

# An auditory oddball (P300) spelling system for brain-computer interfaces

A. FURDEA,<sup>a</sup> S. HALDER,<sup>a</sup> D.J. KRUSIENSKI,<sup>b</sup> D. BROSS,<sup>a</sup> F. NIJBOER,<sup>a</sup> N. BIRBAUMER,<sup>a,c</sup>  
AND A. KÜBLER<sup>a,d</sup>

<sup>a</sup>Institute of Medical Psychology and Behavioural Neurobiology, University of Tübingen, Germany

<sup>b</sup>School of Engineering, University of North Florida, Jacksonville, Florida, USA

<sup>c</sup>Instituto di Ricovero e Cura a Carattere Scientifico, Ospedale San Camillo, Venezia, Italy

<sup>d</sup>Clinical and Health Psychology Research Centre, School of Human and Life Sciences, Roehampton University, London, UK

## Abstract

This study was designed to develop and test an auditory event-related potential (ERP) based spelling system for a brain-computer interface (BCI) and to compare user's performance between the auditory and visual modality. The spelling system, where letters in a matrix were coded with acoustically presented numbers, was tested on a group of healthy volunteers. The results were compared with a visual spelling system. Nine of the 13 participants presented with the auditory ERP spelling system scored above a predefined criterion level control for communication. Compared to the visual spelling system, users' performance was lower and the peak latencies of the auditorily evoked ERPs were delayed. It was concluded that auditorily evoked ERPs from the majority of the users could be reliably classified. High accuracies were achieved in these users, rendering item selection with a BCI based on auditory stimulation feasible for communication.

**Descriptors:** Electroencephalogram, Auditory stimulation, Event-related potential, P300, Brain-computer interface

Individuals with severe degenerative neurological diseases, such as amyotrophic lateral sclerosis (ALS), can use a brain-computer interface (BCI) for communication with a high level of accuracy (Birbaumer et al., 1999; Kübler et al., 2001; Nijboer et al., 2005; Piccione et al., 2006; Sellers & Donchin, 2006).

BCIs are devices that allow users to send messages or commands without using the brain's motor output pathways. After its first description as an input signal for BCI in 1988 by Farwell and Donchin, the P300 component of the event-related potential (ERP) has been recently explored intensively with regards to its triggering, detection, usability for communication, and habituation during long-term use (Nijboer et al., 2008; Sellers & Donchin, 2006). The P300 is a positive deflection of the electroencephalogram (EEG) with a latency of 200 to 700 ms after stimulus onset. The response is elicited when users attend to a random series of stimulus events that contains an infrequently presented set of items, that is, an oddball paradigm, and is typically recorded over central-parietal scalp locations (Fabiani, Gratton, Karis, & Donchin, 1987). Farwell and Donchin showed

that the visual P300 ERP can be successfully used to select letters displayed on a computer monitor (Donchin, Spencer, & Wijesinghe, 2000; Farwell & Donchin, 1988). One reason for developing BCIs is to provide people with severe motor disability with a new output channel independent of the motor system. One such disease that may lead to total motor paralysis is ALS.

Although sparing of neurodegeneration of nerves, which subserve eye movement, has been reported (Hayashi & Kato, 1989; Whitehouse, Wamsley, Zarbin, Price, & Kuhar, 1985), there are now numerous reports about impaired eye movement and slowing of saccades in ALS (Averbuch-Heller, Helmchen, Horn, Leigh, & Buttner-Ennever, 1998; Jacobs, Bozian, Heffner, & Barron, 1981; Leveille, Kiernan, Goodwin, & Antel, 1982; Ohki et al., 1994; Palmowski et al., 1995a; Palmowski et al., 1995b; Szmidt-Salkowska & Rowinska-Marcinska, 2005). Although not commonly seen in neurological practice (because patients usually die before entering the complete locked-in state), eye muscles may become totally paralyzed in ALS rendering the patients completely locked in (complete locked-in state = CLIS) (Cohen & Caroscio, 1983; Harvey, Torack, & Rosenbaum, 1979; Kushner et al., 1984; Palmowski et al., 1995a); our own experience confirms these reports (Hill et al., 2006; Hinterberger, Birbaumer, & Flor, 2005; Kübler & Birbaumer, 2008). For such patients, the visually based P300-BCI would no longer provide a reliable communication channel.

It is now well established that patients in the locked-in state (LIS) are able to successfully control a BCI (Kübler et al., 2001; Kübler, Nijboer, & Birbaumer, 2007; Neuper, Müller, Kübler,

---

The study was supported by the Deutsche Forschungsgemeinschaft (DFG, SFB 550/B5) and the National Institutes of Health (NIH, HD30146). The authors are very grateful to Slavica von Hartlieb for her help and support with recruitment of participants and data collection and Tamara Matuz for the valuable comments.

Address reprint requests to: Adrian Furdea, Institute for Medical Psychology and Behavioral Neurobiology, Gartenstraße 29, 72074 Tübingen, Germany. E-mail: adrian.furdea@uni-tuebingen.de

Birbaumer, & Pfurtscheller, 2003; Piccione et al., 2006), but to date no patient in CLIS has been able to reliably use such a device (Kübler & Birbaumer, 2008). The failure to achieve BCI control might be related to loss of cognitive abilities, as reported in some patients in later stages of ALS (Abrahams, Leigh, & Goldstein, 2005; Schreiber et al., 2005). Another possible explanation, based on operant learning theory, could be that the LIS leads to extinction of goal-directed behavior and goal-oriented thoughts, preventing learning and control of physiological functions (Birbaumer, 2006a; Kübler, Nijboer, & Birbaumer, 2007). To elucidate the learning impediment of BCI control in CLIS patients, a longitudinal study with patients in earlier stages of ALS has been proposed (Birbaumer, 2006a, 2006b). The objective of this study is to determine if the skill to control a BCI transfers to the CLIS provided successful BCI control has been achieved before entering the CLIS.

