as refractory and adjacency distraction effects [3]. Refractory effects occur when a target stimulus is presented more than once in a short time frame, impacting the amplitude and latency of ERPs elicited with the subsequent target stimuli. Adjacency distractions occur when characters in the spatial vicinity of the target character, particularly neighboring characters, are presented.

The RC paradigm, which is the most widely used stimulus presentation paradigm in the BCI literature, is highly susceptible to refractory effects due to the possibility of short time intervals between target stimuli and to adjacency distractions inherent to presenting characters in row/column flash groups [4]. The checkerboard (CB) paradigm was developed to mitigate refractory effects and adjacency errors: a checkerboard overlay is used on the speller grid to design flash groups with nonadjacent characters and impose a minimum time interval between character presentations [4]. However, stimulus presentation schedules in these conventional BCI stimulus paradigms are generated pseudo-randomly, with limited considerations of how to choose the presented stimuli to enhance ERP elicitation and maximize BCI performance.

Recently, we developed a novel algorithm to adapt the BCI stimulus presentation schedule in real-time with an objective criterion that maximizes the amount of information about

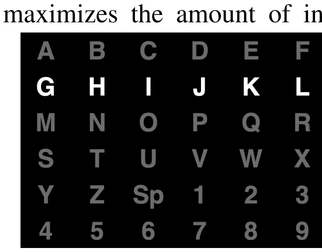

Fig. 1. Example P300 speller interface with a 6 × 6 spelling grid. The highlighted 'flash group' is a row of characters.

*Research supported by the National Institutes of Health (R21DC018347-01).

X. J. Chen, L. M. Collins, and B. O. Mainsah are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC USA.
Corresponding author: boyla.mainsah@duke.edu.

the user's intent that can be obtained with the presented stimuli given the current data [5]. However, preliminary results from an online BCI study showed that performance improvements with the adaptive stimulus selection paradigm were not consistent across users because spatial constraints were not imposed to minimize adjacency distractions during stimulus selection. Furthermore, the objective function used for stimulus selection did not incorporate the temporal dependency of EEG responses arising from refractory effects.

Here, we enhance our adaptive stimulus selection algorithm to mitigate refractory effects and adjacency distractions via modifications to the objective function and stimulus search space, respectively. We present results from simulations comparing BCI performance with our enhanced adaptive stimulus selection algorithm and with conventional stimulus presentation paradigms.

II. Methods

Signal analysis and simulations were performed in MATLAB© (Mathworks, Inc.; Natick, MA, USA). Statistical analysis was performed in R.

A. Bayesian Dynamic Stopping

The P300 speller estimates the character a user intends to spell by using a classifier to distinguish between EEG responses to target and nontarget stimuli. After each stimulus presentation, a time window of the user's EEG data is used to extract a feature vector, to which a classifier is applied to generate a score. The classifier score is used to update a character scoring function that evaluates the probability of each potential BCI character being the user's target character, given the current data. After a certain amount of data is collected, the character that maximizes the scoring function is spelled. In this work, we use a Bayesian dynamic stopping (DS) algorithm [6] as a character scoring function, and maintain a probability distribution over the possible character choices. After each stimulus presentation, the naïve Bayesian algorithm updates character probabilities accordingly:

$$P(c_i = C^*|S_t, \boldsymbol{x}_{t-1}, x_t) = \frac{P(c_i = C^*|S_{t-1}, \boldsymbol{x}_{t-1})p(x_t|c_i = C^*, S_t)}{\sum_m P(c_m = C^*|S_{t-1}, \boldsymbol{x}_{t-1})p(x_t|c_m = C^*, S_t)}, \quad (1)$$

$$p(x_t|c_i = C^*, S_t) = \begin{cases} l1(x_t), c_i \in S_t \\ l0(x_t), c_i \notin S_t \end{cases}, \quad (2)$$

where t is the time index; $P(c_i = C^*|S_{t-1}, \boldsymbol{x}_{t-1}) := P_{t,i}$ is the prior probability of the ith character being the target character C^* given current flash group, S_t, and past classifier scores, $\boldsymbol{x}_{t-1} = [x_1, x_2, ..., x_{t-1}]$; $P(c_i = C^*|S_t, \boldsymbol{x}_{t-1}, x_t)$ is the posterior probability; $p(x_t|c_i = C^*, S_t)$ is the likelihood of generating the current classifier score x_t, conditioned on c_i being the target character and appearing in the flash group S_t; and $l0$ and $l1$ are class-conditional classifier score probability density functions (pdfs) for nontarget and target responses, respectively. A dynamic stopping criterion terminates data collection when a character reaches the probability threshold (p_{th}) or when a data collection limit is reached. Therefore, the amount of data collection prior to character selection varies based on the BCI's level of confidence in estimating the target character.

