A Practical, Intuitive Brain-Computer Interface for Communicating "Yes" or "No" by Listening

N. Jeremy Hill[1,2,*], **Erin Ricci**[1], **Sameah Haider**[1,3], **Lynn M. McCane**[1], **Susan Heckman**[1], **Jonathan R. Wolpaw**[1,4], and **Theresa M. Vaughan**[1,2]

[1]Wadsworth Center, New York State Department of Health (Albany, NY, USA)

[2]Helen Hayes Hospital (West Haverstraw, NY, USA)

[3]Albany Medical College (Albany, NY, USA)

[4]University at Albany, State University of New York (Albany, NY, USA)

Abstract

Objective—Previous work has shown that it is possible to build an EEG-based binary brain-computer interface system (BCI) driven purely by shifts of attention to auditory stimuli. However, previous studies used abrupt, abstract stimuli that are often perceived as harsh and unpleasant, and whose lack of inherent meaning may make the interface unintuitive and difficult for beginners. We aimed to establish whether we could transition to a system based on more natural, intuitive stimuli (spoken words "yes" and "no") without loss of performance, and whether the system could be used by people in the locked-in state.

Methodology—We performed a counterbalanced, interleaved within-subject comparison between an auditory streaming BCI that used beep stimuli, and one that used word stimuli. Fourteen healthy volunteers performed two sessions each, on separate days. We also collected preliminary data from two subjects with advanced ALS, who used the word-based system to answer a set of simple yes-no questions.

Main Results—The N1, N2 and P3 event-related potentials elicited by words varied more between subjects than those elicited by beeps. However, the difference between responses to attended and unattended stimuli was *more* consistent with words than beeps. Healthy subjects' performance with word stimuli (mean 77% ± 3.3 s.e.) was slightly but not significantly better than their performance with beep stimuli (mean 73% ± 2.8 s.e.). The two subjects with ALS used the word-based BCI to answer questions with a level of accuracy similar to that of the healthy subjects.

Significance—Since performance using word stimuli was at least as good as performance using beeps, we recommend that auditory streaming BCI systems be built with word stimuli to make the system more pleasant and intuitive. Our preliminary data show that word-based streaming BCI is a promising tool for communication by people who are locked in.

*Correspondence: Wadsworth Center, C640 Empire State Plaza, Albany, NY 12201, USA. jezhill@gmail.com.

1. Introduction

People may become **locked in**—i.e. "prevented from communicating by word or body movement" [1]—due to a variety of causes. For example, the locked-in state (LIS) may result from degenerative motor-neuron diseases such as amyotrophic lateral sclerosis (ALS), from acute polyneuropathies such as Guillan-Barré syndrome, or from stroke, tumours or trauma that affect the brainstem [2, 3].

Various definitions and sub-categories of LIS allow for some residual voluntary movement [2, 3]. Many people in LIS communicate using eye movements or small muscle movements to give a yes-or-no answer in response to a verbal cue provided by a conversation partner. However, these movements may be very tiring, may weaken or disappear sporadically during the course of a day, and may decline over the course of months or years due to progressive motor-neuron degeneration. Conversation partners may find it difficult, and may vary widely in their ability, to recognize these movements unambiguously.

The current article is the latest in a series in which we aim to optimize, and then translate into everyday use, a simple brain-computer interface (BCI) to address this problem. We aim to develop a communication device that can be cued by a conversation partner to elicit a single reliable yes-or-no answer at a time. We decode the user's intended answer from non-invasive EEG signals.

We have chosen to develop a BCI driven by non-visual stimuli, due to the many problems that paralyzed people may experience regarding eye movement and eye health (see for example [4]). Our previous work [5–7] has shown that effective binary brain-computer interface systems can be driven purely by attention to auditory stimuli. Our design was inspired by the observation of Hillyard et al. [8] that selective attention modulates the early negative (80–100 msec) as well as later positive (300–400 msec) event-related potential components in a dichotic listening task.

We have adopted an approach based on event-related potentials (ERPs) rather than steady-state evoked potentials (SSEPs), because we found ERPs to be more efficient [6]. Auditory SSEPs seem to require longer trials to yield comparable accuracy (see for example [9, 10]).

We have adopted a **streaming** approach in which two simultaneous or interleaved trains of stimuli are presented, and the BCI is driven by the difference in responses between all attended and all unattended stimuli. For interfaces with two target classes, we find this to be more efficient than a **sequential** approach that relies on the difference in responses between standard stimuli and relatively rare targets. For more than two target classes,sequential approaches may allow a potentially larger rate of information transfer due to the fact that the non-target stimuli may be differentiated into multiple classes. This allows the efficient

construction of speller systems, such as the so-called "P300 speller" based on grids of visual stimuli [11] or other spatial arrangements (see for example [12]). It is possible to construct such speller systems using a purely auditory BCI approach [13–16] but so far, such systems impose a high working-memory load, as the subject must memorize, and mentally navigate through, multiple mappings between spatial locations and letters (or groups of letters).

