

UCLA

UCLA Electronic Theses and Dissertations

Title

Characterizing P300 Speller Performance: The Effects of Stimulus Timing Features and Search for a Physiological Biomarker

Permalink

<https://escholarship.org/uc/item/3n76q09f>

Author

Lu, Jessica

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Characterizing P300 Speller Performance: The Effects of Stimulus Timing Features and Search
for a Physiological Biomarker

A thesis submitted in partial satisfaction
of the requirements for the degree Master of Science
in Biomedical Engineering

by

Jessica Lu

2012

ABSTRACT OF THE THESIS

Characterizing P300 Speller Performance: The Effects of Stimulus Timing Features and Search for a Physiological Biomarker

by

Jessica Lu

Master of Science in Biomedical Engineering

University of California, Los Angeles, 2012

Professor Nader Pouratian, Co-chair

Professor Xiao Hu, Co-chair

This thesis presents two studies that focus on improving P300 speller system efficiency by examining fundamental stimulus timing parameters and the possibility of a physiological biomarker to identify user performance. By analyzing four distinct stimulus presentation features, the first study reports that both stimulus-off time and interstimulus interval can significantly affect both accuracy and characters per minute. Optimization of stimulus timing parameters should not only consider accuracy rates, but also bit rates (i.e. characters per minute), considering that reductions in accuracy may be more than offset than the time saved resulting in overall improvements in system efficiency. Subsets of the entire group were compared to show

consistency of performance trends despite great variance among subjects. The next study explored the possibility of an EEG biomarker that distinguished performer differences. Multiple extensive analyses showed no strong correlation between power in the EEG signal and user performance in spelling the temporally associated letter or word.

The thesis of Jessica Lu is approved.

Jack Van Horn

Xiao Hu, Committee Co-chair

Nader Pouratian, Committee Co-Chair

University of California, Los Angeles

2012

Table of Contents

List of Figures	vi
List of Tables	viii
List of Acronyms	ix
Acknowledgments.....	x
CHAPTER 1: Introduction	1
CHAPTER 2: The Effects of Stimulus Timing Features on P300 Speller Performance.....	9
CHAPTER 3: Exploring a Potential Physiological Biomarker of P300 Speller Performance	26
CHAPTER 4: Conclusion.....	42
References.....	44

List of Figures

Figure 1. Average classification accuracy for all subjects grouped by (A) stimulus timing pattern, (B) ISI*, (C) ISI, (D) flash duration, and (E) ratio of flash duration to ISI. SEM error bars are included in (A)..... 15

Figure 2. Average classification accuracies grouped by timing pattern comparing good performers (A) to bad performers (B)..... 16

Figure 3. Average characters per minute for all subjects grouped by (A) stimulus timing pattern, (B) flash duration, (C) ISI*, (D) ISI, and (E) ratio of flash duration to ISI. SEM error bars are included for the largest CPM point of each average trend line..... 18

Figure 4. Average characters/min grouped by timing pattern comparing good performers (A) to bad performers (B)..... 19

Figure 5. Average group waveforms at Pz for each stimulus timing pattern. Waveforms for targets are represented by the solid black line and waveforms for non-targets are represented by the dotted line..... 20

Figure 6. Average classification accuracy (A-D) and CPM (E-H) under the 32-channel set (represented by the solid black line) compared to the 8-channel set (represented by the dashed black line) for each of the four patterns. 21

Figure 7. Accuracy curves for each word spelled by subject A. The area under each curve (AUC) was calculated and used as the accuracy measure for each word. 28

Figure 8a and b. Sections highlighted in gray indicate the time in the EEG signal used to calculate power values. 29

Figure 9. Power spectral density estimate in channel Cz of the top three performers (solid black) compared to the bottom three performers (dotted black). The low gamma (30-70Hz) frequency band is shaded in gray. 30

Figure 10. Gamma power in electrode Cz along the x-axis is plotted against a measurement of accuracy, AUC, for the initial six subjects. The data is fitted to a same-line model, ignoring subject grouping. 33

Figure 11. Gamma power in electrode Cz along the x-axis plotted against AUC along the y-axis for the test set of nine additional subjects. 33

Figure 12. Histogram of Kendall rank correlation coefficients along the x-axis and frequency in each ten percent interval along the y-axis grouped within each subject. 34

Figure 13. Histogram of Kendall rank correlation coefficients along the x-axis and frequency in each ten percent interval along the y-axis grouped within each set of subjects. The line marks the correlation coefficient for the true association. 35

Figure 14. Correlation plot graphing SWLDA classifier score against AUC accuracy measure. 36

Figure 15. ROC curve for SWLDA classifier with poor AUC value = 0.6251. 37

Figure 16. ROC curves for SWLDA classifier after each trial using data from all subjects. 40

List of Tables

Table 1. Stimulus timing patterns used for all six subjects.....	11
Table 2. Features selected (EEG band and electrode) and coefficients of the SWLDA model used in experiment 3.....	36
Table 3. P-values that are highlighted show significant correlations between power in the EEG signal prior to spelling a letter and the classification accuracy of the associated letter after each trial. SWLDA did not work in cells with “NaN” due to perfect classification accuracy.	38
Table 4. Pearson correlation coefficients and p-values after each trial using a leave-one-out SWLDA model built after the first set of flashes.	39

List of Acronyms

ALS	Amyotrophic lateral sclerosis
ANCOVA	Analysis of covariance
AUC	Area under the curve
BCI	Brain-computer interface
CBP	Checkerboard paradigm
CPM	Characters per minute
EEG	Electroencephalography
ERP	Event-related potential
FLD	Fisher's linear discriminant
GSVM	Gaussian kernel support vector machine
GUI	Graphics user interface
ISI	Interstimulus interval = [stimulus-on + stimulus-off time]
ISI*	Interstimulus interval* = stimulus-off time
LSVM	Linear support vector machine
MDS	Multi-dimensional scaling
MMI	Mindfulness meditation induction
PCA	Principal component analysis
PCM	Pearson's correlation method
PSD	Power spectral density
RCP	Row/column paradigm
ROC	Receiver operating characteristic
SCP	Single cell paradigm
SNR	Signal-to-noise ratio
SOA	Stimulus onset asynchrony = [stimulus-on + stimulus-off time], see ISI
SWLDA	Stepwise linear discriminant analysis
VEP	Visual-evoked potential

Acknowledgments

Chapter Two is a version of Lu J, Speier W, Hu X, Pouratian N. The Effects of Stimulus Timing Features on P300 Speller Performance. 2012. Submitted to Clinical Neurophysiology.

William Speier is a Biomedical Engineering graduate student at UCLA and a member of the Neurosurgical Brain Mapping and Restoration Laboratory. The principal investigators, Xiao Hu and Nader Pouratian, are associate and assistant professors, respectively, in the Department of Neurosurgery at UCLA.

CHAPTER 1: Introduction

Originally introduced by Farwell and Donchin (1988), the P300-based brain-computer interface (BCI) is a device that uses brain signals to provide a non-muscular communication channel for people with amyotrophic lateral sclerosis (ALS), high spinal cord injuries, and brainstem strokes. These patients share the devastating fate of the “locked-in” condition, in which cognitive function is maintained, but all voluntary muscle control and ability to communicate is lost.

The current solution to allow non-muscular communication is to enable the task of spelling using a row/column paradigm (RCP) (Farwell and Donchin, 1988). Users spell words one letter at a time by attending to the desired character in a 6x6 matrix of characters displayed on a computer screen. Electroencephalography (EEG) signals are recorded as controlled stimulus events, typically flashes, intensify the rows and columns at random, totaling 12 intensifications for each set. The rare set of events includes the two flashes, one row and one column, that contain the target character, while the frequent set of events includes the 10 flashes not containing the target character, together constituting an oddball paradigm. Items in the rare set of a sequential series of randomly presented stimuli will elicit a P300 event-related potential (ERP), which is a positive deflection in the EEG over the parietal cortex that appears about 300 ms after an “oddball” stimulus and has proven to be a reliable response for controlling a BCI (Townsend et al., 2010).

Much research has aimed to optimize the P300 speller to improve efficiency and increase widespread use of the system. A current status review (Mak et al., 2011) examined issues such as potential users, recording methods, stimulus presentation paradigms, feature extraction and classification algorithms, and applications of the speller. The following will highlight the main points of the review, as well as provide additional research findings specifically emphasizing variations in stimulus presentation paradigms, which will be the focus of the first chapter.

Potential Users.

While patients who require this technology could greatly benefit from the system, most P300-BCI related publications report results from healthy adult subjects leaving limited studies with disabled users. Many studies have addressed variability among different healthy subjects, but the question of whether results from healthy subjects can be generalized to disabled patients still needs further investigation. Though designed to rely on the P300 signal for classification, the ‘P300’ speller that typically uses visual stimuli has been shown to require visual-evoked potentials from the occipital lobe as well as the P300 response from the parietal lobe (Brunner et al., 2010). Patients with ALS present disease progression that could prevent them from being able to use the P300-BCI due to late-stage ALS symptoms such as loss of ocular motor function and ventilator-dependence. In such cases, perhaps a BCI system with auditory or both auditory and visual stimuli (Belitski et al., 2011) would work better. Therefore, instead of selecting subjects based on BCI system capabilities, studies should select devices based on the individual’s physical and cognitive capabilities and needs.

Recording Methods.

Current EEG recording methods allow efficient EEG data collection, but they mostly require inconvenient electrode application, sometimes including scalp skin abrasion, conductive gel applied to the scalp, fully wired EEG cap, and precise positioning of electrodes on the head. An easier, more user-friendly hardware replacement would be dry electrodes or a wireless headset with dry electrodes. Reducing the number of electrodes would also decrease bulkiness and help usability. An ideal solution may be to develop an individualized electrode montage by reducing the set of channels from a full set (10-20 systems) for each user.

Stimulus Presentation Paradigm.

The current visual paradigms used for the P300-BCI include the row-column paradigm (RCP) by Farwell and Donchin (1988), the single cell paradigm (SCP), and the checkerboard paradigm (CBP) by Townsend et al. (2010). The RCP is the standard and most commonly used paradigm that involves stimulus events or flashes occurring in each row and column at random, while the user attends to the target character. The target is then selected by matching the elicited P300 response in the row and column containing that character. The SCP, in contrast, flashes a single character in the matrix at a time, rather than flashing the entire row or column. As expected, reports show that the single cell paradigm results in a larger P300 response, but decreased system accuracy and speed (Guan et al., 2004; Guger et al., 2009).

A newer flashing schema, the checkerboard paradigm (CBP), was developed in order to prevent adjacency-distraction errors and double-flash errors associated with the row-column paradigm. Adjacency-distraction errors result from attention and responses to stimuli in adjacent rows and columns, and double-flash errors involve the target event occurring twice in a row (for example, when the target letter's row flashes right after its column flashes). The double flash

either causes no P300 response or overlapping responses. The CBP involves an 8x9 matrix that is superimposed on a black and white checkerboard, never seen by the subjects. The items in the white boxes make a virtual 6x6 matrix, and the items in the black boxes make another 6x6 matrix. Each row and column in the virtual matrices flashes similar to the RCP paradigm. The corresponding stimuli appear to the user as 6 random boxes flashing at a time rather than a row or column flashing. In eliminating adjacency and double flash errors, the CBP results in better P300-BCI performance. With five complete repetitions, which included ten target flashes, the mean online accuracy improved from 77% using the RCP to 92% using the CBP. Mean bit rate was also significantly increased from 19 bits/min for the RCP up to 23 bits/min for the CBP (Townsend et al., 2010).

Other modifications and aspects of stimulus presentation have also been shown to affect performance. Two methods of structuring stimulus presentation were explored in Allison and Pineda (2006). The first one presented stimuli more quickly, enabling the user to generate more cognemes per minute, by manipulating the stimulus onset asynchrony (SOA), which is the time interval between the start of one stimulus event and the start of the next event, also known as interstimulus interval (ISI). The second method resulted in requiring fewer cognemes to convey a message by altering the pattern of stimulus presentation. (*Cogneme* refers to the subject's response to “/attend to the event/” or “/ignore the event/”.) Specifically, a multiple flash condition, which involves half the columns or rows flashing at a time, was tested in contrast to the standard single flash approach. Results showed that faster SOAs were associated with smaller P300 amplitudes and that the multiple flash condition produced stronger ERPs at faster speeds. The tradeoff between stimulus presentation speed and ERP measures was discussed, but overall

the two general variations tested in this study showed potential to improve information throughput (Allison and Pineda, 2006).

