



Published in final edited form as:

*Clin Neurophysiol.* 2009 July ; 120(7): 1252–1261. doi:10.1016/j.clinph.2009.04.019.

## Toward A High-Throughput Auditory P300-based Brain-Computer Interface

D. S. Klobassa<sup>a,b</sup>, T. M. Vaughan<sup>a</sup>, P. Brunner<sup>a,c</sup>, N. E. Schwartz<sup>d</sup>, J. R. Wolpaw<sup>a</sup>, C. Neuper<sup>b,c</sup>, and E. W. Sellers<sup>a,d</sup>

<sup>a</sup> Wadsworth Center, NYS Dept Health, Albany, NY

<sup>b</sup> Inst Psychology, Karl-Franzens Univ Graz, Austria

<sup>c</sup> Univ Tech, Graz, Austria

<sup>d</sup> East Tennessee State Univ, Johnson City, TN

### Abstract

**Objective**—Brain-computer interface (BCI) technology can provide severely disabled people with non-muscular communication. For those most severely disabled, limitations in eye mobility or visual acuity may necessitate auditory BCI systems. The present study investigates the efficacy of the use of six environmental sounds to operate a 6×6 P300 Speller.

**Methods**—A two-group design was used to ascertain whether participants benefited from visual cues early in training. Group A (N=5) received only auditory stimuli during all 11 sessions, whereas Group AV (N=5) received simultaneous auditory and visual stimuli in initial sessions after which the visual stimuli were systematically removed. Stepwise linear discriminant analysis determined the matrix item that elicited the largest P300 response and thereby identified the desired choice.

**Results**—Online results and offline analyses showed that the two groups achieved equivalent accuracy. In the last session, eight of ten participants achieved 50% or more, and four of these achieved 75% or more, online accuracy (2.8% accuracy expected by chance). Mean bit rates averaged about 2 bits/min, and maximum bit rates reached 5.6 bits/min.

**Conclusions**—This study indicates that an auditory P300 BCI is feasible, that reasonable classification accuracy and rate of communication are achievable, and that the paradigm should be further evaluated with a group of severely disabled participants who have limited visual mobility.

**Significance**—With further development, this auditory P300 BCI could be of substantial value to severely disabled people who cannot use a visual BCI.

### Keywords

Brain-Computer Interface; Brain-Machine Interface; EEG; P300; Event-Related Potential; Rehabilitation

---

Corresponding author: Daniela Klobassa, Attemsgasse 6, 8010 Graz, Austria, daniela\_k@gmx.at, Phone: +43 650 921 6183.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 1. Introduction

Brain-computer interface (BCI) technology can provide severely disabled people with non-muscular communication (Wolpaw et al., 2002). Several different electroencephalographic (EEG) brain signals can be used to operate a BCI. The approach presented here uses the P300 event-related potential (ERP). The P300 is an ERP component that appears as a positive deflection over central and parietal scalp areas approximately 300 ms after the presentation of a rare or salient stimulus. The stimulus can be visual, auditory, or somatosensory. The P300 is an innate response, representing processing activities of stimuli and their context (Sutton et al., 1965).

The P300 speller, first described by Farwell and Donchin (1988), uses P300 responses to choose letters from a matrix presented on a computer screen. This visual P300 speller paradigm is well established and provides reasonable control for a BCI (Donchin et al., 2000; Serby et al., 2005; Allison & Pineda, 2003; Krusienski et al., 2006; Sellers et al., 2006; Hoffmann et al., 2008). It allows the participant to write words and sentences. The operating principle behind the P300 speller is the oddball paradigm (OP), in which target and non-target stimuli are presented in a random series. The participant is then instructed to attend to the target stimulus, which is presented infrequently, and to ignore the non-target stimuli which are presented frequently (Fabiani et al., 1987). The target stimulus elicits a P300 response that can be reliably detected in the EEG (reviewed in Polich, 2007).

The P300 speller is a modified OP. The participant is presented with a 6×6 matrix that includes 36 characters. The individual rows and columns flash in rapid succession. The participant's task is to communicate a specific character by attending to that character and counting the number of times it flashes. The flash of the row or column that contains the desired character elicits a P300 response. By determining which row and which column elicit a P300, the BCI can identify the character the participant wants to select. Counting the number of flashes helps to keep the participant's attention focused on the task.

However, many people with disabilities, including people with late-stage amyotrophic lateral sclerosis (ALS), are unable to maintain gaze. A number of studies report that individuals with ALS can have impaired eye movements and slowing of saccades (Averbuch-Heller et al., 1998; Jacobs et al., 1981; Ohki et al., 1994; Szmids-Salkowska and Rowinska-Marcinska, 2005). Further, individuals with late stage ALS may develop locked-in syndrome (LIS). Individuals with LIS lose all volitional muscle control with the exception of horizontal eye movement. These individuals may eventually enter the totally locked in state (TLIS), where eye muscles are completely paralyzed, and all communication is lost (Harvey et al., 1979; Hayashi and Kato, 1989; Cohen and Caroscio, 1983), while presumably, cognition remains largely intact (Laureys et al., 2004). A BCI system that depends on auditory stimuli may circumvent these visual problems. The current study uses auditory stimuli (Hillyard et al., 1973) to operate the P300 BCI speller.

Recently, several auditory BCI paradigms have been described (Farquhar et al., 2008; Furdea et al., 2009; Hill et al., 2005; Nijboer et al., 2008b; Pham et al., 2005; Sellers and Donchin, 2006). The major limitation of six of these paradigms is that they provide no more than 2–4 alternative choices per trial. In response to this limitation, we designed a paradigm that uses auditory stimuli to operate a 6×6 P300 speller, thereby increasing the number of choices per trial to 36. Recently, Furdea et al (2009) examined a similar approach, where each row or column of a 6 × 6 matrix was coded with the words 1–12. These words were presented as auditory stimuli. Within the same experimental session, the authors compared each user's performance using an auditory paradigm and a visual paradigm (or standard P300 speller). The user's task was to spell the 10-letter word "brainpower." Furdea et al (2009) found lower

accuracy in the auditory modality than in the visual modality. The current results confirm and extend these findings with a 36-choice matrix.

The current study had two goals. The first goal was to demonstrate that people can use an auditory P300-based BCI to make accurate selections at a rate faster than those previously reported. The second goal was to determine if the addition of visual cues early in training would affect final performance. We compared a total of 11 experimental sessions for each of 10 users, where each session consisted of between 16 and 42 character selections. In the current study, there were two groups of five. Group A received only auditory stimuli. Group AV initially received simultaneous auditory + visual stimuli, and then the visual stimuli were gradually withdrawn until only the auditory stimuli remained.

## 2. Methods

### 2.1 Participants

The study enrolled ten volunteer participants (four women and six men) who were modestly compensated for participation. Age ranged from 22 to 68 years (mean=47, SD=16.72). None of the participants had previously participated in a BCI study or had neurological, visual, or auditory impairments. The study was approved by the New York State Department of Health Institutional Review Board, and each participant gave informed consent prior to the first experimental session.

### 2.2 Data acquisition and processing

EEG was recorded using a standard clinical 64-channel electrode cap (Electro-Cap International, Inc.) referenced to the right earlobe and grounded to the right mastoid. The EEG signals were amplified with an SA Electronics amplifier, band-pass filtered between 0.1 Hz and 50 Hz, and digitized at a rate of 256 Hz. Impedances between the cap electrodes and the reference electrode did not exceed 5 k $\Omega$ . The P300 speller was implemented using the BCI2000 software platform (Schalk et al. 2004). Matlab 7.0 was used for offline signal processing.

### 2.3 Experimental paradigm

The participant sat in a reclining chair facing a monitor and two speakers all placed 1.5 m away. The 6 $\times$ 6 square matrix (18.29 cm in height and width) was displayed in the center of the 19-inch screen (Figure 2). The matrix contained the 26 letters of the English alphabet, the numbers 1–9, and a backspace button. Character size was 2.43 cm H  $\times$  2.81 cm W, with equal vertical and horizontal distances between the center of adjacent characters. An auditory P300 speller may be more difficult to use than the standard visual version because it requires mapping sounds to matrix items. Thus, we used a two-group design to ascertain whether participants would benefit from visual cues early in training. The two-group design also allowed us to examine the possibility of interference between the auditory and visual modalities. Each participant was randomly assigned to one of the two experimental groups. Group A, the auditory-only group (N=5), watched a static character matrix on the screen and received auditory stimuli only. Group AV, the auditory + visual group (N=5), watched the character matrix on the screen and received concurrent auditory and visual stimuli. In the AV Group, the visual stimuli were then gradually removed, as described below.