Maximizing intuitiveness, and minimizing cognitive load, is one of the primary motivating factors behind the current study. In our previous studies, our auditory BCIs were driven by abrupt beep stimuli typical of those used to elicit ERPs in psychophysiology research. Since these are devoid of intrinsic meaning, their association with the options "yes" and "no" is arbitrary, and is complicated by the fact that two types of stimuli (standard and target) are associated with each option. We found that experimental subjects and potential users consequently find the system difficult to understand. Many people also reported finding the beeps harsh and mildly unpleasant.

Höhne et al. [17] found that natural stimuli (albeit still standardized, semantically empty syllables) can drive an auditory BCI better than artificial stimuli. Encouraged by this, we wished to know whether it was possible, without loss of performance, to adapt our auditory streaming BCI system to use spoken words instead of harsh, meaningless beeps. The potential advantages of word stimuli are three-fold. First, words can be selected that naturally reflect the options the user can choose (for example, "yes" and "no") making the system much easier to explain to beginning users. Second, the natural stimuli may be more pleasant, as Höhne et al. [17] also reported. Third, voice stimuli are very well segregated by the auditory system into separate perceptual streams, whereas the streaming percept of isolated beeps may more easily break down [18]. If we can improve the streaming quality of the stimuli, it may in future be possible to deliver both streams from the same audio speaker, and thereby remove the reliance on good directional hearing.

Spoken words have variable length and differing temporal distributions of stimulus energy. As a result, we expect that the resulting ERPs may look somewhat different from the textbook waveforms elicited by abrupt unnatural stimuli. And as we shall see, the responses to words also exhibit considerably more variability across subjects. Therefore, before we go any further, it is necessary to test whether we can change from beeps to words without loss of performance. The main laboratory study reported in this paper was a within-subject comparison of our previous stimulus configuration (**Beeps**) against a new stimulus condition involving the words "yes" and "no" (**Words**).

Our study also incorporated various additional aspects that aimed to further the progression of our system towards practical usability. This included a reduction in the number of EEG channels relative to previous studies: here, we use 8 channels to minimize setup time and also so that we could use the 8-channel amplifiers that are part of our standard home BCI system. Furthermore, since EEG signals are highly non-stationary, particularly between sessions, we wished to evaluate the performance of our subjects using fixed classifier weights obtained on a previous session. Therefore our laboratory design included two sessions for every subject. Finally, having established that the Words condition was at least

as effective as the old Beeps condition, we took the first steps towards testing the system (Words condition only) with two potential users who are locked in.

2. Materials and methods

2.1. Subjects

Fourteen healthy subjects, whom we will denote by the letters A–N, took part in the experiment (7 male + 7 female, ages 22–67, or 39 ± 17.8 s.d.). Each subject attended for 2 sessions on separate days. None of them had any history of significant hearing problems or neurological defects. Eleven of the subjects had had previous experience with BCI; 5 of these had had experience with auditory BCI, having been subjects in one of our previous studies [6, 7].

We also tested the system with 2 subjects, whom we refer to as H1 and H3, who had advanced ALS. H1 was a 53-year-old man who had been diagnosed with ALS 14 years before our study and who was dependent on a ventilator. He had previously relied on a visual BCI system for his work for over 4 years, but had stopped using this for free spelling one year before. He could communicate via eye muscle movements which were interpreted by an assistant. H3 was a 64-year-old woman who had been diagnosed with ALS 15 years before our study and who had been dependent on a ventilator for 12 years. She could communicate by using eyebrow movement to make selections on an on-screen keyboard. She had been a participant in our lab's trial of independent home use of a visual BCI, but had not used auditory BCI before.

All subjects gave informed consent. All procedures were approved by the Institutional Review Boards of the New York State Department of Health.

2.2. Hardware and software

EEG recordings were made using a custom 8-channel version of the g.USBamp amplifier (g.tec medical engineering GmbH, Austria) in conjunction with an 8-channel EEG cap (Electrocap Inc.). The cap used gelled 9mm tin electrodes at positions F3, F4, T7, C3, Cz, C4, T8, and Pz of the extended international 10-20 system of Sharbrough et al. [19], with the reference at TP10 (the right mastoid) and the ground electrode at TP9 (the left mastoid). The amplifier performed appropriate anti-alias filtering before digitizing at 24 bits and downsampling to 256 Hz. Data acquisition was performed using the BCI2000 software platform [20, 21] v.3.0; signal-processing, classifier optimization and stimulus presentation were implemented in Python using the 'BCPy2000' add-on to BCI2000 [22]. Stimuli were delivered via the laptop's built-in soundcard to a pair of Sony MDRV600 headphones worn by the subject. The software ran on a Lenovo ThinkPad T61p laptop with a 2.2 Ghz dual-core processor.

2.3. Stimuli and task design

A subject's first session began with a 5-minute pre-recorded audio introduction explaining the experiment. It then consisted of 12 runs: 3 of one condition, 3 of the other condition, 3 more of the first condition, and 3 more of the second. Half the subjects performed the Beeps

condition first and the Words condition second; the other half performed the conditions in the opposite order. Each subject's second session was performed in the opposite order to their first. The pre-recorded instructions were not played in the second session unless the subject asked to hear them.

Each run consisted of 20 trials. Each trial lasted around 15 seconds in total (including a few seconds' rest) and consisted of an attempt to listen to only the stimuli in the left earphone (to select "no") or only the stimuli in the right earphone (to select "yes").