According to clinical definition, no ocular movement is possible in the CLIS. Therefore, the development of non-visual BCIs is necessary for CLIS patients. To date, few studies with non-visually based P300-BCI exist. Hill and colleagues used support vector machines to classify ERPs that occurred in response to auditory stimuli (Hill, Lal, Bierig, Birbaumer, & Schölkopf, 2005). The auditory stimuli consisted of 50 ms square-wave beeps having different frequencies, grouped in two different streams. One stream was presented from a speaker to the left-hand side of the participant and the other from a speaker to the right-hand side. Following an oddball paradigm, each stream contained rare target and frequent non-target beeps, with rare targets presented independently to each ear. In each trial, the participants' task was to focus attention on either the target stimuli presented to the left or right ear. Despite high variation in performance between participants, results suggested that it was possible for users to direct attention to one of the stimulus streams and that the BCI could detect which of the two targets the user was attending.

A study including healthy volunteers and patients with ALS was presented by Sellers and Donchin (2006). The authors used a four-choice P300-BCI, which presented users with either visual or auditory stimuli, or both. The stimuli were the words 'yes,' 'no,' 'pass,' and 'end' presented in a random sequence. The participants' task was to count the number of times the target, either 'yes' or 'no,' was presented. The authors showed that a target probability of 25% was low enough to reliably detect a P300 and that the response remained stable over a period of ten sessions in healthy volunteers as well as in ALS patients; although accuracy was lower in ALS patients.

To date, a spelling system (herein referred to as a "speller"), which uses ERPs elicited by means of auditory stimulation as an input signal, has not been tested. The objective of the current explorative study is to determine whether BCI users would be able to select characters from an auditorily presented letter matrix. On the behavioral level, the communication speed, that is, written symbols per minute, at a given number of stimulus presentations is determined. Secondly, the level of accuracy in spelling that could be achieved with an auditory ERP-BCI is evaluated. On a physiological level, the difference between the morphology of the ERPs elicited by auditory vs. visual stimulation is assessed. Finally, the well established Stepwise Linear Discriminant Analysis (SWLDA) classification method for visually evoked potentials (Krusienski et al., 2006) was applied to determine whether it would deliver comparable results if applied to auditorily evoked ERPs. In the present study, an auditory and a visual spelling system were both tested using healthy volunteers.

## Methods

### Participants

Fifteen healthy participants (ten women and five men, mean age 26.4 years, SD 6.3, range 17–42) were presented with the auditory and the visual ERP spelling tasks (described below). The study was approved by the Ethical Review Board of the Medical Faculty, University of Tübingen. Each participant was informed about the purpose of the study and signed informed consent prior to participation.

### Procedure

The *visual support matrix* (used in the auditory ERP speller and described in the section Auditory ERP Speller) and the visual spelling matrix (used in the visual ERP speller and described in the Visual ERP Speller section) were displayed to the participants on a 19-inch monitor. Auditory stimuli were presented using two separate speakers approximately 1 meter apart, positioned in front and directed toward the participants. Participants were comfortably seated approximately 1.2 meters from speakers and monitor. All participants took part in one experimental session, which consisted of three experimental runs. Data from the first run were used as a learning set to determine the coefficients of the discriminant function (see Data Analysis). Within the first run, the participants did not receive any feedback. The discriminant function derived from the first run was then used for online classification for the next run. After the second run, another offline analysis followed, which produced a new discriminant function trained on the data from the first two runs in order to obtain a better suited discriminant function for the third run. Data from runs two and three were collected in the copy spelling mode of the spelling system (Krusienski et al., 2006; Kübler et al., 2001). In this mode the software prompts participants to spell a prescribed word, character by character (Kübler, Kotchoubey, Kaiser, Wolpaw, & Birbaumer, 2001), and feedback of the selected letter is provided after each selection.

In the current experiment, each run of the task consisted of spelling an entire word prescribed by the investigator. For one trial the user had to select a single letter from a  $5 \times 5$  matrix by attending to target stimuli from a random series of either visual or auditory stimuli, depending on the paradigm. The stimuli (row and column intensifications or presentations of auditory stimuli) were presented in random order. For each trial, 15 sequences of stimuli were presented, where each sequence contained ten stimuli (one for each row and one for each column). According to this setup, a trial (needed for one character selection) consisted of 150 stimulus presentations, that is, 30 target and 120 non-target stimuli. A short break followed each run; its duration was determined by the participant, but was typically around 3 min.

Rather than the standard  $6 \times 6$  matrix, a  $5 \times 5$  matrix was used in both visual and auditory ERP spellers to reduce the trial times as a result of the longer stimulus presentation times required for the auditory modality. It has been demonstrated by Sellers and colleagues that matrix size manipulation does not compromise classification accuracy (Sellers, Krusienski, McFarland, Vaughan, & Wolpaw, 2006). In the following sections we will explain the visual and auditory spelling systems in detail.

### Visual ERP Speller

The visual ERP speller, as described initially by Farwell and Donchin and later by Sellers and Donchin, presents to the user a  $6 \times 6$  matrix of characters (Farwell & Donchin, 1988; Sellers &

Donchin, 2006). Each row and column is intensified in random order. Throughout one trial, the participant has to focus attention on one of the 36 characters of the matrix. The random sequence of six row and six column flashes constitutes an oddball paradigm, with the row and the column containing the desired character constituting the rare set. For a thorough description of the method, see (Sellers & Donchin 2006).

As aforementioned, the present study uses a  $5 \times 5$  character matrix to reduce the trial times. In each run the user's task was to spell the English word "brainpower" by passively counting how often the target character was intensified. The stimulus presentation time, that is, rows or column flashing, was 62.5 ms and the inter-stimulus interval (ISI) 125 ms. Additional 8.75 s were given to each participant to view the feedback and to identify the location of the next character in the spelling matrix. The duration of a sequence was 1.87 s, each character selection totaled 36.87 s.

