

This article was downloaded by: [Sellers, Eric W.]

On: 29 December 2010

Access details: Access Details: [subscription number 931650943]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## International Journal of Human-Computer Interaction

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t775653655>

### Predictive Spelling With a P300-Based Brain-Computer Interface: Increasing the Rate of Communication

D. B. Ryan<sup>a</sup>; G. E. Frye<sup>a</sup>; G. Townsend<sup>b</sup>; D. R. Berry<sup>a</sup>; S. Mesa-G<sup>a</sup>; N. A. Gates<sup>a</sup>; E. W. Sellers<sup>a</sup>

<sup>a</sup> East Tennessee State University, Johnson City <sup>b</sup> Algoma University, Sault Ste. Marie, Ontario, Canada

Online publication date: 28 December 2010

**To cite this Article** Ryan, D. B. , Frye, G. E. , Townsend, G. , Berry, D. R. , Mesa-G, S. , Gates, N. A. and Sellers, E. W.(2011) 'Predictive Spelling With a P300-Based Brain-Computer Interface: Increasing the Rate of Communication', International Journal of Human-Computer Interaction, 27: 1, 69 – 84

**To link to this Article: DOI:** 10.1080/10447318.2011.535754

**URL:** <http://dx.doi.org/10.1080/10447318.2011.535754>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## ***Predictive Spelling With a P300-Based Brain-Computer Interface: Increasing the Rate of Communication***

**D. B. Ryan<sup>1</sup>, G. E. Frye<sup>1</sup>, G. Townsend<sup>2</sup>, D. R. Berry<sup>1</sup>, S. Mesa-G<sup>1</sup>,  
N. A. Gates<sup>1</sup>, and E. W. Sellers<sup>1</sup>**

<sup>1</sup>East Tennessee State University, Johnson City

<sup>2</sup>Algoma University, Sault Ste. Marie, Ontario, Canada

This study compared a conventional P300 speller brain-computer interface (BCI) to one used in conjunction with a predictive spelling program. Performance differences in accuracy, bit rate, selections per minute, and output characters per minute (OCM) were examined. An 8 × 9 matrix of letters, numbers, and other keyboard commands was used. Participants ( $N = 24$ ) were required to correctly complete the same 58 character sentence (i.e., correcting for errors) using the predictive speller (PS) and the nonpredictive speller (NS), counterbalanced. The PS produced significantly higher OCMs than the NS. Time to complete the task in the PS condition was 12 min 43 s as compared to 20 min 20 sec in the NS condition. Despite the marked improvement in overall output, accuracy was significantly higher in the NS paradigm. P300 amplitudes were significantly larger in the NS than in the PS paradigm, which is attributed to increased workload and task demands. These results demonstrate the potential efficacy of predictive spelling in the context of BCI.

### **1. INTRODUCTION**

Brain-computer interface (BCI) technology can help people with severe neuromuscular disease communicate (Wolpaw & Birbaumer, 2006). For example, amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that may eventually cause people to become completely paralyzed, or locked-in to their bodies, and typically causes death within 2 to 5 years (Kunst, 2004). Until recently, it was assumed that cognitive function remains intact even in advanced stages of ALS; however, current research shows that some people with ALS experience some type of cognitive impairment, although the actual number of people affected is still

---

This article is not subject to copyright law.

This work has been supported by NIBIB & NINDS, NIH (EB00856), NIDCD, NIH (1 R21 DC010470-01), NIDCD, NIH (1 R15 DC011002-01). We thank Juliane Armstrong, James Bailey, Chris Hauser, Tiffany Lewis, and Kayla Winnen for data collection. We thank Peter Brunner, Geoff Bashore, and Steve Carmack for software that supports BCI2000.

Correspondence should be addressed to E. W. Sellers, Department of Psychology, East Tennessee State University, P.O. Box 70649, Johnson City, TN 37614. E-mail: sellers@etsu.edu

debated (Murphy et al., 2007). Nonetheless, people with advanced ALS have little or no means of effective communication given existing alternative and augmentative communication (AAC) devices. For these people, a BCI may be the only option for independent communication.

The P300 BCI is based on event-related potentials (ERPs). An ERP is a time-locked electrophysiological brain response to a meaningful stimulus. The P300 ERP is a positive going deflection occurring approximately 300 ms postevent. P300 BCI has received much attention because it requires little training due to the P300 ERP being elicited by meaningful attended stimuli (Picton, 1992; Ritter & Vaughan, 1969), and as compared to other BCIs it produces high bit rates (e.g., Serby et al., 2005). The first P300 BCI was described by Farwell and Donchin (1988). Since that time, approximately 90 articles addressing the topic have been published. Moreover, people have now begun to use the P300 BCI in their homes on a daily basis (Sellers, Vaughan, & Wolpaw, in press) and Vaughan et al. (2006) have described a research program focused on placing BCIs in numerous homes of people with severe communication disorders. The system uses BCI2000 software (Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004), and can provide icon selection, alphanumeric character selection, and multiple menus. These components can provide input to other software and even environmental control. It is now clear that a P300 BCI can be an effective method of communication for ALS patients (e.g., Kubler et al., 2005; Nijboer et al., 2008; Sellers & Donchin, 2006; Sellers et al., in press).