While most studies have continued to work on improving the P300 speller using visual stimuli, some have also explored different modalities, auditory and tactile stimuli, to elicit the P300 response. This BCI system by description was designed to rely primarily on the P300-evoked potential and minimally on other evoked potentials, such as the visual-evoked potential (VEP). Brunner et al. (2010) aimed to determine whether the 'P300' speller depends on eye gaze, which would have definite clinical implications. The P300-BCI's target population includes patients with amyotrophic lateral sclerosis, a disease which can often lead to complete paralysis including impaired or lost eye movement. The experiment consisted of two conditions: 1) the 'letter' condition, where subjects focused their eye gaze on the target letter, and 2) the 'center' condition, where subjects focused their eye gaze on a fixation cross in the center of matrix. The results showed that people performed significantly better under the 'letter' condition and that under the 'center' condition, accuracy declined as the distance of the target character from the center increased. Excluding VEP components during the 150 to 350ms post stimulus greatly decreased classification accuracy and showed that performance depends on VEPs as well as the P300 response.

Proving that the 'P300' speller does depend on eye gaze strongly encourages non-visual P300-BCI systems. Mak et al. (2011) reviewed several non-visual P300-BCI studies and concluded that while the research is promising, their development is still quite recent and not as advanced as the visual P300 speller system. Furdea et al. (2009) and Schreuder et al. (2010) tested using auditory evoked responses in the P300-BCI, and both showed either lower accuracy

or bit rate. Brouwer and van Erp (2010) designed a tactile P300-BCI that also resulted in lower accuracy results compared to visual systems.

The audio-visual speller proposed by Belitski et al. (2011) involves flipping the classic matrix to allow use of the same interface for visual, auditory, or simultaneous visual and auditory stimuli, whichever is necessary for the user's current needs, such as those defined by the progression of ALS. For each condition that was tested (visual, audio, and audio-visual), the initial columns were presented, and then the matrix was rotated 90 degrees to make the rows into columns, so that the horizontal spatial setup of the speakers used to give auditory cues corresponded with the horizontal arrangement of the character columns. The audio-visual condition resulted in significantly better performance compared to the other conditions. This study provides a novel paradigm for stimulus presentation and a system that can be used throughout a patient's increasing disability due to ALS.

Feature Extraction and Classification Algorithms.

Feature extraction and classification algorithms are needed in the P300-BCI system in order to detect the elicited ERP, which can be difficult due to large signal-to-noise ratio (SNR). The goal of feature extraction is to extract the most distinctive components of the ERP representing the user's intent. These components or features are then classified by a specific algorithm into an appropriate output. Krusienski et al. (2006) tested four linear classification methods: Pearson's correlation method (PCM), Fisher's linear discriminant (FLD), stepwise linear discriminant analysis (SWLDA) and a linear support vector machine (LSVM); as well as one nonlinear method: Gaussian kernel support vector machine (GSVM) for optimal P300 speller classification. While all five methods provided satisfactory results, SWLDA and FLD showed the best overall performance compared to other methods. By using the univariate statistics of a

fixed model, the simplicity of PCM allowed rapid training and implementation. PCM, however, showed the lowest average performance because it does not account for covariance between features, which is considered in FLD, improving its performance over PCM. SVMs can perform better than other methods with less training data, but their disadvantages include a complex algorithm and slower training. Added complexity of the nonlinear method (GSVM) also proved to be unnecessary. SWLDA remains to be the standard classification technique used for the P300 speller with potential advantages over FLD in its ability to selectively limit the input feature space.

Applications.

Further suggestions for applications of this P300-BCI include adapting the system for logographic writing systems such as Chinese and using the BCI as a switch for assistive technology software or environmental control devices (Mak et al., 2011). Innovative applications could greatly advance the usability of P300-BCI technology.

Limitations to the P300-BCI system can be found within many factors. After exploring many studies with various approaches to improving stimulus presentation, this study aims to investigate the fundamental level of stimulation presentation. Despite numerous examinations of these factors affecting P300 speller performance, the impact of stimulus presentation parameters remains incompletely understood. Maintaining the original RCP paradigm, this study begins with manipulating elementary parameters involved in stimulus presentation to optimize the performance of the P300 system. Currently, the training process remains repetitive and slow. Altering variables such as the number of trials, or sets of flashes, and the interstimulus interval to decrease the time and strain of the training process can enhance the system's efficiency and

practicality. Chapter 1 investigates the effects of four distinct stimulus presentation parameters (stimulus-off time [ISI*], interstimulus interval [ISI], flash duration, and flash-duration: ISI ratio) on the accuracy and efficiency of the P300 speller performance. Additional analyses, such as comparing differences between user performances and electrode sets, were completed. Motivated by the differences seen in accuracy results among users, chapter 2 studies the possibility of a physiological biomarker that is associated with spelling accuracy. Using multiple methods of analysis, the possible correlation between power in different EEG bands in the signal preceding the start of the BCI spelling task and resulting classification accuracies was thoroughly investigated.

CHAPTER 2: The Effects of Stimulus Timing Features on P300 Speller Performance

1. Introduction

A primary goal in BCI research is to increase accuracy and speed. Traditionally, a fixed, or preset, number of evoked responses are averaged in order to increase the signal-to-noise ratio (SNR). In the most commonly used “rows and columns” version (RCP) of the P300 speller, each row and column flashes once in every trial (i.e., 12 flashes for a 6x6 character matrix) with up to 15 trials repeated for each selection. This sequence results in relatively slow typing speed, motivating many studies to optimize system efficiency by changing grid size (Sellers et al., 2006), interstimulus interval (ISI) (McFarland et al., 2003; McFarland et al., 2011), signal processing methods (Krusienski et al., 2006; Kaper et al., 2004), and flashing paradigm (Townsend et al., 2011; Jin et al., 2011).

Despite various methods of improving user performance, the effect of stimulus presentation parameters, including flash duration and interstimulus intervals, on the efficiency and accuracy of the original RCP remain relatively under-characterized. Modest improvements in RCP bit rate have been achieved by modifying stimulus presentation parameters and thereby reducing the time required per trial (e.g., shorter ISI) while maintaining or improving system accuracy. Most such studies, however, have used fixed numbers of trials or sets of flashes. We hypothesize that greater improvements in efficiency could be achieved by determining stimulus presentation parameters that optimize bit rate (or characters per minute, CPM) rather than focusing on accuracy exclusively. This approach theoretically allows selection of stimulus presentation

parameters that may in fact require more time per trial (e.g., longer ISI), but could still improve system efficiency by requiring fewer trials per character selection.

We identified four potentially important and modifiable stimulus presentation parameters: flash duration, ISI (time from the onset of one flash to the next flash), ISI* (stimulus-off time or [ISI – flash duration]), and ratio between flash duration and ISI [Ratio]. Whereas longer ISI has been associated with improved accuracy, flash duration and ISI* have been reported to not affect accuracy (McFarland et al., 2011). However, the impact of these factors on system performance without a fixed number of trials has not been studied systematically. Using the standard stepwise linear discriminant analysis (SWLDA) and standard row/column paradigm (RCP) in an offline analysis, this study examined effects of different stimulus presentation parameters on speed and accuracy by using four specifically designed stimulus timing patterns (Krusienski et al, 2006).

2. Methods

Six adult males ages 22 to 35 participated in this IRB-approved study. Only one subject had prior BCI experience.

2.1 Data Collection

The experiment data was collected in a single two-hour session with each subject. EEG signals were recorded with a 32-electrode cap (Fpz, Fz, FC1, FCz, FC2, FC4, FC6, C4, C6, CP4, CP6, FC3, FC5, C3, C5, CP3, CP5, CP1, P1, Cz, CPz, Pz, POz, CP2, P2, PO7, PO3, O1, Oz, O2, PO4, PO8) and referenced to the right or left mastoid. The signals were amplified with two g.tec (Guger technologies) 16-channel USB biosignal amplifiers, digitized at 256 Hz, and filtered between 0.1 and 60 Hz. The BCI2000 general-purpose system for BCI research was used for data acquisition and experimental design (Schalk et al., 2004).

Pattern	Flash duration (ms)	ISI* (ms)	ISI (ms) = Flash dur + ISI*	Ratio = Flash dur: ISI
(A)	31.25	31.25	62.5	1:2
(B)	31.25	62.5	93.75	1:3
(C)	31.25	93.75	125	1:4
(D)	62.5	62.5	125	1:2

Table 1. Stimulus timing patterns used for all six subjects.

Table 1 shows the four timing patterns used to vary the parameters of flash duration (ms), ISI* (ms), ISI (ms), and ratio between flash duration and ISI. In order to identify the effect of flash duration, pattern D was designed to have twice as long of a flash duration as patterns A, B, and C. The effects of ISI* and ISI can be seen by comparing patterns A, B, and C, in which the flash duration is held constant. ISI* is held the same in patterns B and C, while ISI is the same in patterns C and D, to allow comparisons between ISI and ISI* effects. The ratio between flash duration and ISI equaled 1:2 for patterns A and D and differed in patterns B and C, allowing assessment of performance based on ratio. While the planned comparison of effects are not completely independent and confound interpretation to some extent, patterns were selected to enable efficient data collection while providing constraints across each timing feature that might provide insight into each feature's impact on system performance.

Each subject was instructed to stare at an LCD monitor about 1 m away with a 6x6 matrix display of characters and focus on the target character, attending to the number of times it flashed. Each character flashed within an entire row or column in the RCP introduced by Farwell and Donchin (1988). Each word to be spelled consisted of five letters and was presented above the matrix with the current target letter shown in parentheses at the end of the word. Every row

and column flashed once randomly (12 flashes) per trial. Fifteen trials (or repetitions) were conducted per letter selection. No feedback was given during these trials. In order to ensure attentiveness, subjects were asked to count each time the target letter flashed. Subjects spelled a total of 15 words in the single session, three words under each pattern except for pattern C with a flash duration of 31.25ms and ISI* of 93.75ms, which had six words.

2.2 Data Analysis

BCI2000 was used for data acquisition and analysis was performed offline using MATLAB (Schalk et al., 2004). SWLDA, a classification algorithm that selects a set of signal features to include in a discriminant function, was used to extract the P300 event-related potentials (ERPs) (Krusienski et al., 2006). We used leave-one-out cross-validation by letter to assess accuracy. Under each pattern, the linear classifier was trained on 14 or 29 letters and tested on the excluded letter, and then repeated in the set of letters spelled in each stimulus timing pattern. The signals in the training set are assigned one of two class labels based on whether the signals corresponded to flashes containing the target character or non-target characters. Each new signal is reduced to a score that reflects how similar it is to the target class. The algorithm uses ordinary least-squares regression to predict class labels for the training set. It adds the features that are most significant in the forward stepwise analysis and removes the least significant features in the backward analysis step. These steps are repeated until either a set number of features is met or it reaches a state where no features are added or removed (Krusienski et al., 2006).

To assess whether the observed results could be generalized to the previously described and published reduced 8-electrode set described by Sharborough et al. (1991), results were also produced for a subset of eight channels (Fz, Cz, PO3, Pz, PO4, PO7, PO8, Oz) in an identical

manner as that described for the entire channel set, in order to compare results with that obtained with the larger 32-channel set.

The number of characters per minute (CPM) was computed for each trial as the difference between the number correct (P) and the number wrong (1-P) divided by the total time in minutes, excluding the time between trials (McFarland et al., 2010). Since $P-(1-P) = 2P-1$,

$$CPM = \begin{cases} \frac{2P-1}{time}, & CPM > 0 \\ 0, & CPM \leq 0 \end{cases}.$$

CPM, also known as practical bit rate from Townsend et al. (2011), accounts for error correction, which requires a minimum of two additional selections: a backspace then a correct selection.