With a naïve assumption (1), the data likelihood function in (2) does not model the temporal dependency of target classifier scores. We therefore consider a nonnaïve Bayesian algorithm where the target classifier scores are dependent on the interval between character presentations. Let [...TNN**T**...] represent a target character's presentation pattern with a character-to-character interval (CCI) of 3, where N denotes a nontarget stimulus; T denotes a target stimulus; and **T** denotes the target stimulus under consideration. Extending (2), the data likelihoods in a nonnaïve Bayesian algorithm are assigned based on the character's CCI accordingly:

$$p(x_t|c_i = C^*, S_t) = \begin{cases} l1(x_t, CCI_{i,t}), c_i \in S_t \\ l0(x_t), c_i \notin S_t \end{cases}, \quad (3)$$

where $CCI_{i,t}$ is the CCI of character c_i present in flash group S_t, assuming that c_i is the target character at time index t, and $l1(x_t, CCI_{i,t})$ is a CCI-specific pdf.

CCIs can take any positive integer value. To reduce model complexity, we grouped CCIs to estimate binned CCI-specific pdfs for target classifier scores. Fig. 2a shows estimated CCI-naïve pdfs obtained from a BCI user, and Fig. 2b shows CCI-specific target pdfs from the same user. In Fig. 2b, the pdf of the bin with the lowest CCIs is closest to the nontarget pdf, reflecting refractory effects. These effects may be mitigated if CCI-specific pdfs assign a higher likelihood of generating low scores to target stimuli with low CCIs, e.g., for a classifier score $x = 0$ and $CCI = 1$, the associated likelihood for the nonnaïve case, $l1(x = 0, CCI = 1)$ (Fig. 2b) is greater than for the naïve case, $l1(x = 0)$ (Fig. 2a).

B. Adaptive Stimulus Selection

Objective Function: After each stimulus presentation, the BCI extracts information about the target character from the user's EEG response to the presented stimulus, which is condensed into a classifier score. In our prior work, we developed an algorithm to optimize BCI stimulus selection by maximizing the amount of information about the target character that can be obtained from a future flash group given the current data [5]. Mutual information (MI) is a measure in information theory that quantifies the average reduction in uncertainty about one random variable that can be obtained by knowledge of a second random variable [8]. A future flash

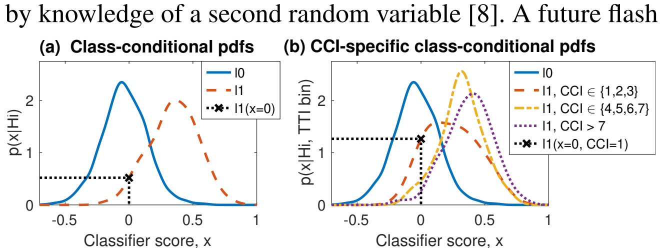

Fig. 2. Probability density functions (pdfs) of nontarget ($l0$) and target ($l1$) classifier scores estimated from training data from a BCI user [7], assuming the Bayesian algorithm is: (a) character-to-character interval (CCI)-naïve; and (b) CCI-nonnaïve, i.e., separates target classifier scores by CCI.

group is adaptively selected by optimizing a MI function [5]:

$$S_{t+1}^{*} = \arg_{S_{t+1}^{h}} \max I(X_{t+1}^{h}; C^{*}|\boldsymbol{x}_t, \boldsymbol{S}_{t+1}^{h}), \quad (4)$$

where $I(X_{t+1}^{h}; C^{*}|\boldsymbol{x}_t, \boldsymbol{S}_{t+1}^{h})$ is the MI between the target character C^* and the classifier score X_{t+1}^{h} generated by a hypothetical future flash group, S_{t+1}^{h}, conditioned on the future sequence of flash groups, $\boldsymbol{S}_{t+1}^{h} = [S_1, ..., S_t, S_{t+1}^{h}]$ and the current sequence of classifier scores, $\boldsymbol{x}_t$; and S_{t+1}^{*} is the flash group that maximizes the MI function.