The P300 BCI first models a given participants response to attended stimuli and then uses that information to try and determine which of the items being presented is the one that the subject wishes to select. Typically, the P300 BCI can provide between three and eight selections per minute; this study examines how a predictive speller can transform these selections into additional output characters and the predictive speller's effects on performance measures. Previous studies have measured performance through accuracy (percentage correct), selections per minute (total selections correct or incorrect in a minute), and bit rate (formulated from accuracy, number of possible choices, and time to complete a task). In this study, we introduce a new performance measure "output characters per minute," or OCM. OCM was calculated by taking the total selections to complete a session (including spaces, and a selection to end the session) and dividing it by the total time to complete the task. This new measure was used to calculate the contribution of the predictive speller program. It is important to include a performance measure such as OCM when examining the effectiveness of a BCI system because it provides more useful information than accuracy and/or bit rate alone. That is, OCM provides information about how "powerful" each selection is in terms of what it can accomplish. In other words, OCM is more or less independent of accuracy and bit rate. In addition, OCM is certainly more important to the BCI user than bit rate because it provides a realistic assessment of the system output, which bit rate cannot.

Predictive spelling applications have previously been examined in the context of AAC devices. Typically these comparisons use interfaces such as manual typing (Venkatagiri, 1994), mouth stick typing (Koester & Levine, 1994b, 1996), or touch screen typing (Trnka, McCaw, Yarrington, McCoy, & Pennington, 2009). A primary

goal of this research is to examine and maximize the benefits of word prediction by reducing user effort and maximizing output; however, these studies have produced conflicting results regarding the efficacy of predictive spelling applications (Garay-Vitoria & Abascal, 2004, 2006). Some researchers have suggested that keystroke savings as high as 50 to 60% is a realistic limit of the benefits of delayed word prediction with an AAC user (Copestake, 1997; Leshner & Rinkus, 2002). Conversely, it has been noted that significant cognitive demands occur with the use of word prediction programs and that savings in keystrokes do not necessarily lead to an increase in the rate of communication (Koester & Levine, 1994a; Venkatagiri, 1994).

Because a predictive speller may enable the user to produce more information with fewer selections, it has the ability to enhance communication for those who depend on a P300 BCI. Although predictive spellers have been used in-home with ALS patients (Sellers et al., in press), a formal comparison between the use of a predictive P300-speller and a conventional P300 speller has never been conducted. Therefore, we integrated a predictive speller software package into a P300 BCI and compared its performance to a nonpredictive (i.e., conventional) system.

### **1.1. The Present Study**

To approximate in-home use, participants were required to accurately copy a sentence and stop the session once complete. This is the first study to hold the participant to the same simple, yet tedious, demands of an in-home user. To make P300 BCIs more viable for everyday home use by individuals who rely on communication devices, the program must be able to quickly output words without sacrificing accuracy. Conventional performance measures (i.e., accuracy, bit rate) were not designed for an additional output from a second program such as a predictive speller. These performance measures are based only on single selections made by the user; they do not encompass the potential output of a selection. Thus, OCM was used to accurately measure the advantage or disadvantage of the predictive speller.

We predicted that the predictive spelling (PS) paradigm will improve performance, in terms of OCM, as compared to the nonpredictive spelling (NS) paradigm because the same number of selections per minute (or bit rate) should allow participants to select several items at a time (i.e., words).

We also predict that in the PS paradigm, P300 amplitude may be reduced and P300 latency may be lengthened due to increases in workload or dual task interference in the PS paradigm (e.g., Isreal, Chesney, Wickens, & Donchin, 1980; Isreal, Wickens, Chesney, & Donchin, 1980; Kramer, Wickens, & Donchin, 1985; Wickens, Kramer, Vanasse, & Donchin, 1983). It is reasonable to assume that using a PS in addition to a BCI is more cognitively demanding than using a conventional BCI. In the conventional method, other than attending to the desired item, the only task of the participant is to evaluate the feedback between selections and determine what to select next, either backspace or the next character. Using a PS requires more attentional resources than the conventional method. An individual using a PS must (a) evaluate whether an item is correct; (b) decide if an incorrect item must

be corrected; (c) evaluate the list of suggested words from the predictive speller; and (d) determine whether the next selection will be a backspace, an undo, a word from the list, or the next character of a word. Indeed, predictive spellers used in non-BCI context have shown an increase in cognitive demand (Koester & Levine, 1994; Venkatagiri, 1994). These cognitive effects will become evident in the performance measures, but any negative effects will be overshadowed by the increase in communication rate.

## **2. METHODS**

### **2.1. Participants**

Twenty-nine able-bodied adults were recruited from the East Tennessee State University undergraduate subject pool. Twenty-four (10 men, 14 women; age range = 18–47) completed the experiment. All were naive to BCI use, and none had uncorrected visual impairments or any known cognitive deficit. The study was approved by the East Tennessee State University Institutional Review Board, and each subject gave informed consent.