### Auditory ERP Speller

When developing this new stimulus presentation paradigm, the objective was to keep the arrangement of the characters in a matrix structure, as described by (Farwell & Donchin, 1988) and (Nijboer et al., 2008). Because the stimuli are auditory rather than visual, the flashes are replaced with presentation of auditory stimuli that are coded to particular rows and columns of the matrix. These auditory stimuli consisted of spoken numbers (male voice), which were assigned to each row and column; see Figure 1.

Thus, each character's position in the matrix was coded by two auditorily presented number words: one corresponding to the row and one corresponding to the column. To select a particular target character, the participant had to attend to the two target stimuli representing the coordinates of the character in the matrix. In addition to auditory presentation of numbers, the matrix, referred to as *visual support matrix* (Figure 1), was displayed on a monitor. This visual presentation was solely intended to help the participants to remember the coordinates for the target letter, and no visual stimulation occurred corresponding to the task, that is, rows and columns did not flash. Users were required to focus their attention on the numbers coding the target



**Figure 1.** The  $5 \times 5$  visual support matrix for the auditory ERP speller. The numbers surrounding the matrix aimed at facilitating finding the target coordinates. Each number corresponded to an auditory stimulus. For instance, to select the letter "B" the user was required to focus attention on auditory stimulus "one" during the first interval of the trial and on auditory stimulus "seven" during the second interval of the trial.

character by counting how often the numbers were presented. This design constitutes an oddball paradigm, because in one trial of random stimuli presentations, only two stimuli refer to the target letter. The two rare events in the context of the other "irrelevant" stimuli were expected to elicit a P300-like event-related response.

In contrast to the visual speller, to simplify the complexity of the task, the row and column stimuli presentation occurred in a sequential manner, that is, first only the row stimuli were presented to determine the row, followed by only the column stimuli to determine the column. Because auditory stimuli presentation requires a larger amount of time than visual stimuli presentation, the stimulus presentation time and ISI were increased (as compared to the visual ERP speller) to 450 ms and 175 ms, respectively. Consequently, one trial consisted of first, presenting the numbers from one to five for row selection, and then, of presenting the numbers six to ten for column selection. As with the visual speller, in each run the users' task was to spell the English word "brainpower" by passively counting how often the target auditory stimulus was presented. To ensure that the participant had enough time to locate in the spelling matrix the coordinates of the next character to select, an interval of 3.75 s was provided between two consecutive trials. This resulted in a sequence length of 6.25 s and character selection time of 97.5 s.

### Data Acquisition

Stimulus presentation and data collection were controlled by the BCI2000 software (Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004) (<http://www.bci2000.org/>). The EEG was recorded using a tin electrode cap (Electro-Cap International, Inc., Eaton, OH) with 16 channels (F3, Fz, F4, T7, T8, C3, Cz, C4, Cp3, Cp4, P3, Pz, P4, PO7, PO8, and Oz) based on the modified 10–20 system of the American Electroencephalographic Society (Sharbrough et al., 1991). Each channel was referenced to the right and grounded to the left mastoid. The EEG was amplified using a g-tec 16-channel amplifier, sampled at 256 Hz, band-pass filtered between 0.01–30 Hz. Fifty Hertz noise was filtered using the notch filter implemented in the BCI2000 software. Data processing, storage, and on-line display of the participants' EEG were conducted using an IBM ThinkPad laptop.

### Data Analysis

We used the stepwise linear discriminant analysis method (SWLDA) for classification and weight generation (Donchin, Spencer, & Wijesinghe, 2000; Farwell & Donchin, 1988). The method, an extension of the Fisher's Linear Discriminant (FLD), is well established as a successful classification method for EEG data in general, and more recently for BCI data, for which rapid classification is essential. Previous studies of classification methods have demonstrated that SWLDA provides good overall performance in classifying the visually evoked P300 (Donchin, Spencer, & Wijesinghe, 2000; Fabiani, Gratton, Karis, & Donchin, 1987; Farwell & Donchin, 1988; Sellers & Donchin, 2006).

The algorithm seeks an optimal discriminant function by adding spatiotemporal features (amplitude values from particular channel locations and time samples) to a linear equation. A combination of forward and backward stepwise analysis was used as described by Krusienski et al., 2006. The input features were weighted using least-squares regression to predict the stimulus type (target or standard). Initially the discriminant function does not contain any features. In each following step, the algo-

rithm evaluates every input feature, determines the statistically most significant single feature having a  $p$ -value  $< .1$ , and adds it to the discriminant function. Each new entry is then followed by a backward stepwise analysis to remove any features that no longer meet the predetermined criterion to remain in the discriminant function (a  $p$ -value  $> .15$  was used in this study). This procedure eliminates features that no longer account for a significant amount of unique variance after additional features have entered the model. The process is repeated until no features satisfy the criteria for entering in or being removed from the model, or both, or until the discriminant function contains a predefined number of features, 60 in the present study.

For each of the recorded channels, 1200 ms data segments were extracted after each stimulus presentation. The data segments were then averaged according to the corresponding row/column stimulus for each trial. A moving average filter was then applied to the segments, which were further decimated by a factor of 20. The data segments were concatenated by channel to create the feature vectors used to train the classifier. The classification coefficients, derived using the stepwise discriminant function, were then applied to the averaged values of the data at critical time points as selected by the algorithm.

A possible method for evaluating communication with a BCI is the amount of information that is conveyed per time unit, also known as data transfer rate or bit rate. This measure, as initially derived from Shannon and Weaver (1963), incorporates speed and accuracy in a single value; in the ERP-BCI the accuracy depends on the number of sequences. However, in the context of a P300 speller, the bit rate alone does not provide a realistic measure of the communication speed. Additionally, correction of erroneously selected letters has to be taken into account. Assuming that the user attempts to correct all errors, the written symbol rate ( $WSR$ ) can be determined by first computing the bits ( $B$ ) per trial and then the symbol rate ( $SR$ ; see below) (McFarland & Wolpaw, 2003).