In the Words condition, the left stream consisted of a synthesized male voice saying "*no*" seven times at a rate of 2 per second. Randomly on each trial, 1, 2 or 3 out of the last 5 "*no*" stimuli were instead target stimuli in which the voice said "*nope*". The right stream began 250 msec later than the left stream but also consisted of 7 stimuli at a repetition frequency of 2 per second (so the streams were in constant anti-phase). It was a synthesized female voice saying "*yes*" with 1, 2 or 3 target stimuli ("*yep*"). A third synthesized voice gave a spoken instruction at the beginning of each trial: either "*Listen to 'yes' and count the number of 'yep'*," or "*Listen to 'no' and count the number of 'nope'*." At the end of each trial the voice gave feedback according to the number of targets—for example, "*The correct answer was two*." No overt response was required from the subject. After each trial and before the voice feedback, the subject heard a bell ring if the trial had been classified correctly online.

The Beeps condition was conceptually identical to the Words condition, but the standard stimuli were 150-msec beeps at 512 Hz (left) or 768 Hz (right), and the target stimuli were amplitude-modulated versions of the standard beeps. The stimuli were thus identical to those of Experiment II of Hill et al. [7], where they are described in greater detail. The synthesized vocal cue on each trial was, for example, "*Listen to <LATERALIZED BEEP> to say 'yes'*."

There was no need for subjects to look at a screen. We asked them to keep their eyes still during stimulus presentation by fixating a spot marked on the wall.

Our two subjects with ALS performed one session each, in the Words condition only. The session consisted of 3 runs of 20 cued-decision trials to provide training data for the classifier, followed by a series of natural-language questions. Some of these questions were personal (for example, "Were you born in New York State?") and some were questions to which we assumed the answer would be obvious (for example, "Is today Monday?"). For subject H1, we asked 9 yes-or-no questions to which we did not necessarily know the answer. The subject gave responses to all the questions using the BCI system. We then ascertained the correct answers by asking the questions again, and the subject responded via his normal communication channel (eye movement, interpreted by an assistant). For subject H3, we asked 40 questions based on our personal knowledge of the subject. To allow application of a response-verification analysis [23], the questions were presented in pairs with opposite answers (for example, "Do you have a grandson?" and "Do you have only granddaughters?").

2.4. Online signal-processing and classification

The signal-processing methods were identical to those of Hill et al. [7]. First the signal was band-pass filtered between 0.1 and 8 Hz using a Butterworth filter of order 6. Features were computed by taking the right-left difference of the within-trial averages of 600-msec epochs. We applied symmetric spatial whitening—i.e. premultiplication by the inverse matrix square root of the 8×8 spatial covariance matrix averaged across all training epochs—as recommended and explained by Farquhar and Hill [24]. We then computed classifier weights in the whitened space using L2-regularized linear logistic regression, and transformed the weights from the whitened to the original space as described by Hill and Schölkopf [6]. The BCI output signal, computed as the sum of the feature values weighted by these classifier weights, was updated every time a new epoch became available (roughly 4 times per second) and the sign of its final value was used as the prediction for each trial.

Separate subject-specific classifiers were maintained for the two conditions: one for each subject's Beeps data and one for each subject's Words data. In the first session, classifiers were re-trained after every new run of 20 trials. In the second session, the final classifier weights from the first session were used and kept fixed.

3. Results

3.1. Performance

The main question of our laboratory study was whether there was any significant difference in performance between the Words and the Beeps stimulus conditions. Figure 1 addresses this question by showing the results of the within-subject comparison of the two conditions. Each data-point represents the data from one subject: points that lie above the diagonal indicate better performance in the Words condition than in the Beeps, and points below the diagonal indicate better performance on Beeps than on Words.

On average, Words narrowly beat Beeps. On day 1, the mean and standard deviation of % correct scores across subjects was 76.9% ± 11.1 for Words, and 73.0% ± 10.6 for Beeps. The same was also true on day 2 (70.18% ± 11.9 for Words vs. 66.6% ± 10.4 for Beeps). Neither of these gains was significant at the 5% level in a Wilcoxon signed rank test.

Whether a subject performs better in Words or Beeps depends very much on the individual. Note that the directions of the individuals' respective performance advantages stay consistent from day to day. Of the 14 subjects, only 2 show a reversal of the order of the two conditions' performance in the transition from day 1 to day 2: Subject D performed slightly better with Words on day 1, but slightly better with Beeps on day 2; Subject N shows the opposite pattern. We also note that when subjects perform better with Words, the advantage is often large (for example, on day 1, it is 13.1 percentage points ± 7.5 s.d. across 7 subjects) whereas the subjects who perform better with Beeps do so with a smaller margin (5.4 pp ± 3.5 s.d. across the other 7 subjects).