### **2.2. Experimental Paradigm**

Each participant completed two experimental sessions on separate days within a 1-week period. Participants completed one PS and one NS session; sessions were counterbalanced to control for order effects. Each session consisted of a calibration phase and an online test phase using an identical  $8 \times 9$  matrix. Classification coefficients (described next) were generated with data collected during the calibration phase and subsequently applied during the online test phase. In each phase, participants were provided target items to select. In the calibration phase, items were displayed at the top of the monitor with the next item-to-spell (the target item) indicated in parentheses at the end of the word. As shown in Figure 1A, if the assigned word was “DRIVING,” it would appear at the beginning of the run as: DRIVING (D). The participant’s task was to attend to (or count) the number of times the item in parentheses flashed. After the first item, there was a 3.5-s pause before the next target appeared in parentheses (e.g., DRIVING (R)). This process repeated until the word was complete (one run). Data were collected from five such runs (four words and one numeric string). For both the PS and NS, each set of items flashed for 62.5 ms. This was followed by a 62.5-ms interstimulus interval. Thus, a flash occurred every 125 ms (i.e., eight flashes/second). For each of the 36 calibration items, five complete sequences (i.e., including 10 flashes of the target item) occurred. The flashes were presented using the checkerboard paradigm, which presented items in a quasi-random format. The checkerboard paradigm allows neither adjacent items to flash in the same group nor any item to flash without a minimum of six intervening flashes (for more details, see Townsend et al., 2010).



**FIGURE 1** A) The  $8 \times 9$  matrix used during the calibration phase and online spelling phase of the experiment. In this example, the target item “D” is noted by the letter in parentheses at the end of the word. Participants are instructed to count the number of time the target item flashes. After all items have flashed a predetermined number of times, there is a 3.5-s pause in which the item in parentheses changes to the next letter of the word to indicate the next target item. B) The  $8 \times 9$  matrix and additional windows used during the online spelling phase of the experiment. Right: the flashing matrix used to make item selections. Left top: the sentence target window. Left middle: the sentence output window. Left bottom: the predictive spelling window used in the PS condition (see text for details).

During the online test phase of the NS paradigm, participants copied a sentence from a Notepad “target window” to a blank Notepad “output window” (Figure 1B, top and middle left). The target sentence consisted of 58 selections, including spaces between words, a period, and a Sleep command to end the session. At the beginning of the test phase the output window was blank and the participant’s task was to copy the entire sentence correctly; lowercase letters were used for the output window to reduce possible confusion between the target and output windows. After each item selection feedback was presented to the participant (as a translucent character that filled approximately 30% of the screen), and the keystroke was entered into the Notepad output window. In the event of an incorrect selection, the participant was required to use the Backspace command to erase the error and then correct the selection. After each selection a 6-s pause was provided before the next set of sequences began to flash. This pause was provided to ensure that the participant had sufficient time to evaluate the feedback presented by the BCI, decide what the next item selection should be, and find the correct item in the  $8 \times 9$  matrix.

The online test phase of the PS paradigm was identical to that of the NS except for the addition of the Quillsoft WordQ2 (version 2.5) predictive spelling program (Figure 1B, left bottom). BCI2000 (Schalk et al., 2004) includes a user datagram protocol (UDP) that can send output to peripheral programs. The interface between WordQ2 and BCI2000 was achieved using the BCIKeyboard, a program written and supported by the BCI2000 software project. Once an item had been selected and appeared in the output window, the WordQ2 window would populate with seven words, each preceded by a number. In the event that participants desired to select a word from the list, they could “select” the corresponding number in the  $8 \times 9$  matrix on the next selection by attending to the flashes of the desired number. In Figure 1B, once the “y” had been selected, the WordQ2 window generates the word “your” as choice 1. Thus, to select the word *your*, the participant would select the number 1 from the matrix. Upon selecting the 1 from the matrix, WordQ2 would type the remaining characters “our” and a space, thus completing the word in the output window. At this time, WordQ2 would populate with the seven most probable words. If the participant’s target word did not appear in the WordQ2 list, it was necessary to provide additional characters until the word appeared in the predictive window or it was completed. As every participant was spelling the same sentence, the learning vocabulary feature of WordQ2 was disabled to prevent the program from listing each target word after a single selection. In the event that a word was incorrectly selected (e.g., 2 was selected instead of 1), the participant could select Escape (Esc) from the matrix and WordQ2 would undo the selection. Thus, returning the participant to the previous location in the sentence. However, if a participant was attending to Esc and the resulting selection was incorrect, the participant was required to backspace all of the incorrect characters individually (a limitation of WordQ2 for the current application). In this way, a predictive speller can provide powerful correct selections with time savings and powerful errors with time losses.

Not all errors required a correction. Under certain conditions, the predictive speller also corrected misspelled words. For example, if the output window read “plos” the predictive speller would still list “please” as one of the options and

would correct the errors if “please” was selected. If End or RtArw was selected the cursor in the output window would not move; it only cost the participant a single selection. Participants were not required to correct an error if F5 was selected. In this case, a date/time stamp would appear in the Notepad window. The participant was asked ignore the mistake and attend to the next selection. Once this error was observed it was addressed by changing F5 to F6 in the matrix, which has no output in Notepad thus keeping the error and correct selection count consistent across all participants.

### **2.3. Sentence Selection**

The length of the sentence is typical of a moderately easy sentence in English, the selected words are representative of the mean length of words in English, and five of the 10 words are in the 200 most common English words (Brysbaert & New, 2009). Thus, the sentence used in the online test phase was made up of 50% of the 200 most commonly used words in the English language.

### **2.4. Data Acquisition, Processing**

Participants were seated in a chair approximately 1 m from a computer monitor that displayed an  $8 \times 9$  matrix of letters, numbers, and other keyboard commands. A 72-item speller matrix was used because it is similar to the one designed for home use (Sellers et al., 2010). Moreover, larger matrices have been shown to increase P300 amplitude as the probability of the desired item is reduced (Allison & Pineda, 2003; Sellers, Krusienski, McFarland, Vaughan, & Wolpaw, 2006).