3. Results

3.1 Effect of Stimulus Timing Patterns on Accuracy

Analysis of the entire dataset using a mixed model revealed a significant effect of trials or sets of flashes ($F(14,354) = 69.002, p < 0.0001$) and pattern used ($F(3,354) = 117.512, p < 0.0001$), as well as a significant interaction between these factors ($F(3,354) = 3.931, p < 0.009$). Performance under each of the four patterns was further investigated using pairwise comparisons, showing that each pair showed a significant difference in performance ($p < 0.0001$) except for patterns C and D, which both had an ISI of 125ms with differing flash durations of 31.25ms and 62.5ms. Average accuracy for the pattern with the shortest flash duration of 31.25ms and ISI of 62.5ms was lower than performance for the pattern with a 31.25ms flash duration and 93.75ms, which had lower accuracy than the two patterns with highest accuracy values (Figure 1A).

3.1.1 Effect of Stimulus Presentation Parameters on Accuracy

*ISI**. The parameter *ISI** ($F(2,264)=157.321$, $p<0.0001$) significantly affected accuracy across the fifteen trials. Pairwise comparisons showed that performance under each of the three *ISI** values: 31.25ms, 62.5ms, and 93.75ms, were significantly different. The longest *ISI** of 93.75ms was associated with the highest accuracy rates, while the shortest *ISI** of 31.25ms was associated with the lowest accuracy values (Figure 1B).

ISI. Variations in *ISI* ($F(2,264)=52.931$, $p<0.0001$) were also significantly associated with accuracy across all trials. Similar to the *ISI** parameter, pairwise comparisons again showed that performance under the separate *ISI* values statistically differed, with the longest *ISI* and shortest *ISI* resulting in the highest and lowest accuracy rates, respectively (Figure 1C).

Flash Duration. Flash duration ($F(1,174)=105.041$, $p<0.0001$) was also a significant factor in determining accuracy. The stimulus timing pattern with the longer flash duration of 62.5ms demonstrated significantly higher accuracy rates than the other three patterns with shorter flash durations of 31.25ms (Figure 1D).

Ratio. Ratio of flash duration to *ISI* ($F(2,264) = 54.982$, $p < 0.0001$) was also found to be significantly associated with accuracy. Stimulus timing patterns A and D with a flash duration to *ISI* ratio of 1:2 and pattern B with a ratio of 1:3 showed statistically identical performance. Pattern C with a flash duration of 31.25ms and *ISI* of 125ms, giving a ratio of 1:4 showed significantly higher accuracy values (Figure 1E).

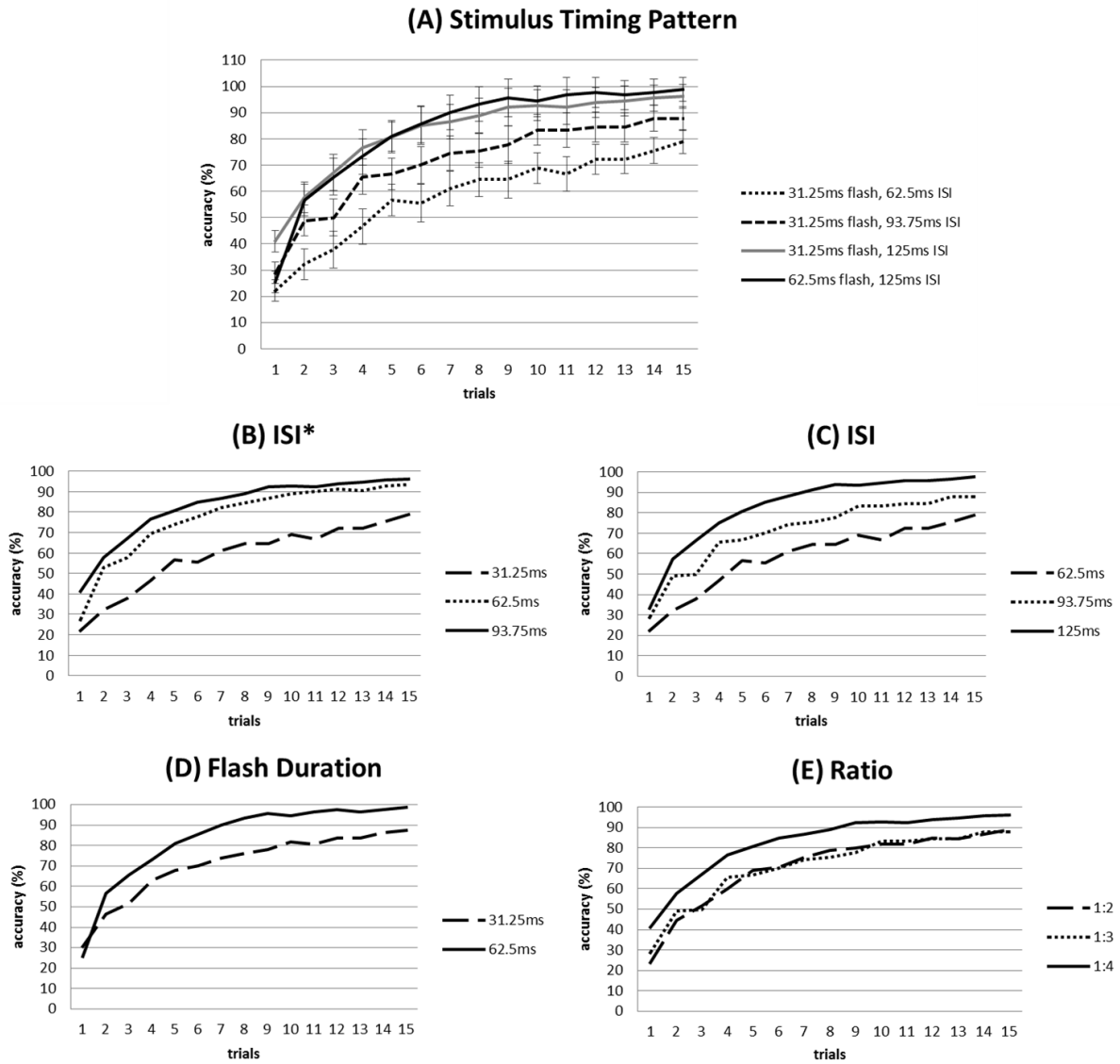


Figure 1. Average classification accuracy for all subjects grouped by (A) stimulus timing pattern, (B) ISI*, (C) ISI, (D) flash duration, and (E) ratio of flash duration to ISI. SEM error bars are included in (A).

3.1.2 Consistency of Accuracy between Good and Bad Performers

To assess the effect of stimulus timing pattern on subsets of the subjects, we performed secondary analyses. The six subjects were divided into the top three performers (Figure 2A)

and the bottom three performers (Figure 2B). Overall results from the entire group, good performers, and bad performers remained consistent as the lowest accuracy rates in each subset resulted from the pattern with the shortest 31.25ms flash duration and 62.5ms ISI. The two patterns associated with the best accuracy (both with an ISI of 125ms) were statistically indistinguishable, but the two patterns with shorter ISIs at 62.5ms and 93.75ms showed significantly different results from each other and from the former two patterns.

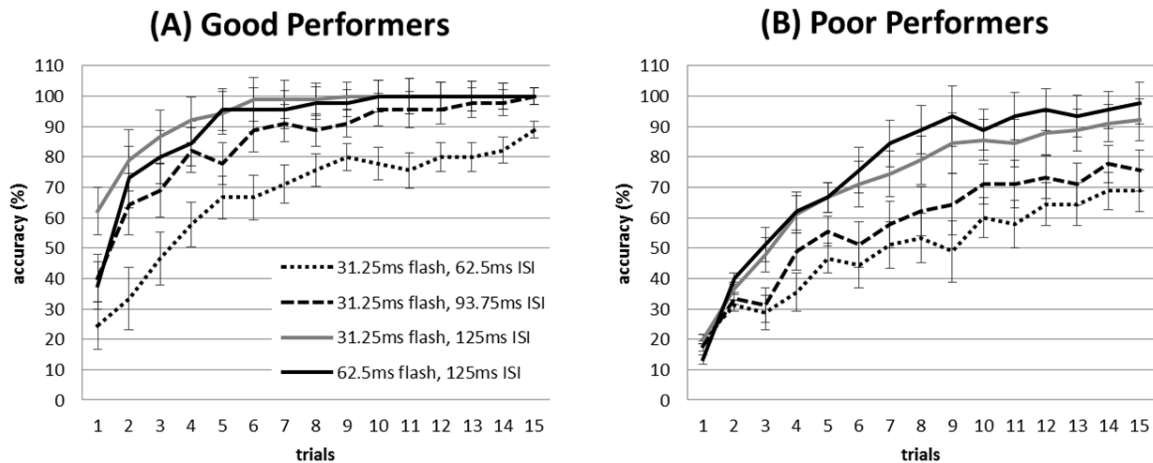


Figure 2. Average classification accuracies grouped by timing pattern comparing good performers (A) to bad performers (B).

3.2 Effect of Stimulus Timing Patterns on CPM

Statistical analysis using a mixed model revealed significant effects of stimulus timing pattern on CPM rates ($F(3,354) = 2.888, p < 0.036$). Pairwise comparisons showed that accuracy under pattern A with flash duration of 31.25ms and ISI of 62.5ms was significantly different from patterns C and D ($p < 0.008$ and $p < 0.027$, respectively), which both had 125ms long ISIs.

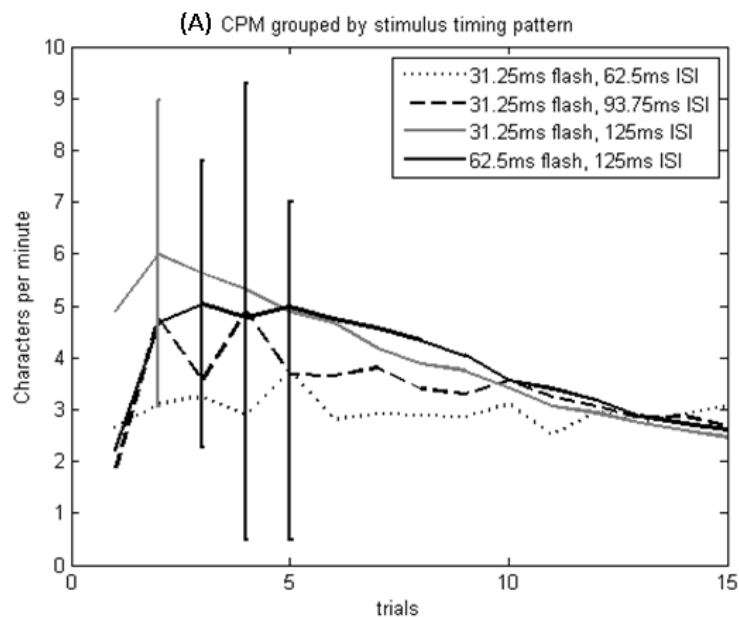
Patterns C and D, however, were not statistically significant. The timing pattern with a flash duration of 31.25 ms and an ISI of 125 ms resulted in the highest average group CPM values within the first five flash repetitions after which, the timing pattern with a flash duration of 62.5ms and ISI of 125ms was associated with higher CPM (Figure 3A).

3.2.1 Significance of Stimulus Presentation Parameters on CPM

ISI*. ISI* was significantly affected observed CPM ($F(2,264) = 3.950, p < 0.02$). Consistent with accuracy results, high CPM rates resulted from the longest ISI*, which were significantly different from CPM rates from the shortest ISI* ($p < 0.006$) (Figure 3B).

ISI. The effect of ISI ($F(2,264) = 2.995, p < 0.052$) on CPM was marginally significant. Post-hoc analysis revealed that performance under the longest ISI of 125ms was significantly better than under the shortest ISI of 62.5ms ($p < 0.015$) (Figure 3C).

Flash duration and ratio between the flash duration and ISI did not significantly affect observed CPM rates.