We also wished to know whether, having trained a classifier on one session, we could expect good performance on future sessions without the necessity to perform additional supervised trials for re-calibration. Performance on day 2, relative to day 1, was 6.7 percentage points

worse in the Words condition on average (Wilcoxon signed rank: $p = 0.005$) and 6.4 points worse in the Beeps condition ($p = 0.048$). Figure 2 examines this effect in greater detail by plotting performance as a function of the number of trials performed in the session. Panel A shows the performance obtained in the online session (using incremental classifier training within the session on day 1, and classifying day-2 data using fixed classifier weights optimized on day-1 data), averaged across all subjects and both conditions. Note that there is a steady decline in performance over the course of the session on day 2. This decline is significant (Spearman's $\rho = -0.25$, $p < 0.001$, $N = 168$). We examined this further by repeating the analysis offline. Panel B shows the results of applying incremental within-session classifier training to both sessions offline. The results for day 1 are essentially unchanged, but now the decline in day-2 performance is eliminated—although without an increase in performance on average across the session. Panel C shows the results of training a classifier on one complete session, then transferring the classifier weights to the other session and keeping them fixed to classify all the trials of that session. Again, we replicate the online results (the steady decline on day 2) but we do not see any such decline in the data from day 1.

Having established that we can adopt a design based on Words stimuli without loss of performance relative to our older Beeps design, we then wished to verify that the Words approach could work for potential users who were locked in. Figure 3 shows the performance results from our two locked-in subjects H1 and H3, with the comparable results from healthy subjects for comparison (Words condition, day 1, sleeping subject M excluded). The performance of H1 and H3 at answering natural-language questions with the BCI (single-hatched bars) is roughly at the same level as the performance of the healthy subjects in performing cued selections, when the questions were considered in isolation from each other. For subject H3, we can increase the performance estimate by asking what would happen if a response verification method had been used. There were 19 valid question pairs (one question pair had to be removed from the analysis because the correct answer turned out to be "yes" to both halves). Of these 19, the subject delivered consistent answers (either "no" followed by "yes", or *vice versa*) to 15 pairs. We received inconsistent answers (either two positive answers or two negative answers) to the other 4 pairs, which would therefore be considered inconclusive results—in a real usage scenario the question would have to be repeated. Thus we could say that the system had an estimated efficiency of 15/19, or 79%. Of the 15 conclusive answers, 13 were correct (87% accuracy, as shown by the cross-hatched bar in Figure 3).[‡]

3.2. Features

Figure 4 shows the patterns that we can see in the responses to Beeps stimuli. Each panel is an 8-channel × 600 msec image of the epoch following stimulus onset. Column (i) is simply the averaged response to all stimuli: left stimulus onset is at time 0 and right stimulus onset is at 250 msec. The colour scale from -4 to $+4$ μV is common to all plots in column (i). The primary sensory response, the N1 at a latency of 100 msec, appears as a dark stripe at 100

[‡]We should note that both subjects scored considerably lower (both 67.5% correct) on the 40 cued selection trials on which they were given online feedback during calibration. It is possible that the subjects found cued selection less motivating than making free selections to answer real questions, and that this affected their performance.

msec (elicited by left stimuli) and at 350 msec (elicited by right stimuli). This feature is very consistent across subjects—even subject M, who was not consistently awake throughout the experiment. For some subjects (F and G) it is possible that the response in the two ears is different due to an difference between the two ears' sensitivity to sound.

Columns (ii)–(iv) of Figure 4 are all contrast plots showing **d-prime** measures of signal-to-noise ratio, i.e. the mean voltage differences divided by standard-deviations of the differences, across stimulus epochs [25]. They share a common colour scale of −0.6 to +0.6σ. The column (ii) shows the difference between standards and targets in the unattended stream, and column (iii) shows the same for the attended stream. In both, the N2a component (known as the mismatch negativity, or MMN [26]) is visible as a dark stripe at around 200–250 msec. To some extent in both, although to a greater extent in the attended stream (column iii) the P3 response is visible as a light feature immediately following the MMN.

Column (iv) contrasts all attended vs. all unattended stimuli (onset at time 0) and is therefore most closely related to BCI performance. Note that this set of features is very variable from subject to subject. Although some subjects show similar patterns to each other (for example, subjects B and D), there are some subjects whose patterns appear to have opposite polarity to each other (for example, D and G). Such individual differences underscore the need for subject-specific classifier training.

The picture looks very different for the Words condition, shown in Figures 5 and 6. The format and colour scaling are identical to those of Figure 4. The subjects in Figure 5 are now reordered according to their performance in the Words condition on day 1. Note that the average response in column (i) is now much more variable from subject to subject, as we mentioned in the Introduction. The MMN-P3 pattern is no longer consistently visible in the unattended stream (column ii) and, although it is visible in the attended stream results (column iii), it is also more variable from subject to subject. Note, however, that despite the increased variability of the primary sensory responses, the attended-vs.-unattended contrast shows *less* inter-individual variation than it did for Beeps. Many of the subjects, particularly the better-performing ones, show the same negative feature 300 msec after stimulus onset (most visible in subjects B, J, K, G and C) although there are still subjects (e.g. D and I) who appear to show the opposite polarity while still performing above chance. Figure 6 shows the features from the two subjects with ALS. The negative feature at 300 msec is visible during H3's free-selection trials (although not during the cued-selection trials, during which performance was lower).

4. Discussion

Our results show that an online binary auditory-streaming BCI can be built with as few as 8 EEG channels, and use single trials in which the critical EEG segment is less than 4.5 seconds long, and still achieve 77% correct on average (93% for the best subject). We also find that there is no significant disadvantage, but rather some non-significant tendency towards improved performance, in switching to more-intuitive natural stimuli (voices saying "*yes*" and "*no*") instead of the abstract, unpleasant beeps used in previous studies.