Electroencephalograph (EEG) was recorded with a 32-channel electrode cap embedded with tin electrodes (Electro-Cap International, Inc., Eaton, OH). All channels were referenced to the right mastoid and grounded to the left mastoid. Impedance on each channel was reduced below  $10.0 \text{ k}\Omega$  before testing began. Two g.tec (Guger Technologies, Graz, Austria) 16-channel biosignal amplifiers (version 2) were used. The amplifiers have a  $\pm 250 \text{ mV}$  input sensitivity and are amplified to  $\pm 2 \text{ V}$  before the ADC converts the signals to digital format. Signals were sampled at a rate of 256 Hz, high-pass filtered at 0.5 Hz, and low-pass filtered at 30 Hz. Before analyses EEG data were moving average filtered and downsampled to 20 Hz. Thirty-two channels were collected for the possibility of future analysis, but only electrodes Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz (Sharbrough, Lesser, Lüders, Nuwer, & Picton, 1991) were used for BCI operation (Krusienski, Sellers, McFarland, Vaughan, & Wolpaw, 2008).

Due to the P300s low signal-to-noise ratio, each item must be flashed multiple times and the results averaged (Cohen & Polich, 1997). During calibration, the number of target item flashes was constant across participants and presentation methods. Item sets of six were flashed in quasi-random groups, with two flashes of each of the 72 items of the matrix flashing twice per sequence, and 10 times in the five sequences of each selection. In the calibration phase for the PS and NS



conditions, 36 target items were presented; each of the 36 item selections contained 120 flashes (360 targets and 3,960 nontargets).

## 2.5. Classification

Classification coefficients were determined with a stepwise linear discriminate analysis (SWLDA) algorithm (Draper & Smith, 1981) implemented in MATLAB (version 7.6 R2008a, `stepwisefit` function). The SWLDA algorithm performs forward and backward partial regression procedures to select the spatiotemporal features (i.e., features determined by the combination of electrode location and specific time points during the recording epoch) that account for the most unique variance. Initially, the single feature that accounts for the most unique variance is added to the model (forward regression), then the feature accounting for the most unique remaining variance is added (forward regression). The model is then tested to determine if each feature of the two-feature model still accounts for a significant amount of unique variance (backward regression); if so, both features remain in the model and a third is selected. This forward and backward process continues until the model includes the maximum number of features (set to 60) or until no additional features reach the criteria for entry or removal from the model ( $p < .10$  for entry and  $p > .15$  for removal). SWLDA outputs a set of spatiotemporal classification coefficients that are subsequently applied to the averaged ERP responses during the online phase.

Before the online phase, the number of sequences was optimized for each participant using the maximum written symbol rate (or symbols/minute; Furdea et al., 2009; Townsend et al., 2010). This metric determines the number of item selections a participant can correctly make in 1 min, taking into account error correction. Using the written symbol rate, nearly all participants were presented with fewer than five sequences during the online test phase. In theory, the calibration phase should yield equal numbers of sequences for each participant in each paradigm because the calibration tasks are identical for each session. Given our goal of comparing the PS and NS in an unbiased means, we sought to match the number of sequences in the PS and NS conditions. Thus, five of the participants were removed from the study due to having a difference in optimal sequences equal to or greater than two after calibration. Each sequence of flashes requires three seconds; thus, a difference of two or more sequences yields a minimum of 6 additional seconds per selection. Such a large difference would have confounded the primary goal of the study. By eliminating these five participants the two paradigms were better matched for time and accuracy.

After the matrix flashed the predetermined number of times during online testing, ERPs were averaged for each channel and each cell of the 72-matrix item locations, and then the spatiotemporal coefficients were multiplied by the amplitude value of each model feature. The matrix item with the highest summed score was selected by the classifier and presented to the participant as feedback. The method used was analogous to that used by Krusienski et al. (2008), with the exception that eight channels were used.

The present experimental paradigm derived a classifier for each session independently because within participant differences between sessions could influence

performance. For example, if a participant has had a variable amount of sleep or caffeine it is possible that such variables would affect attentional processes and waveform morphology. In addition, removing and replacing the cap may result in electrodes being located at slightly different locations, contributing to deleterious effects on classification performance in the subsequent session. Thus, performing two calibration sessions should have provided classifiers best suited for a given session.

## 2.6. Dependent Measures

Accuracy was measured by taking the number of correct selections (i.e., feedback matched the character to which the participant was attending) and dividing this value by the total number of selections per session. The formula for calculating bit rate described by Pierce (1980) incorporates the number of possible targets ( $N$ ) and the probability that the target is accurately classified ( $P$ ):

$$\text{Bit Rate} = \log_2 N + P + (1 - P) \log_2(1 - P/N - 1) \quad (1)$$

The result is divided by number of minutes in a session yields bits per minute. The calculation “selections per minute” was performed by taking the total number of selections and dividing by the total time of the session. OCM was calculated by taking the 58 total selections in each session (including sleep) and dividing it by the total time of the PS session. OCM was used to calculate the contribution of the predictive speller program. This calculation includes the time it took for the participant to correct errors while the number of correct target selections (58) remained static. Therefore, the more errors a participant made, the more time it took to finish the session, resulting in lower output characters per minute. However, PS and NS selections per minute were a direct result of sets per sequence and time, thus not affected by error correction.