The standard P300 6 $\times$ 6 matrix paradigm (Donchin et al., 2000; Sellers et al., 2006) was modified to simplify the task. While the standard paradigm flashes the columns and rows in a mixed random order, the current study presented the columns and rows separately. First, the six columns were presented a pre-defined number of times in random order, and then the six rows were presented a pre-defined number of times in random order. Six environmental sounds (bell, bass, ring, thud, chord, buzz) were used to represent the six columns and six rows, with

each sound corresponding to a particular row and to a particular column. The auditory stimuli were presented via two speakers for 110 ms at constant volume. The six environmental sounds were easily distinguished from one another, and showed comparable spectral envelopes. In the auditory + visual condition, the corresponding row or column was flashed simultaneously with the presentation of the auditory stimuli for the same duration. Each auditory or visual stimulus was presented for 110 ms and stimulus offset was 390 ms, so that the inter-stimulus interval was 500 ms (Fig. 1).

Prior to the start of the experiment, participants were given both written and oral instructions. The participant's task was to attend to the sound that represented the column or row of the item designated by the experimenter and to ignore the other sounds. The task was explained and demonstrated using a PowerPoint slide show and any remaining questions were answered verbally by the experimenter. Before the beginning of the first session, one or more test runs were performed to ensure that each participant fully understood the task.

Figure 2 illustrates the task. The characters to be selected, the targets, are presented above the matrix. In this case the targets are "A" and "W". In this example, the participant has already correctly selected the letter "A." Thus, a second "A" is displayed in the feedback field directly under the target. In Figure 2A, when the column containing the letter "W" is to be selected, the participant can select the correct column by attending to the "chord" sound. To simplify the task, the word "chord" is displayed next to the target character "W." During each column selection (e.g., chord), the six stimuli were presented in a randomized block design (i.e., in randomized sequences) a predefined number of times. In each sequence of sounds, each sound was only presented once. Then, after a 2-second pause, the name of the sound (e.g., chord) printed on the screen changed to the appropriate row stimulus (e.g., thud), as indicated in Figure 2B. In the subsequent set of stimuli (Figure 2B), when a row is to be selected, the participant attends to the "thud" sound to select the row containing the letter "W." Again, the word "thud" is presented next to the target character "W," so that there is no need to remember the corresponding sounds.

All six auditory stimuli were presented once without replacement in each stimulus-presentation sequence. During the first session, there were eight sequences for each column selection, followed by a two-second pause, followed by eight sequences for a row selection. Thus, one selection presented a total of 96 stimuli (48 for columns and 48 for rows). Following all 16 sequences, the BCI chose the column and row that produced the highest classification coefficients (described below). That is, on correct trials, the BCI selected the character to which the participant was attending.

All participants completed 12 experimental sessions over a 4–6 week period. Each session lasted 60 minutes including set-up and clean-up time. The first session consisted of 8 runs of 2 selections each for a total of 16 selections. Data from the first session were used to develop classification coefficients applied during subsequent sessions. The results describe the remaining 11 sessions.

Participants did not receive classification feedback during the first experimental session. The first session (Session 0) was used to collect data to train a classifier (described below). In the subsequent 11 sessions, classification feedback was provided to the participant after each character selection. At the beginning of each session, the task and sounds were reintroduced using PowerPoint slides.

As already mentioned, participants did not need to remember the positions of the sounds or characters because the names of the target sounds corresponding to the target characters were presented next to the target characters above the matrix (Fig. 2). However, to determine if the participants were memorizing the sound-to-target character relationship, they were asked

before and after sessions 2–12, to write down the sounds corresponding to the columns and rows of the matrix. All of the participants were able to correctly recall the corresponding column and row sounds after one or two sessions. However, we do not have data from the beginning of the experiment because we started this protocol after some initial sessions had been conducted. However, it is clear that the columns and rows corresponding to the sounds were easily remembered.

For Group AV, the visual stimuli (i.e., the flashing of the rows and columns) were presented in all runs in Sessions 1 and 2. Beginning in Session 3, visual stimuli were gradually and systematically removed according to the protocol shown in Table 1. In Session 3, visual stimuli were presented in 75% of the runs of the session; in subsequent sessions, the percentage decreased by 12.5% each session so that in Sessions 9–11 only auditory stimuli were presented (see Table 1). Overall, Group AV received 42.05% concurrent auditory and visual stimuli and 57.95% auditory stimuli, whereas Group A received 100% auditory stimuli.

Target characters were chosen in such a way that the sounds were approximately equally common in each session. The chance rate to select one letter out of 36 was very low (2.8%). Initial sessions were comprised of eight runs of two character selections. Each character selection was comprised of 16 stimulus-presentation sequences, 8 for column selection, and 8 for row selection (48 sec/character), with a 2-sec pause between column and row stimulus-presentation sequences, and a 3-sec pause following the first character selection. This 3-second interval was inserted to allow the participant to orient their attention to the next character. Thus, including the 96 sec of flashing (48 sec/character), 2 sec between the columns and rows, and the 3 sec between each character, the first selection of the run lasted 101 seconds, and the second lasted 99 seconds. Therefore, initial sessions consisted of eight 3.33 min runs, or 26.67 min/session. The time between successive runs was determined by the participant, but was typically 30–60 sec. Thus, the total time per experimental session was approximately 30 minutes. When offline classification (described below) rates reached 90% accuracy or above, the number of sequences per trial was reduced and the number of selections made was increased accordingly. Data collection time per session was held constant at about 30 min. The purpose of this adjustment was to define mean and maximal bit rates.

#### 2.4 Stepwise linear discriminant analysis (SWLDA): classification coefficients

Stepwise linear discriminant analyses (SWLDA) (Draper and Smith, 1981) was used to determine the column and row that elicited the P300 response and thereby to identify the desired character. Several previous studies have shown that SWLDA provides a robust classifier for P300 signals (Donchin et al., 2000; Farwell and Donchin, 1988; Krusienski et al, 2006; 2008; Sellers and Donchin, 2006; Sellers et al., 2006). In this study, it served as the learning algorithm to find a linear combination of features to use in a linear classifier that identified the row and column of the item desired by the participant.

SWLDA determined the optimal linear discriminant function by adding the spatiotemporal features of the EEG to a linear equation in a stepwise fashion. Starting with no model terms, the feature accounting for the most unique variance was entered into the model. After each entry to the model, a backward stepwise regression was performed to remove any variables that no longer met the criterion to remain in the model. The entry criterion was set to a p-value of <0.10 and the removal criterion was set to a p-value of >0.15. This process was repeated until no additional features met the entry/removal criteria, or until a predefined number of features were entered into the model (Draper and Smith, 1981; Krusienski, et al. 2006; 2008). In the current study, the maximum number of features allowed in the model was 60. For each row and column, the brain signals were time locked to the stimulus onset and averaged over all repetitions. The row and column that provided the highest coefficient values in the linear regression determined the selected character.

Before the SWLDA procedure was performed, data were moving-average filtered and decimated to 20 Hz. The response window for the analysis was 800 ms, starting at stimulus onset. The SWLDA procedure derived coefficients from two pre-defined channel sets. Set 1 included eight channels (F<sub>Z</sub>, C<sub>Z</sub>, P<sub>Z</sub>, P3, P4, PO7, PO8 and O<sub>Z</sub>). Krusienski et al (2008) showed that this eight-channel, ear-referenced, subset with a moving average window, and a decimation factor of 20 Hz provided the best performance for a group of 8 participants who performed multiple P300 BCI sessions. Set 2 included 30 Channels (F<sub>Z</sub>, F1-F4, FC<sub>Z</sub>, FC1-FC4, C<sub>Z</sub>, C1-C4, CP<sub>Z</sub>, CP1-CP4, P<sub>Z</sub>, P1-P4, PO<sub>Z</sub>, PO3, PO4, PO7, PO8, O<sub>Z</sub>). Set 2, which included Set 1, was chosen to determine if a different set of electrode locations would generalize to the auditory matrix better than Set 1 (which had been optimized using the visual matrix).