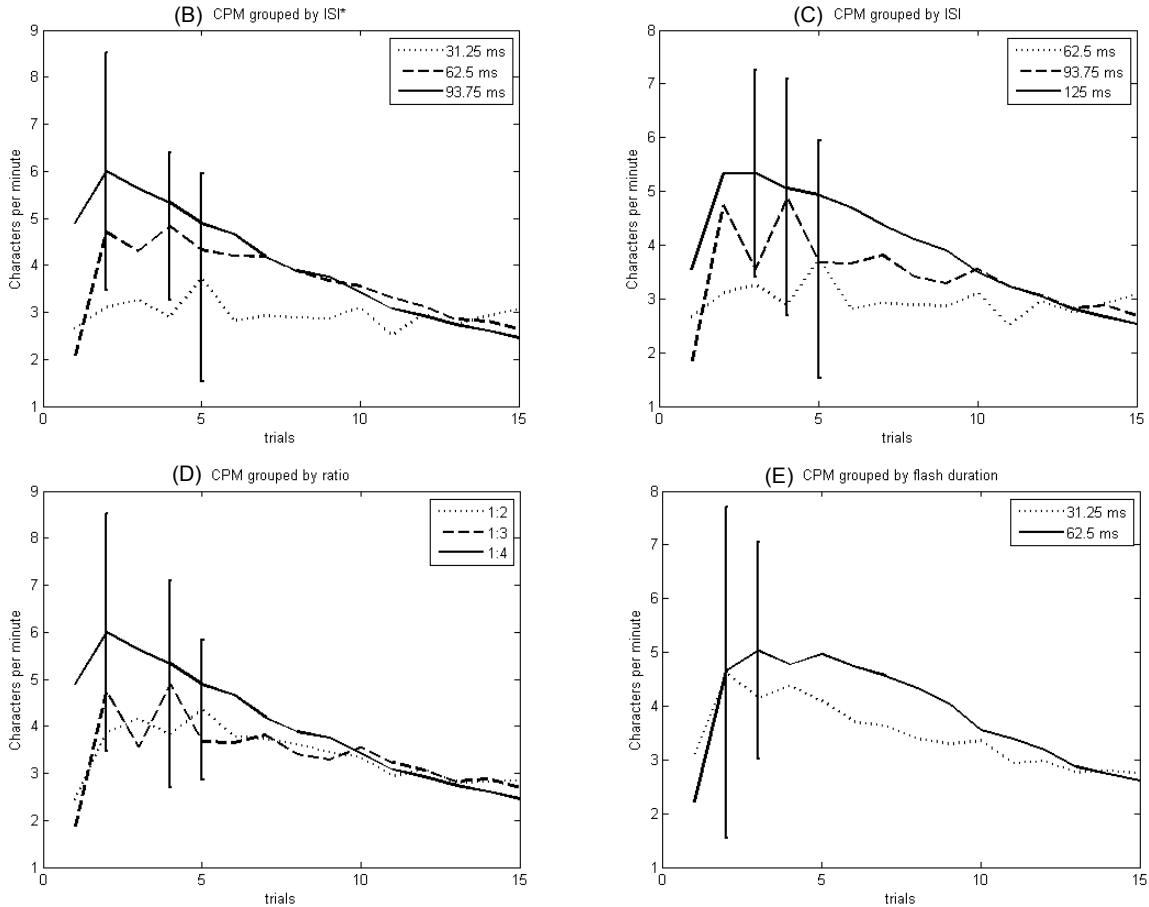


Figure 3. Average characters per minute for all subjects grouped by (A) stimulus timing pattern, (B) flash duration, (C) ISI*, (D) ISI, and (E) ratio of flash duration to ISI. SEM error bars are included for the largest CPM point of each average trend line.

3.2.2 Consistency of CPM on Good and Bad Performers

As with the analysis of accuracy rates, the effect of each factor was also assessed in subsets of the subjects studied: good and poor performers (Figure 4A and 4B, respectively). The effect of stimulus timing patterns in “good performers” mirrored the overall results (Figure 3A). Interestingly however, the effect of stimulus timing patterns on CPM in poor performers was

different, with timing pattern D (with the longest flash duration of 62.5ms and ISI of 125ms) consistently displaying superior results. This is in contrast to the good performers in which pattern C (with a 31.25ms flash duration and 125ms ISI) tended to have superior CPM rates for the first five trials.

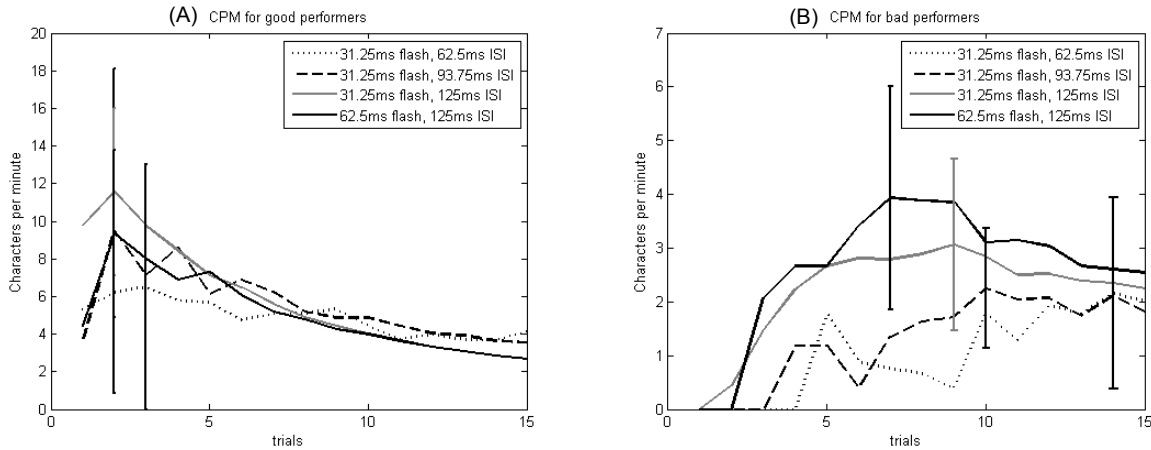


Figure 4. Average characters/min grouped by timing pattern comparing good performers (A) to bad performers (B).

3.3 P300 Waveforms

Average amplitudes and latencies for the P300 waveforms at electrode Pz during each stimulus timing pattern are shown in Figure 5. The P300 peaks begins at about 300 ms and peaks at approximately 400ms for every pattern, and the magnitude of the peak increases with greater ISI, consistent with our accuracy and CPM results.

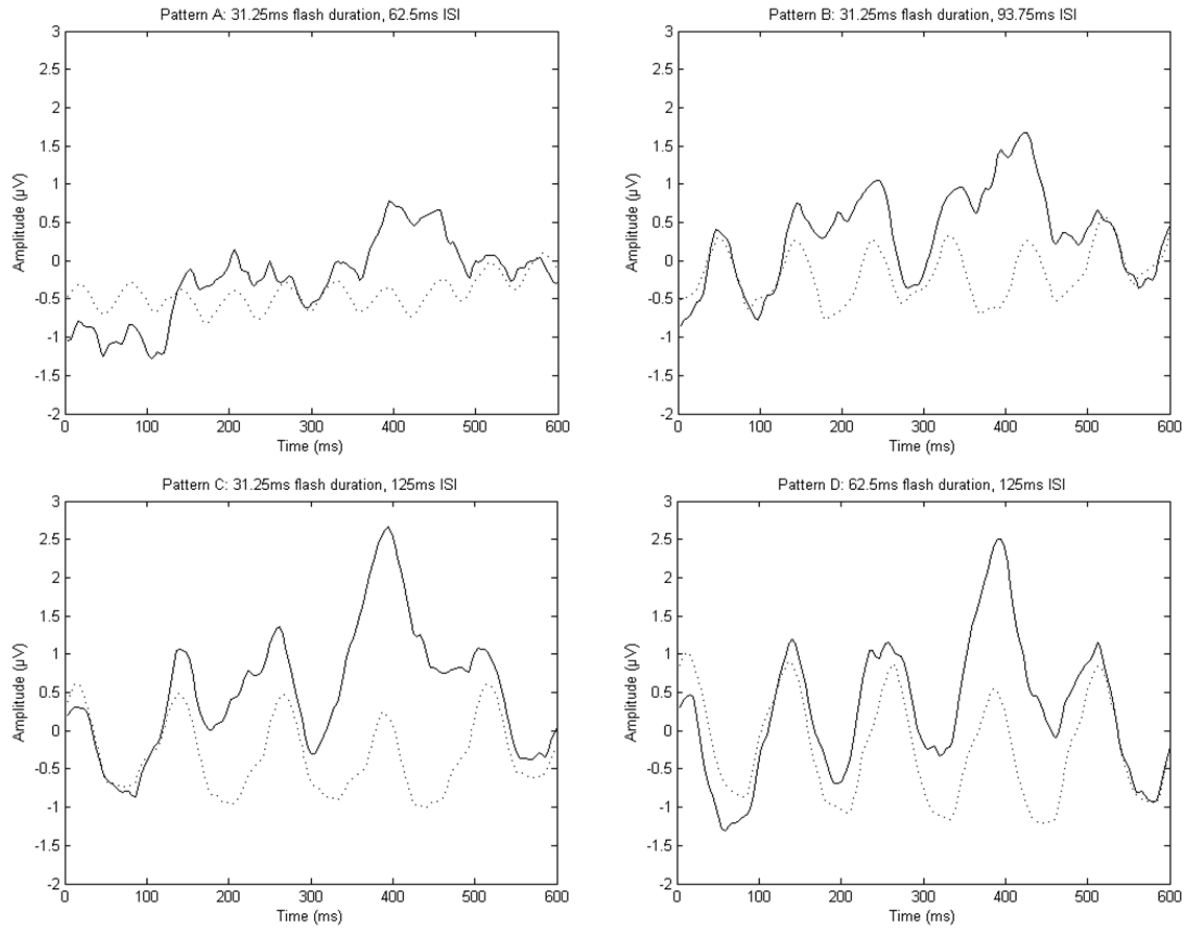


Figure 5. Average group waveforms at Pz for each stimulus timing pattern. Waveforms for targets are represented by the solid black line and waveforms for non-targets are represented by the dotted line.

3.4 8 Channel vs. 32 Channel Analysis

A subset of the eight channels from the 32-channel set used for EEG data collection, based on that described by Sharborough et al. (1991), was used to classify accuracy and determine CPM values. The standard set including Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz (Sharborough et al., 1991) was slightly altered due to the constraints of our original 32-channel set. The reduced channel set that produced the results in Figure 6 contained the eight channels: Fz, Cz, PO3, Pz, PO4, PO7, PO8, and Oz. The results demonstrated that a reduced channel set did not

significantly affect the accuracy or CPM values for three of the four stimulus timing patterns (Figure 6).

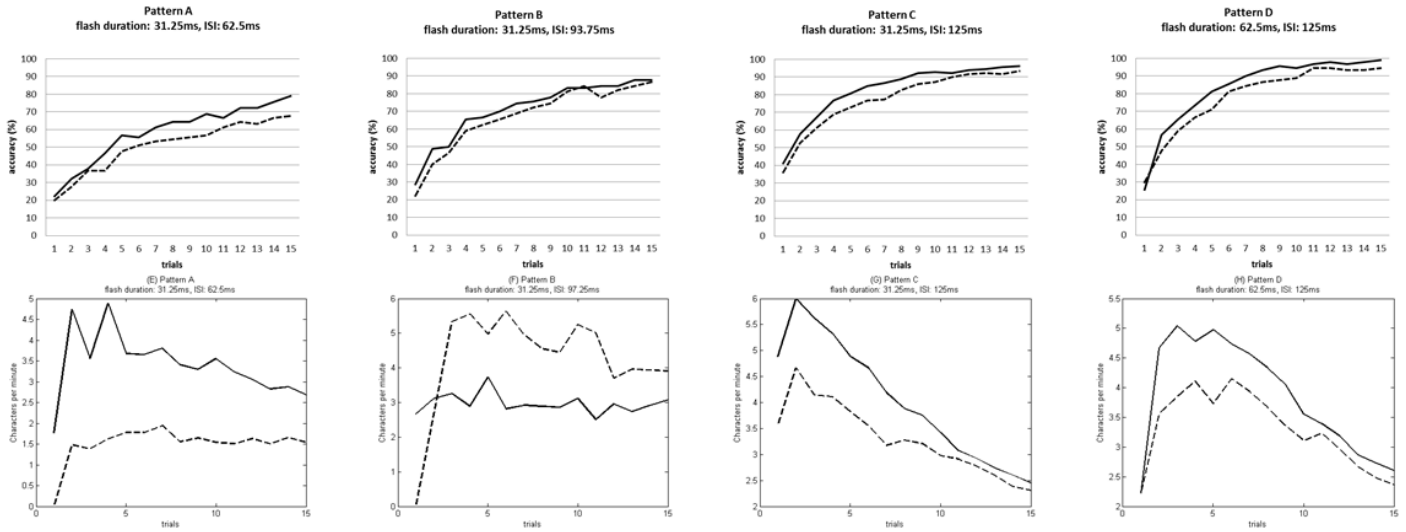


Figure 6. Average classification accuracy (A-D) and CPM (E-H) under the 32-channel set (represented by the solid black line) compared to the 8-channel set (represented by the dashed black line) for each of the four patterns.