## 3. RESULTS

A  $2 \times 2$  mixed model analysis of variance—Order (NS first vs. PS first)  $\times$  Condition (NS vs. PS)—was used to examine if an order effect was present in the data. The results provided insufficient evidence to reject the null hypothesis,  $F(1, 22) = 0.185, p = .671$ . Thus, we collapsed across the conditions and analyzed the data using paired  $t$  tests to examine the differences between the PS and NS conditions on the measure of mean accuracy, selections per minute, bit rate, theoretical bit rate, output characters per minute, and waveform latency and amplitude.

### 3.1. Online Accuracy, Bit Rate, and Theoretical Bit Rate

Table 1 shows raw scores and means for accuracy, bit rate, and theoretical bit rate. Online accuracy was significantly higher for NS ( $M = 89.80\%$ ,  $SD = 7.78$ ) than

**Table 1: Online Test Phase Accuracy (Acc), Bit Rate (BR), and Theoretical Bit Rate (Theo BR) for the Predictive Speller (PS) and Nonpredictive Speller (NS)**

<i>Subject</i>	<i>PS Acc</i>	<i>NS Acc</i>	<i>PS BR</i>	<i>NS BR</i>	<i>PS Theo BR</i>	<i>NS Theo BR</i>
1	96.88	95.31	23.70	28.26	39.33	56.09
2	88.89	87.50	19.93	19.54	32.62	32.38
3	70.00	88.16	11.48	16.46	17.11	24.58
4	79.59	89.86	18.78	20.41	33.52	33.82
5	91.89	92.65	17.71	15.39	26.33	21.50
6	87.18	95.31	21.73	22.58	38.70	37.39
7	91.67	100.00	21.21	24.85	34.96	41.13
8	81.13	87.50	15.79	21.72	24.66	38.86
9	80.95	70.83	17.35	17.60	28.64	35.05
10	82.35	98.33	22.28	29.98	44.12	59.45
11	80.00	91.18	12.11	14.91	16.87	20.79
12	77.59	82.50	11.55	12.69	16.10	17.70
13	82.22	93.94	17.61	22.00	28.91	36.45
14	94.29	77.17	22.01	14.57	36.00	22.81
15	91.18	95.31	19.10	20.51	29.70	32.05
16	94.29	85.25	18.52	15.62	27.51	23.29
17	72.50	77.23	8.18	11.45	10.48	15.98
18	91.89	100.00	26.69	31.12	52.65	61.70
19	100.00	100.00	25.00	24.85	41.13	41.13
20	96.88	91.18	21.25	19.00	33.01	29.70
21	86.67	91.43	16.06	14.95	23.92	20.82
22	57.58	83.67	5.02	15.13	6.19	22.62
23	67.07	96.77	11.80	16.55	18.46	23.06
24	94.44	84.15	20.27	15.28	31.54	22.82
<i>M</i>	84.88	89.80	17.71	19.39	28.85	32.13
<i>SD</i>	10.59	7.78	5.38	5.39	10.95	12.83
<i>SE</i>	2.16	1.59	1.10	1.10	2.24	2.62

for the PS ( $M = 84.88\%$ ,  $SD = 10.59$ ),  $t(23) = 2.15$ ,  $p = .04$ ,  $d = 0.40$ . We suspect that lower accuracy in PS is attributed to the higher workload and/or dual task processing requirements of the PS paradigm. In addition, we found marginal differences between PS bit rate and NS bit rate ( $M = 17.71$ ,  $SD = 5.38$ ;  $M = 19.39$ ,  $SD = 5.37$ , respectively),  $t(23) = 2.04$ ,  $p = .053$ ,  $d = 0.39$ . Theoretical bit rate (i.e., bit rate with the time between selections removed) is presented for comparison to studies that report bit rate with the time between selections removed; in this study 6 s were provided between each item selection.

### 3.2. Selections per Minute

Table 2 shows raw scores and means for PS and NS sets per sequence, time to complete the sentence, selections per minute, and OCM. We compared means of PS selections per minute against NS selections per minute ( $M = 3.71$ ,  $SD = 0.75$ ;  $M = 3.76$ ,  $SD = 0.75$ , respectively) and found no difference between groups,

