A P300-based brain–computer interface: Initial tests by ALS patients

Eric W. Sellers [a,*], Emanuel Donchin [b,1]

[a] *New York State Department of Health, Wadsworth Center, E1001 Empire State Plaza, Albany, NY 12201, USA*
[b] *Department of Psychology, University of South Florida, PCD 4118G, 4202 Fowler, Tampa, FL 33620, USA*

See Editorial, pages 479–483

Abstract

Objective: The current study evaluates the effectiveness of a brain–computer interface (BCI) system that operates by detecting a P300 elicited by one of four randomly presented stimuli (i.e. YES, NO, PASS, END).

Methods: Two groups of participants were tested. The first group included three amyotrophic lateral sclerosis (ALS) patients that varied in degree of disability, but all retained the ability to communicate; the second group included three non-ALS controls. Each participant participated in ten experimental sessions during a period of approximately 6 weeks. During each run the participant's task was to attend to one stimulus and disregard the other three. Stimuli were presented auditorily, visually, or in both modes.

Results: Two of the 3 ALS patient's classification rates were equal to those achieved by the non-ALS participants. Waveform morphology varied as a function of the presentation mode, but not in a similar pattern for each participant.

Conclusions: The event-related potentials elicited by the target stimuli could be discriminated from the non-target stimuli for the non-ALS and the ALS groups. Future studies will begin to examine online classification.

Significance: The results of offline classification suggest that a P300-based BCI can serve as a non-muscular communication device in both ALS, and non-ALS control groups.

© 2005 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

Keywords: Amyotrophic lateral sclerosis; Electroencephalogram; Brain–computer interface; P300; Event-related potentials; Rehabilitation

1. Introduction

Farwell and Donchin (1988) have shown that a P300-based brain–computer interface (BCI) can by used by able-bodied young adults, to input a string of characters to a computer using the P300 as a substitute typing finger. Further demonstrations and an assessment of the communication speed achieved by such a system were provided by Donchin et al. (2000) who tested the BCI with wheelchair-bound healthy adults as well as with able-bodied subjects. Similar results, with able-bodied subjects, were obtained by Allison and Pineda (2003). While the P300-BCI was designed to serve the needs of locked-in patients (i.e. people who have lost all voluntary

muscle control), no subject who participated in these initial demonstrations was locked-in. In the study reported here, and in pilot testing, we have conducted an extensive test of the P300-BCI with several amyotrophic lateral sclerosis (ALS) patients and a control group of able-bodied adults. ALS is one of the most common causes of the locked-in syndrome; it is a neurodegenerative disease leading to paralysis and death, typically within 2–5 years (Kunst, 2004). A BCI system will provide locked-in ALS patients whose cognitive abilities are left intact a means of communication that does not depend on neuromuscular control. The focus of the current study is on ALS patients because they are the most likely population to benefit from a BCI.

The current study examines the efficacy of a P300-BCI, with an emphasis on a number of aspects of the system's function not addressed in previous studies. We assess the effect of the BCI across three different stimulation modes (auditory, visual, and auditory + visual). By testing across

* Corresponding author. Tel.: +1 518 474 7958; fax: +1 518 486 4910.
E-mail addresses: esellers@wadsworth.org (E.W. Sellers), donchin@shell.cas.usf.edu (E. Donchin).
[1] Tel.: +1 813 974 0466; fax: +1 813 974 8617.

multiple sessions, we examined the accuracy with which P300s are detected, and hence the systems reliability across multiple experimental sessions. We also tested different detection and classification algorithms that allowed the system to discriminate between stimuli that did, and did not, elicit a P300.

A BCI allows a user to communicate with devices without voluntary muscle activity (i.e. using only the electrical activity of the brain; Wolpaw et al., 2002). A BCI is not a mind-reading device. Rather, its primary function is to provide the subject with a virtual keyboard, whose keys are pressed by aspects of brain activity. Each key press constitutes a choice of an item from the set of items contained in the keyboard, and the subject's choice is indicated through the control of electrical brain activity. Such a system may be used by a completely paralysed or locked-in individual.

1.1. The P300 speller

The P300-BCI, the 'P300 Speller,' described by Farwell and Donchin (1988) presents a 6×6 matrix of characters. Each row and each column are intensified and the intensifications are presented in a random sequence. The subject focuses attention to one of the 36 cells of the matrix. The sequence of 12 flashes, six rows and six columns, constitutes an 'Oddball Paradigm' (Fabiani et al., 1987) with the row, and the column, containing the character to be communicated constituting the rare set, and the other 10 intensifications constituting the frequent set. Items that are presented infrequently (the rare set) in a sequential series of randomly presented stimuli will elicit a P300 response if the observer is attending to the stimulus series. Thus, the row and the column containing the target character will elicit a P300, because when the target stimulus flashes this constitutes a rare event in the context of all other character flashes. Donchin et al. (2000) evaluated the P300 Speller data offline, using stepwise discriminate analysis (SWDA) or a combination of discrete wavelet transformation (DWT) and SWDA. The SWDA and DWT/SWDA algorithms produce classification coefficients that classified responses similarly. Donchin et al. (2000) demonstrated that such coefficients classify the character correctly in an online mode 56% of the time. Additionally, 92% of the time either the row or column is correctly classified (i.e. the classification is half-correct).

1.2. P300 Speller pilot data and the current four-choice paradigm

Preliminary studies of the P300 Speller with ALS patients indicated that some patients found it difficult to communicate by spelling text character by character (Sellers et al., 2003); even though a P300 component was elicited from these patients in a standard oddball task. Sellers et al. (2003) reported that ALS patients' responses were more variable, and the patients may have more difficulty using a matrix that includes many items because of uncontrollable eye movements and the rapid presentation rate of the 6×6 matrix. Although eye movements are not required to orient attention (Posner, 1980; Yantis et al., 2002), involuntary eye movements may make it difficult to orient attention to a specific location. For this reason, we chose to focus the present study on a paradigm based on a four-choice oddball to provide users with the ability to answer simple questions. In practice, this paradigm is similar to the manner of communication used by nearly locked-in patients who retain some rudimentary muscle control. For example, caregivers ask binary questions and the patient may respond YES by looking to the right, or raising an eyebrow, and do nothing for a NO response.