Before each session, and for each channel set, the classifier was optimized using data from previous sessions. For example, before session 5, we used all combinations of sessions 1, 2, and 3 (i.e., 1 only, 2 only, 3 only, 1 and 2, 1 and 3, 2 and 3, and 1, 2, and 3) to determine which set optimally classified session 4. Subsequently, we used that set for online classification of session 5. Our rationale for generating coefficients in this manner is that it most closely resembles the way coefficients are derived to maximize subsequent online performance. We feel that using the last session as a test set, should, in theory, be most similar to the next subsequent session because people are more or less likely to change from one day to the next (e.g., progression of ALS). If we were to combine all available data and analyze it using a leave-one-out method, the coefficients producing the highest level of classification may not generalize very well to the next subsequent session. Offline accuracy can be affected by over-fitting effects when data is analyzed using cross validation or leave-one-out methods. While it is perfectly valid to use these methods in offline analyses, they do not guarantee optimal online performance. In contrast, the offline analyses used in the present study reflect coefficients derived a priori. Thus, in theory, they could be used online and would be expected to yield the same results online that they did offline.

Classification coefficients for online operation were chosen prior to each session and remained constant during the session. Offline, classification using different feature sets (i.e., classification coefficients) were compared to determine the best possible classification accuracy. This best possible classification accuracy is referred to here as offline accuracy. Often, the offline accuracy was higher than the online accuracy

For all analyses, the probability of a Type I error was maintained at 0.05. Statistical tests were conducted using analysis of variance.

### 3. Results

#### 3.1 Accuracy

Accuracy was calculated as the percentage of total correct letter selections. Figure 3 presents accuracy for both the auditory-only condition in Group A and both the auditory and auditory-visual conditions in Group AV for all sessions.

Accuracy between groups did not differ significantly in initial sessions (Session 1 and 2), although Group A received auditory stimuli and Group AV received concurrent auditory + visual stimuli, nor did it differ significantly in the final sessions (Session 10 and 11) when both groups received auditory-only stimuli. Although the two groups were statistically equivalent in initial and final performance, the variability among individual participants was high (Fig. 3). Because sample size was small (N=5) for each group, the group comparison should be viewed as explorative and may not allow broader generalizations.

Group A achieved higher accuracy in the final sessions than in the initial sessions [ $F_{1,4} = 15.011, p < 0.05$ ], with all participants increasing their performance over sessions. In contrast, Group AV achieved the same accuracy in initial sessions (concurrent auditory + visual stimuli) as in final sessions (auditory-only stimuli). Looking at individual performances in Group AV, four of five participants increased their performance both in auditory-only stimulus presentation and in combined presentation, but all participants of Group AV reached higher maximum accuracies in auditory + visual stimulus presentations than in auditory-only stimulus presentations.

Variability in accuracy was high both among participants and in individual sessions for each participant. For example, in the auditory + visual mode, the best participant of Group AV (Participant A) achieved online mean accuracy of 95.12% (SD = 7.89) and offline mean accuracy of 98.88% (SD = 3.18). The best participant in auditory-only mode (Participant I, Group A) achieved online mean accuracy of 77.15% (SD = 20.33) and offline mean accuracy of 84.25% (SD = 18.46).

Table 2 shows each participant's performance in the last session. Online, eight of ten participants achieved 50% or more in this session and four scored 75% or more. In the offline analysis, all ten participants performed 50% or more, with six of them achieving 75% or more, and four of them more than 85%. However, the last session was not necessarily the best one for every participant. All participants except one (Participant J = 68.75%) achieved 75% or more accuracy in their best session.

### 3.2 Bit rate

The number of stimulus presentation sequences (i.e., time per selection) was reduced for those who performed well on the task. This allowed us to define mean and maximal bit rates. For all participants, the initial session had 16 stimulus presentation sequences (total of 96 sec) per character. If offline classification accuracy of 90% or higher was achieved, the number of stimuli presentations per sound selection was reduced. Therefore, more characters could be selected (i.e., more trials could occur) in the same amount of time. Depending on offline classification accuracy, the number of stimulus presentation sequences was reduced to a minimum of 6 sequences per character selection or 36 sec per trial. Rate of communication (bit rate) is defined by the formula

$$B = \log_2(N) + P \cdot \log_2(P) + (1 - P) \cdot \log_2[(1 - P)/(N - 1)]$$

(Wolpaw et al., 2002; Serby et al., 2005) in which  $B$  is the number of bits per decision,  $N$  is the number of possible targets, and  $P$  is the accuracy probability. The bit rate  $R$  (in bits/min) can be calculated by multiplying  $B$  and  $M$ , when  $M$  is the average number of decisions per minute.

Figure 4 shows mean bit rate (+/- standard error) for the sessions of each group. It shows the high level of variability across participants. Statistical analyses showed no differences in bit rate between initial (Session 1 and 2) or final sessions (Session 10 and 11) across all participants. It should be noted that in Group AV, when the protocol requirement indicated the removal of visual stimuli, participants started again with the maximum number of stimulus presentation sequences in auditory-only runs. Thus, bit rate is low in these sessions. However, by the last session, bit rate for Group AV was as high as that for Group A. Although there were no statistically significant differences in mean bit rates between groups, Group A achieved higher maximum bit rates (5.64 bits/min) than did Group AV in auditory stimuli presentation (2.56 bits/min). However, the highest bit rates of all were produced in the auditory and visual

conditions of Group AV (8.61 bits/min). The bit rate results are consistent with the classification accuracy results (Section 3.3). To express it in terms of time per selection, the best participant in the AV mode (Participant A from Group AV) made a selection every 36 sec (8.61 bits/min), and the best participant in the A mode (Participant I from Group A) made a selection in 48 sec (5.64 bits/min).

Figure 5 shows the mean bit rates as a function of the number of stimulus presentation sequences for all participants of Groups A and AV. A minimum of six sequences were presented online. Therefore, the results for stimulus presentations below 6 sequences are estimated. These values estimate performance after each sequence of presentations. It must be noted that the actual effect of reducing the number of presentations on the participant's performance cannot be discerned from offline analyses such as these; however, subsequent online testing could verify whether such estimates are accurate. Figure 5 also shows classification accuracies (averaged over all sessions) as a function of the number of stimulus presentation sequences. In auditory + visual mode, bit rates were quite high after a few stimulus presentation sequences (1–4), but predicted accuracy was poor. While bit rate is a standard measure used for estimating performance, high bit rate can be associated with levels of accuracy that may be insufficient for BCI control (e.g., Sellers and Donchin, 2006). Increasing the number of stimulus presentation sequences resulted in accuracies high enough for effective communication. More time, i.e., more stimulus presentations, was needed to produce higher accuracy in both groups and modes. Group AV achieved the highest accuracy and highest bit rate in the auditory + visual mode.

### 3.3 Amplitude and latency of P300

The P300 component is described by its amplitude and latency within a certain time window. In the present study, P300 amplitude was defined as the voltage difference ( $\mu\text{V}$ ) between a negative peak and the largest positive peak (peak-to-peak measurement). Peak amplitude (which defines P300 amplitude) was measured from 200 to 600 ms after stimulus onset and referenced to the most negative peak between 100 and 200 ms. Latency (ms) was defined as the period between stimulus onset and the time when the maximum positive peak amplitude was reached. In the present task, the largest difference between target and non-target was observed at location Pz, after the first positive peak around 300ms or later.

Across all participants, auditory stimuli evoked higher peak amplitudes after target (5.65  $\mu\text{V}$ ) than after non-target stimuli (4.06  $\mu\text{V}$ ) [ $F_{1,9} = 6.66$ ,  $p < 0.05$ ] and longer latencies after target (406.40 ms) than after non-target stimuli (295.70 ms) at Pz [ $F_{1,9} = 24.73$ ,  $p < 0.01$ ]. In Group AV, higher peak amplitudes were measured after combined auditory and visual presentation (7.544  $\mu\text{V}$ ) than after auditory-only presentation (5.55  $\mu\text{V}$ ) [ $F_{1,4} = 23.54$ ,  $p < 0.001$ ]. No differences in peak amplitude were found between the two modes after non-target stimuli.

Table 3 presents peak amplitudes for target and non-target stimuli and latencies averaged over all sessions for each participant. For users H and A, amplitudes for non-target stimuli were larger than amplitudes for target stimuli in the auditory-only condition, but the difference was not significant. Fig 6 presents the physiological responses of two users to auditory or combined auditory and visual stimulation.