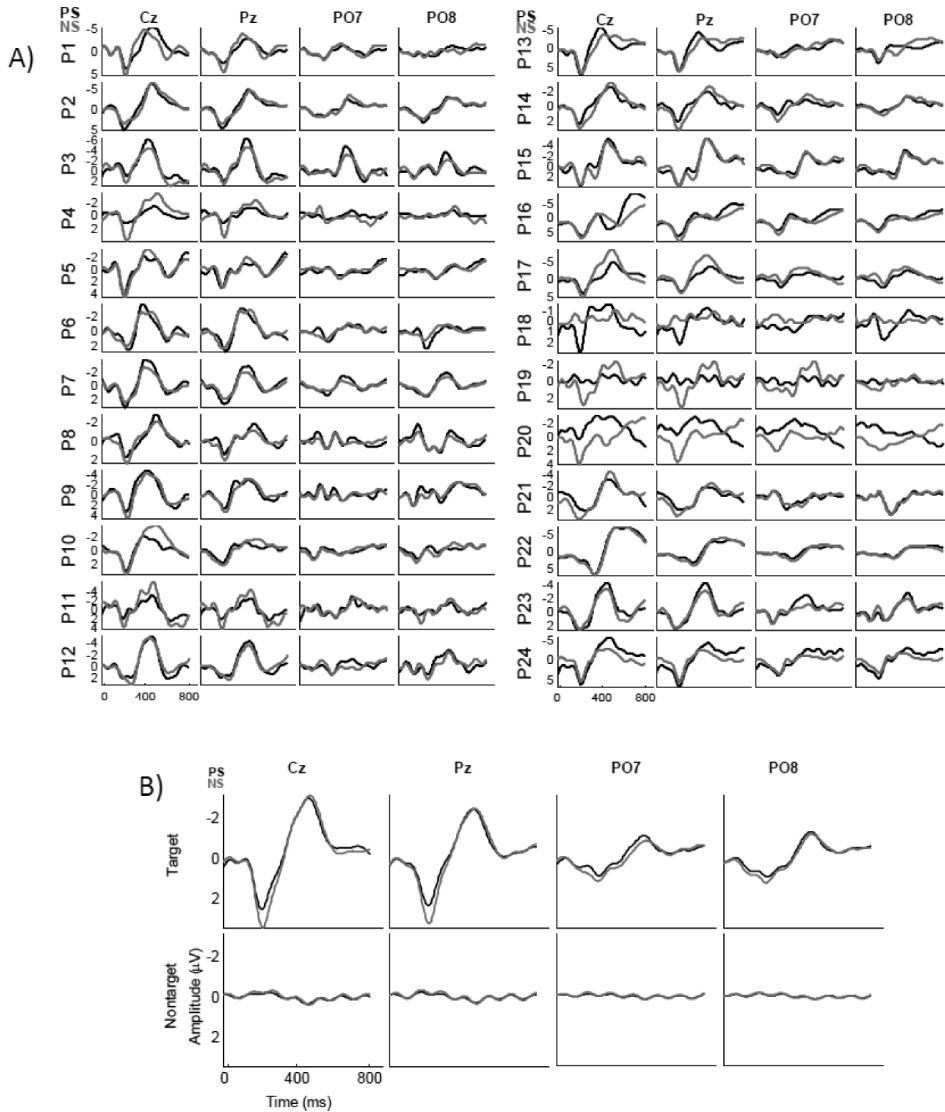
**Table 2: Online Test Phase Sets per Sequence (Sets/Seq), Time to Complete the Sentence (Comp(min)), and selections per Minute (Sel/min) in the Predictive Speller (PS) and Nonpredictive Speller (NS) Paradigms, and the Predictive Output Characters per Minute (OCM)**

<i>Subject</i>	<i>PS Sets/Seq</i>	<i>NS Sets/Seq</i>	<i>PS Comp(min)</i>	<i>NS Comp(min)</i>	<i>PS Sel/min</i>	<i>NS Sel/min</i>	<i>PS OCM</i>
1	3.00	2.00	7.80	12.70	4.10	5.04	7.44
2	3.00	3.00	9.00	17.90	4.00	4.02	6.44
3	4.00	4.00	24.00	22.70	3.33	3.35	2.42
4	2.50	3.00	10.92	17.15	4.49	4.02	5.31
5	4.00	5.00	11.00	23.58	3.36	2.88	5.27
6	2.50	3.00	8.67	15.90	4.50	4.03	6.69
7	3.00	3.00	8.90	14.40	4.04	4.03	6.52
8	3.50	2.50	14.47	16.10	3.66	4.47	4.01
9	3.00	2.00	10.40	23.90	4.04	5.02	5.58
10	2.00	2.00	10.10	11.90	5.05	5.04	5.74
11	5.00	5.00	19.15	23.70	2.87	2.87	3.03
12	5.00	5.00	20.20	27.90	2.87	2.87	2.87
13	3.00	3.00	11.25	16.40	4.00	4.02	5.16
14	3.00	3.50	8.75	25.20	4.00	3.65	6.63
15	3.50	3.50	9.25	17.50	3.68	3.66	6.27
16	4.00	4.00	10.40	18.20	3.37	3.35	5.58
17	5.00	5.00	17.75	35.25	2.25	2.87	3.27
18	2.00	2.00	7.30	11.50	5.07	5.04	7.95
19	3.00	3.00	7.65	14.40	4.05	4.03	7.58
20	3.50	3.50	8.70	18.60	3.68	3.66	6.67
21	4.00	5.00	13.40	24.45	3.36	2.86	4.33
22	3.50	4.00	16.95	29.30	1.95	3.34	3.42
23	3.50	5.00	22.45	21.60	3.65	2.87	2.58
24	3.50	4.00	9.80	24.50	3.67	3.35	5.92
<i>M</i>	3.42	3.54	12.43	20.20	3.71	3.76	5.28
<i>SD</i>	0.830	1.062	4.963	5.978	0.745	0.749	1.666
<i>SE</i>	0.169	0.217	1.013	1.220	0.152	0.153	0.340

$t(23) = 0.49, p = .62, d = 0.10$ . Although this comparison provided null findings, when compared to OCM significant differences were revealed. OCM was significantly higher than PS selections per minute ( $M = 5.28, SD = 1.67$ ),  $t(23) = 6.05, p < .001, d = 0.78$ . Similarly, OCM was significantly higher than NS selections per minute,  $t(23) = 5.61, p < .001, d = 0.76$ . Moreover in total time to complete the sentence (in minutes), the PS was significantly faster than the NS paradigm ( $M = 12.43, SD = 4.96; M = 20.20, SD = 5.98$ , respectively),  $t(23) = 7.52, p < .001, d = 0.84$ .