The formula described in Pierce (1980) was used to compute the number of bits transmitted per trial:

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 \left( \frac{1 - P}{N - 1} \right) \quad (1)$$

where  $N$  is the number of possible targets and  $P$  is the probability that the target is accurately classified. Then from equation (1) the symbol rate is determined as,

$$SR = \frac{B}{\log_2 N} \quad (2)$$

If  $T$  is the trial duration in minutes, the  $WSR$  can be determined as follows

$$WSR = \begin{cases} \frac{2SR - 1}{T} & SR > 0.5 \\ 0 & SR \leq 0.5 \end{cases} \quad (3)$$

By only accounting for the written symbols (i.e., excluding backspaces) that were correctly selected per time interval (i.e., per minute), the  $WSR$  provides a more realistic measure of the actual speed of written communication. Assuming that the user attempts to correct all errors, a  $SR$  of less than or equal to 0.5 would indicate that the user will make, on average, more errors than she/he is capable of correcting, and the final message will contain more errors than correctly selected letters and likely be indecipherable. Therefore, the range of  $SR$  from 0 to 0.5 is assigned a  $WSR$  of zero.

## Results

Two participants were excluded from the study because the algorithm was unable to detect a reliable difference in their EEG between the relevant visual/auditory stimuli (oddball) and the task-irrelevant (standard) visual/auditory stimuli. Thus, 13 volunteers are included in the analysis.

### Classification Performance, Bit Rate, and $WSR$

We defined classification accuracy as the percentage of characters (per runs two and three) correctly classified by the SWLDA. Mean classification accuracy, bit rate, peak amplitude, and latency for both spellers are presented in Table 1. When the interval between characters was included, the average bit rate was 1.54 bits/min for the auditory and to 6.8 bits/min for the visual ERP speller. The values were computed based on the actual performance achieved at the end of the two copy spelling runs.

Table 2 presents mean classification accuracy across participants, bit rate (bits/minute), and  $WSR$  achieved at the end of the session. With 15 sequences of stimulus presentations a nonzero  $WSR$  was achieved by 9 of 13 users presented with the auditory ERP speller and by all users of the visual ERP speller. A nonzero  $WSR$  indicates that the paradigm has the potential for practical written communication.

Similarly, the assessment of the minimum number of sequences that have to be averaged to achieve a nonzero  $WSR$  (approximately 70% accuracy for this task) has important practical implications for reliable communication. Therefore, offline analysis was performed using the data collected in the second and third run (the first run was disregarded because the participants did not receive any feedback about their performance).  $WSR$  was computed for each number of sequences (ranging from 1 to 15). The number of sequences that yielded the highest  $WSR$  was considered as optimal. With the optimal number of sequences, 12 of 13 users of the auditory speller would achieve a nonzero  $WSR$  (see Table 2).

Figure 2 depicts accuracy as a function of the number of sequences for each user in the auditory and visual spelling mode. With the exception of user 11, the highest possible accuracy is achieved with fewer sequences in the visual modality as compared to the auditory.

### ERP Analysis

The ERPs averaged across three runs within one experimental session elicited with the auditory ERP speller ( $N = 13$ ) at the electrode position with the highest  $r^2$  coefficient, are depicted in Figure 3a; likewise for the visual ERP speller ( $N = 13$ ) in Figure 3b.

In the auditory speller (Figure 3a), a typical P300 waveform can be seen in participants 1, 2, 3, 4, 9, 10, 11, 13 with the peak

**Table 1.** Group Mean Values of Classification Accuracy, Bit Rate, Amplitude of the ERP (in Microvolts), and Peak Latency (in Milliseconds)

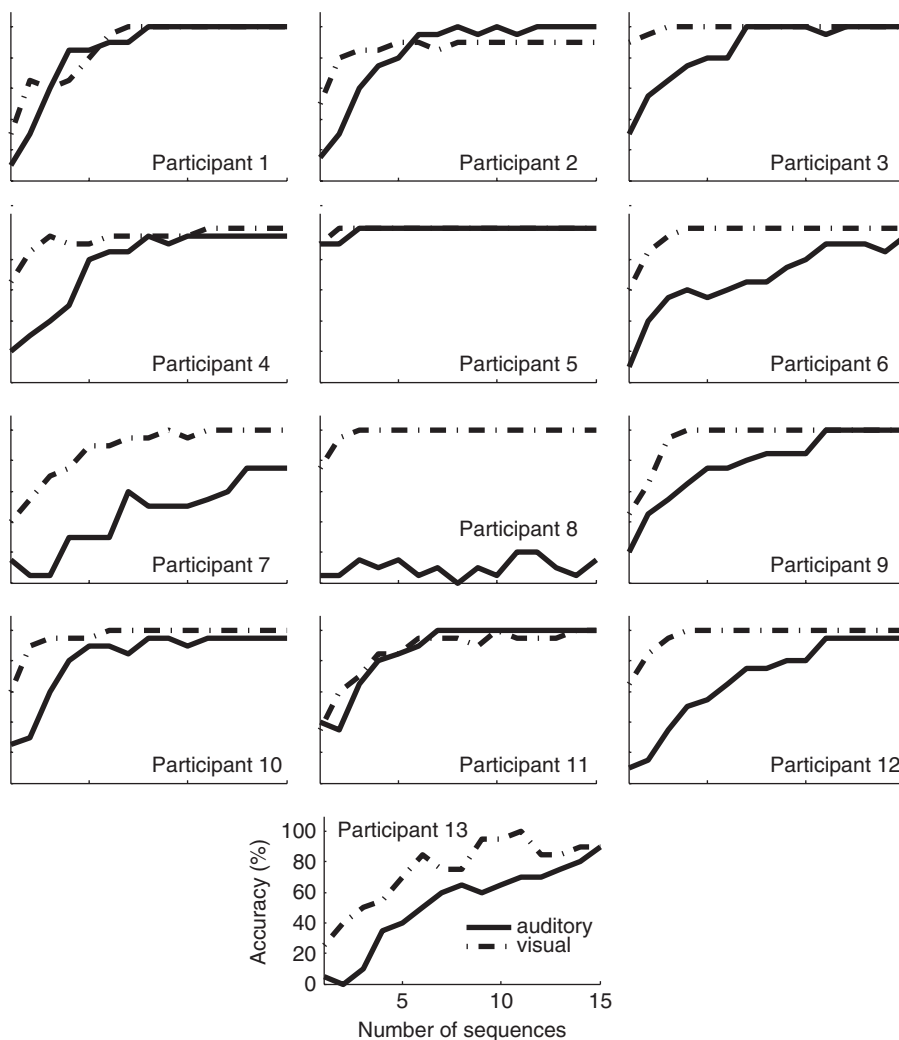
ERP speller	Classification accuracy (%)	Bit rate (bits/min)	Peak amplitude ( $\mu V$ )	Peak latency (ms)
Auditory	65.00	1.54	4.65	575.62
Visual	94.62	6.80	4.34	349.96