The finding that natural, intuitive stimuli are at least as effective as abstract stimuli spurred us to conduct a preliminary assessment of the effectiveness of this system when used by people in the locked-in state (LIS). We presented simple yes-or-no questions to two people who could only make small, limited muscle movements because of advanced amyotrophic lateral sclerosis (ALS). Both subjects were able to use the system to answer questions correctly at roughly the same level of accuracy that we observed among healthy volunteer subjects. This is an encouraging sign that the system will translate well to the target user group.

The paradigm was suitable for eliciting and examining a number of well-known ERP components in a relatively short time, and may therefore be valuable for developing a profile of a new user. Such a profile will yield useful information about the user's ability to hear the stimuli, hear equally well with both ears, discriminate differences between stimuli, and follow instructions for attending to one side or another. It may therefore allow preliminary inferences to be made about the subject's state of consciousness and perceptual and cognitive state.

While abrupt beep stimuli produced N1, MMN and P3 ERP components that were very consistent across subjects, the difference between responses to attended and unattended beeps was very variable, even among subjects who performed well. Since this difference drives the BCI, it is clearly very important to tailor classifier weights individually to each subject. By contrast, word stimuli produced less-consistent N1, MMN and P3 responses, but the difference between attended and unattended stimuli was actually *more* consistent across subjects. The most consistent feature of the difference wave was a negative peak around 300 msec after stimulus onset. The energy of each word stimulus was spread out over about 350 msec with a relatively slow attack, rather than concentrated within the first 150 msec with a very sudden attack, as with the beep stimuli. As a result, the effective latency of the crucial negative component should perhaps be considered to be smaller than 300 msec—but, for the same reason, it is difficult to be certain which of the well-known ERP components, if any, most closely corresponds to it.

Finally, we show that there is a significant loss of performance when classifier weights are transferred from session to session, which remains as a challenge for the development of signal-processing algorithms. We had hoped to be able to use these data to develop and examine adaptive semi-supervised methods that might help us to transition from session to session with minimal loss of performance. However, it seems that the current data were unsuitable for an offline examination of this issue. The steady decline in performance over the course of the second session can be eliminated by updating the classifier after each run, but even with supervised re-training, *average* performance on day 2 does not increase. So it seems that our day-2 data are inherently noisier than the day-1 data. This is not purely a result of using fixed classifier weights, since neither the 6–7 percentage-point drop in performance, nor the decline as a function of time, occurs when evaluating day-1 data using fixed weights transferred from day 2 instead of incrementally re-trained weights. We examined trends in classifier bias over time in the incrementally re-trained data, and also the variability of the classifier weights over time, but found no obvious difference between day-1 classifier solutions and day-2 classifier solutions that might explain the differing

character of the two days' results. Rather, we suspect that the drop in performance from day to day, and the decline as a function of time, must result from an online interaction between the use of fixed weights and other factors intrinsic to the subject's state (for example, motivation). It seems that even an early feasibility test of an adaptive semi-supervised algorithm designed to overcome this session-to-session transfer problem may require collection of BCI data performed with the candidate algorithm running online.

While the current results are in the appropriate performance range for comparing accuracy across conditions and avoiding ceiling effects, for future practical use we will want to increase the absolute level of accuracy above 90% for the majority of users. This may potentially be accomplished by lengthening the trials, or ideally by using trials whose length is determined by a dynamic stopping mechanism [27, 28]. Further performance improvements might be expected if future studies investigate the effect of changing the speed of stimulus delivery, and the effect of user training—the results of Hill and Schölkopf [6] suggest that the latter in particular may be possible because BCI performance is predicted by subjects' accuracy in giving an overt behavioural response to the target-counting task.

We conclude that our paradigm is a promising candidate for a simple, practical BCI system that could be used by a locked-in person to communicate "yes" or "no" via an intuitive non-visual interface during a conversation with a human partner. The value of such non-visual approaches may go beyond application to people who cannot see. One of our subjects with ALS, who can see a computer screen well and is able to use an on-screen keyboard via eyebrow movement, welcomed this novel non-visual access method, telling us, "My eyes get tired, but never my ears." This suggests that the approach may provide an attractive ergonomic alternative for some tasks, even for users who can see well.

Acknowledgments

References

[1]. Plum, F.; Posner, JB. The Diagnosis of Stupor and Coma. 1st edition. Philadelphia, PA, USA; FA Davis Co.: 1966.

[2]. Bauer G, Gerstenbrand F, Rumpl E. Varieties of the locked-in syndrome. Journal of Neurology. 1979; 221(2):77–91. [PubMed: 92545]

[3]. Hochberg, LR.; Anderson, KD. BCI Users and their Needs. In: Wolpaw, JR.; Wolpaw, EW., editors. Brain-Computer Interfaces: Principles and Practice. 1st edition. Oxford University Press; Oxford, UK: 2012. p. 317-323.chapter 19

[4]. Sharma R, Hicks S, Berna CM, Kennard C, Talbot K, Turner MR. Oculomotor Dysfunction in Amyotrophic Lateral Sclerosis. Archives of Neurology. 2011; 68(7):857–861. [PubMed: 21747027]

[5]. Hill NJ, Lal TN, Bierig K, Birbaumer N, Schölkopf B. Saul LK, Weiss Y, Bottou L. An Auditory Paradigm for Brain-Computer Interfaces. Advances in Neural Information Processing Systems. 2005; ume 17:569–576.