It is impractical to base a P300-BCI on a binary choice, as the probability of the correct item would be 0.50 minimizing the difference in P300 amplitude between the target and non-target stimuli (Duncan-Johnson and Donchin, 1977). For this reason we adopted a four-choice stimulus paradigm. Using four stimuli (YES, NO, PASS, END) in the sequence, subjects would focus on either the 'YES' or 'NO' during each series of flashes. The probability of the target event (i.e. the event the subject wished to select) was 0.25, a probability likely to yield a detectable P300 (Duncan-Johnson and Donchin, 1977; Johnson and Donchin, 1978). Although the main focus of the current paradigm is to test a system to answer YES/NO questions, four choices are presented for experimental purposes. However, practically speaking, if a system is as efficient with twice as many choices, more effective and faster communication will be possible.

1.3. Summary and goals of the current study

Kubler et al. (2001) have shown that severely disabled patients can use slow cortical potentials to operate a BCI. More recently, Kubler et al. (2005) have reported that severely disabled patients can operate a BCI using EEG rhythms recorded over sensorimotor cortex. However, a P300-based BCI has never been tested with an ALS population; therefore, a main impetus of the current study is to determine whether such a system may be a viable communication option for ALS patients. In a locked-in state, ALS patients are thought to be cognitively intact; however, they have lost the ability to communicate by traditional means (see Hanagasi et al., 2002; Lomen-Hoerth et al., 2003; Paulus et al., 2002; for studies that question the cognitive abilities of ALS patients). Unfortunately, effective communication with a completely locked-in patient has not been demonstrated. An essential first step is to determine if the system can be effective with ALS patients who are not yet locked-in. It is important to be able to communicate with the participants because they must be able to provide their intended message.

In the current study, we examine how several different variables affect the classification rates of the P300-BCI.

Extended use of a P300-BCI, mode of presentation, task manipulations, and methods of deriving SWDA weights are all examined. Each of these factors is discussed below.

2. Methods

2.1. Subjects

The subjects were recruited from the local community and with the help of the Amyotrophic Lateral Sclerosis Association, Florida Chapter. Three control participants were included in the study (one man, age 33, and two women, age 31 and 37). Three ALS patients who were able to communicate with the experimenter were also included in the study (two men, age 37 and 44, and one woman, age 50). The ALS patients varied in their degree of mobility. ALS participant one was wheelchair bound but retained the use of upper extremities. ALS participant two could use all extremities, but found it very difficult to walk, and choose to use a wheelchair. ALS participant three retained only head movement and some speech ability. All subjects signed an informed consent approved by the University of South Florida Institutional Review Board, IRB approval 100650.

2.2. Data acquisition and processing

The EEG was recorded using a cap (Electro-Cap International, Inc.) embedded with 16 electrodes covering left, right, and central scalp locations (Fz, Cz, Pz, Oz, Fp1, Fp2, F3, F4, C3, C4, P3, P4, P7, P8, T7, T8) based on the modified 10–20 system of the International Federation (Sharbrough et al., 1991). The recordings were referenced to the right earlobe, and grounded to the right mastoid. The EEG was amplified with a SA Electronics amplifier, digitized at a rate of 160 Hz, high-pass filtered at 0.1 Hz, and low-pass filtered at 50 Hz. The electrode impedance did not exceed 5 kΩ. All aspects of data collection and experimental control were controlled by the BCI2000 system, developed at the Wadsworth Center, New York State Department of Health (Schalk et al., 2004). Signal processing (i.e. time domain averaging, generation of classification algorithms, etc.) was conducted offline using Matlab 7.0 and SPSS 11.0. Before offline analyses were performed, a moving-average filter of four samples and a decimation factor of four samples were applied to the data. Both of these procedures can also be performed online before a classification algorithm is applied to the data.

2.3. Task, procedure, and design

Each subject initially participated in a typical oddball experiment that presented black and white line drawings of a zebra and an elephant as stimuli. The stimuli were presented randomly, at fixation, with a probability of 0.25 for the zebra and a probability of 0.75 for the elephant. Each

stimulus was presented for 600 ms and the ISI was set to 1400 ms (the same presentation characteristics were used in the subsequent four-choice paradigm sessions). The subject's task was to attend to the zebra, by passively counting or noting when it appeared. Because motor responses are not possible for locked-in patients, the current study has adopted this method of attending to target stimuli.

After the initial oddball experiment subjects began testing with the four-choice oddball paradigm. Each stimulus was presented with a probability of 0.25 and the subjects were asked to attend to a stimulus, either YES or NO. This was achieved in one of two ways, depending on task condition. One task was to focus on the target stimulus (i.e. YES or NO as defined by the experimenter at the beginning of each run). The other task was to focus on the stimulus (YES or NO) that correctly answered a question provided by the experimenter. The questions were constructed such that the answer was always unambiguous (e.g. Is today Monday?). Stimulus duration was 600 ms for the visual stimulus, and 600 ms for the auditory stimulus, the ISI was 1400 ms. Each experimental run consisted of 100 stimulus presentations (75 non-target, 25 target). The stimuli were presented randomly in blocks of four (one of each stimulus type), 25 times, for a total of 100 presentations. After each run a short break ensued. The duration of the break was determined by the participant, but was typically around 1 min. All subjects participated in 10 experimental sessions that lasted approximately 1 h. The six subjects took from 4 to 6 weeks to complete the 10 experimental sessions. Each session was composed of 12 runs that were counter-balanced across the following variables, mode of presentation, task, and target item. In total, each session included four runs in each of the three modes, six runs for the target YES, six runs for the target NO, and six runs in each of the two task conditions. See Table 1 for an example run sheet for one session.