### 3.4 Sound comparison

The sound selection results were analyzed to detect any sound bias and to document that the six different sounds were equally salient. The relation between correct selections (*hits*) and false selections (*errors*) for each sound was examined for each participant across sessions. The total number of target presentations and selections was not the same for every participant, and

because the sound distribution was not exactly equal, the results were normalized before the ratio of hits and errors was determined (Table 4).

Across all participants and across the two groups, there was no bias in hits or errors towards any one of the six sounds. However, the AV condition produced more hits ( $M = 81.53$ ,  $SD = 11.37$ ) than the auditory condition ( $M = 64.09$ ,  $SD = 9.11$ ) [ $F_{1,8} = 7.176$ ,  $p < 0.05$ ] and fewer errors ( $M = 17.05$ ,  $SD = 12.17$ ) than the auditory condition ( $M = 36.32$ ,  $SD = 9.03$ ) [ $F_{1,8} = 8.081$ ,  $p < 0.05$ ].

## 4. Discussion

### 4.1. Accuracy

We hypothesized that naive participants would achieve an adequate level of accuracy using the auditory P300 speller within 11 or fewer sessions. The results show that most but not all participants performed well. In the last session, eight of ten participants achieved at least 50% (online accuracy), and four achieved 75% or more overall (with 2.8% expected by chance).

As indicated in the results, classification coefficients representing different feature sets were applied to the data after the fact to determine maximum offline accuracy. Based on these results, we can conclude that improved understanding of feature selection could improve online performance. On average, there was greater variability in offline versus online classification rates for the study participants who performed less well; i.e., classification coefficients derived from their data failed to generalize. Thus, it was more difficult to predict which classification coefficients would produce optimal online results.

Participants in the current study reported that the auditory-only paradigm required them to pay more attention than the auditory + visual paradigm. Hill et al. (2005) and Nijboer et al. (2008b) report high error rates in their auditory paradigms. For the BCI to be effective, it is essential that the participant be alert and attentive, as drowsiness can decrease the amplitude of the P300 (Misulis and Fakhoury, 2001). Hillyard (1984) outlined the dependence of the P300 response on attention and postulated that the P300 is a correlate of stimulus classification and decision. Lower alertness and arousal yield a lower (or absent) P300 amplitude, and performance is lowered. Thus, it is possible that those participants who did not achieve high accuracy were not sufficiently attentive. Another possibility is that having both presentation modes causes more attentional interference or placed more attentional demand on the task. For example, Donchin, Kramer and Wickens (1982) demonstrated that the amplitude of the P300 is related to the demands placed on processing resources. Simultaneous presentation of both auditory and visual stimuli may tax the same group of resources and reduce performance. In addition, Squires et al. (1977) found large differences in P300 latency for auditory and visual stimuli. When the stimuli were combined (as in this study) the P300 response followed the pattern of the auditory P300. Thus, our initial assumption was that the auditory signal would dominate the P300 response.

The current study also sought to determine if auditory training would be enhanced by visual cues in early training. The AV Group started with concurrent auditory and visual stimuli, while the A Group experienced only auditory stimuli throughout the course of the study. Nevertheless, the two groups showed statistically equivalent accuracy in initial and final sessions. However, sample size was small ( $N=5$  for each group) and variability among participants was high, so group comparisons may not generalize to a larger sample. Our primary reason for choosing a small sample size with a high number of sessions is related to our target population, people who will be using the system day after day, for long periods of time. We recognize that using a relatively small number of participants may not be optimal for parametric

statistical analyses; however, we feel that it is more representative of how the system will be used for actual communication purposes.

A comparison of the two modes showed higher accuracies in concurrent auditory and visual stimulus presentation than in auditory stimulus presentation. It is possible that the visual flashes were more dominant, and that participants attended to the visual flashes more than to the sounds. In the final two sessions, after visual stimuli had been removed, participants of Group AV performed as well as they had in the initial two sessions with concurrent auditory and visual stimulation. Group A performed significantly better in the final sessions than in the initial ones. Hence, Group A participants improved over the course of the study.

The fact that concurrent auditory + visual stimuli produced the highest accuracies implies that the auditory/visual combination did not interfere with P300 generation. Rather, the dual mode stimulation enhanced the response. This is consistent with the findings of Teder-Sälejärvi et al. (2002) who showed in their ERP study that combined auditory and visual stimuli produced higher accuracies and shorter latencies than did an auditory-only or visual-only modality. Pham et al. (2005) found contrary results when they tested BCIs with visual, auditory, and combined auditory and visual conditions using slow cortical potentials (SCPs) (potential changes below 1 Hz lasting up to several seconds). The visual condition produced the highest accuracy, followed by auditory condition; in the combined visual and auditory stimuli condition, participants were not able to control SCP amplitude above chance level. Pham et al. (2005) concluded that performance in the auditory condition was lower than performance in the visual condition because auditory stimuli constitute a higher degree of task difficulty and require increased selective attention. Because our main interest was in developing an auditory paradigm, we did not test a visual-only condition. Thus, we cannot conclude that the Pham et al. (2005) results are applicable to the current study. In addition, they were using SCPs rather than a P300 paradigm. However, comparing the results of the current study to other visual-only P300 speller studies suggests that this is a reasonable conjecture.

Evidence for this is provided by that fact that the concurrent auditory + visual stimuli in the current study produced the highest mean accuracy levels of approximately 70% (see Figure 3 Group AV: auditory and visual (sessions 1 – 8)). This is 10% to 20% lower than visual-only studies where typical accuracy is in the 80% to 90% range (e.g., Krusienski et al, 2006, 2008; Nijboer et al, 2008a,b; Sellers et al, 2006; Serby et al, 2005). This is possibly due to the demands placed on processing discussed above (Donchin et al, 1982). It is also worth noting that Group A started at a mean accuracy level of approximately 30% and concluded at a mean accuracy level of approximately 60% in session 12 (see Figure 3 Group A: auditory). Similarly, Group AV auditory-only condition started at a mean accuracy rate of about 30% beginning in Session 7, when 75% of the stimuli were auditory-only, and concluded at a mean accuracy rate of about 65% in session 12 (see Figure 3 Group AV: auditory). These results are very similar to those of Group A. A possible explanation for this is that the increased task demands of attending to the auditory stimuli require learning and that with additional training sessions accuracy continues to improve. These results are in contrast to the visual-only P300 speller, where accuracy typically asymptotes quickly and remains at a relatively constant level throughout many sessions (Nijboer et al., 2008a,b; Sellers et al., 2006).

#### 4.2. Bit rate

Schloegl et al (2007) refers to four conditions that must be fulfilled for a correct application of the basic bit-rate formula (Wolpaw et al. 2002): (1) M Selections (classes) are possible, (2) each class has the same probability, (3) the specific accuracy is the same for each class, (4) each undesired selection has the same probability of selection. In the present auditory paradigm, all four conditions appear to be fulfilled. In contrast to the visual P300 speller, where columns or rows next or near to the target character are more like to be chosen, we assume that this is

not the case for the auditory P300 speller. Condition 1 and 2 were incorporated in the design of the paradigm. Condition 3 is fulfilled, as there was no difference in hits or errors among the 6 sounds. Condition 4 is fulfilled as the auditory mapping did not lead to a difference in the probability of undesired selections, since the 6 sounds were easily distinguishable from another.

Within Group AV, higher bit rates could be achieved with auditory + visual stimuli than for auditory-only stimuli. In auditory + visual mode, the maximal bit rate was 8.61 bits/min (online and offline), whereas in auditory-only mode the maximal bit rate was 5.64 bits/min (offline 6.01 bits/min).

Two studies using online analysis and a visual-only matrix are comparable to the present work in the number of possible selections (i.e., 36) (Donchin et al., 2000 and Serby et al., 2005). The number of possible selections is important because it is a factor in the calculation of bit rate. Serby et al. (2005) described offline bit rates of 23.75 bits/min and online of 15.3 bits/min. Donchin et al. (2000) reported offline bit rates of 20.1 bits/min and online bit rates of 9.23 bits/min. In comparison to these bit rates for visual P300 spellers, the auditory P300 speller in this study achieved lower communication rates (online 5.64 bits/min and offline 6.01 bits/min). However, this study's bit rate is the highest reported thus far for the auditory systems (c.f., for comparison, Hill et al., 2005; Nijboer et al., 2007; Pham et al., 2005, Sellers & Donchin, 2006). For example, participants of the four-choice paradigm in the auditory-only mode of Sellers & Donchin (2006) achieved a maximum bit rate of 1.8 bits/selection whereas participants in the current study achieved a maximum bit rate of 5.17 bits/selection.