*Note:* Data were averaged across the 2 runs in which participants received feedback of results (copy spelling) separated by speller type.

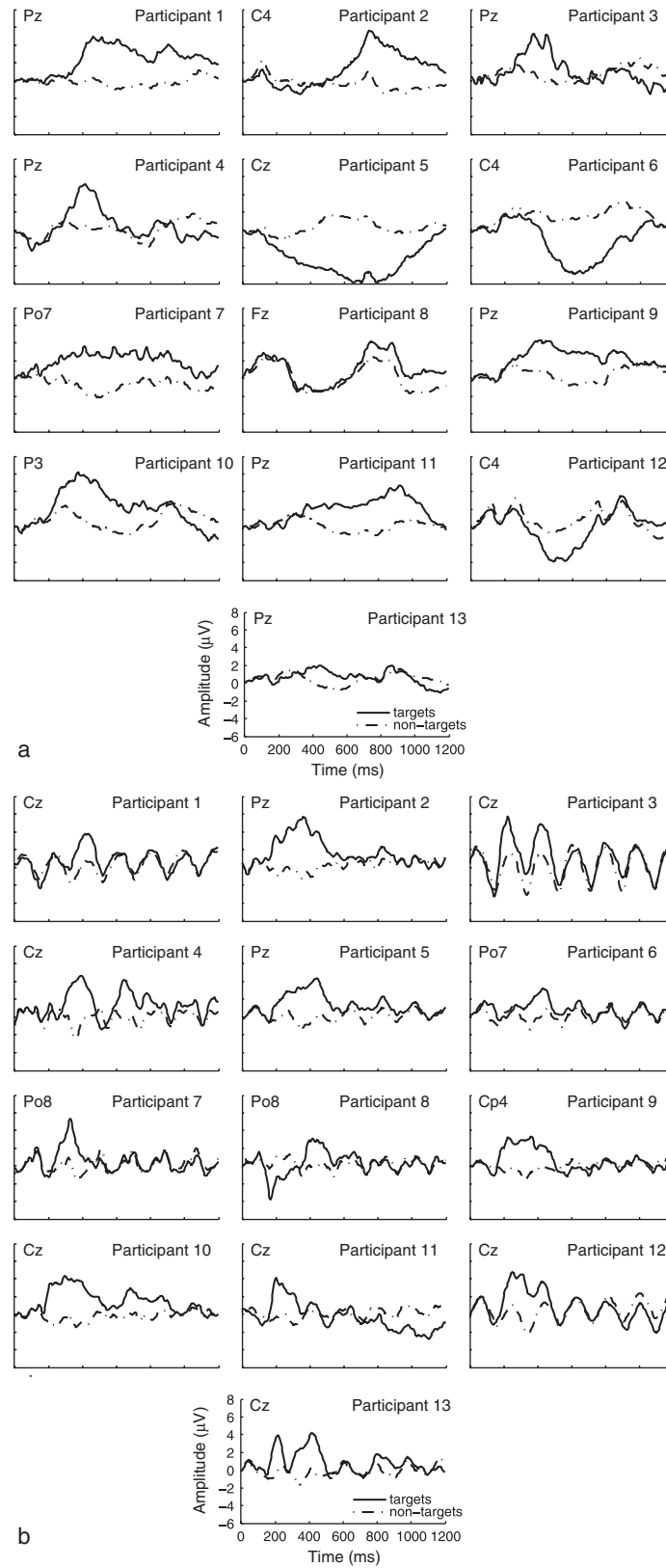
**Table 2.** Group Values of Bit Rate (measured in bits/min) and Written Symbol Rate Derived from Runs Two and Three

Participant	Classification accuracy	Auditory speller					Classification accuracy	Visual speller				
		15 sequences		Optimal nr of sq				15 sequences		Optimal nr of sq		
		Bits/min	WSR	sq. #	Bits/min	WSR		Bits/min	WSR	sq. #	Bits/min	WSR
1	100%	2.85	0.61	8	5.18	1.11	100%	7.55	1.67	7	12.74	2.74
2	70%	1.47	0.01	6	6	1.13	75%	4.37	0.25	5	12.29	1.98
3	80%	1.84	0.18	7	5.86	1.26	100%	7.55	1.67	3	19.37	4.17
4	75%	1.65	0.10	8	4.6	0.86	95%	6.71	1.26	3	17.22	3.24
5	100%	2.85	0.61	3	12.37	2.66	100%	7.55	1.67	2	22.28	4.79
6	60%	1.13	0.00	11	3.07	0.49	95%	6.71	1.26	4	17.13	3.68
7	15%	0.08	0.00	15	1.65	0.09	95%	6.71	1.26	9	10.86	2.34
8	0%	0	0.00	15	0.08	0.00	100%	7.55	1.67	3	19.37	4.17
9	80%	1.84	0.18	11	3.84	0.82	100%	7.55	1.67	4	17.13	3.68
10	80%	1.84	0.18	5	6.36	1.02	100%	7.55	1.67	3	17.22	3.24
11	75%	1.65	0.10	7	5.86	1.26	95%	6.71	1.26	6	12.38	2.33
12	35%	0.45	0.00	11	3.41	0.64	100%	7.55	1.67	4	17.13	3.68
13	75%	1.65	0.10	15	2.28	0.36	75%	4.37	0.25	11	9.48	2.04