The experimental sessions deviated in two important ways from how the system would be implemented in a clinical setting. First, online performance feedback was

Table 1
Example experimental session run sheet: Answer, answer a question; Focus, attend to item provided by experimenter (YES or NO); A, auditory; AV, auditory + visual; V, visual

Run	Task	Mode	Target Item
1	Answer	A	YES
2	Answer	V	NO
3	Answer	AV	YES
4	Answer	A	NO
5	Answer	V	YES
6	Answer	AV	NO
7	Focus	A	YES
8	Focus	V	NO
9	Focus	AV	YES
10	Focus	A	NO
11	Focus	V	YES
12	Focus	AV	NO

random. Random feedback was used so that each participant's performance would not be influenced by how well (s)he was performing (ideally no feedback would have been provided; however, due to a system limitation this was not possible). Second, the number of trials per run was held constant for all subjects and all runs. This was a necessary control so that each participant would have an equal amount of experimental data. In practice, the fewest number of trials that allow a desired level of classification would be used. For example, in the current case, 100 trials were used in each run; thus, 25 instances of each item are available to be averaged together. If only five trials per stimulus are needed to make sufficiently accurate classifications, only 20 trials per run would be necessary. This would allow the subject to make five times the number of classifications in the same amount of time. Obviously, in a clinical setting where both speed and accuracy are at a premium, the minimum necessary number of trials per run should be used.

2.4. SWDA analysis

We applied the step-wise discriminant analysis (SWDA) method of classification for the current analysis. Previous studies of the P300-BCI (e.g. Donchin et al., 2000; Farwell and Donchin, 1988) have demonstrated that SWDA performs equally well or better than several other methods of operationally defining P300 waveforms (Fabiani et al., 1987). Previous research has not, however, examined how deriving SWDA weights from different subsets of data affected classification accuracy. Therefore, the analyses were conducted using three different methods of deriving SWDA weights. Deriving weights from a subset of data and applying them to independent data sets for testing allows us to directly compare performance across the different methods of derivation. In general, two steps are involved in the process. The first step is concerned with deriving weights. The second step applies the weights to data sets that are created from independent samples via bootstrapping.

2.4.1. Deriving weights

Stepwise Discriminant Analysis seeks an optimal discriminant function by adding variables (in this case, time points) to the equation until an optimality criterion is satisfied. In the present study the criteria were set to a minimum of four steps and a max of 10 steps, or a P value of <0.10 for adding a variable, and a P value of >0.15 to eliminate variables. In some cases it was necessary to relax these constraints (e.g. Method 1 (see below) uses only 100 stimuli and in some instances it was necessary to increase the P value to enter variables into the solution, otherwise no features would be selected). All SWDA methods in the current study derived weights from data sets that were independent from the test sets.

We recorded data from 16 electrodes at a sample rate of 160 Hz; therefore, we have 16×160 channel-by-time variables available to derive the classification weights.

Although we have 2560 variables at our disposal it is not necessary to include all of this information in the analysis for two primary reasons. First, P300 amplitude is typically largest along the midline electrodes. Second, using contiguous time points adds a substantial amount of redundant information to the analysis. This may result in the solution 'over-fitting' the data. Over-fitting the training data may reduce the generalizability of the solution to a subsequent data set. Accordingly, three midline electrodes (Fz, Cz, and Pz) were used in the analysis, and the data were decimated by a factor of four. A moving-average of four samples was also applied to the data. The time epoch used for all analyses was 900 ms, beginning at stimulus onset. The use of these parameters results in a total of 108 spatial location $\times$ time features for each analysis.

Each set of SWDA weights was derived separately for each mode of presentation (auditory, visual, and auditory + visual), and applied to data sets that included only same mode data. In other words, each analysis was conducted three times, once for each presentation mode. In addition, three different methods were used to derive SWDA weights. Method 1 used only the first run of a session. One run consisted of 100 stimulus presentations; SWDA weights were derived using 25 attended and 75 non-attended single-trial stimulus presentations. The purpose of method 1 is to simulate a mode in which the system could be calibrated at the beginning of a session and the derived weights could be used online following the calibration period. Method 2 used the data from an entire session to derive the weights; in this case 100 attended and 300 non-attended stimuli were available for the analysis. Method 3 used data from two sessions to derive the weights; in this case 200 attended stimuli and 600 non-attended stimuli were available for the analysis. Method 2 and 3 simulate modes in which weights would be determined offline, following a session (or multiple sessions). The derived weights would then be used in a subsequent session online.

2.4.2. Applying weights

In general, the rare events in the oddball sequence elicit a P300 with considerable reliability, and the P300 can be readily detected provided the number of trials averaged is sufficient to allow the extraction of the P300 'signal' from the EEG 'noise' (Farwell and Donchin, 1988). Of course, the required number of trials determines the speed with which the BCI can operate. Each trial adds 1400 ms to the detection time in the current four-choice paradigm. The critical value for assessing the performance of a BCI is, therefore, the smallest number of trials that allows the detection of the P300 at a given level of accuracy. Farwell and Donchin (1988) used a bootstrapping approach (Efron and Tibshirani, 1993) to estimate the smallest number of trials that would yield specified detection accuracies. We employed the same approach in the current study.

Each data set consisted of 400 trials from which the bootstrapping samples could be extracted (100 trials for

each stimulus type). We created 1000 sets of trials at each of the four stimulus types, for each sample size of *N*, *N* ranging from 1 to 31, incremented by 3. This results in 4000 total cases for each bootstrapped data set. For each mode of presentation and each experimental session, the following steps were executed (modeled after Donchin et al., 2000): (1) obtain a random sample of *N* trials (stimulus presentations) for each of the four stimulus types (YES, NO, PASS, and END) by sampling with replacement from the set of 400 trials; (2) compute the average for each stimulus type; (3) apply SWDA weights to the appropriate features and select the stimulus with the maximum discriminant score; (4) if the selected stimulus is defined as the target, count a hit, if a non-target stimulus is selected, count a miss; (5) record the percentage of hits among the 1000 sets of samplings. The final result is the percent accuracy at each level of *N* trials. The results can be used to determine the value of *N* at which a given accuracy level is reached for each subject, each condition, and each modality of presentation.