One major problem for an auditory P300-based BCI is its slow information transfer rate. Schalk (2008) discusses the BCI communication problem. One issue concerns the low information transfer of the nervous system, which ranges from 1 bit/sec to not more than 50 bits/sec, while computers can process information at a rate exceeding 1 terabit/sec. He proposes that direct communication between the brain and the computer can overcome this low rate. His concept is that improvement of the interaction between technologies and areas of the brain with higher fidelity will increase communication rates between the brain and the computer.

### 4.3. Amplitude and latency of P300

The main difference between the peak amplitudes resulting from target vs. and non-target presentations was observed after the first positive peak around 300 ms or later. For some participants, the largest difference between target and non-target amplitudes appeared around 400 ms after stimulus onset. These longer latencies are described for more complex stimuli (Fabiani et al., 1987; Combs and Polich, 2006; Hagen et al., 2006).

Since even the waveforms for each participant varied greatly, it is difficult to draw meaningful conclusions about the difference in waveform for the two presentation modes. Indeed, Fabiani et al. (1987) already noted individual differences in P300 latency and amplitude. However, overall statistical analysis in this study showed that the target responses were characterized by higher amplitudes and longer latencies. Within Group AV, a comparison of both stimuli modes indicates that auditory and visual stimuli evoked higher amplitudes than auditory stimuli, but only for target stimuli. This could be a result of the fact that visual stimuli generate larger P300 amplitudes than auditory stimuli (Fabiani et al., 1987; Polich & Heine, 1996; Katayama & Polich, 1999).

### 4.4. Sound comparison

Hits and errors were equally distributed among the six different sounds. Therefore, salience of the sounds was comparable and significant sound bias was not present. Thus, it appears that

the six environmental sounds used are appropriate auditory stimuli for the auditory P300 speller.

#### 4.5. Summary and Conclusions

One goal of this study was to determine whether an auditory P300 speller can achieve accuracy and bit rate levels high enough for useful communication. This study shows that acceptable levels can be achieved with an auditory P300 speller. At this time, however, the speed and accuracy of the auditory speller is somewhat lower than that of the visual version. Average accuracy for the 6×6 36-item matrix is typically 80–90% for the visual P300 speller (e.g., Krusienski et al, 2006, 2008; Nijboer et al, 2008a,b; Sellers et al, 2006; Serby et al, 2005), whereas in this study the mean online accuracy of the auditory P300 speller for the last several sessions was about 66%. Compared to other auditory P300-based BCI systems, the results reported here are the highest to date. One other study reports a slightly lower average bit rate of 1.54 bits/min, and a maximum of 2.85 bits/min (Furdea, et. al., 2009), whereas the present study reports an average bit rate of 1.86 bits/min and a maximum of 5.64 bits/min.

As mentioned above, the complexity of the task in the auditory system may contribute to the lower performance relative to the visual system. With the visual P300 speller, the participant's task is to attend to a character of the matrix and to count how many times it flashes; the participant can simply ignore the other flashes. In contrast, with the auditory P300 speller, the participant must listen to all of the sounds and note the occurrence of the target sound. Higher task complexity may require greater attention and cognitive capability (Kramer et al., 1983; Johnson, 1986, 1993). In his neural model of the P300, Polich (2004) distinguished between the P3a and the P3b. The P3a is thought to reflect activity of the anterior cingulate when new stimuli are processed into working memory. The P3b is thought to reflect subsequent activation of the hippocampal formation when frontal lobe mechanisms interact with the temporal/parietal lobe connection. High task difficulty increases focal attention and enhances P3a amplitude by constraining other memory operations that reduce P3b amplitude and increase P3b latency (Hagen et al., 2006). It is not possible to estimate whether P300 amplitudes of the present study are increased or decreased since a benchmark (e.g., a standard oddball paradigm) was not used to compare task difficulty and amplitudes.

A second goal of this study was to examine if visual cues early in training had an effect on ultimate performance with the auditory speller. The results show that visual cues early in training do not facilitate performance after the visual cues have been removed. That is, when Group AV was presented with auditory-only stimuli, its performance was equivalent to that of Group A. This finding suggests that the introduction and use of an auditory P300 speller could remain entirely within the auditory modality, but may require more training sessions than the visual P300 speller. The implications are that the auditory speller could be useful for people who have vision impairment when they begin to use a BCI.

The results of the present study are encouraging and suggest that further development is worthwhile, and that a system as robust as the visual P300 speller might be achieved. In addition, work is under way to test the auditory paradigm with participants who have limited visual function.

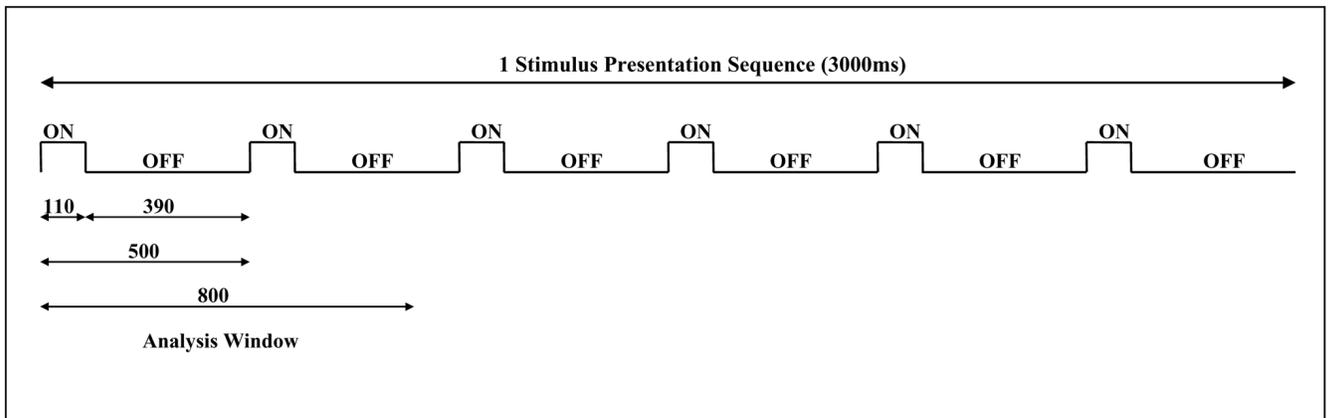
#### Acknowledgments

This work was supported in part by the National Institutes of Health (Grants NICHD HD30146 and NIBIB/NINDS EB00856), the James S. McDonnell Foundation, the ALS Hope Foundation, and the NEC Foundation. The authors would like to thank Dr. Elizabeth Winter Wolpaw for her critical comments on the manuscript.