4. Discussion

4.1 Significance of Stimulus Presentation Parameters

The P300 speller is potentially a robust communication system for individuals with severely impairing neuromuscular disease that prevents verbal or manual communication. Yet, studies investigating the factors that can optimize system efficiency are few. In this analysis, we examine four potentially modifiable stimulus presentation parameters to determine their effect on system performance, not only on accuracy, but also CPM.

We confirm prior reports of the significant impact of ISI on accuracy and CPM rates (McFarland et al., 2011; Allison and Pineda, 2006; Sellers et al., 2006, Gonsalvez and Polich,

2002; Croft et al., 2003). Gonsalvez and Polich (2002) reported that P300 amplitudes are positively correlated with target-to-target intervals (TTIs). Since longer ISIs necessarily lead to longer TTIs, our results indicating longer ISIs increase accuracy reflect the same conclusion. While Sellers et al. (2006) found apparently opposite results that a shorter 175ms ISI compared to a longer 350ms ISI yielded highest accuracies, the two ISIs tested were both longer than any of the ISI values used in the current study. Taken together, these findings suggest a limit to the beneficial effects of increasing ISI on observed accuracy.

In contrast to previous reports however, we report the new finding that longer ISI* may also significantly improve accuracy and CPM rates (McFarland et al., 2011). While longer ISI* and ISI were both associated with superior CPM rates, comparing the two patterns with a 125ms ISI shows that the timing pattern with a 93.75ms ISI* performs better than that with a 62.5ms ISI*, at least up to five trials where the last CPM curve peaks.

Likewise, our results indicate that flash duration can impact accuracy, which is inconsistent with previous reports (McFarland et al., 2011). While this contrast in results may be due to differences in experiment paradigms or specific flash lengths tested, flash duration has proved to be important in certain cases, and therefore, cannot be completely ignored.

4.2 Accuracy vs. CPM

In optimizing the P300-based BCI system, the goal is to maximize both accuracy and CPM, which are correlated, but do not necessarily improve together. In this study, CPM rates peaked around two or five trials (except for poor performers – figure 5B), and accuracy rates plateaued around eight trials. Both results suggest that 15 trials are more than what is necessary to achieve current peak accuracy and CPM rates, and adversely add time and decrease speed. Reducing the number of trials would increase CPM, optimizing system performance even if accuracy rates

have not yet been optimized. The number of trials could either be set to a reduced number, or a dynamic system changing the number of trials as individually needed could be implemented (Speier et al., 2012). Serby et al. (2005) demonstrated a dynamic system that would run an additional trial each time the set threshold was not reached. The study improved accuracy and symbols/min results compared to Farwell and Donchin (1988).

The preference between higher accuracy at a slower rate compared to a faster rate with lower accuracy may depend on the user. Usability of the P300 speller is highest in the 70-95% accuracy range (Citi et al., 2010), proposing that perfect accuracy is not necessary to convey a user's message, allowing for mistakes and a faster bit rate.

4.3 Good vs. Poor Performers

Subjects were divided into two groups due to visible disparity between their performances. Despite large differences in accuracy and CPM, both subsets showed the same effects by the significant parameters, ISI and ISI*, demonstrating that these results are applicable to all subjects regardless of performance. As shown in the CPM results, while an ISI of 62.5ms results in poor performance across all subjects, good performers can still do well with a middle length ISI of 93.75ms (Figure 5A). The pattern with the longest flash duration of 62.5ms and ISI of 125ms, which shows the greatest accuracy, actually gives the lowest CPM values by trial 15. While the cause of performance variance among subjects has yet to be identified, it is important to know that performance does not affect parameter analysis. Possible factors contributing to performance variance include attention and motivation. Future research interests lie in determining or predicting the ability of an individual to perform well based on certain measures of attention or motivation.

4.4 32-Channel Set vs. 8-Channel Set

Results from the reduced electrode set did not differ significantly compared to results from using 32 electrodes. Despite consistent results from this comparative analysis using the historically reported reduced 8-electrode set, a more comprehensive analysis of the effect of spatial parameters on system performance is likely needed in future studies. The selection of electrodes should be dictated either by (1) physiological assumptions about the origin of the signal or (2) data-driven approaches that objectively identify which electrodes provide the greatest discriminatory pattern between attended and non-attended stimuli, using methods such as principle or independent component analysis. By identifying the essential components need for target identification, the goal would be to find a reduced channel set that does not diminish accuracy or CPM, but hopefully increases both.

4.5 Limitations

For purposes of this investigation, the timing parameters used in this study were designed to contrast as many factors as possible in order to gain insight into the impact of these timing factors on accuracy and CPM rates. As a result, the timing parameters used were not completely independent of one another, thereby confounding results and interpretation to some extent, possibly resulting in false-negative or false-positive results. Moreover, the current analyses were done in offline analyses and remain to be further characterized and validated in online analyses. Nonetheless, the current results can be used as a starting point for the design of more rigorous and comprehensive analyses of these identified factors. Despite these limitations, this study highlights the need to optimize stimulus presentation parameters not only with respect to accuracy, but also CPM. Since reducing the number of trials would increase CPM values by

greatly decrease the amount of time needed, future designs must consider using fewer trials, possibly in a dynamic manner.

4.6 Conclusion

ISI and ISI* were shown to be the most significant parameters in determining P300 speller system accuracy and efficiency. The accuracy and CPM of the P300 speller increased as ISI* and ISI increased. The flash duration and ratio parameters proved to minimally modulate performance. To optimize CPM, which is a surrogate of system performance, a system with fewer trials or a dynamic system must be implemented. The discrepancy between good and bad performers remains to be investigated. Advancement from the standard reduced channel set would be an optimal channel set that will be explored in further research.

CHAPTER 3: Exploring a Potential Physiological Biomarker of P300 Speller Performance

1. Introduction

Brain-computer interfaces such as the P300 speller can restore communication and functionality to severely paralyzed patients, but require a certain level of performance to work efficiently. The previous chapter highlighted differences in user performance by splitting the subjects into two groups to determine whether subsets of performers showed the same stimulus timing effects as the entire group. The results showed that performance does not affect parameter analysis, yet raises the question of what causes performance variance. Attention and motivation are likely factors that contribute to variance and were examined by Lakey et al. (2011), who specifically investigated the effects of mindfulness meditation induction (MMI) on improving P300 speller performance. The study concluded the MMI subjects performed significantly better than normal subjects with larger P300 amplitudes at electrodes Cz and PO7. Exploring the idea that brain states can affect a BCI user's level of performance, this chapter examines the possibility of a physiological biomarker that is associated with spelling accuracy.

2. Methods

One female and fourteen males ages 22 to 35 participated in this IRB-approved study. Only one subject had prior BCI experience.

2.1 Data Collection

The experimental settings were the same as in the study described in chapter 1 using a 32-electrode cap referenced to the right or left mastoid, the BCI2000 general-purpose system, and signals amplified with two g.tec (Guger technologies) 16-channel USB biosignal amplifiers, digitized at 256 Hz, and filtered between 0.1 and 60 Hz (Schalk et al., 2004). The stimulus presentation paradigm utilized was the original RCP by Farwell and Donchin, flashing each row and column once randomly, totaling 12 flashes per trial, and repeating for 15 total trials for each letter selection.

The initial set of six subjects each spelled a total of 15 words in the single session under four different flash timing patterns. A set of six words spelled under a single flash timing pattern with a flash duration of 31.25ms and an interstimulus interval (ISI) of 125ms were used. An additional nine subjects were added to the study and each spelled ten words in a single session under the same flash timing pattern with a flash duration of 31.25ms and ISI of 125ms.

2.2 Data Analysis

The BCI2000 P300 Classifier GUI (graphics user interface) was used for accuracy classification offline. SWLDA was used to generate feature weights, which were used to train and test the linear classifier to detect event-related potentials collected with BCI2000 as described in the data analysis section of chapter 1 (Schalk et al., 2004). We used leave-one-out cross-validation to assess accuracy: the linear classifier was trained on up to nine words and tested on the excluded word, and then repeated in the set of up to ten words, so that each word was excluded once and used to test the classifier.

The classification accuracy for each five-letter word spelled and each letter individually was recorded. The classification accuracy for each word equaled an average of the six or ten accuracy values obtained by the leave-one-out method previously described. The measurement of accuracy used for each word was obtained by plotting an accuracy curve, which represents the percent accuracy of spelling one five-letter word over 15 trials, and calculating the area under the curve (AUC) (Figure 7). A binary code was recorded for classification accuracy by letter with '1' representing that the letter was classified correctly and '0' representing that the letter was classified incorrectly.

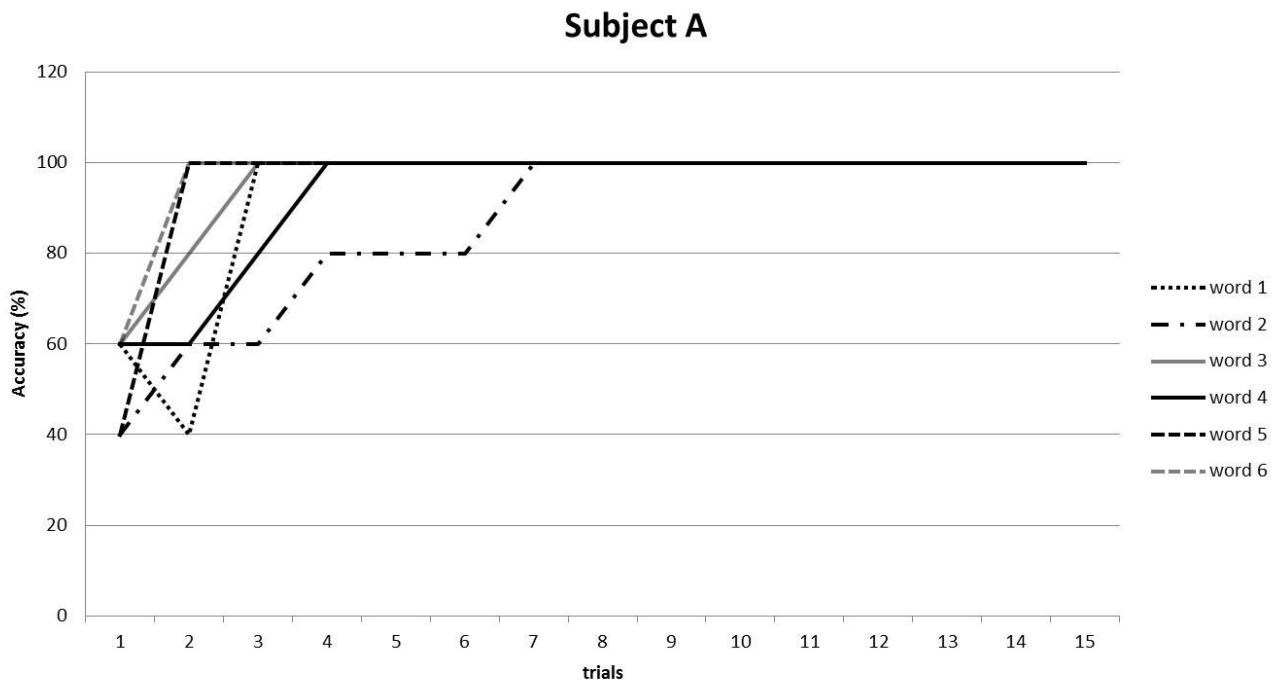


Figure 7. Accuracy curves for each word spelled by subject A. The area under each curve (AUC) was calculated and used as the accuracy measure for each word.

Each run consisted of a subject spelling an entire five-letter word. The EEG data used to calculate the power spectrum density was taken from the time, varying from 1.25 seconds to 2.5 seconds, before the stimulus for the task began. For experiment 1, we used the first two seconds of EEG signal in the beginning of each run, prior to the beginning of the spelling task (Figure 8a). For experiment 3, we used the time before each letter was spelled, so the EEG signal used was taken during the run, but between sequences of stimuli (Figure 8b).