3. Results

3.1. Waveform analysis

Fig. 1 presents an example of the waveform morphology for one subject, on a run that resulted in a correct classification. Each of the four stimuli (YES, NO, PASS, END) were presented 25 times and the resulting averaged waveforms are presented. The target item for the run was YES; the figure demonstrates a robust P300 elicited by the target item and very similar non-P300 responses for the remaining three stimuli. The ERPs elicited in the oddball experiment are displayed in Fig. 2. The subjects were presented with 200 stimuli, 50 of which were targets. The results of the oddball experiment suggest that all of the participants can potentially use a BCI based on an oddball sequence because all subject's elicited responses were
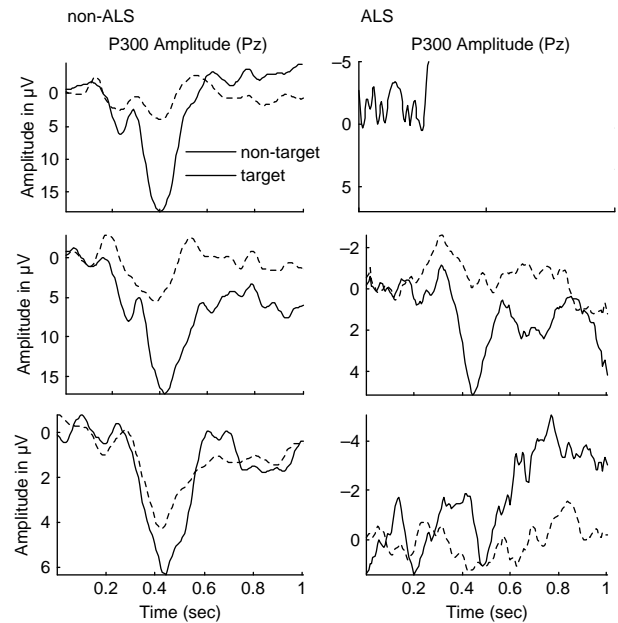
Fig. 2. Visual oddball experiment data for all six participants. The target waveforms (solid lines) represent an average of 50 stimulus presentations, and the non-target waveforms represent 150 stimulus presentations.

differential for the rare and frequent stimulus. Waveform data for the four-choice paradigm is presented in Figs. 3 and 4. Each figure shows, for each of the three modes, averaged data from Session 1 and Session 10 for each of the six participants. Fig. 3 shows data from the non-ALS participants and Fig. 4 shows data from the ALS participants. The waveforms demonstrate that variability exists across modes and sessions for both groups of participants. In general, as corroborated by the classification data discussed below, the ALS participant's responses are more variable across mode and session than the non-ALS participants.

3.2. SWDA classification performance

There were no significant differences based on the user's task, that is, whether they focused on a given word or whether they answered a question by focusing on the word that correctly answered a question posed by the experimenter. Given that user task had no significant effects, the data were collapsed across task before the present analysis was conducted. Classification accuracy was entered into a mixed design factorial ANOVA using the between groups variable Group (non-ALS vs. ALS), and the within groups variables of Session (3–10 (session 1 and 2 were not included because Method 3 required 2 previous sessions data to derive SWDA weights)), Mode of Presentation (Auditory vs. Visual vs. Auditory + Visual), Method (1 Run vs. 1 Session vs. 2 Sessions), and Number of Stimuli ((i.e. the number of presentations for each stimulus) 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31). Table 2 shows classification accuracy for each of the three modes and each of the three methods, averaged for all sessions, and all subjects.
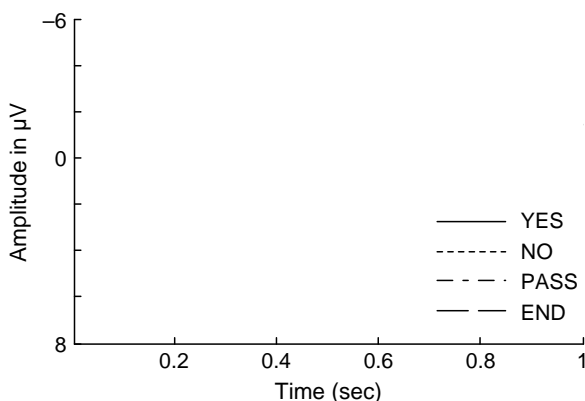

Fig. 1. Averaged waveforms for a target (YES) and three non-target stimuli (Pz electrode). Each waveform represents 25 stimulus presentations (one experimental run of 100 stimuli). The data was recorded from non-ALS participant 1, visual only mode.
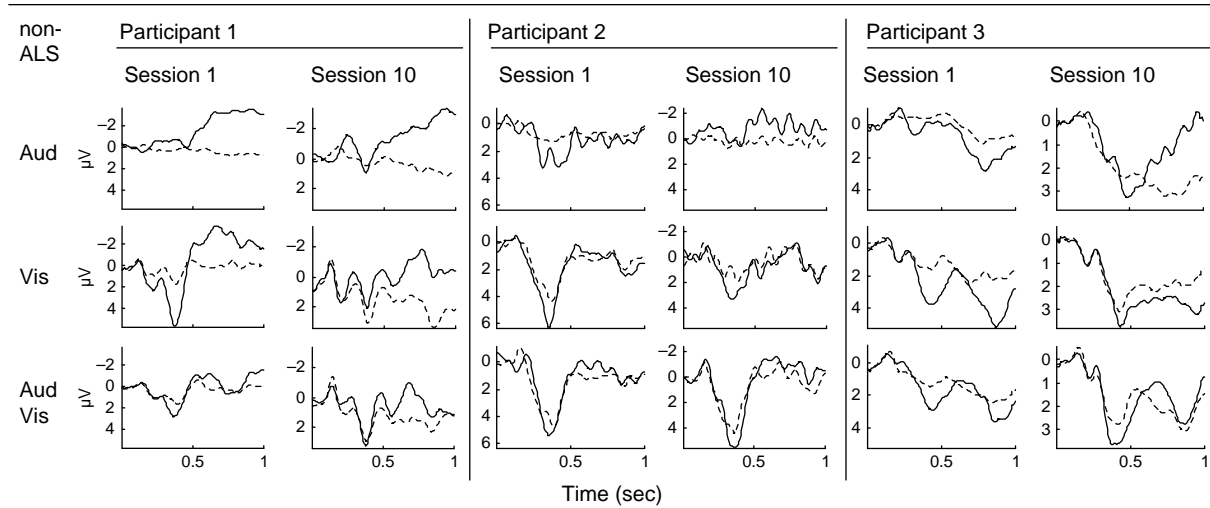
Fig. 3. Non-ALS participant's average waveforms for each presentation mode in Session 1 and Session 10. Target waveforms (solid lines) represent an average of 100 stimuli and non-target waveforms (dashed lines) represent an average of 300 stimuli (Pz electrode).
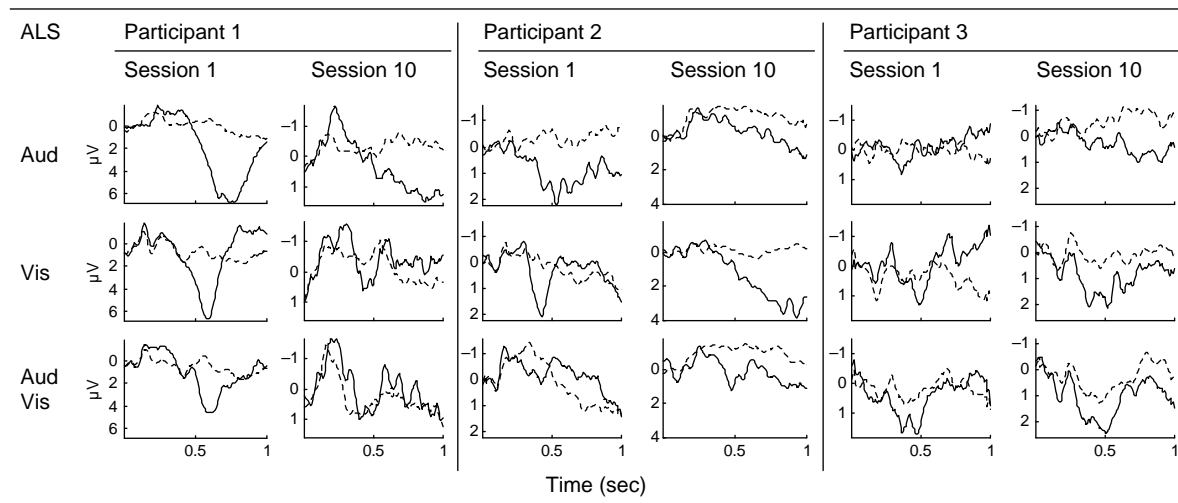