## References

- Averbuch-Heller L, Helmchen C, Horn AK, Leigh RJ, Buettner-Ennever JA. Slow Vertical Saccades in Motor Neuron Disease: Correlation of Structure and Function. *Ann Neurol* 1998;44:641–648. [PubMed: 9778263]
- Allison BZ, Pineda JA. ERP's Evoked by Different Matrix Sizes: Implications for a Brain Computer Interface (BCI) System. *IEEE Trans Neural Syst Rehabil Eng* 2003;11(2):1–4. [PubMed: 12797719]
- Cohen B, Carosio J. Eye movements in amyotrophic lateral sclerosis. *J Neural Transm Suppl* 1983;19:305–315. [PubMed: 6583314]
- Combs LA, Polich J. P3a from auditory white noise stimuli. *Clin Neurophysiol* 2006;117:1106–1112. [PubMed: 16564203]
- Donchin, E.; Kramer, AF.; Wickens, C. Probing the cognitive infrastructure with event-related brain potentials. *Proceedings of the AIAA Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics*; Washington, DC: U.S. Government Printing Office; 1982. p. 371-387.
- Donchin E, Spencer KM, Wijesinghe R. The Mental Prosthesis: Assessing the Speed of a P300-Based Brain-Computer Interface. *IEEE Trans Neural Syst Rehabil Eng* 2000;8(2):174–179.
- Draper, N.; Smith, H. *Applied Regression Analysis*. Vol. 2. New York: John Wiley and Sons; 1981. p. 307-312.
- Fabiani, M.; Gratton, G.; Karis, D.; Donchin, E. Definition, identification and reliability of measurement of the P300 component of the event-related brain potential. In: Achles, PK.; Jennings, R.; Coles, MGH., editors. *Advances in Psychophysiology*. New York: JAI; 1987. p. 1-78.
- Farquhar, J.; Blankespoor, J.; Velk, R.; Desain, P. Towards a Noise-Tagging Auditory BCI- Paradigm. In: Müller-Putz, GR.; Brunner, C.; Leeb, R.; Pfurtscheller, G.; Neuper, C., editors. *Proceedings of the 4th Int BCI Workshop and Training Course 2008*. Graz: Verlag der TU Graz; 2008. p. 50-55.
- Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroenceph clin Neurophysiol* 1988;70:510–523. [PubMed: 2461285]
- Hagen GF, Gatherwright JR, Lopez BA, Polich J. P3a from visual stimuli: Task difficulty effects. *Int J Psychophysiol* 2006;59:8–14. [PubMed: 16253363]
- Furdea A, Halder S, Krusienski DJ, Bross D, Nijboer F, Birbaumer N, Kuebler A. An auditory oddball (P300) spelling system for brain-computer interfaces. *Psychophysiology* 2009;46.10.1111/j.1469–8986.2008.00783.x
- Harvey DG, Torack RM, Rosenbaum HE. Amyotrophic lateral sclerosis with ophthalmoplegia. A clinicopathologic study. *Arch Neurol* 1979;36(10):615–617. [PubMed: 485889]
- Hayashi H, Kato S. Total manifestations of amyotrophic lateral sclerosis: ALS in the totally locked-in state. *J Neurol Sci* 1989;93:19–35. [PubMed: 2809628]
- Hill, NJ.; Lal, TN.; Bierig, K.; Birbaumer, N.; Schölkopf, B. An Auditory Paradigm for Brain-Computer Interfaces. In: Saul, LK.; Weiss, Y.; Bottou, L., editors. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2005. p. 569-576.
- Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical signs of selective attention in the human brain. *Science* 1973;182:177–180. [PubMed: 4730062]
- Hillyard, SA. Event-related potentials and selective attention. In: Donchin, E., editor. *Cognitive Psychophysiology: Event-Related Potentials and the Study of Cognition*. New Jersey: Hillsdale; 1984. p. 51-72.
- Hoffmann U, Vesin J-M, Ebrahimi T, Diserens K. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods* 2008;167(1):115–125. [PubMed: 17445904]
- Jacobs L, Bozian D, Heffner RR Jr, Barron SA. An eye movement disorder in amyotrophic lateral sclerosis. *Neurology* 1981;31(10):1282–1287. [PubMed: 7202138]
- Johnson R Jr. A Triarchic Model of P300 Amplitude. *Psychophysiology* 1986;23:367–384. [PubMed: 3774922]
- Johnson R Jr. On the neural generators of the P300 component of the event-related potential. *Psychophysiology* 1993;30:90–97. [PubMed: 8416066]
- Katayama J, Polich J. Auditory and visual P300 topography from a 3 stimulus paradigm. *Clin Neurophysiol* 1999;110:463–468. [PubMed: 10363770]

- Kramer AF, Wickens CD, Donchin E. An Analysis of the Processing Requirements of a Complex Perceptual-Motor Task. *Human Factors* 1983;26(6):597–621. [PubMed: 6671646]
- Krusiński DJ, Sellers EW, Cabestaing F, Bayouduh S, McFarland DJ, Vaughan TM, Wolpaw JR. A comparison of classification techniques for the P300 Speller. *J Neur Eng* 2006;3(4):299–305.
- Krusiński DJ, Sellers EW, Mc Farland DJ, Vaughan TV, Wolpaw JR. Toward enhanced P300 Speller performance. *J Neurosci Methods* 2008;167(1):15–21. [PubMed: 17822777]
- Laureys S, Owen AM, Schiff ND. Brain function in coma, vegetative state, and related disorders. *Lancet Neurol* 2004;3(9):537–46. [PubMed: 15324722]
- Misulis, KE.; Fakhoury, T. Spehlmann's Evoked Potential Primer. Vol. 3. MA: BH; 2001. p. 199-203.
- Nijboer F, Sellers EW, Mellinger J, Jordan MA, Matuz T, Furdea A, et al. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin Neurophysiol* 2008a;119:1909–1916. [PubMed: 18571984]
- Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland DJ, Birbaumer N, Kübler A. An auditory brain-computer interface (BCI). *J Neurosci Methods* 2008b;167(1):43–50. [PubMed: 17399797]
- Ohki M, Kanayama R, Nakamura T, Okuyama T, Kimura Y, Koike Y. Ocular abnormalities in amyotrophic lateral sclerosis. *Acta Otolaryngol Suppl* 1994;511:138–142. [PubMed: 8203216]
- Pham M, Hinterberger T, Neumann N, Kübler A, Hofmayer N, Grether A, Wilhelm B, Vatine J, Birbaumer N. An Auditory Brain-Computer Interface Based on the Self-Regulation of Slow Cortical Potentials. *Neurorehabil Neural Repair* 2005;19(3):206–218. [PubMed: 16093411]
- Polich, J. Neuropsychology of P3a and P3b: A theoretical overview. *Brainwaves and mind: Recent developments*. Moore, NC.; Arikan, K., editors. Wheaton, IL: Kjellberg Inc; 2004. p. 15-29.
- Polich J. Updating P300: An integrative theory of P3a and P3b. *Clin Neurophysiol* 2007;118(10):2128–2148. [PubMed: 17573239]
- Polich J, Heine MRD. P300 topography and modality effects from a single-stimulus paradigm. *Psychophysiology* 1996;33:747–752. [PubMed: 8961797]
- Schalk G. Brain-computer symbiosis. *J Neural Eng* 2008;5(1):1–15. [PubMed: 18310806]
- Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Trans Biomed Eng* 2004;51:1034–1043. [PubMed: 15188875]
- Schloegl, A.; Kronegg, J.; Huggins, JE.; Mason, SG. Evaluation criteria for BCI research. In: Dornhege, G.; Millan, J.; Hinterberger, T.; McFarland, DJ.; Müller, K., editors. *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press; 2007. p. 327-342.
- Sellers EW, Donchin E. A P300-based brain-computer interface: Initial tests by ALS patients. *Clin Neurophysiol* 2006;117:538–548. [PubMed: 16461003]
- Sellers EW, Krusiński DJ, McFarland DJ, Vaughan TM, Wolpaw JR. A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biol Psychol* 2006;73(3):242–252. [PubMed: 16860920]
- Serby H, Yom-Tov E, Inbar G. An Improved P300-Based Brain-Computer Interface. *IEEE Trans Neural Syst Rehabil Eng* 2005;13(1):89–98. [PubMed: 15813410]
- 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. [PubMed: 864401]
- Szmidt-Salkowska E, Rowinska-Marcinska K. Blink reflex in motor neuron disease. *Electromyogr Clin Neurophysiol* 2005;45:313–317. [PubMed: 16218200]
- Sutton S, Braren M, Zublin J, John E. Evoked potential correlates of stimulus uncertainty. *Science* 1965;150:1187–1188. [PubMed: 5852977]
- Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn Brain Res* 2002;14:106–114.
- Wolpaw JR, Birbaumer N, McFarland D, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol* 2002;113:767–91. [PubMed: 12048038]