Given class-conditional pdfs estimated from a user's training data, the MI function with the naïve Bayesian algorithm can be parameterized solely by the prior probabilities of characters in a future flash group [5]. Fig. 3a shows the MI functions associated with the pdfs from Fig. 2a: the stimulus selection process favors flash groups with characters whose sum of prior probabilities are close to 0.5.

To generalize for the nonnaïve case using (3), we define an *enhanced* MI function that is parameterized by the prior probabilities of characters in a future flash group, which are grouped by CCI accordingly:

$$\begin{aligned} I(X_{t+1}^{h}; C^{*}|x_t, \mathbf{S}_{t+1}^{h}) = \int_{-\infty}^{\infty} \Sigma_{n=1}^{N} P1_{t,n} l1(x_{t+1}^{h}, n) \\ \times \log\left(\frac{l1(x_{t+1}^{h}, n}{l0(x_{t+1}^{h})(1-\sum_{n=1}^{N} P1_{t,n}) + \sum_{n=1}^{N} P1_{t,n} l1(x_{t+1}^{h}, n)}\right) \\ + (1 - \Sigma_{n=1}^{N} P1_{t,n}) l0(x_{t+1}^{h}) \\ \times \log\left(\frac{l0(x_{t+1}^{h})}{l0(x_{t+1}^{h})(1-\sum_{n=1}^{N} P1_{t,n}) + \sum_{n=1}^{N} P1_{t,n} l1(x_{t+1}^{h}, n)}\right) dx_{t+1}^{h}, \end{aligned} \quad (5)$$

$$P1_{t,n} := P1_{t,n}(S_{t+1}^{h}) = \Sigma_{\forall i: c_i \in S_{t+1}^{h} \cap CCI_{i,t} = n} P_{t,i}, \quad (6)$$

where N is the number of binned CCI-specific pdfs; and $P1_{t,n}(S_{t+1}^{h})$ is the sum of probabilities at time index t for characters in the flash group S_{t+1}^{h} with a CCI of n.

By accounting for CCIs in (6), the stimulus selection process is inherently biased towards flash groups with characters whose CCI values are likely to elicit ERPs with high SNRs. Fig. 3b shows the MI function of the pdfs in Fig. 2b. In this case, stimulus selection favors flash groups with characters that have a sum of prior probabilities close to 0.5, particularly those with higher CCIs falling in the second and third bins.

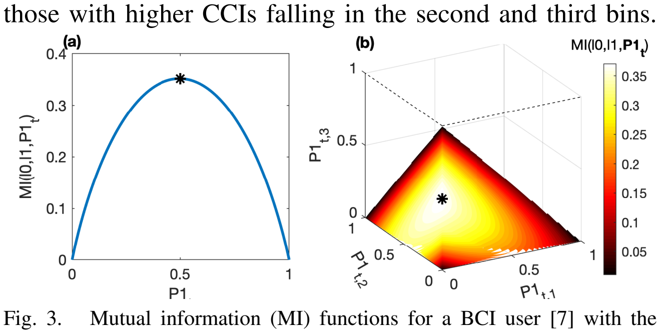

Fig. 3. Mutual information (MI) functions for a BCI user [7] with the class-conditional probability density functions in Fig. 2. (a) Target classifier scores are grouped; $P1_t$ denotes the sum of prior probabilities of characters in a hypothetical future flash group. (b) Target classifier scores are binned by character-to-character interval (CCI); $P1_{t,n}$ denotes the sum of prior probabilities of characters in a hypothetical future flash group for CCI bin n. The sums that maximize MI in each case are denoted by an asterisk (*).

Adaptive stimulus selection paradigms with flash groups selected using (4) were implemented for the P300 speller by defining the stimulus search space accordingly.

1) Adaptive Row-Column: The search space is restricted to row or column flash groups. Character selection occurs in two stages: the selection of the row flash group, followed by the selection of the column flash group, or vice versa. The stopping probability during each selection stage is set to $p_{th}^{*} = \sqrt{p_{th}}$, where p_{th} is the stopping probability.

2) Adaptive Diffuse: A greedy approach with spatial constraints is used to select flash groups with up to M characters. 8-adjacent characters to existing flash group members are excluded from the search space.