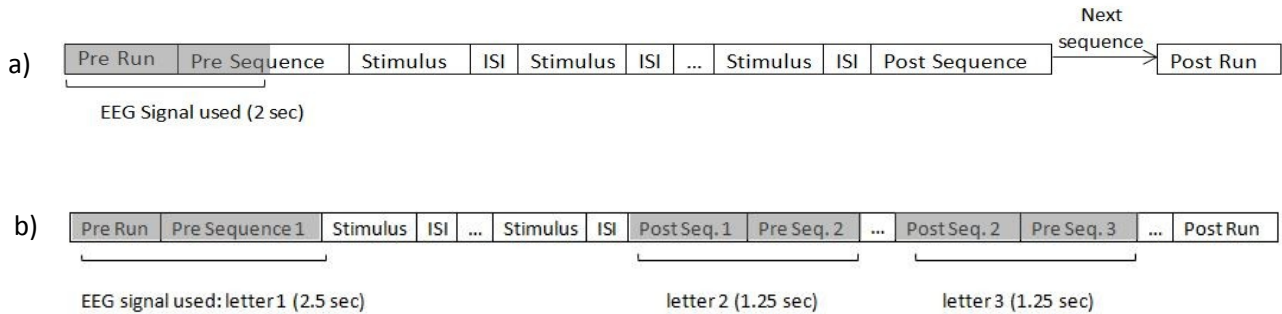


Figure 8a and b. Sections highlighted in gray indicate the time in the EEG signal used to calculate power values.

2.2.1 Experiment 1.

Initial exploratory analysis included dividing the first six subjects to identify differences in spectra between the three best and three worst performers. The power spectral density (PSD) was used to examine between-group differences in signal power in central electrode Cz (Figure 9).

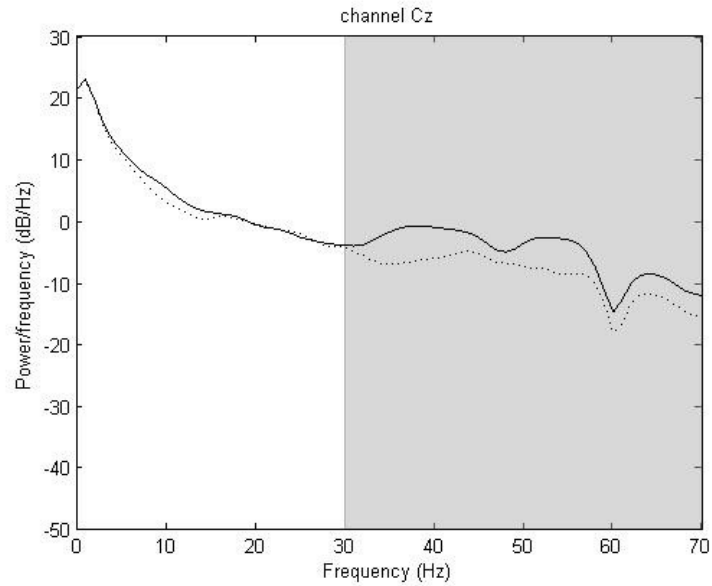


Figure 9. Power spectral density estimate in channel Cz of the top three performers (solid black) compared to the bottom three performers (dotted black). The low gamma (30-70Hz) frequency band is shaded in gray.

Having identified differences across subjects, we evaluated the reliability of potential biomarkers within subjects across words. An analysis of covariance analysis (ANCOVA) was performed with the continuous outcome variable as accuracy, and two predictor variables, subject and power in five frequency bands (delta: 0-4Hz, theta: 4-7Hz, alpha: 8-12Hz, beta: 13-30Hz, and gamma: 30-70Hz) normalized to total power from 1 to 70Hz in electrode Cz, to find a linear model to predict the effects of subject and different EEG band powers on accuracy. The gamma band was highlighted in figure 9 to show the observable difference in comparing the two groups of performers. All five EEG bands were tested, and gamma was chosen to use in the model. This analysis began with creating a model from data from six subjects, followed by adding nine subjects in order to test the model. The PSD was calculated from the first two seconds of EEG signal of each run or word spelled. The measurement of accuracy, therefore, was

the area under the curve (AUC) that represented the percent accuracy of spelling the entire five-letter word after each of the 15 trials.

2.2.2 Experiment 2.

A negative experiment was performed to test the hypothesis that if there is a significant temporal correlation between the band power taken from the EEG signal prior to spelling a word and the resulting classification accuracy for that word, randomly associating the power values and the accuracy values should yield insignificant correlations. All possible permutations ($n = 720$) of associating the six accuracy values with the six different power values were tested within each subject. For the nine subjects with ten words, a random selection of 720 permutations was used out of the total ($n = 10!$) possible permutations. The true correlation coefficient between the correct power and accuracy association was plotted on top of a histogram of the Kendall rank correlation coefficients resulting from random associations.

2.2.3 Experiment 3.

The next experiment tested the hypothesis of whether a linear combination of power values correlated with accuracy. The previously described experiments focused on power values in one frequency band (low gamma) and one electrode (Cz) at a time. In this analysis, a leave-one-out cross validation using stepwise linear discriminant analysis (SWLDA) was performed using 160 features, including power values in the five frequency bands (delta, theta, alpha, beta, and gamma) at each of the 32 recording electrodes. The AUC accuracy values were used to create a binary class label using the mean AUC value (0.73) as a threshold. AUC values greater than 0.73 equaled '1' and less than 0.73 equaled '0'. The classifier output scores generated using leave-one-out cross-validation were then correlated to the target class labels.

Similar to the previous experiments, this experiment attempted to associate the power in the signal prior to spelling each word to the resulting accuracy of spelling that word. Extending the leave-one-out cross validation SWLDA analysis, the next analysis examined the correlation between the power in the signal prior to each letter spelled and the accuracy of spelling each associated letter after each of the 15 trials or sets of flashes. A new SWLDA model was built for each set of flashes. Binary classification accuracy was used for the accuracy measure by letter. Both an intrasubject (sample size = 30 or 50) and intersubject (sample size = 630) analysis were performed.

3. Results

3.1 Experiment 1

Using an analysis of covariance same-line model, separate EEG band powers (delta, theta, alpha, beta, and gamma) were correlated with the AUC accuracy measure for each of the six words spelled by the first six subjects, ignoring subject grouping. Among the five different EEG bands, analysis in the low gamma band (30 to 70Hz) showed a significant correlation between gamma power and accuracy in electrode Cz, with a Pearson correlation coefficient of 0.38641 ($p = 0.0199$). Figure 10 illustrates that greater power in the gamma band is associated with higher AUC and classification accuracy value. While the correlation proved to be statistically significant, studying the clustering of data points within each subject led to further investigation. Subjects 3 and 5 showed more widespread distribution of data points with poorer performance or AUC values while the remaining subjects showed clustered data points near the top of the graph at higher AUC values. This linear model associating gamma power and accuracy was tested on

nine additional subjects in the same electrode, Cz. The correlation in the test set was not found to be significant with a Pearson correlation coefficient of 0.0407 ($p = 0.703$) (Figure 11).

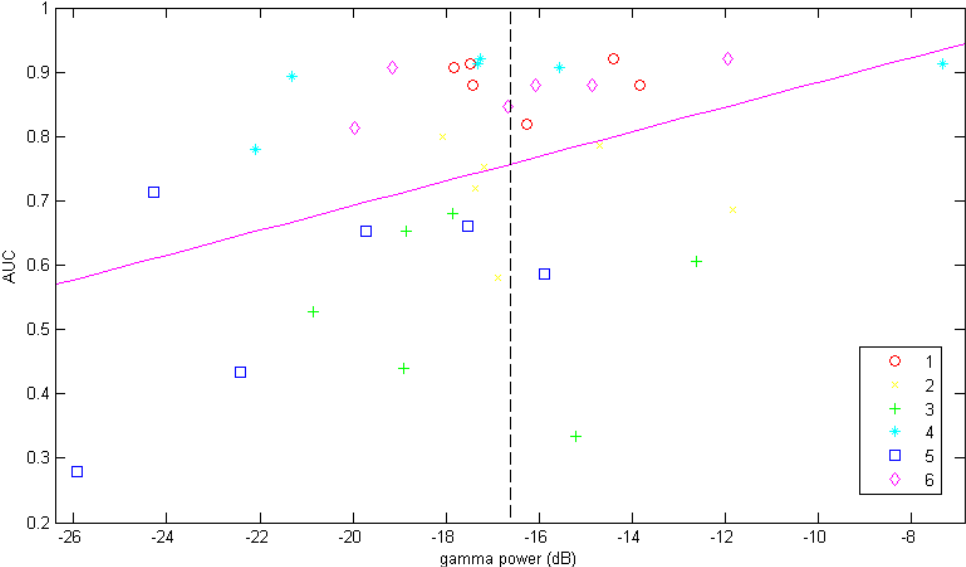


Figure 10. Gamma power in electrode Cz along the x-axis is plotted against a measurement of accuracy, AUC, for the initial six subjects. The data is fitted to a same-line model, ignoring subject grouping.

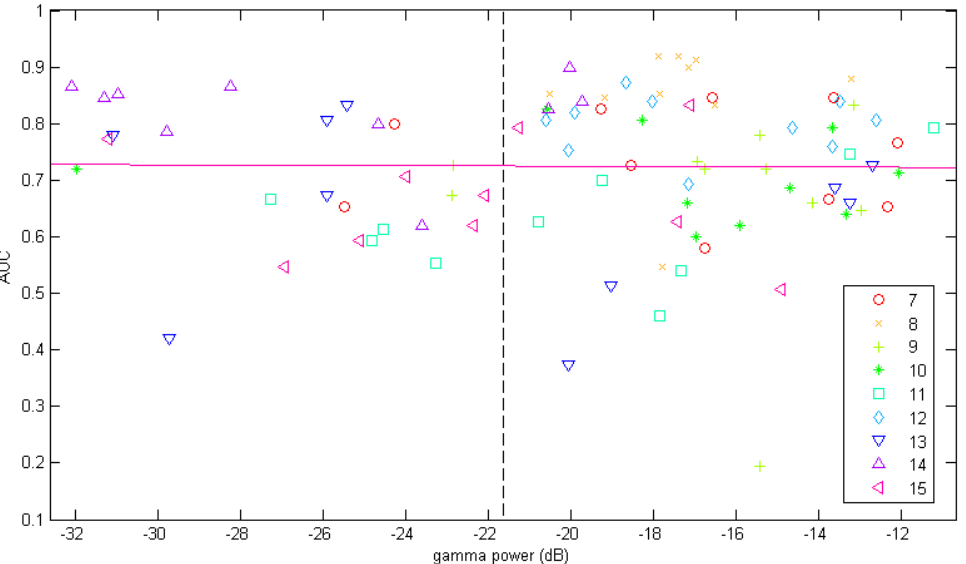


Figure 11. Gamma power in electrode Cz along the x-axis plotted against AUC along the y-axis for the test set of nine additional subjects.

3.2 Experiment 2

A negative experiment was done to confirm the results of the ANCOVA. The histograms of Kendall rank correlation coefficients were produced by correlating gamma power values with random accuracy values, using 720 permutations and grouped within each subject (figure 12) and within the two groups of subjects (figure 13). The true correlation coefficient between the correct power and accuracy association was plotted as a line on top of each histogram. In each histogram, the true correlation coefficient was marked near the center and never landed in the top ten percent, and therefore, showed an insignificant association between gamma power and accuracy.