**Fig. 1.**

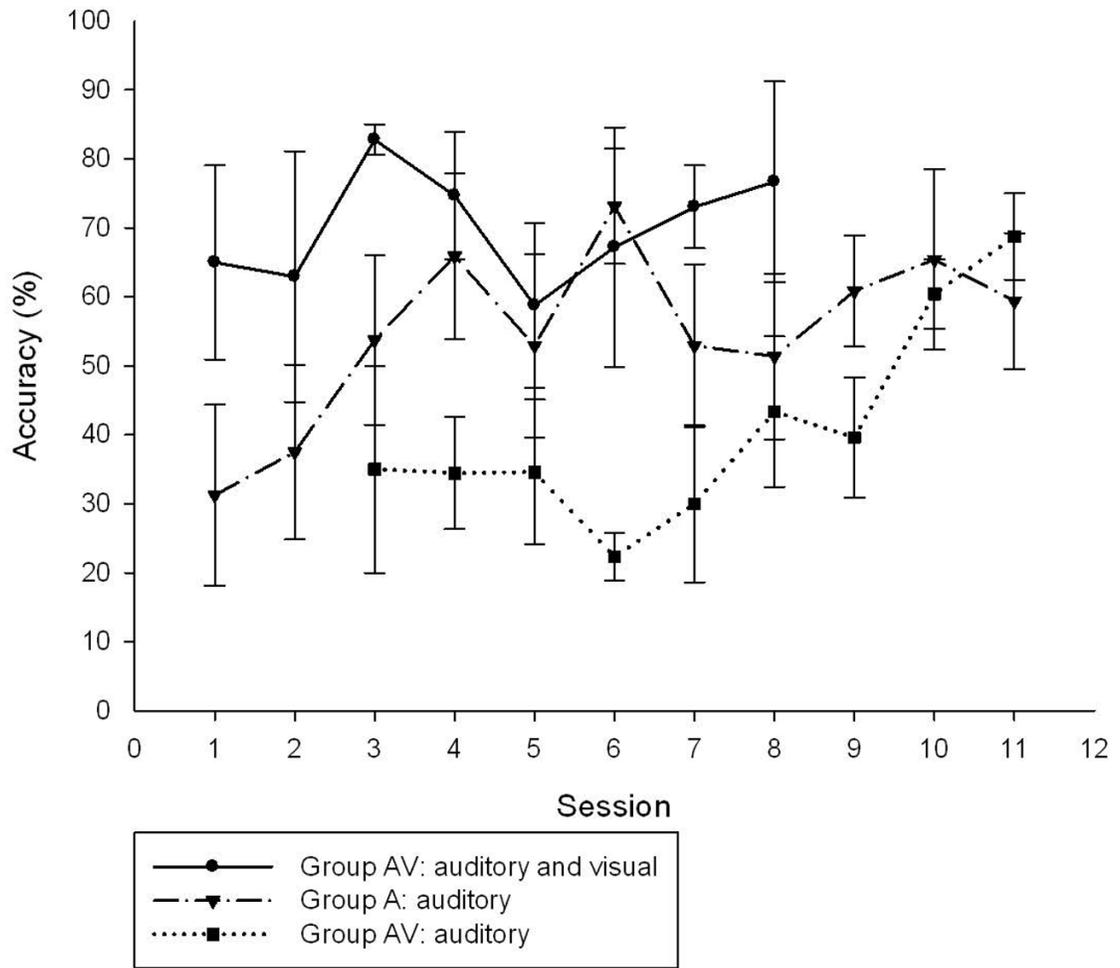
A timeline of one stimulus-presentation sequence. Visual and auditory stimuli lasted 500 ms each: 110 ms (ON) followed by a 390 ms inter stimulus interval (OFF). Thus, six stimuli were presented, in block-random order, without replacement, every 3 seconds. The response window used for analysis (bottom line) began with the stimulus onset and lasted for 800 ms.

AW (W) = (Chord)					
A					
A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	—

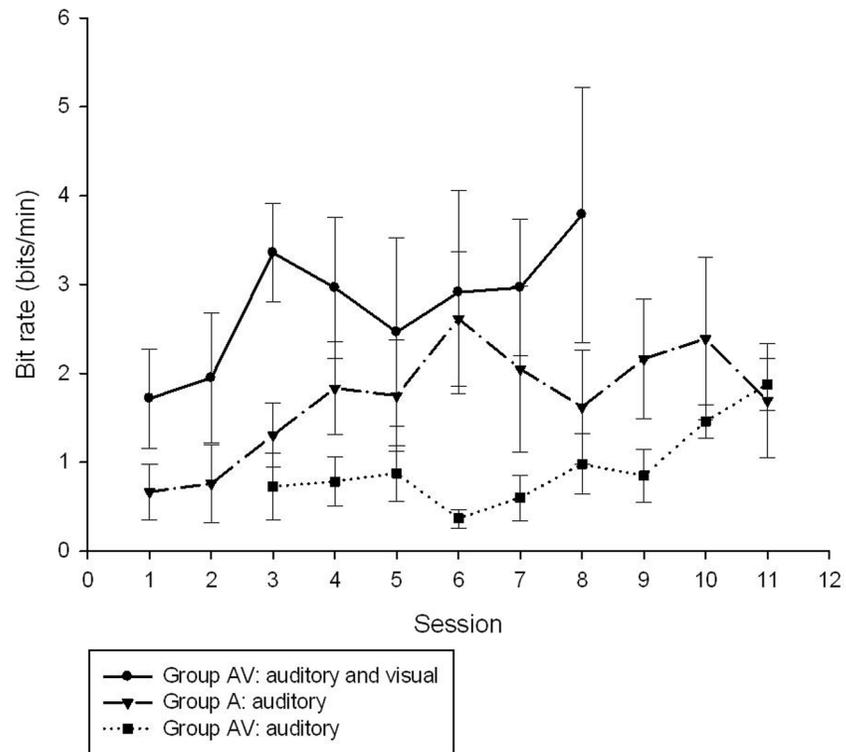
AW (W) = (Thud)					
A					
A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	—

**Fig. 2.**

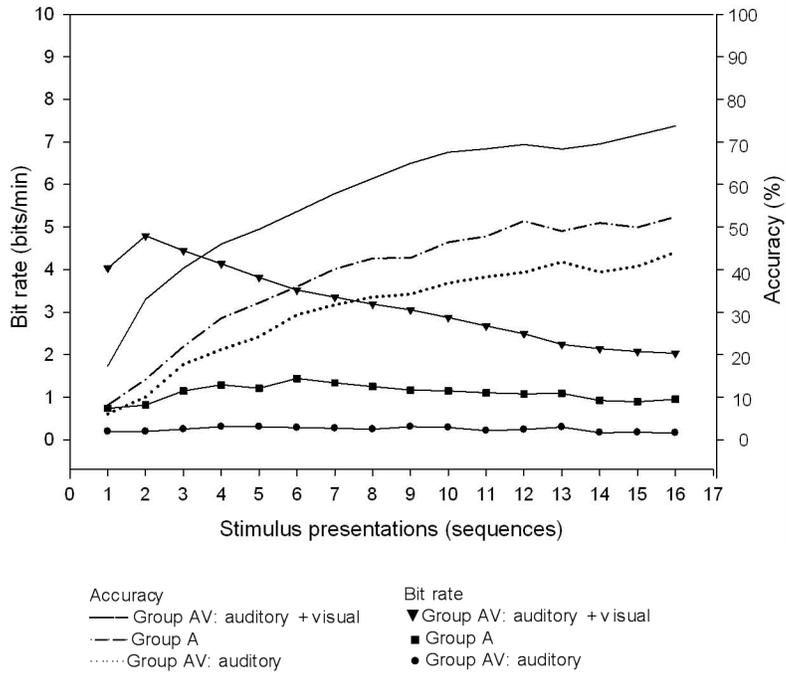
The 6×6 matrix. The six sounds: bell, bass, ring, thud, chord, and buzz, are each associated, in turn, with one of six columns and then, one of six rows. To select the target letter “W” from the 36-character matrix on the screen, the participant’s task was to (A) attend to the “chord” sound while the column stimuli were presented. Then, after a 2-second pause, the participant’s task was to (B) focus on the “thud” sound. For Group AV, in addition to the auditory stimuli, the column or row corresponding to each auditory stimulus was highlighted simultaneously with the corresponding column or row.