Fig. 4. ALS participant's average waveforms for each presentation mode in Session 1 and Session 10. Target waveforms (solid lines) represent an average of 100 stimuli and non-target waveforms (dashed lines) represent an average of 300 stimuli (Pz electrode).

Table 2
Mean accuracy for each subject, each mode of presentation, and each method of classification, averaged across all experimental sessions

	Non-ALS			ALS			ModexMethod
	Ss1	Ss2	Ss3	Ss1	Ss2	Ss3	Mean
Aud							
1 Run	65.1	70.2	56.8	65.4	73.2	59.1	65.0
1 Ses	62.6	60.0	53.9	57.7	52.7	39.2	54.4
2 Ses	64.5	72.7	58.4	79.6	56.5	35.5	61.2
Vis							
1 Run	80.4	76.1	49.3	73.8	62.6	53.9	66.0
1 Ses	92.6	77.1	43.3	55.5	58.5	57.8	64.1
2 Ses	97.0	82.1	58.9	63.2	67.9	47.0	69.4
Aud + Vis							
1 Run	78.1	64.6	70.4	64.9	68.9	61.6	68.1
1 Ses	86.2	80.3	70.3	56.5	56.1	43.9	65.6
2 Ses	89.6	76.7	73.2	55.4	46.5	50.8	65.4
Ss Mean	79.6	73.3	59.4	63.6	60.3	49.9	64.3

The values in Table 2 reflect the number of stimuli that produced the highest level of accuracy. In some cases, 28 stimuli produced higher accuracy than 31 stimuli.

Several effects are statistically significant. The main effect of Method yielded significant effects, $F(2,8)=9.64$, $MSe=109.5$, $P=.007$. Method 1, using the first run of a session (mean $=58.98$), and Method 3, using the aggregate of two previous sessions (mean $=57.05$), classified significantly better than Method 2, data from the previous session (mean $=53.44$, Newman-Keuls, $P<.01$, and $P<.05$, respectively). The main effect of Number of Stimuli was also significant, $F(10,40)=135.57$, $MSe=171.4$, $P=.0001$. As the number of stimuli before averaging was increased, classification accuracy also increased. Including more stimuli in the average reduces the amount of variability contributed by any given item.

Although the main effect for Group was not statistically significant, the Group x Method interaction was significant, $F(2,8)=16.95$, $MSe=145.2$, $P=.0013$. The ALS group classification accuracy was highest for Method 1 (mean $=56.83$), Method 3 classification accuracy was highest for the non-ALS group (mean $=65.91$). This result indicates more response variability across sessions in the ALS group. The Group $\times$ Number of Stimuli interaction was also significant, $F(10,40)=2.29$, $MSe=22.3$, $P=.031$. As can be seen in Fig. 5, as the number of stimuli averaged increases, the Non-ALS group accuracy increased at a faster rate than that of the ALS group, after 7 stimuli the performance curves remain relatively parallel.

Two other interactions were significant. First, the Session $\times$ Number of Stimuli interaction was significant, $F(70,280)=1.46$, $MSe=6.67$, $P=.019$. Obtained classification accuracy in Session 3 and Session 10 was highest, and classification accuracy in Sessions 4–9, although similar in terms of the performance curves, was slightly lower. In addition, as the number of samples averaged together increases the differences between each sessions classification accuracy increase. The Mode of Presentation $\times$ Number of Stimuli interaction was also significant, $F(20, 80)=1.97$, $MSe=10.1$, $P=.018$. Accuracy levels for all
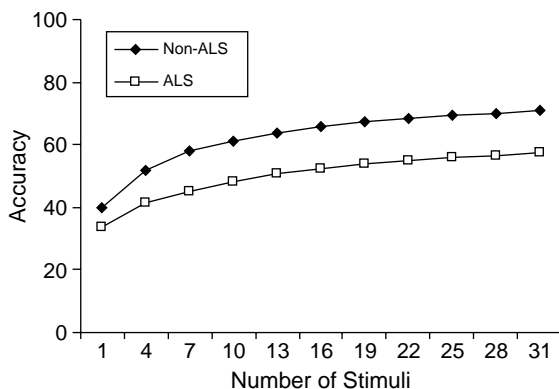
Table 3
Best classification accuracy for each participant

Group	Ss1	Ss2	Ss3
Non-ALS	97.0 (1.80)	82.1 (1.04)	73.2 (0.74)
ALS	79.6 (0.97)	73.2 (0.74)	61.6 (0.43)

Each value represents the best combination of presentation method (auditory, visual, or aud + vis) and classification method (1 Run, 1 Session, or 2 Sessions) averaged across all experimental sessions. Corresponding bit rate/selection in parentheses.

three modes begin at the same level, and as the number of stimuli averaged increases, the accuracy increases more for the visual and auditory + visual modes than it does for the auditory mode. Table 3 shows the combination of presentation method and SWDA classification that classified most accurately for each of the six subjects. The table shows that the classification rates for non-ALS subjects 2 and 3 are nearly identical to that of ALS subjects 1 and 2.