[6]. Hill NJ, Schölkopf B. An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. Journal of Neural Engineering. 2012; 9(2):026011. [PubMed: 22333135]

[7]. Hill NJ, Moinuddin A, Häuser A-K, Kienzle S, Schalk G. Communication and Control by Listening: Toward Optimal Design of a Two-Class Auditory Streaming Brain-Computer Interface. Frontiers in Neuroscience. 2012; 6:181. [PubMed: 23267312]

[8]. Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical Signs of Selective Attention in the Human Brain. Science. 1973; 182(4108):177–180. [PubMed: 4730062]

[9]. Kim D-W, Hwang H-J, Lim J-H, Lee Y-H, Jung K-Y, Im C-H. Classification of Selective Attention to Auditory Stimuli: Toward Vision-free Brain-Computer Interfacing. Journal of Neuroscience Methods. 2011; 197(1):180–185. [PubMed: 21335029]

[10]. Lopez M-A, Pomares H, Pelayo F, Urquiza J, Perez J. Evidences of cognitive effects over auditory steady-state responses by means of artificial neural networks and its use in braincomputer interfaces. Neurocomputing. 2009; 72(16-18):3617–3623.

[11]. Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology. 1988; 70(6):510–523. [PubMed: 2461285]

[12]. Williamson J, Murray-Smith R, Blankertz B, Krauledat M, Muller K-R. Designing for uncertain, asymmetric control: interaction design for brain-computer interfaces. International Journal of Human-Computer Studies. 2009; 67(10):827–841.

[13]. Furdea A, Halder S, Krusienski D, Bross D, Nijboer F, Birbaumer N, Kübler A. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology. 2009; 46(3): 617–625. [PubMed: 19170946]

[14]. Klobassa DS, Vaughan TM, Brunner P, Schwartz NE, Wolpaw JR, Neuper C, Sellers EW. Toward a high-throughput auditory P300-based brain-computer interface. Clinical Neurophysiology. 2009; 120(7):1252–61. [PubMed: 19574091]

[15]. Schreuder M, Rost T, Tangermann M. Listen, You are Writing! Speeding up Online Spelling with a Dynamic Auditory BCI. Frontiers in Neuroscience. 2011; 5(112):112. [PubMed: 22016719]

[16]. Höhne J, Schreuder M, Blankertz B, Tangermann M. A novel 9-class auditory ERP paradigm driving a predictive text entry system. Frontiers in Neuroprosthetics. 2011; 5:99.

[17]. Höhne J, Krenzlin K, Dähne S, Tangermann M. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. Journal of Neural Engineering. 2012; 9(4):045003. [PubMed: 22831919]

[18]. Bregman, AS. Auditory Scene Analysis. MIT Press; Cambridge, MA, USA: 1994.

[19]. Sharbrough F, Chatrian GE, Lesser PR, Lüders H, Nuwer M, Picton TW. American Electroencephalographic Society guidelines for standard electrode position nomenclature. Journal of Clinical Neurophysiology. 1991; 8(2):200–202. [PubMed: 2050819]

[20]. Schalk G, McFarland D, Hinterberger T, Birbaumer N, Wolpaw J. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Transactions on Biomedical Engineering. 2004; 51(6):1034–1043. [PubMed: 15188875]

[21]. Schalk, G.; Mellinger, J. A Practical Guide to Brain-Computer Interfacing with BCI2000. Springer-Verlag London Limited; 2010.

[22]. Hill NJ, Schreiner T, Puzicha C, Farquhar J. BCPy2000. 2007 http://bci2000.org/downloads/BCPy2000.

[23]. Wolpaw JR, Ramoser H, McFarland DJ, Pfurtscheller G. EEG-based communication: improved accuracy by response verification. IEEE Transactions on Rehabilitation Engineering. 1998; 6(3): 326–33. [PubMed: 9749910]

[24]. Farquhar J, Hill NJ. Interactions Between Pre-Processing and Classification Methods for Event-Related-Potential Classification : Best-Practice Guidelines for Brain-Computer Interfacing. Neuroinformatics. 2012; 11(2):175–192. [PubMed: 23250668]

[25]. Simpson AJ, Fitter MJ. What is the best index of detectability? Psychological Bulletin. 1973; 80(6):481.