Note: For each user, the minimum number of sequences that have to be averaged to achieve a nonzero WSR (approximately 70% accuracy for this task) was computed along with the corresponding bit rate. (nr = number; sq = sequences; WSR = written symbol rate).



**Figure 2.** Predicted spelling accuracy as a function of number of sequences for each user. The solid line represents predicted accuracy for the auditory and the dashed line for the visual ERP speller.



**Figure 3.** Averaged waveforms of 13 participants presented with the auditory (a) and the visual (b) ERP speller. The target waveforms (solid line) were obtained by averaging 900 stimulus presentations; non-target waveforms (dashed line) by averaging 4500 stimulus presentations during one session. The electrode location is provided in the upper left-hand corner of each plot and corresponds to the location of the ERP with the highest  $r^2$  value.

**Table 3.** Individual Peak Amplitude (in microvolts), Peak Latencies (in milliseconds)

Participant	Auditory speller			Visual speller		
	Peak amplitude ( $\mu$ V)	Peak latency (ms)	Localization	Peak amplitude ( $\mu$ V)	Peak latency (ms)	Localization
1	4.93	464.75	Pz	3.77	433.50	Cz
2	5.60	746.00	C4	5.66	363.19	Pz
3	5.26	371.00	Pz	5.75	222.56	Cz
4	5.15	413.97	Pz	4.65	394.44	Cz
5	5.88	652.25	Cz	4.35	437.41	Pz
6	4.89	613.19	C4	3.20	429.59	PO7
7	3.62	410.06	PO7	5.30	328.03	PO8
8	4.19	753.81	Fz	3.07	417.88	PO8
9	4.33	410.06	Pz	3.33	363.19	Cp4
10	6.14	371.21	P3	4.33	296.78	Cz
11	4.72	925.69	Pz	4.08	199.13	Cz
12	3.86	488.19	C4	4.75	249.91	Cz
13	2.01	863.19	Pz	4.18	413.97	Cz

Note: Data presented across three runs with the auditory and visual ERP speller, electrode positions of the ERP with the highest  $r^2$  value.

latencies given in Table 3 (average latency is presented in Table 1). User 5 had a negative ERP to targets with a latency of 652 ms. Accuracy above 70% could be achieved in 9 out of 13 participants. Note that not all participants who had a robust response to targets (i.e., a clearly visible difference between targets and standards) were able to achieve accuracy above 70%. For example, in users 6 and 12, who both exhibited a negative ERP to targets at 613 ms and 488 ms, respectively, accuracy was only 60% and 35%, respectively, even though a robust ERP was elicited.

In the visual speller (Figure 3b), all participants exhibit typical P300 waveforms with an average latency of 349 ms (individual data is presented in Table 3). The frequency of the flashings can be observed in the waveform averages across target and non-target stimuli in all users.

## Discussion

The present study investigated whether it was possible for healthy participants to use an auditory ERP-BCI for letter selection, and the results were compared to those of a visual spelling system. Nine users were able to focus their attention such that the ERPs occurring in response to the auditorily presented stimuli could be reliably detected and classified, thus rendering item selection from a  $5 \times 5$  matrix with an auditory BCI possible. Offline analysis also revealed that, using an optimal number of sequences, 12 out of 13 participants would be able to achieve a nonzero *WSR*.

Users of the auditory ERP speller achieved an average accuracy of 65% and an average bit rate of 1.54 bits/min. Users presented with the visual speller were able to select characters with an average accuracy of 94.62% and an average bit rate of 6.8 bits/min.

Mean latencies in the auditory ERP speller were higher than mean latencies in the visual ERP speller, these findings being in line with other reports in the literature (Kotchoubey & Lang, 2003). Auditory latencies of late components are always delayed compared to visual because of longer synaptic delays on the cortex and more relays (Robles & Delano, 2007). Squires and colleagues found large differences in P300 latency for auditory and visual stimuli, the auditory ERPs occurred earlier compared to the visual ERPs (Squires, Donchin, Squires, & Grossberg,

1977). The authors have also shown that the latency of the auditory P300s varied with the discriminability of the relevant stimuli, that is, the greater the latency of the evoked response, the more difficult it is to identify the stimuli. In the present study, the stimuli were not matched for complexity or discriminability, and the participants reported no difficulty in discriminating them. However, it is very likely that the stimuli were recognized with different time courses.

It was demonstrated that increasing memory load and concurrent performance of multiple tasks reduced the amplitude of the P300 (Gopher & Donchin, 1986; Kramer, Schneider, Fisk, & Donchin, 1986; Wintink, Segalowitz, & Cudmore, 2001). However, in the current study the amplitudes of the elicited responses were not different in the two conditions, indicating that remembering the auditory codes for the letters in the matrix did not impose any additional working memory load. However, it has to be taken into account that the visual support matrix was provided to the participants throughout auditory stimulation. Similar findings were reported by Sellers and colleagues who tested three healthy participants and three ALS patients with a four-choice speller (Sellers & Donchin, 2006). Four stimuli (Yes, No, Stop, and Pass) were presented auditorily, visually, or in both modalities with a probability of 0.25. The participants were asked first to attend to one stimulus, specified by the investigator, and disregard the other three. A second task was to focus on the stimulus that answered correctly a question provided by the investigator. The results indicated that the amplitudes of the elicited responses were not different in the two task conditions. However, as the answer to the investigator's question could be determined immediately after presentation, no additional work load was imposed during presentation of stimuli. In line with our results, the authors concluded that the results achieved with auditory stimulus presentation appeared to be good enough, such that a user with a compromised visual system might still be able to use a BCI (Sellers & Donchin, 2006).