The above results provide an objective measure of how accurately the system will classify responses at each level of number of stimuli averaged. The following example describes the speed with which the P300-BCI could function using the current SWDA classification procedure. The following example is based on non-ALS subject 1 data. Assuming that a performance level of 75% was the target, this level of accuracy can be attained using four samples; thus, the system would have to present four sequences of the four stimuli (16 presentations) before averaging. Given the ISI of 1400 ms used in the current study, the result corresponds to one classification every 22.4 s. If target performance level is set at 92%, 19 samples must be obtained before averaging (76 presentations). Classification time at this classification level increases to 106 s/selection. In practice the speed/accuracy tradeoff will be dependent upon the individual user's performance preferences. Some users will choose to communicate at a higher rate, accepting more errors, while others may wish to be more accurate but proceed at a slower pace.

4. Discussion

The primary question of the current study was whether a P300-BCI could function as an alternative method of communication for an ALS population. The results indicate that ALS patients who are losing the ability to communicate by traditional means could communicate with a P300-BCI system. The current study examined variables related to practical, theoretical, and methodological issues within the framework of the current design. Several of these factors are discussed below.

4.1. Waveform differences

Although there are some similarities between the oddball waveforms presented in Fig. 2 and the four-choice

Fig. 5. Mean accuracy for the ALS and non-ALS groups as a function of the number of stimuli averaged before applying SWDA weights.

waveforms presented in Figs. 3 and 4, it is not entirely surprising that they are somewhat different for each subject and each condition. Squires et al. (1977) found large differences in P300 latency for auditory and visual stimuli, auditory P300s occurred 140 ms earlier. Squires et al. (1977) also demonstrated that the differences in latency were due to stimulus complexity and difficulty of discrimination. In the current study, stimuli were not matched for complexity and discriminability across the three modalities. All subjects reported no difficulty is discriminating the auditory or visual stimuli; however, it is likely that stimuli were recognized with different time courses in the different presentation modalities.

Studies have also indicated that some ALS patients may have abnormal evoked potentials. Paulus et al. (2002) reported that 12 of 16 ALS patients displayed abnormal P300 patterns, as compared to an age-matched control group. They suggest that ALS may cause damage beyond motor areas. In addition, Lomen-Hoerth et al. (2003) found a higher level of fronto-temporal lobar dementia in ALS patients as compared to an age-matched control group. Individual differences in P300 latency and amplitude have also been reported (Fabiani et al., 1987), and the elicited response changes with age (Fabiani and Friedman, 1995; Goodin et al., 1978; Miller et al., 1987). For example, latency is slightly longer and amplitude is reduced for older adults as compared to young adults (Donchin et al., 1986). It would appear, from the analyses by Spencer et al. (1999) that these changes in the pattern of ERPs elicited in the elderly may be due to a more frequent elicitation of a Novelty-P3 in the elderly. In addition, the current groups were not age-matched which may exacerbate differences between the groups.

4.2. Mode of presentation

It is important to determine if a BCI device can function effectively using different presentation modalities, because it is quite possible that a user may have an auditory or visual deficiency. Previous research has shown that auditory oddball tasks, and visual oddball tasks, both elicit large P300 responses (Donchin et al., 1982; Fabiani et al., 1987; Squires et al., 1977). In addition, McDonald et al. (2000) and Teder-Salejarvi et al. (2002) reported higher accuracy and larger ERP amplitude when auditory and visual stimuli were presented simultaneously, than when either modality was presented in isolation. In the current study, the spoken stimuli may be recognized with different time courses, this would make classification more difficult for the spoken stimuli. In fact, the mode of presentation by number of stimuli interaction shows that the auditory mode did not reach the same level of classification accuracy as the visual or the auditory + visual mode. However, the auditory mode results do appear to be high enough that a user with a compromised visual system might still use a BCI.

4.3. Task manipulation

Previous research has reported that the effect of concurrently performing multiple tasks reduces P300 amplitude (Gopher and Donchin, 1986; Isreal et al., 1980a,b; Kramer et al., 1983; Sirevaag et al., 1989), and that increasing memory load also reduces P300 amplitude (Kramer et al., 1986; Wintink et al., 2001). The task manipulation in the current study is relevant to this issue. Focusing on a stimulus, or focusing on a stimulus that answers a specific question can be thought of as performing two tasks and it may also increase memory load. The results indicated that the elicited response were not different in the two task conditions. This is important because users do not have to answer a series of questions when data is being collected to derive classification weights. On the other hand, if a difference was present it would be important to maintain a strict correspondence between the conditions used for system calibration and actual use.

4.4. Multiple sessions

Collecting data across ten sessions provided a large enough sample of data to examine how repeated use affects classification accuracy. The classification data showed that significant classification differences are present across sessions. However, the differences are not large enough to cause the system to be ineffective. For example, classification accuracy for all sessions starts at approximately the same level, and is relatively stable as the number of averaged stimuli is increased, Sessions 3 and 10 notwithstanding. The slight reduction in performance during the intermediate sessions may be, in part, due to a mild habituation effect. In support of this proposal Kinoshita et al. (1996) found similar effects. In an initial session P300 amplitude was at the highest level, and in subsequent session the amplitude slightly decreased, while latency remained relatively constant. When participants were asked to return for an unexpected final experimental session P300 amplitude spontaneously returned to the same level as the initial session. In practice, after session 1 a user will always be in the intermediate range of sessions because use may be expected to continue indefinitely. Ravden and Polich (1998) showed that the P300 amplitude, but not latency, decreases across a 10-block (approximately 60-min) session of trials. They interpret this variation in terms of ultradian rhythm variation, which is thought to underlie oscillations in vigilance performance. On the other hand, the P300 of individual trials has been shown to be relatively stable (Cohen and Polich, 1997; Polich, 1989). In addition, Fabiani et al. (1987) have reported the within and between session reliability, for a given subject, on a given task, to be .70 or higher (Fabiani et al., 1987). Given the current level of classification, across multiple sessions, response habituation does not seem to be a likely source of poor performance.