### 3.3. Waveform Morphologies

The PS and NS produced virtually identical waveforms. Our analyses focused on the electrodes Cz, Pz, Po7, and Po8 because most of the P300 amplitude change in BCI applications is captured in these four electrodes (Kaper, Meinicke,



**FIGURE 2** A) Target waveforms for electrode locations Cz, Pz, Po7, and Po8 for each of the 24 participants; PS paradigm data are presented in black and NS paradigm data are presented in gray. (Amplitude units are  $\mu\text{V}$ .) B) Grand mean waveforms for all 24 participants at electrode locations Cz, Pz, Po7, and Po8. The top row consists of target responses for both paradigms, and the bottom row consists of nontarget responses for both paradigms. PS data are presented in black and NS data are presented in gray.

Grossekathoefer, Lingner, & Ritter, 2004; Krusienski et al., 2008). Figure 2A shows average target waveforms for each of the 24 participants. Figure 2B shows the grand mean waveforms for the target waveforms (top row) and the nontarget waveforms (bottom row). The difference in the positive peak at electrode location

Cz around 200 ms was marginally higher in the NS than in the PS paradigm ( $M = 3.45$ ,  $SD = 1.47$ ;  $M = 2.82$ ,  $SD = 1.71$ , respectively),  $t(23) = 2.06$ ,  $p = .051$ ,  $d = 0.39$ . In addition, the NS peak at electrode location Pz around 200 ms was significantly larger than the PS peak ( $M = 3.82$ ,  $SD = 1.49$ ;  $M = 3.24$ ,  $SD = 1.81$ , respectively),  $t(23) = 2.34$ ,  $p = .028$ ,  $d = 0.43$ .

#### 4. DISCUSSION

The primary goal of this study was to test the efficiency of a predictive speller program in conjunction with a P300 BCI. The main hypotheses were that the predictive speller should improve overall character output and possibly affect waveform morphology. The first hypothesis was supported, even though accuracy was significantly lower in the PS paradigm, and bit rate and selections per minute were statistically equivalent in both paradigms. Despite the NS advantage in accuracy, the PS showed an average time advantage of 7 min 37 s over the NS, and OCM were significantly higher for the PS than the NS by 1.51 characters/minute. Given the current maximum character selection rate of approximately four selections per minute in P300 BCIs (also see Lenhardt, Kaper, & Ritter, 2008; Townsend et al., 2010), these results impressively convert to an additional 91.2 output characters per hour, or nearly 1.5 per minute. These results suggest that a predictive speller can provide a substantial advantage to an individual communicating via a P300 Speller in an online environment.

The significant difference in accuracy between the two paradigms may be a result of increased workload and/or task difficulty associated with the PS. This hypothesis is indirectly supported by the finding of lower amplitude responses in the PS condition at the Cz and Pz electrode locations. Previous P300 research has shown that workload (i.e., the measure of the interaction between task difficulty and an individual's ability to perform a given task; Gopher & Donchin, 1986), and dual task interference can significantly reduce P300 amplitude and increase P300 latency (Gopher & Donchin, 1986; Isreal, Chesney, et al., 1980; Isreal, Wickens, et al., 1980; Kramer, Wickens, & Donchin, 1983; Kramer et al., 1985; Wickens et al., 1983). The relatively small amplitude differences in the current study may be due to the fact that the increase in workload was discontinuous (i.e., increased during the time in which target stimuli were not flashing). This is in contrast to studies investigating workload that typically use continuous increases in task demands (e.g., tracking a stimulus). In addition, the AAC literature also suggests that cognitive demand is increased when a predictive speller is used (Koester & Levine, 1994; Venkatagiri, 1994a).

As this study used naive participants, we believe that with training PS accuracy will increase, thus increasing OCM. Gopher and Donchin (1986) suggested that the effects of workload decrease with practice. In addition, the predictive speller can learn to adapt to the individual over time, which we did not allow in the current study.

Further support of the inefficiency of the naive participants to use a predictive speller is shown by the number of selections required for an ideal user to complete the sentence; only 31 selections were necessary using the untrained

predictive speller. However, many participants failed to select a word from the predictive speller at the first opportunity, leading to additional unnecessary selections.

## 5. CONCLUSIONS

These results demonstrate the potential efficacy of predictive spelling in the context of BCI. Future research should be conducted in an ALS population to determine if similar improvements in output character selections are obtained.