A more recent study of Nijboer and colleagues on regulation of sensorimotor rhythms (SMR) by means of auditory feedback showed that the number of training sessions had a positive influence on the level of accuracy (Nijboer et al., 2008). Auditory and visual SMR feedback was compared in a group of sixteen healthy participants. Although at the beginning performance in the visual feedback group was better than in the auditory group, after three training sessions performance was the same in both

groups. Although the ERP-BCI does not require learning to regulate a brain response, we may speculate that an increase of performance after more training sessions would also occur within the auditory ERP speller, because with further automatization participants may get used to the setup and to rely on the auditory rather than on the visual system.

We were also interested in whether the SWLDA, an ERP classification method well established for the visual P300-BCI, would yield similar results for classification of auditorily elicited ERPs. We were able to demonstrate that the ERPs found in our participants could be reliably detected and classified with the SWLDA.

For practical considerations, the assessment of the smallest number of sequences that have to be averaged to achieve a given level of accuracy was considered. We showed that, based on offline analysis, the number of sequences could be reduced to an optimal level such that the written symbol rate would increase (or remain constant), corresponding to an accuracy above 70% for this task. The results show that on average, users of the visual ERP speller would need five sequences to achieve the above-

mentioned criteria, whereas the users of the auditory ERP speller needed nine. Whether these results can also be achieved online awaits investigation.

To summarize, users presented with the visual ERP speller were able to achieve a better average performance than when presented with an auditory ERP speller. However, albeit slower, letter selection was also possible with the auditory ERP speller, and some users were able to achieve 100% accuracy. It is believed that results of this study are encouraging enough to further develop the auditory BCI. Future work should investigate whether different auditory stimuli or different modalities of stimulus presentation would help participants to achieve a higher accuracy, faster item selection, and a better *W/SR*. Further, it is necessary to investigate whether disabled participants with impaired vision can achieve satisfactory (for themselves) results with the auditory ERP speller. One might speculate that, due to their condition of restricted or lost eye movement, ALS patients may develop a better sense of hearing, as seen in blind people (Stevens, Snodgrass, Schwartz, & Weaver, 2007), which may enable them to focus attention more easily on auditory stimulation.

## REFERENCES

- Abrahams, S., Leigh, P. N., & Goldstein, L. H. (2005). Cognitive change in ALS: A prospective study. *Neurology*, *64*, 1222–1226.
- Averbuch-Heller, L., Helmchen, C., Horn, A. K., Leigh, R. J., & Butner-Ennerv, J. A. (1998). Slow vertical saccades in motor neuron disease: Correlation of structure and function. *Annals of Neurology*, *44*, 641–648.
- Birbaumer, N. (2006a). Brain-computer-interface research: Coming of age. *Clinical Neurophysiology*, *117*, 479–483.
- Birbaumer, N. (2006b). Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, *43*, 517–532.
- Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., et al. (1999). A spelling device for the paralysed. *Nature*, *398*, 297–298.
- Cohen, B., & Carosio, J. (1983). Eye movements in amyotrophic lateral sclerosis. *Journal of Neural Transmission. Supplementum*, *19*, 305–315.
- Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *8*, 174–179.
- Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in Psychophysiology*, *2*, 1–78.
- 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*, 512–523.
- Gopher, D., & Donchin, E. (1986). Workload—An examination of the concept. In L. Kaufman & K. Boff (Eds.), *Handbook of Perception and Human Performance* (pp. 41.1–41.49). New York: Wiley.
- Harvey, D. G., Torack, R. M., & Rosenbaum, H. E. (1979). Amyotrophic lateral sclerosis with ophthalmoplegia. A clinicopathologic study. *Archives of Neurology*, *36*, 615–617.
- Hayashi, H., & Kato, S. (1989). Total manifestations of amyotrophic lateral sclerosis. ALS in the totally locked-in state. *Journal of the Neurological Sciences*, *93*, 19–35.
- Hill, N. J., Lal, T. N., Bierig, K., Birbaumer, N., & Schölkopf, B. (2005). An auditory paradigm for brain-computer interfaces. *Advances in Neural Information Processing Systems* (pp. 569–576). Cambridge, MA: MIT Press.
- Hill, N. J., Lal, T. N., Schroder, M., Hinterberger, T., Wilhelm, B., Nijboer, F., et al. (2006). Classifying EEG and ECoG signals without subject training for fast BCI implementation: Comparison of non-paralyzed and completely paralyzed subjects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*, 183–186.
- Hinterberger, T., Birbaumer, N., & Flor, H. (2005). Assessment of cognitive function and communication ability in a completely locked-in patient. *Neurology*, *64*, 1307–1308.
- Jacobs, L., Bozian, D., Heffner, R. R. Jr., & Barron, S. A. (1981). An eye movement disorder in amyotrophic lateral sclerosis. *Neurology*, *31*, 1282–1287.
- Kotchoubey, B., & Lang, S. (2003). Parallel processing of physical and lexical auditory information in humans. *Neuroscience Research*, *45*, 369–374.
- Kramer, A., Schneider, W., Fisk, A., & Donchin, E. (1986). The effects of practice and task structure on components of the event-related brain potential. *Psychophysiology*, *23*, 33–47.
- Krusienski, D. J., Sellers, E. W., Kabestaing, F., Bayouth, S., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A comparison of classification techniques for the P300 Speller. *Journal of Neural Engineering*, *3*, 299–305.
- Kübler, A., & Birbaumer, N. (2008). Brain-computer interfaces for communication in paralysis: Extinction of goal directed thinking in completely paralyzed patients? *Clinical Neurophysiology*, *119*, 2658–2666.
- Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain-computer communication: Unlocking the locked in. *Psychological Bulletin*, *127*, 358–375.
- Kübler, A., Neumann, N., Kaiser, J., Kotchoubey, B., Hinterberger, T., & Birbaumer, N. P. (2001). Brain-computer communication: Self-regulation of slow cortical potentials for verbal communication. *Archives of Physical Medicine and Rehabilitation*, *82*, 1533–1539.
- Kübler, A., Nijboer, F., & Birbaumer, N. (2007). Brain-computer interfaces for communication and motor control—perspectives on clinical application. In G. Dornhege, del R. Millan J., T. Hinterberger, D. J. McFarland, & K. R. Müller (Eds.), *Towards Brain-Computer Interfacing (Vol. 1)*, pp. 373–391. Cambridge, MA: MIT Press.
- Kushner, M. J., Parrish, M., Burke, A., Behrens, M., Hays, A. P., Frame, B., & Rowland, L. P. (1984). Nystagmus in motor neuron disease: Clinicopathological study of two cases. *Annals of Neurology*, *16*, 71–77.
- Leveille, A., Kiernan, J., Goodwin, J. A., & Antel, J. (1982). Eye movements in amyotrophic lateral sclerosis. *Archives of Neurology*, *39*, 684–686.
- McFarland, D. J., & Wolpaw, J. R. (2003). EEG-based communication and control: Speed-accuracy relationships. *Applied Psychophysiology and Biofeedback*, *28*, 217–231.
- Neuper, C., Müller, G. R., Kübler, A., Birbaumer, N., & Pfurtscheller, G. (2003). Clinical application of an EEG-based brain-computer interface: A case study in a patient with severe motor impairment. *Clinical Neurophysiology*, *114*, 399–409.