C. P300 Speller Simulations

P300 speller simulations were performed with classifier scores generated synthetically or from bootstrapped data from a previous study [7] using the framework developed in [9], which has been validated with data from online BCI studies. For each simulation, 10,000 runs of character selections were performed. In each run, the target character was uniformly sampled from a 6×6 grid and flash groups were selected based on the specified stimulus paradigm. The maximum flash group size in the adaptive diffuse paradigm was 6 characters. The Bayesian DS algorithm was used for character selection, with character probabilities initialized uniformly and a stopping threshold, $p_{th} = 0.9$. Data collection was limited to 72 stimulus presentations. Statistical significance was tested using the paired, two-sided Wilcoxon signed-rank test ($p < 0.05$).

For simulations with synthetic data, classifier scores were assumed to be normally distributed with parameters $d = \frac{\mu_1 - \mu_0}{\sigma}$, where d is the detectability index (DI), which quantifies the between-class discriminability of the scores; μ_0 and μ_1 are the means of the target and nontarget pdfs; and σ is the assumed common standard deviation. To reflect varying performance levels, a range of DI values was used.

To obtain more realistic performance estimates, we conducted simulations with data from a prior P300 speller study [7] that recruited 8 subjects with ALS. EEG data were recorded from 8 centro-parietal electrode channels [10]. The stimulus duration was 187.5 ms and the inter-stimulus interval was 62.5 ms. Each subject spelled 60 characters using the RC paradigm on a 6×6 grid, corresponding to 5040 observations, split evenly into training and testing data. Features were extracted as in [11] from the training data and used to train a stepwise linear discriminant analysis (SWLDA) classifier. The classifier's scores for the training data were used to generate kernel density estimates of target and nontarget pdfs [6]. During simulations, the trained classifier's scores for the testing data were bootstrapped.

Constraints were imposed in simulations to reflect real-world physiological and system conditions. During P300 speller use, EEG data processing causes a delay between stimulus presentation and classifier score generation. Thus, an observation delay, δ, was imposed; data from time step $t - \delta$ were used to select the flash group at time step t. In

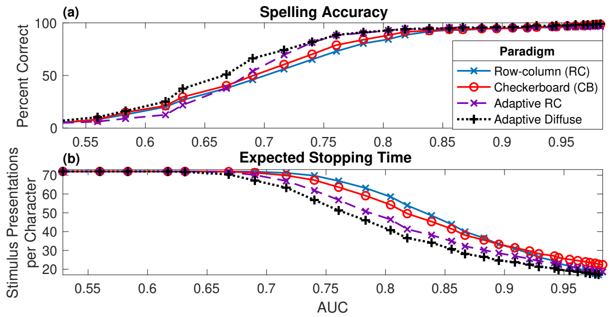

Fig. 4. Results from simulations with synthetic data. (a) Spelling accuracy and (b) expected stopping times of conventional and adaptive stimulus presentation paradigms assuming a naïve Bayesian algorithm. Performance is presented as a function of the area under the receiver operating characteristic curve (AUC). Minimum CCIs were imposed for the adaptive paradigms.

simulations with BCI data, refractory effects were modelled by sampling target classifier scores binned by CCI (Fig. 2b).

III. Results

In our P300 speller simulations, we evaluated overall spelling accuracy and speed (expected stopping time, in stimulus presentations/character). Fig. 4 shows results from simulations with synthetic classifier scores, presented as a function of the area under the receiver operating characteristic curve (AUC). Overall, the use of naïve adaptive paradigms improved spelling accuracy and speed relative to conventional RC and CB. Of the adaptive paradigms, the adaptive diffuse (AD) paradigm achieved the best performance, likely due to its larger stimulus search space.

Fig. 5 shows simulation results for the RC and AD paradigms with bootstrapped data from a BCI study [7], sorted by the classifier AUCs associated with each subject's testing data. Statistical measures and analyses are summarized in Table I. To assess the ability of the enhanced MI function to mitigate refractory effects at low CCIs, the AD paradigm was simulated with *no* imposed minimum CCIs. Compared to RC, a notable decrease in accuracy was observed with the naïve AD paradigm in 4 subjects; however, the decrease in accuracy was not significant ($p = 0.055$). Further analysis revealed that the naïve AD paradigm favors the inclusion of potential target characters, even at low CCIs, in flash groups. Thus, a decrease in spelling accuracy may arise from nontarget classifier scores being less discriminable from target scores at low CCIs than from target scores at high CCIs (e.g., Fig. 2b). Refractory effects appeared to be mitigated by the nonnaïve AD paradigm, which yielded comparable or improved spelling accuracy at significantly faster speeds ($p < .01$) relative to RC. During flash group selection, our nonnaïve algorithm incorporates character probabilities and presentation history to maximize MI. Consequently, characters with high CCIs are often favored (Fig. 3b), and performance improved overall, relative to naïve AD and RC.