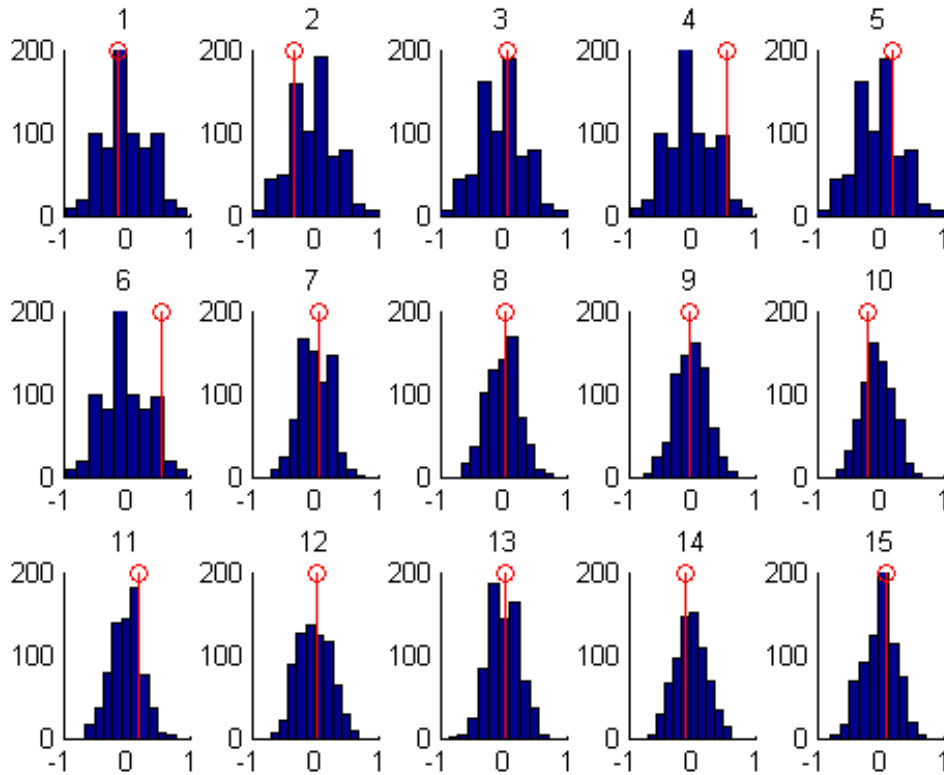


Figure 12. Histogram of Kendall rank correlation coefficients along the x-axis and frequency in each ten percent interval along the y-axis grouped within each subject.

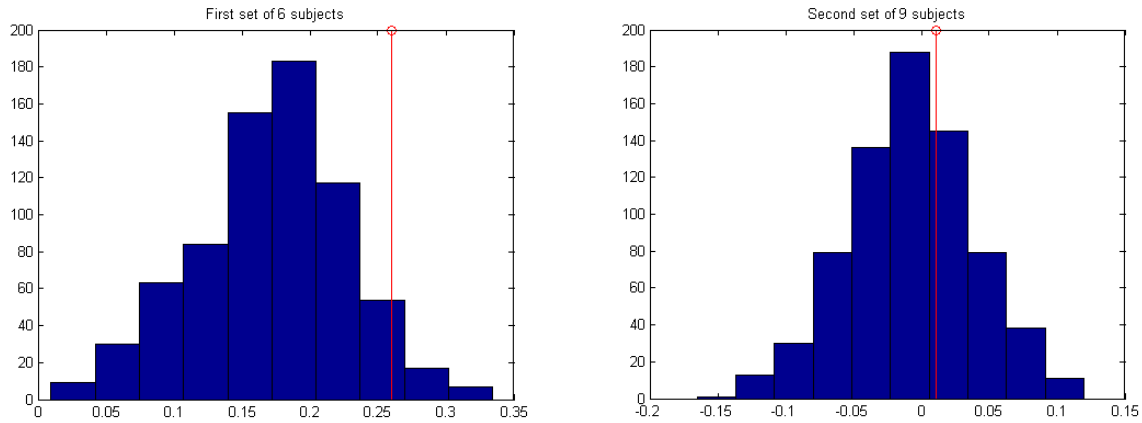


Figure 13. Histogram of Kendall rank correlation coefficients along the x-axis and frequency in each ten percent interval along the y-axis grouped within each set of subjects. The line marks the correlation coefficient for the true association.

3.3 Experiment 3

The previous analyses examined power and accuracy association within one frequency band, low gamma, and one electrode, Cz. Using SWLDA, the next analysis investigated whether classification accuracy could be proven to be associated with a linear combination of 160 features, including all the power values from each of the five EEG bands (delta, theta, alpha, beta, and gamma) at each of the 32 electrodes from the two seconds of EEG signal preceding the spelling task for each word. The features selected and coefficients integrated into the SWLDA model are shown in Table 2. The classifier score was plotted against the AUC accuracy measure in figure 14, again showing no significant correlation between power in the EEG signal before the start of a BCI task and the resulting classification accuracy (Pearson correlation coefficient = 0.1409, $p = 0.1155$). Additionally, the ROC curve showed poor classifier performance with a low AUC value of 0.6251 (Figure 15).

EEG band	Electrode	Coefficient
Delta	CP3	-0.23856
Delta	CP2	0.444611
Alpha	FC2	0.039128
Alpha	POz	0.040358
Alpha	O1	-0.04572
Beta	Fz	-0.04145
Beta	C6	-0.03688
Gamma	CP6	0.029047
Gamma	Pz	0.026602

Table 2. Features selected (EEG band and electrode) and coefficients of the SWLDA model used in experiment 3.

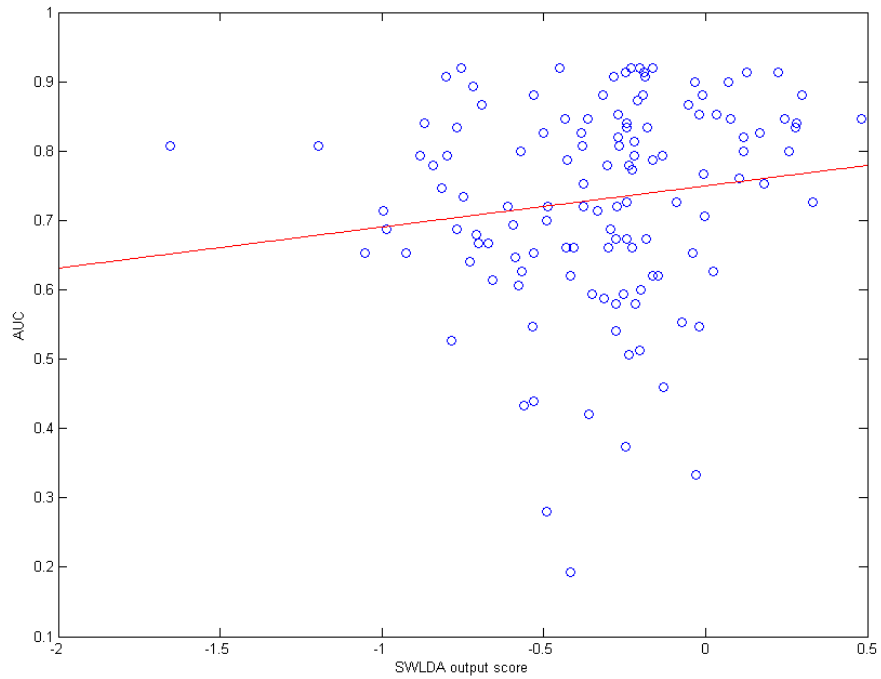


Figure 14. Correlation plot graphing SWLDA classifier score against AUC accuracy measure.

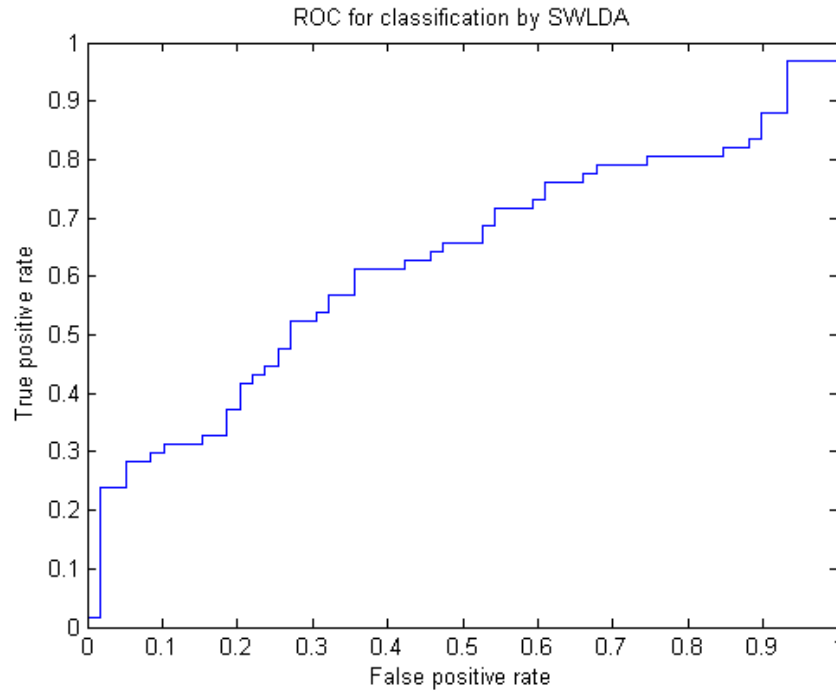


Figure 15. ROC curve for SWLDA classifier with poor AUC value = 0.6251.

Stemming from the previous experiments, further investigation inquired as to whether a closer temporal association between power in the EEG signal and classification accuracy may result in a significant relationship. Rather than attempting to correlate the EEG signal preceding the spelling task of an entire five-letter word with the corresponding accuracy results, the subsequent analysis examined the potential correlation between the EEG signal preceding the spelling of each letter with the result of classifying that letter as correct or incorrect after each of the 15 sets of flashes. In an intrasubject analysis, a number of correlation coefficients showed significance (Table 3). A paired t-test between the average classifier scores for correct and incorrect classifications after each of the 15 trials showed significant distinction between correct and incorrect classifications after the eighth ($p = 0.0364$), eleventh ($p = 0.0441$), and twelfth ($p = 0.0024$) set of flashes. The intersubject analysis showed significant correlation coefficients after

every trial when using a SWLDA model built after the first trial, however, the ROC curves showed poor classifier performance (Table 4 and Figure 16).

	Subj: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Trials: 1	0.1761	0.6027	0.4309	0.5011	0.6400	0.1351	0.0920	0.1479	0.4259	0.9168	0.7552	0.9334	0.5816	0.7185	0.0199
2	0.0627	0.1746	0.4777	0.0817	0.5553	0.4196	0.0386	0.4105	0.3218	0.0519	0.8785	0.7995	0.3229	0.0097	0.2575
3	0.1923	0.3710	0.8709	0.4084	0.8833	0.0014	0.8222	0.6385	0.6480	0.6391	0.3546	0.4696	0.1251	0.8362	0.0268
4	0.1825	0.6948	0.9898	0.8545	0.0565	0.1356	0.8365	0.9933	0.3936	0.3732	0.1344	0.3690	0.8085	0.7538	0.4814
5	0.1825	0.2632	0.5776	0.7542	0.0514	0.0014	0.0224	0.1121	0.9832	0.2939	0.3029	0.3396	0.6061	0.0455	0.6087
6	0.1825	0.1631	0.1459	0.7542	0.0178	0.1838	0.3973	0.9486	0.9170	0.8762	0.3442	0.0274	0.5774	0.2529	0.0486
7	NaN	0.3017	0.1806	NaN	0.9719	NaN	0.0692	0.1121	0.4321	0.3736	0.3969	0.6056	0.0668	0.0236	0.2183
8	NaN	0.0175	0.5296	NaN	0.0003	NaN	0.1639	0.3909	0.7458	0.0383	0.8401	0.0343	0.8617	0.9098	0.3651
9	NaN	0.8290	0.2673	NaN	0.2517	NaN	0.1639	0.3909	0.6812	0.0694	0.5065	0.0674	0.1200	0.9098	0.0048
10	NaN	0.8290	0.1397	NaN	0.9323	NaN	0.0008	0.1121	0.6812	0.9634	0.8239	0.0674	0.0002	0.3823	0.2382
11	NaN	0.8290	0.6679	NaN	0.0845	NaN	0.0287	0.1121	0.6812	0.8786	0.7974	0.0358	0.0034	0.0269	0.1876
12	NaN	0.8290	0.0177	NaN	0.2103	NaN	0.1607	0.1121	0.0065	0.5921	0.9017	0.0113	0.3018	0.3001	0.0258
13	NaN	0.8290	0.0674	NaN	0.7616	NaN	NaN	0.1121	0.3775	0.1078	0.0109	0.0113	0.0000	NaN	0.4553
14	NaN	0.8290	0.9596	NaN	0.3984	NaN	NaN	0.1121	0.3775	0.0001	0.3983	0.0113	0.0000	NaN	0.0419
15	NaN	0.0794	0.6665	NaN	0.3984	NaN	NaN	0.1121	0.3775	0.8386	0.3430	0.0674	0.0000	NaN	NaN

Table 3. P-values that are highlighted show significant correlations between power in the EEG signal prior to spelling a letter and the classification accuracy of the associated letter after each trial. SWLDA did not work in cells with “NaN” due to perfect classification accuracy.