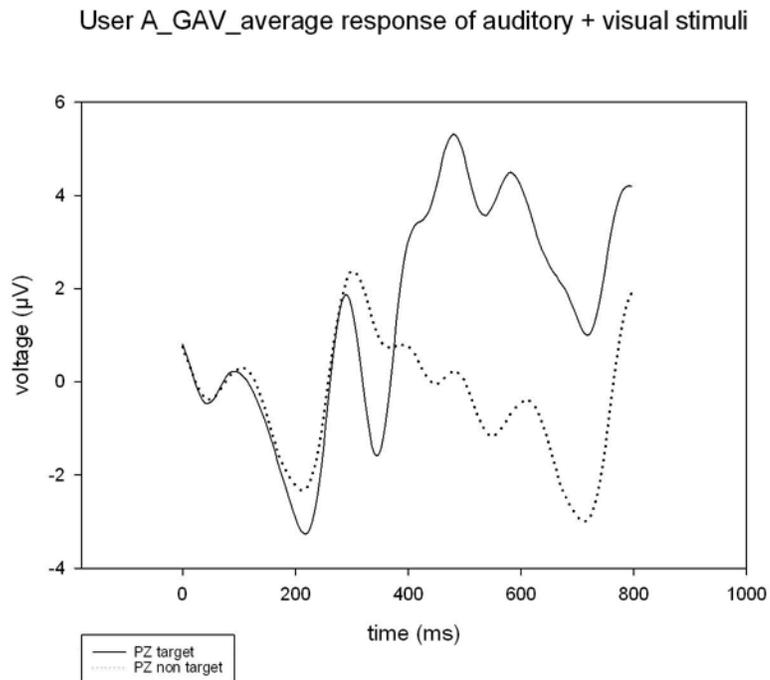
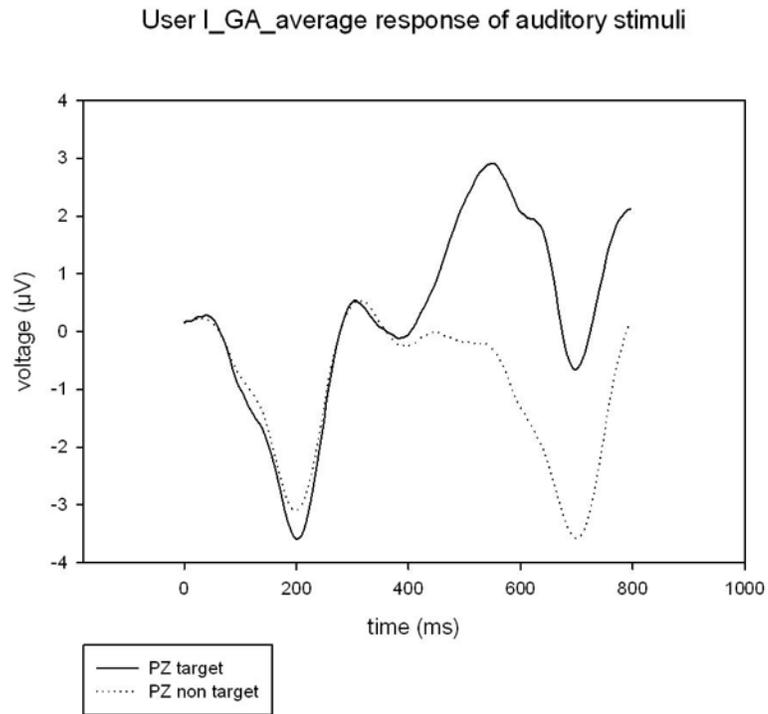
**Fig. 3.**  
Mean online accuracy  $\pm$  S.E. for Group A and AV for all sessions.



**Fig. 4.**  
Mean bit rates  $\pm$  S.E. for Groups A and AV for all sessions.



**Fig. 5.** Mean bit rates (left axis) and mean accuracies (right axis) averaged over all sessions as a function of the number of stimulus presentations for Group A (square and dash-dot) and Group AV (AV: triangle and solid line; A: bullet and dotted line). With fewer repetitions, bit rate is highest (left axis); more repetitions are required to produce high accuracy (right axis).



**Fig 6.** Average waveforms recorded at Pz, across all sessions for user I of group A (Fig 6A); and user A of group AV (Fig 6B). The solid line shows the average target response at Pz whereas the dotted line shows the non-target response. They begin to differ from another at the second positive peak about 300 ms after stimulus onset.

**Table 1**

Design of stimuli presentation in Group AV: During Sessions 1 and 2, Group AV received simultaneous visual and auditory stimuli. Starting in Session 3, visual stimuli were systematically removed in single runs: in Session 3, 75% of the runs presented visual and auditory stimuli, while the remaining 25% presented auditory-only stimuli. From Session 3 to Session 9, the % of runs presenting visual + auditory stimuli was decreased until only auditory stimuli were presented in session 9 to 11.

Session #	% visual + auditory runs	% auditory-only runs	Run #
1-2	100.0 (8 runs)	0.0 (0 runs)	N/A
3	75.0 (6 runs)	25.0 (2 runs)	3, 6
4	62.5 (5 runs)	37.5 (3 runs)	2, 5, 8
5	50.0 (4 runs)	50.0 (4 runs)	2, 4, 6, 8
6	37.5 (3 runs)	62.5 (5 runs)	2-3, 5-6, 8
7	25.0 (2 runs)	75.0 (6 runs)	2-4, 6-8
8	12.5 (1 run)	87.5 (7 runs)	1-3, 5-8
9-11	0.0 (0 runs)	100.0 (8 runs)	N/A

**Table 2**

Accuracies for final session (session 11) with auditory stimuli alone for all participants.

<b>GROUP A Participants</b>	<b>Online accuracy (%)</b>	<b>Offline accuracy (%)</b>
B	50.00	63.00
D	87.50	100.00
E	43.75	50.00
H	37.50	50.00
I	78.13	88.00
Mean Group A	59.38	70.20
Standard deviation	22.10	22.76
Standard error	8.27	7.86
<b>GROUP AV Participants</b>	<b>Online accuracy (%)</b>	<b>Offline accuracy (%)</b>
A	87.50	94.00
C	62.50	75.00
F	50.00	69.00
G	75.00	83.00
J	68.75	88.00
Mean Group AV	68.75	81.80
Standard deviation	13.98	9.99
Standard error	8.27	7.86

Mean peak amplitudes and latencies averaged over all sessions for each participant of Group A and Group AV, measured at Pz. For each participant, an ANOVA was computed to examine differences between target and non-target responses.

**Table 3**

Group A	Auditory-only (averaged over all sessions)	
	Amplitude ( $\mu$ V)	Latency (ms)
<b>B</b>	Target	412*
	Non-target	257*
<b>D</b>	Target	482**
	Non-target	324**
<b>E</b>	Target	300
	Non-target	289
<b>H</b>	Target	333
	Non-target	287
<b>I</b>	Target	552**
	Non-target	386**

Group AV	Auditory-only (averaged over all sessions)		Auditory + visual (averaged over all sessions)	
	Amplitude ( $\mu$ V)	Latency (ms)	Amplitude ( $\mu$ V)	Latency (ms)
<b>A</b>	Target	502**	8.844**	496**
	Non-target	288**	5.082**	306**
<b>C</b>	Target	405	6.343**	466**
	Non-target	312	4.135**	314**
<b>F</b>	Target	439**	5.322*	309
	Non-target	270**	4.159*	300
<b>G</b>	Target	306*	6.680**	320*
	Non-target	279*	3.758**	310*
<b>J</b>	Target	333	10.529**	287**

Group A	Auditory-only (averaged over all sessions)	
	Amplitude ( $\mu$ V)	Latency (ms)
Non-target	8,466	8,748 <sup>*</sup>
		275 <sup>**</sup>

\* indicates a significance level of 0.05, whereas

\*\* indicates a significance level of 0.01.

**Table 4**

Hit rate and error rate of each sound. The diagonal in the table shows the absolute values of all correct selections (hits), whereas the other cells display the absolute values of all wrong selections (errors). The columns labeled “hits” and “errors” provide normalized values for a comparison among the six sounds. For example, from all correct selections (72.78%) the bell was correctly chosen in 12.36%, whereas from all wrong selections (27.22%) the bell was incorrectly chosen in 5.24%.

target\selected	bell	bass	ring	thud	chord	buzz	hits	errors
bell	<b>545</b>	36	59	44	33	45	<b>12.36%</b>	<b>5.24%</b>
bass	61	<b>499</b>	33	56	38	51	<b>11.32%</b>	<b>4.22%</b>
ring	51	31	<b>631</b>	25	29	39	<b>14.31%</b>	<b>4.97%</b>
thud	45	46	34	<b>505</b>	36	38	<b>11.46%</b>	<b>4.06%</b>
chord	40	28	44	23	<b>534</b>	40	<b>12.11%</b>	<b>3.90%</b>
buzz	34	45	49	31	36	<b>494</b>	<b>11.21%</b>	<b>4.83%</b>
							$\Sigma=72.78\%$	$\Sigma=27.22\%$