[26]. Näätänen R. The mismatch negativity: a powerful tool for cognitive neuroscience. Ear & Hearing. 1995; 16(1):6–18. [PubMed: 7774770]

[27]. Schreuder M, Höhne J, Blankertz B, Haufe S, Dickhaus T, Tangermann M. Optimizing event-related potential based braincomputer interfaces: a systematic evaluation of dynamic stopping methods. Journal of Neural Engineering. 2013; 10(3):036025. [PubMed: 23685458]

[28]. Lopez-Gordo MA, Pelayo F, Prieto A, Fernandez E. An Auditory Brain-Computer Interface With Accuracy Prediction. International Journal of Neural Systems. 2012; 22(03):1250009–1. [PubMed: 23627625]
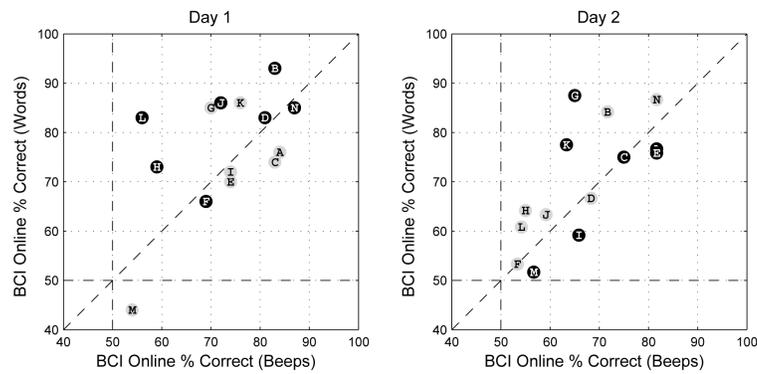
Figure 1.

The left panel shows the results from 100 online BCI trials in each subject's first session. The right panel shows the results from 120 online trials in each subject's second session, when transferred classifier weights were used on a later day. The online BCI system's accuracy (% correct) in the Words condition is plotted against its accuracy in the Beeps condition. Chance performance, marked by the dashed horizontal and vertical lines, was 50% in both cases. Subject identity is coded by letters A–N. Letters in black-on-grey indicate that Beeps were performed first on a given day, whereas letters in white-on-black indicate that Words were performed first that day. On day 2, the data-point for Subject A is mostly obscured by that of Subject E. One subject (M) had to be woken up repeatedly in both sessions, and performed near chance in both stimulus conditions on both days.
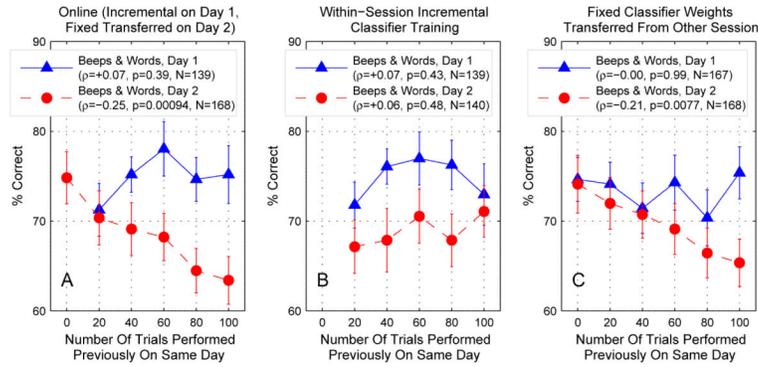
Figure 2.
Each panel shows BCI classification performance, averaged across all subjects and across both stimulus conditions, as a function of time (measured in trials) during the session. Triangles denote performance on the day-1 data and circles denote performance on the day-2 data. Error bars show ±1 standard error, across subjects and conditions, of the percentage accuracy scores. In panel A, we show performance that was obtained in the online session: this meant that the classifier was re-trained after every run of 20 trials on day 1, but then the day-1 classifier weights were fixed and used unchanged throughout day 2. In panel B, we re-create the day-1 method offline for both days—i.e. we show offline classification performance on each run, having trained the classifier on all runs performed so far within the same session. In panel C, we re-create the day-2 method offline for both days—i.e. we show offline classification performance on each run, having trained the classifier on all the data from the *other* day. The legends show the non-parametric correlation coefficient (Spearman's ρ) and its significance value ρ for each of the lines.
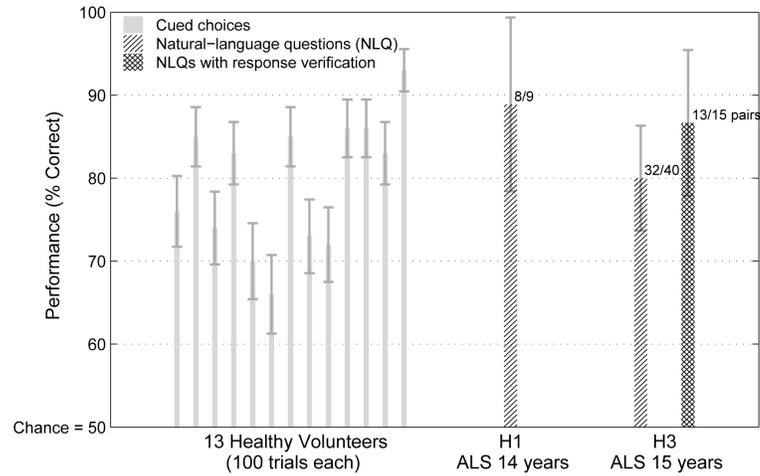
Figure 3.