Trial	Pearson Correlation coefficient	p-value
1	0.245182901	4.45E-10
2	0.153226227	1.13E-04
3	0.160445518	5.22E-05
4	0.159295836	5.92E-05
5	0.149454433	0.000167
6	0.196833479	6.38E-07
7	0.177666955	7.25E-06
8	0.116921182	0.003293
9	0.124510685	1.74E-03
10	0.17808014	6.90E-06
11	0.157110398	7.49E-05
12	0.136226698	6.07E-04
13	0.104717987	0.008528
14	0.088140761	0.026951
15	0.095064317	0.016997

Table 4. Pearson correlation coefficients and p-values after each trial using a leave-one-out SWLDA model built after the first set of flashes.

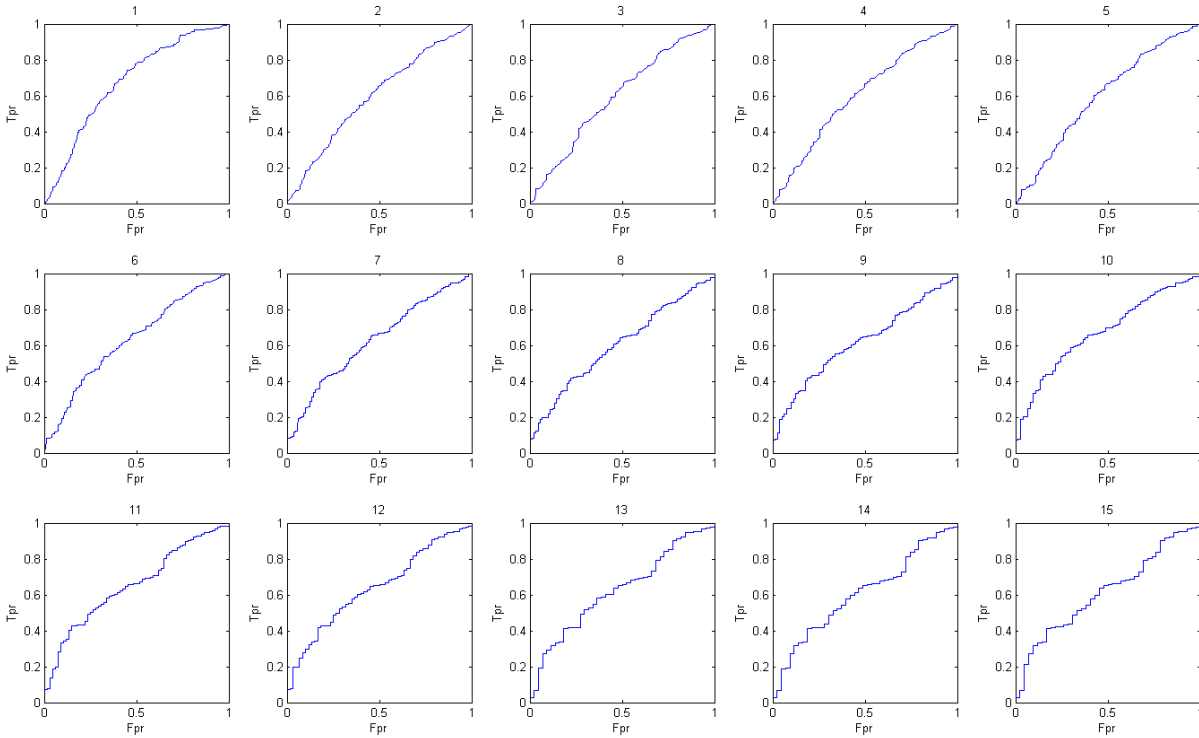


Figure 16. ROC curves for SWLDA classifier after each trial using data from all subjects.

4. Discussion

The significant correlation found in the first experiment with the initial six subjects drove the subsequent analyses to seek a better understanding of the relationship between power in the EEG signal and P300 speller performance. However, the following experiments testing the model on a new set of subjects and creating a negative experiment to test significance of the true correlation coefficients both proved that a strong temporal correlation between power in the initial EEG signal and the resulting classification accuracy of the spelled word did not exist when looking at one frequency band and one central electrode. The next experiments served to extend the original analysis by examining a linear combination of power values in all frequency bands and electrodes rather than a single power value in a single frequency band and electrode. Temporal

correlation was investigated at two different levels: first, by comparing power in the first two seconds of the EEG signal to the accuracy of spelling a word, and second, by comparing power in the EEG signal immediately prior to the spelling of each letter to the accuracy of classifying the letter correctly after each set of flashes. While at the word level there was no significant correlation, examining a shorter time association between power and accuracy of each letter resulted in a select number of significant correlations. Despite these results, the paired t-test showed that the difference between the average scores of correct compared to incorrect classifications is only significant after certain sets of flashes. The final analysis of the power and accuracy association specifically by letter suggested some potential relation between the two values, however, the results were not convincing enough to state a solid correlation. Although the low p-values suggested significant correlation coefficients, they may have simply been a result of the greatly increased number of data points in the intersubject analysis, since the associated ROC curves implied poor classifier performance.

In conclusion, by exhausting multiple data analysis approaches, this study found no relationship between the power in the EEG signal and classification accuracy from the P300 speller. Future directions to potentially find a correlation include using unsupervised learning methods, such as principal component analysis (PCA), multi-dimensional scaling (MDS), and manifold learning, to reduce the dimensions of the total feature matrix, which has size [630x160] in an intersubject analysis. The paired t-test provides a one-dimensional representation of the results whereas other methods could provide a larger dimensional interpretation. Other supervised methods aside from SWLDA could also be productive approaches.

CHAPTER 4: Conclusion

Current P300-BCI research continues to look for avenues to innovate and improve usability. The two studies presented here sought to improve system efficiency by investigating fundamental stimulus timing parameters and the possibility of a physiological biomarker to identify user performance.

The first study confirmed previous studies (McFarland et al., 2011; Allison and Pineda, 2006; Sellers et al., 2006, Gonsalvez and Polich, 2002; Croft et al., 2003) that proved the ISI stimulus timing parameter to be significant in determining P300 speller system accuracy and efficiency. Contrary to the results of a previous study (McFarland et al., 2011), stimulus-off time or ISI* also showed significant effects on speller performance. Both aspects of P300 speller performance that were investigated, accuracy and CPM, increased as ISI* and ISI increased. A system with fewer trials or a dynamic system was also suggested in order to optimize CPM, which was highlighted in this study to be as important as accuracy to focus on when improving performance. Recording methods were also briefly addressed in the 32-electrode versus 8-electrode subset analysis, showing that while the standard reduced set works well, a more systematically reduced electrode set may be more optimal.

The discrepancy between good and bad performers was pointed out in the first chapter and served as the motivation for the second study. Unfortunately through multiple extensive analyses of the data, there was no correlation to be found between power in the EEG signal and user performance in spelling the temporally associated letter or word. Further studies may want to explore different methods, perhaps following the research looking into motivation and attention effects (Lakey et al., 2011), to identify differences in P300 performance.

The main goals in P300-BCI research are to increase accuracy and speed and to improve clinical practicality. As mentioned in the introduction, many aspects, such as potential users, recording methods, stimulus presentation paradigms, feature extraction and classification algorithms, and applications of the speller need to be further addressed to help optimize the system (Mak et al., 2011). System modifications that could be implemented include a minimal or optimal electrode set designed systematically or specific to an individual, various novel flash paradigms that eliminate adjacency errors and reduce the number of total flashes to decrease time and increase bit rate, and a dynamic classification system perhaps using semi-supervised or adaptive learning. While healthy subjects are more accessible and easier to recruit, future studies should focus on designing and testing devices based on the physical and cognitive capabilities and needs of specific patients.

While there are many imperfections in the P300-BCI system that drive the motivation to enhance system performance and efficiency, the current P300 speller benefits many patients who suffer from the locked-in syndrome and would otherwise not be able to communicate to others. New BCI research and innovations continue to advance the technology helping disabled patients every day.

References

- Allison BZ, Pineda JA. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: Implications for a BCI system. *Psychophysiol* 2006;59:127-140.
- Belitski A, Farquhar J, Desain P. P300 audio-visual speller. *J Neural Eng* 2011;8:025022.
- Brouwer AM and van Erp JB. A tactile P300 brain-computer interface *Front Neurosci* 2010;4:19.
- Brunner P, Joshi S, Briskin S, Wolpaw JR, Bischof H, Schalk G. Does the 'P300' speller depend on eye gaze? *J Neural Eng* 2010;7:056013.
- Citi L, Poli R, Cinel C. Documenting, modeling and exploiting P300 amplitude changes due to variable target delays in Donchin's speller. *J Neural Eng* 2010;7:056006.
- Croft RJ, Gonsalvez CJ, Gabriel C, Barry RJ. Target-to-target interval versus probability effects on P300 in one- and two-tone tasks. *Psychophysiol* 2003;40:322-8.
- Farwell LA, Donchin E. "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials". *EEG Clin Neurophysiol* 1988;70:510-23.
- Furdea F, Halder S, Krusienski D, Nijboer F, Bross D, Birbaumer N et al. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiol* 2009;46:617-625.
- Gonsalvez CJ, Polich J. P300 amplitude is determined by target-to-target interval. *Psychophysiol* 2002;39:388-396.
- Guan C, Thulasidas M, Wu J. High performance P300 speller for brain-computer interface. 2004 IEEE Int. Workshop on Biomedical Circuits and Systems 2004;S3/5/INV-S3/13-6.

- Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R et al. How many people are able to control a P300-based brain–computer interface (BCI)? *Neurosci Lett* 2009;462:94–8.
- Jin J, Horki P, Brunner C, Wang X, Neuper C, Pfurtscheller G. A new P300 stimulus presentation pattern for EEG-based spelling systems *Biomedizinische Technik* 2010;55:203–10.
- Kaper M, Meinicke P, Grossekhoefer U, Lingner T, Ritter H BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng* 2004;51:1073-6.
- Krusienski D, Sellers EW, Cabestaing F, Bayouh S, McFarland DJ, Vaughan TM, et al. A comparison of classification techniques for the P300 speller. *J Neural Eng* 2006;3:299–305.
- Lakey CE, Berry DR, Sellers EW. Manipulating attention via mindfulness induction improves P300-based brain–computer interface performance. *J Neural Eng* 2011;8:025019.
- Mak JN, Arbel Y, Minett JW, McCane LM, Yuksel B, Ryan D, et al. Optimizing the P300-based brain–computer interface: current status, limitations and future directions. *J Neural Eng* 2011;8:025003.
- McFarland DJ, Sarnacki WA, Wolpaw JR. Brain-computer interface (BCI) operation: optimizing information transfer rates. *Biol Psychol* 2003;63:237–51.
- McFarland DJ, Sarnacki WA, Townsend G, Vaughan TM, Wolpaw JR. The P300-based brain-computer interface (BCI): Effects of stimulus rate. *Clin Neurophysiol* 2011;122:731–37.

- Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng* 2004;51:1034-43.
- Schreuder M, Blankertz B, Tangermann M. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PLoS One* 2010;5:e9813.
- Sellers EW, Krusienski DJ, McFarland DJ, Vaughan TM, Wolpaw JR. A P300 event-related brain-computer interface (BCI): The effects of matrix size and interstimulus interval on performance. *Biol Psychol* 2006;73:242-52.
- Serby H, Yom-Tov E, Inbar G. An improved P300-based brain-computer interface. *IEEE Trans Rehabil Eng* 2005;13:89-98.
- Sharbrough F, Chatrian GE, Lesser RP, Luders H, Nuwer M, Picton W. AEEGS guidelines for standard electrode position nomenclature. *Clin Neurophysiol* 1991;8:202-4.
- Speier W, Arnold C, Lu J, Taira RK, Pouratian N. Natural language processing with dynamic classification improves P300 speller accuracy and bit rate. *J Neural Eng* 2012;9: 016004.
- Townsend G, LaPallo BK, Boulay CB, Krusienski DJ, Frye GE, Hauser CK, et al. A novel P300-based brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns. *Clin Neurophysiol* 2010;121:1109-20.