4.5. Classification methodology

Three methods of deriving SWDA weights were compared. The first method, using the first run of a session, was selected to simulate classification weights that could be derived online at the beginning of a session. This method examined how well the system could function using a minimal amount of calibration data, and requires less off-line data processing. Interestingly, the ALS group performed best with weights derived from method 1. This result indicates that the ALS group's responses are more variable than those of the non-ALS group's. In contrast, for the non-ALS group, classification accuracy was highest with method 3. Method 3 used data aggregated from the previous two sessions. This result suggests that, over time, the responses are stable, so that using more data to derive classification weights may be an optimal method for some users.

Classification methods in addition to SWDA may also be promising. Serby et al. (2005) has shown that Independent Components Analysis (ICA) can be used to classify P300 Speller data online with 79.5% accuracy. Ritter and colleagues have demonstrated that support vector machines produce very robust classification rates as well, approaching 100% (Kaper et al., 2004; Meinicke et al., 2002); however, the method has not been implemented online, and it is not clear how this would be achieved. Furthermore, there has not been a direct comparison of the methods to determine method superiority.

4.6. Rate of item selection

The focus of the current study was to determine whether a P300-system could be a viable option for an ALS population; not to optimize the speed of the system. Therefore, the rate of stimulus selection is not impressively rapid. One reason for this is because spoken stimuli were used and this reduced the speed with which the system could operate. Each stimulus was presented for 600 ms (the duration of the vocalized stimulus). This constraint would not be necessary with visual only presentations, or tone bursts. The ISI can also be adjusted to increase the number of stimulus presentations in a given time period; however, the current study did not examine this variable. Farwell and Donchin (1988) examined the effects of ISI and they found that a longer ISI provided a higher rate of communication for three of the four participants tested. Thus, it is possible that the current configuration did not optimize stimulus presentation rate, and, in turn, performance speed for the participants.

The information transfer rate (bits/selection) in the current study ranged from 0.43 to 1.80 (see Serby et al., 2005 for a summary of studies reporting bit rate). While this value is lower than most published studies, some notable differences are present. The item selection in the current paradigm is selecting an answer to a question. In the other

systems several selections may be necessary to answer a single question. For example, a spelling system would require three correct selections to spell the word 'YES.' Obviously, one could use the letter Y to stand for YES, but this protocol would necessarily need to be established between the user and the person whom communication is with. In addition, participants in the current study as well as their caregivers (and others not included in the current study) have indicated that 1–5 min is not an unreasonable amount of time to wait for an answer to a question.

Bit rate can also be misleading. Calculations are made after necessary time between selections is deleted (e.g. Kaper et al., 2004; Meinicke et al., 2002; Serby et al., 2005). It is also important not to overshadow the importance of accuracy as illustrated by Wolpaw et al. (2000a,b, 2002). An increase in accuracy from 75 to 90%, given four choices, nearly doubles the bit rate. Meinicke et al. (2002) report a maximum bit rate of 84.7 bits/min using a 6×6 matrix speller; unfortunately, the bit rate is accompanied with an accuracy level of less than 50%. The net result is a system that will not be able to correctly spell a single word, as it would enter an iterative loop of backspace operations.

The measure of bit rate as used, for example, by Serby et al. (2005) is, therefore, of doubtful value in the assessment of a BCI. The metaphor of a 'communication channel' may provide the comfort of worked out engineering solutions and a convenient mathematical structure. However, the metaphor does not survive the realities encountered at the patient's bed side. A key issue here is that 'communicate' while seemingly accurate is actually quite vague in the context of the ALS person. There is a huge difference between being able to respond to YES/NO questions, to spell out arbitrary text, to converse at acceptable rates with a relative, to communicate choices to a caretaker, to order services, and to control devices. The notion of bit rate, for example, seems to lose all meaning in the context of this broad range of meanings of 'communicate' as each of these present different communicative challenges.

4.7. Conclusions

BCI devices are beginning to allow people to communicate through non-muscular means (Birbaumer et al., 1999, 2000; Kubler et al., in press; Pfurtscheller et al., 1996; Wolpaw and McFarland, 2004; Wolpaw et al., 2000a,b, 2002). People suffering from neuromuscular disabilities may soon be using BCI communication on a daily basis. With any new technology, as more research is conducted, the systems will be refined and performance will increase accordingly. In addition, it will become clear as to which systems and interfaces work best in which circumstances and conditions. The current study has demonstrated two important points: (1) A P300-BCI can be effective with an ALS population; (2) auditory and/or visual stimuli can function as a P300-BCI control signal. Thus, we have

established the efficacy of pursuing P300-BCIs for the severely disabled population.

Acknowledgements

References

Allison BZ, Pineda JA. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. IEEE Trans Neural Syst Rehabil Eng 2003;11:110–3.

Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H. A spelling device for the paralyzed. Nature 1999;398:297–8.

Birbaumer N, Kubler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, Iversen I, Kotchoubey B, Neumann N, Flor H. The thought translation device (TTD) for completely paralyzed patients. IEEE Trans Rehabil Eng 2000;8:190–3.

Cohen J, Polich J. On the number of trials needed for P300. Int J Psychophysiol 1997;25:249–55.

Donchin E, Kramer AF, Wickens C. Probing the cognitive infrastructure with event-related brain potentials. In: Proceedings of the AIAA workshop on flight testing to identify pilot workload and pilot dynamics. Washington, DC: US Government Printing Office; 1982. p. 371–87.