These results show that adaptive stimulus selection can potentially improve BCI performance even under temporal and spatial constraints. During stimulus selection, psychophysical effects, i.e., refractory effects and adjacency distractions, can be mitigated by using a CCI-nonnaïve algorithm and by selecting flash groups without adjacent characters, respectively. Future work will involve testing the enhanced adaptive stimulus selection algorithm in online experiments.

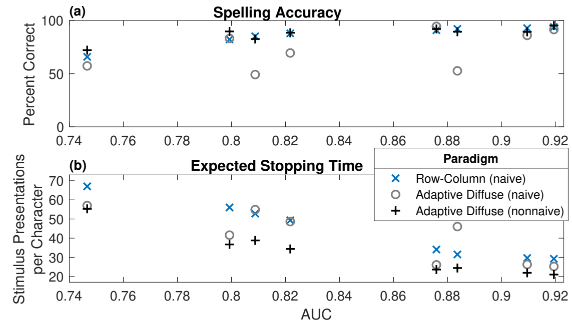

Fig. 5. Results from simulations with bootstrapped data obtained from 8 BCI users with ALS recruited by a prior study [7]. (a) Spelling accuracy and (b) expected stopping times for row-column and adaptive diffuse (AD) paradigms using naïve and nonnaïve Bayesian algorithms. Subject performance is represented by AUCs of classifier scores for their testing data. Minimum CCIs were *not* imposed for the AD paradigms.

TABLE I

Summary of Results from Simulations with BCI User Data

Paradigm	Accuracy	Expected Stopping Time
RC	0.86±0.09	43.71±14.43
AD (naïve)	0.73±0.18 (p=0.055)	40.72±13.20 (p=0.313)
AD (nonnaïve)	0.88±0.07 (p=0.742)	**32.03±11.76 (p <.01)**

The mean (μ) and standard deviation (σ^2) of each performance measure are reported as $\mu \pm \sigma^2$. For the adaptive diffuse (AD) paradigm, p-values are reported from comparisons with the row-column (RC) paradigm. Statistically significant performance improvements are bolded.

References

[1] J. Wolpaw and E. W. Wolpaw, "Brain-Computer Interfaces: Something New under the Sun," in *Brain-Computer Interfaces: Principles and Applications*. New York, NY: Oxford University Press, 2012, ch. 1.

[2] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr Clin Neurophysiol*, vol. 70, no. 6, pp. 510-23, 1988.

[3] S. Lees et al., "A review of rapid serial visual presentation-based brain–computer interfaces," *J Neural Eng*, vol. 15, no. 2, p. 021001, 2018.

[4] G. Townsend et al., "A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clin Neurophysiol*, vol. 121, no. 7, pp. 1109-20, 2010.

[5] B. Mainsah, D. Kalika, L. Collins, S. Liu, and C. Throckmorton, "Information-based Adaptive Stimulus Selection to Optimize Communication Efficiency in Brain-Computer Interfaces," in *Adv Neural Inform Process Sys*, Montréal, Canada, Dec. 2018, vol. 31.

[6] C. S. Throckmorton, K. A. Colwell, D. B. Ryan, E. W. Sellers, and L. M. Collins, "Bayesian approach to dynamically controlling data collection in P300 spellers," *IEEE Trans Neural Syst Rehabil Eng*, vol. 21, no. 3, pp. 508-17, 2013.

[7] D. B. Ryan, K. A. Colwell, C. S. Throckmorton, L. M. Collins, K. Caves, and E. W. Sellers, "Evaluating Brain-Computer Interface Performance in an ALS Population: Checkerboard and Color Paradigms," *Clin EEG Neurosci*, vol. 49, no. 2, pp. 114-121, 2018.

[8] T. M. Cover and J. A. Thomas, "Entropy, Relative Entropy, and Mutual Information," in *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006, ch. 2.

[9] B. O. Mainsah, L. M. Collins, and C. S. Throckmorton, "Using the detectability index to predict P300 speller performance," *J Neural Eng*, vol. 13, no. 6, p. 066007, 2016.

[10] D. J. Krusienski et al., "A comparison of classification techniques for the P300 Speller," *J Neural Eng*, vol. 3, no. 4, pp. 299-305, 2006.

[11] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J Neurosci Methods*, vol. 167, no. 1, pp. 15-21, 2008.