This figure shows online BCI classification accuracy in the Words condition from both healthy subjects and the two subjects with ALS. The solid grey bars on the left show the performance of subjects A-N (this time omitting the sleeping subject M) in the Words condition on day 1. The hatched bars show the proportions of natural-language questions answered correctly by subjects H1 and H3. The cross-hatched bar shows the performance for H3 when the answer to each question was inferred from the responses to both its positively and negatively phrased version. Error bars indicate ±1 binomial standard error for each dataset.
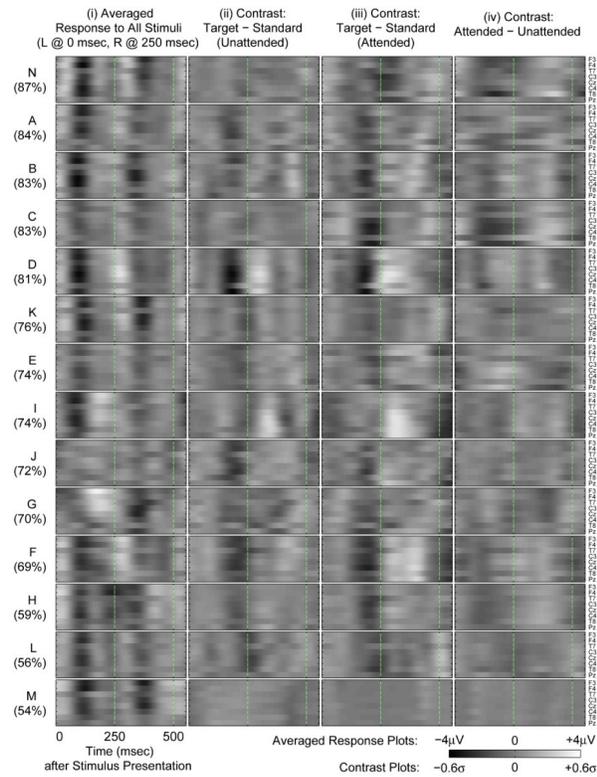
Figure 4.
Each panel shows an 8-channel × 600 msec image of the epoch following the onset of stimuli in the Beeps condition (cf. Figure 5 for the corresponding Words results). Each row of panels corresponds to one subject—they are ordered from top to bottom according to their online performance in the Beeps condition on day 1, which is given in % on the left of the figure. The four columns show four different representations—see text, Section 3.2, for details.
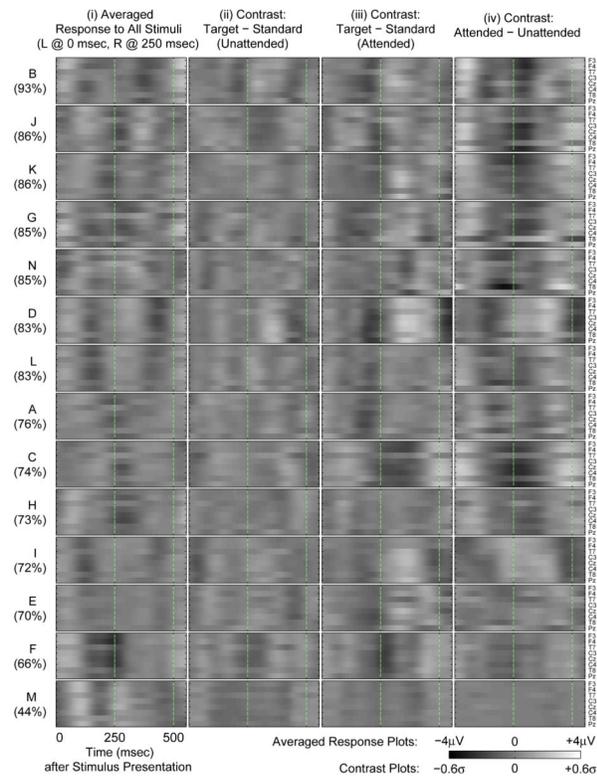
Figure 5.
Each panel shows an 8-channel × 600 msec image of the epoch following the onset of stimuli in the Words condition (cf. Figure 4 for the corresponding Beeps results). Each row of panels corresponds to one subject—they are ordered from top to bottom according to their online performance in the Words condition on day 1, which is given in % on the left of the figure. The four columns show four different representations—see text, Section 3.2, for details.
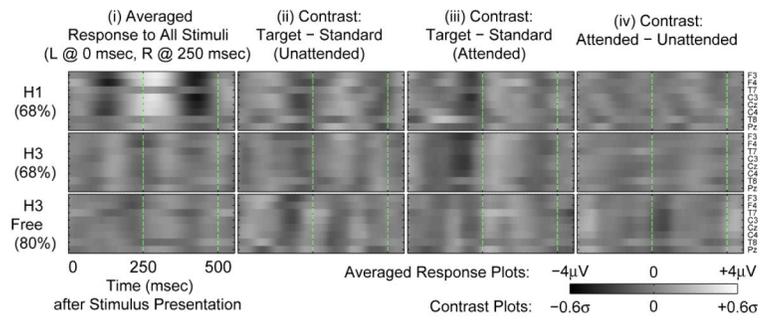
Figure 6.

This figure shows ERPs in the same format and colour scaling as for Figure 5, but for the subjects with ALS. The first two rows show the online performance of H1 and H3 during the 40 cued trials during which classifier weights were loaded. The last row shows the results from H3 during free selection (answering real questions).