Donchin E, Miller GA, Farwell LA. The endogenous components of the event-related potential—a diagnostic tool?. In: Swaab DF, Fliers E, Mirmiran M, Van Gool WA, Van Haaren F, editors. Progress in brain research, vol. 70. Amsterdam: Elsevier; 1986. p. 87–102.

Donchin E, Spencer KM, Wijesinghe R. The mental prosthesis: assessing the speed of a P300-based brain–computer interface. IEEE Trans Rehabil Eng 2000;8:174–9.

Duncan-Johnson C, Donchin E. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology 1977;14:456–67.

Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.

Fabiani M, Friedman D. Changes in brain activity patterns in aging: the novelty oddball. Psychophysiology 1995;32:579–94.

Fabiani M, Gratton G, Karis D, Donchin E. Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. Adv Psychophysiol 1987;2:1–78.

Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 1988;70:510–23.

Goodin D, Squires K, Henderson B, Starr A. Age related variations in evoked potentials to auditory stimuli in normal human subjects. Electroencephalogr Clin Neurophysiol 1978;44:447–58.

Gopher D, Donchin E. Workload—an examination of the concept. In: Kaufman L, Boff K, editors. Handbook of perception and human performance. New York: Wiley; 1986. p. 41.1–41.49.

Hanagasi HA, Gurvit IH, Ermutlu N. Cognitive impairment in amyotrophic lateral sclerosis: evidence from neuropsychological investigation and event-related potentials. Brain Res Cogn Brain Res 2002;14:234–44.

Isreal JB, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. Psychophysiology 1980a;17:259–73.

Isreal JB, Wickens CD, Chesney GL, Donchin E. The event-related brain potential as an index of display monitoring workload. Hum Factors 1980b;22:212–24.

Johnson R, Donchin E. On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalogr Clin Neurophysiol 1978;44:424–37.

Kaper M, Meinicke P, Grossekathoefer 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.

Kinoshita S, Inoue M, Maeda H, Nakamura J, Morita K. Long-term patterns of change in ERPs across repeated measurements. Physiol Behav 1996; 60:1087–92.

Kramer AF, Wickens CD, Donchin E. An analysis of the processing requirements of a complex perceptual-motor task. Hum Factors 1983; 25:597–621.

Kramer AF, Schneider W, Fisk A, Donchin E. The effects of practice and task structure on components of the event-related brain potential. Psychophysiology 1986;23:33–47.

Kubler A, Kotchoubey B, Kaiser J, Wolpaw JR, Birbaumer N. Brain–computer communication: unlock the locked-in. Psychol Bull 2001; 127:358–75.

Kubler A, Nijboer F, Mellinger J, Vaughan TM, Pawelzik H, Schalk G, McFarland DJ, Birbaumer N, Wolpaw JR. Patients with ALS can learn to operate a sensorimotor-rhythm based brain computer interface (BCI). Neurology 2005; 64: 1775–7.

Kunst CB. Complex genetics of amyotrophic lateral sclerosis. Am J Hum Genet 2004;75:933–47.

Lomen-Hoerth C, Murphy J, Langmore S. Are amyotrophic lateral sclerosis patients cognitively normal? Neurology 2003;60:1094–7.

McDonald JJ, Teder-Salajarvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. Nature 2000;407:906–8.

Meinicke P, Kaper M, Hoppe F, Huemann M, Ritter H. Improving transfer rates in brain computer interface: a case study. Neural Inf Proc Syst 2002;1107–14.

Miller GA, Bashore TR, Farwell LA, Donchin E. Chapter 1: research in geriatric psychophysiology. In: Schaie KW, Eisdorfer C, editors. Annual review of gerontology and geriatrics, vol. 7. New York: Springer; 1987. p. 1–27.

Paulus KS, Magnano I, Piras MR, Solinas MA, Solinas G, Sau GF, Aiello I. Visual and auditory event-related potentials in sporadic amyotrophic lateral sclerosis. Clin Neurophysiol 2002;113:853–61.

Pfurtscheller G, Flotzinger D, Pregenzer M, Wolpaw JR, McFarland D. EEG-based brain computer interface. Med Prog Technol 1996;21:111–21.

Polich J. Habituation of P300 from auditory stimuli. Psychobiology 1989; 17:19–28.

Posner MI. The orienting of attention. Q J Exp Psychol 1980;32:3–25.

Ravden D, Polich J. Habituation of P300 from visual stimuli. Int J Psychophysiol 1998;30:359–65.

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.

Sellers E, Schalk G, Donchin E. The P300 as a typing tool: tests of brain computer interface with an ALS patient. Psychophysiology 2003;40:S77.

Serby H, Yom-Tov E, Inbar GF. An improved P300-based brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 2005;13:89–98.

Sharbrough F, Chatrian GE, Lesser RP, Lüders H, Nuwer M, Picton TW. American electroencephalographic society guidelines for standard electrode position nomenclature. J Clin Neurophysiol 1991;8:200–2.

Sirevaag EJ, Kramer AF, Coles MGH, Donchin E. Resource reciprocity: an event-related brain potentials analysis. Acta Psychol 1989;70: 77–97.

Spencer KM, Dien J, Donchin E. A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology 1999;36:409–14.

Squires NK, Donchin E, Squires KC, Grossberg S. Bisensory stimulation: inferring decision-related processes from the P300 component. J Exp Psychol Hum Percept Perform 1977;3:299–315.

Teder-Salejarvi WA, McDonald JJ, Di Russo F, Hillyard SA. An analysis of audio–visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 2002;14: 106–14.

Wintink AJ, Segalowitz SJ, Cudmore LJ. Task complexity and habituation effects on frontal P300 topography. Brain Cogn 2001; 46:307–11.

Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans. Proc Natl Acad Sci USA 2004;101:17849–54.

Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM. Brain–computer interface technology: a review of the first international meeting. IEEE Trans Rehabil Eng 2000a;8:164–73.

Wolpaw JR, McFarland DJ, Vaughan TM. Brain–computer interface research at the Wadsworth Center. IEEE Trans Rehabil Eng 2000b;8: 222–5.

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clin Neurophysiol 2002;113:767–91.

Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM. Transient neural activity in human parietal cortex during spatial attention shifts. Nat Neurosci 2002;5:995–1002.