- Nijboer, F., Furdea, A., Gunst, I., Mellinger, J., McFarland, D. J., Birbaumer, N., & Kübler, A. (2008). An auditory brain-computer interface (BCI). *Journal of Neuroscience Methods*, *167*, 43–50.
- Nijboer, F., Mellinger, J., Matuz, T., Mochty, U., Eitel-Braitsch, Y., Sellers, E., et al. (2005). Brain-computer interface (BCI) use with mu rhythm, slow cortical potentials and P300s by patients with amyotrophic lateral sclerosis; A within subject comparison. *Psychophysiology*, *42*(Suppl 1), 29.
- Nijboer, F., Sellers, E., Mellinger, J., Matuz, T., Mochty, U., Jordan, M., et al. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, *119*, 1909–1916.
- Ohki, M., Kanayama, R., Nakamura, T., Okuyama, T., Kimura, Y., & Koike, Y. (1994). Ocular abnormalities in amyotrophic lateral sclerosis. *Acta Oto-laryngologica. Supplementum*, *511*, 138–142.
- Palmowski, A., Jost, W. H., Osterhage, J., Prudlo, J., Kasmann, B., Schimrigk, K., & Ruprecht, K. W. (1995a). [Disorders of eye movement in amyotrophic lateral sclerosis—report of 2 patients]. *Klinische Monatsblätter für Augenheilkunde*, *206*, 170–172.
- Palmowski, A., Jost, W. H., Prudlo, J., Osterhage, J., Kasmann, B., Schimrigk, K., & Ruprecht, K. W. (1995b). Eye movement in amyotrophic lateral sclerosis: a longitudinal study. *German Journal of Ophthalmology*, *4*, 355–362.
- Piccione, F., Giorgi, F., Tonin, P., Priftis, K., Giove, S., Silvoni, S., et al. (2006). P300-based brain computer interface: Reliability and performance in healthy and paralysed participants. *Clinical Neurophysiology*, *117*, 531–537.
- Pierce, J. R. (1980). *An introduction to information theory*. New York: Dover Publications.
- Robles, L., & Delano, P. H. (2007). Efferent system. In A. I. Basbaum, C. M. Bushnell, & D. V. Smith (Eds.), *The senses: A comprehensive reference* (Vol. 3, pp. 413–445). Elsevier.
- 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.
- Schreiber, H., Gaigalat, T., Wiedemuth-Catrinescu, U., Graf, M., Utner, I., Muche, R., & Ludolph, A. C. (2005). Cognitive function in bulbar- and spinal-onset amyotrophic lateral sclerosis. A longitudinal study in 52 patients. *Journal of Neurology*, *252*, 772–781.
- 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.
- Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication*. Illinois: University of Illinois Press.
- Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200–202.
- Squires, N. K., Donchin, E., Squires, K. C., & Grossberg, S. (1977). Bisensory stimulation: Inferring decision-related processes from P300 component. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 299–315.
- Stevens, A. A., Snodgrass, M., Schwartz, D., & Weaver, K. (2007). Preparatory activity in occipital cortex in early blind humans predicts auditory perceptual performance. *Journal of Neuroscience*, *27*, 10734–10741.
- Szmidt-Salkowska, E., & Rowinska-Marcinska, K. (2005). Blink reflex in motor neuron disease. *Electromyography and Clinical Neurophysiology*, *45*, 313–317.
- Whitehouse, P. J., Wamsley, J. K., Zarbin, M. A., Price, D. L., & Kuhar, M. J. (1985). Neurotransmitter receptors in amyotrophic lateral sclerosis: possible relationship to sparing of eye movements. *Annals of Neurology*, *17*, 518.
- Wintink, A. J., Segalowitz, S. J., & Cudmore, L. J. (2001). Task complexity and habituation effects on frontal P300 topography. *Brain and Cognition*, *46*, 307–311.

(RECEIVED December 17, 2007; ACCEPTED June 23, 2008)