## REFERENCES

- Allison, B. Z., & Pineda, J. A. (2003). ERPs evoked by different matrix sizes: Implications for a brain computer interface (BCI) system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*, 110–113.
- Brysaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.
- Cohen, J. D., & Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, *25*(3), 6.
- Copestake, A. (1997). *Augmented and alternative NLP techniques for augmentative and alternative communication*. Paper presented at the Natural Language Processing for Communication Aids. Stanford, CA: Stanford University.
- Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York, NY: Wiley.
- Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *70*, 510–523.
- Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., & Kübler, A. (2009). An auditory oddball (P300) spelling system for brain–computer interfaces. *Psychophysiology*, *46*, 617–625.
- Garay-Vitoria, N., & Abascal, J. (2004). A comparison of prediction techniques to enhance the communication rate. *User-Centered Interaction Paradigms for Universal Access in the Information Society: 8th ERCIM Workshop on User Interfaces for All*, 3196, 400–417. doi:10.1007/b95185
- Garay-Vitoria, N., & Abascal, J. (2006). Text prediction systems: A survey. *Universal Access in the Information Society*, *4*, 20.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. K. Boff, T. Lloyd, & P. James (Eds.), *Handbook of perception and human performance*, Vol. 2: *Cognitive processes and performance* (pp. 1–49). Oxford, UK: Wiley & Sons.
- Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology*, *17*, 259–273.
- Isreal, J. B., Wickens, C. D., Chesney, G. L., & Donchin, E. (1980). The event-related brain potential as an index of display-monitoring workload. *Human Factors*, *22*, 211–224.
- Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T., & Ritter, H. (2004). BCI Competition 2003—Data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng*, *51*, 1073–1076.

- Koester, H. H., & Levine, S. P. (1994a). Learning and performance of able-bodied individuals using scanning systems with and without word prediction. *Assistive Technology*, 6(1), 42–53.
- Koester, H. H., & Levine, S. P. (1994b). Modeling the speed of text entry with a word prediction interface. *IEEE Transactions on Rehabilitation Engineering*, 2(3), 10.
- Koester, H. H., & Levine, S. P. (1996). Effect of a word prediction feature on user performance. *Augmentative and Alternative Communication*, 12(3), 23.
- Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing requirements of a complex perceptual-motor task. *Human Factors*, 25, 597–621.
- Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: evidence for dual-task integrality. *Journal of Experimental Psychology. Human Perception and Performance*, 11, 393–408.
- Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167(1), 15–21.
- Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T. M., Pawelzik, H., Schalk, G. . . . Wolpaw, J. R. (2005). Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64, 1775–1777.
- Kunst, C. B. (2004). Complex genetics of amyotrophic lateral sclerosis. *American Journal of Human Genetics*, 75, 933–947.
- Lenhardt, A., Kaper, M., & Ritter, H. J. (2008). An adaptive P300-based online brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16, 121–130.
- Lesh, G., & Rinkus, G. (2002). *Domain-specific word prediction for augmentative communication*. Paper presented at the RESNA Annual Conference. Spencerport, NY: Enkidu Research, Inc.
- Murphy, J. M., Henry, R. G., Langmore, S., Kramer, J. H., Miller, B. L., & Lomen-Hoerth, C. (2007). Continuum of frontal lobe impairment in amyotrophic lateral sclerosis. *Archives of Neurology*, 64, 530–534.
- Nijboer, F., Sellers, E. W., Mellinger, J., Jordan, M. A., Matuz, T., Furdea, A., . . . Kübler, A. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119, 1909–1916.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9, 456–479.
- Pierce, J. R. (1980). *An introduction to information theory* (pp. 145–165). New York: Dover.
- Ritter, W., & Vaughan, H. G., Jr. (1969). Averaged evoked responses in vigilance and discrimination: A reassessment. *Science*, 164(3877), 326–328.
- Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51, 1034–1043.
- Sellers, E. W., & Donchin, E. (2006). A P300-based brain-computer interface: Initial tests by ALS patients. *Clinical Neurophysiology*, 117, 538–548.
- Sellers, E. W., Krusienski, D. J., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*, 73, 242–252.
- Sellers, E. W., Vaughan, T. M., & Wolpaw, J. R. (in press). A brain-computer interface for long-term independent home use. *Amyotrophic Lateral Sclerosis*.
- Sellers, E. W., Vaughan, T. M., Wolpaw, J. R. (2010). A brain-computer interface for long-term independent home use. *Amyotrophic Lateral Sclerosis*, 11(5), 449–455.



- Serby, H., Yom-Tov, E., Inbar, G.F., (2005). An improved P300-based brain-computer interface. *IEEE Trans Neural Syst Rehabil Eng.*; 13(1): 89–98.
- Sharbrough, F. C. G., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, W. (1991). AEEGS guidelines for standard electrode position nomenclature. *Clinical Neurophysiology*, 8, 202–204.
- Townsend, G. T., LaPallo, B. K., Boulay, C., Krusienski, D. J., Frye, G. E., Hauser, C. K., . . . Sellers, E. W. (2010). A novel P300-based brain–computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clinical Neurophysiology*, 121, 1109–1120. doi:10.1016/j.clinph.2010.01.030
- Trnka, K., McCaw, J., Yarrington, D., McCoy, K. F., & Pennington, C. (2009). User interaction with word prediction: the effects of prediction quality. *ACM Transactions on Accessible Computing*, 1(3), 34.
- Vaughan, T. M., McFarland, D. J., Schalk, G., Sarnacki, W. A., Krusienski, D. J., Sellers, E. W., & Wolpaw, J. R. (2006). The Wadsworth BCI Research and Development Program: At home with BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14, 229–233.
- Venkatagiri, H. S. (1994). Effect of window size on rate of communication in a lexical prediction AAC system. *AAC Augmentative and Alternative Communication*, 10, 8.
- Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*, 221(4615), 1080–1082.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*. 113(6):767–791.