An Improved Auditory Streaming BCI with Voice Stimuli

E. Ricci[1], S. Haider[1], T. M. Vaughan[1], N. J. Hill[1]
[1]*Wadsworth Center, Albany, NY, USA*

Correspondence: N. J. Hill, Wadsworth Center, C640 Empire State Plaza Albany, NY 12201, USA. E-mail: jezhill@gmail.com

Abstract. We have previously shown that it is possible to build EEG brain-computer systems based on voluntary shifts of covert attention between simultaneous streams of auditory stimuli. The system exploits not only the later event-related potential components (P3) which are strongest in response to rare "oddball" stimuli, but also the early (N1 or Nd) components that are attention-modulated on every stimulus occurrence. So far, however, such systems have been based on abrupt artificial stimuli (short discrete beeps or pulses). This creates two problems. First, many subjects find the stimuli annoying, intrusive or otherwise unpleasant. Second, the abstract nature of the stimuli makes the system unintuitive to many users. We would like to build a system in which the stimuli are natural and also semantically indicative of the purpose of the corresponding interface selection. This greatly simplifies the instructions for the users: when they want to say "yes" using the BCI, they must listen to a voice repeating the word "yes", and when they want to say "no", they listen to a voice repeating the word "no". Encouraged by recent findings that natural stimuli can improve auditory BCI performance, we set out to assess, in a within-subject design, whether these new voice stimuli or the beep stimuli of previous studies allow better performance in an 8-channel EEG-based BCI. Despite increased between-subject variability in the ERPs, we find no significant penalty (in fact, a non-significant advantage) for the voice stimuli (mean ± std online performance 76% ± 11) in comparison with the beep stimuli (73% ± 11). We also show, for the first time, that the system can be used on a subsequent day without retraining the classifier. However, without applying transfer learning methods to the problem, this entails a large, significant drop in classification performance (7–10 percentage points on average).

Keywords: EEG, auditory ERPs, attention, auditory streaming BCI, N100, P300, dichotic listening, natural stimuli

1. Introduction

Our previous work [Hill and Schölkopf, 2012; Hill et al., 2012] has shown that binary EEG BCIs can be driven purely by attention to auditory stimuli, which we believe will be an advantage for severely paralyzed users for whom spatial vision is often poor. Höhne et al. [2012] 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 beeps' lack of inherent meaning has previously made it difficult to explain the use of the BCI to beginning users, so we wanted to move to stimuli that naturally reflect the options being chosen (spoken words "yes" and "no"). To assess the impact of this change, we ran a laboratory experiment to compare a Beeps condition with a Words condition in a within-subject design. In the same experiment, we also wished to assess, for the first time, the impact of session-to-session transfer on auditory BCI performance.

2. Material and Methods

14 healthy subjects took part in the experiment (7 male, 7 female, ages 22–67, or 39 ± 17.8). Each subject attended for 2 sessions on separate days. With a reference electrode at the right mastoid (TP10) and a ground electrode at the left mastoid (TP9) we recorded EEG from 8 channels of the extended international 10-20 system (F3, F4, T7, C3, Cz, C4, T8, and Pz) using a g.USBamp (g.tec, Graz, Austria) sampling and digitizing at 256 Hz. Signal acquisition, processing and stimulus presentation were implemented using BCI2000 and BCPy2000.

Each 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 run consisted of 20 trials. Each trial lasted around 15 s 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 say "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 ms 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: "*The correct answer was {one|two|three}.*" No overt response was required from the subject. In the Beeps condition, the standard stimuli were 150-ms beeps at 512 Hz (left) or 768 Hz (right), and the target beeps were amplitude-modulated versions of the standards: the stimuli were thus identical to those of Experiment II of [Hill et al., 2012]. The synthesized vocal cue on each trial was *"Listen to <LATERALIZED BEEP> to say {yes|no}"*.

The signal-processing methods were identical to those of [Hill et al., 2012]. Separate subject-specific classifiers were maintained for Beeps and for Words. In the first session, the classifier was 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. After each trial and before the voice feedback, the subject heard a bell ring if the trial had been classified correctly online.

3. Results

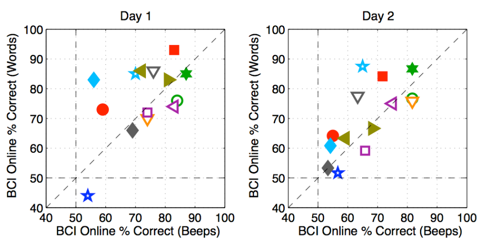

Figure 1. *The left panel shows the results from each subject's first session, and the right shows the results from each subject's second session, when transferred classifier weights were used. Each different symbol shape/color represents 100 online trials from a different subject. The online BCI system's % correct in the Words condition is plotted against its % correct in the Beeps condition. One subject (dark blue star) had to be repeatedly woken up, and performed near chance on both days.*

Fig. 1 shows the results of the within-subject comparison of Words vs Beeps. Words narrowly beat Beeps on day 1 (76.9% ± 11.1 vs. 73.0% ± 10.6) and also on day 2 (70.18% ± 11.9 vs. 66.6% ± 10.4). 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 on the individual. We note, however, that when subjects prefer Words, the preference is often large (for example, on day 1, it is 13.1 percentage points ± 7.5 across 7 subjects) whereas a preference for Beeps tends to be smaller (5.4 pp ± 3.5 across the other 7 subjects).

With 100 trials and Bonferroni correction for 56 simultaneous comparisons, under the null hypothesis that results are binomially distributed with generating probability 0.5, performance must be > 66% to be significantly above chance at the $\alpha = 0.05$ level. On their first day, two subjects (one awake, one asleep) failed to exceed this level with Words; with Beeps, 3 subjects failed. The transfer of weights from day 1 to day 2 caused a 6.7 percentage-point drop in Words' performance on average (Wilcoxon signed rank: $p = 0.005$) and a 6.4 point drop in Beeps' performance ($p = 0.048$). Now 6 subjects fail to reach the threshold with Words, and 8 fail with Beeps.

4. Discussion and Summary

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 s 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 abstract, unpleasant beeps). Finally, we show that there is a large and significant loss of performance when classifier weights are transferred from session to session.

Acknowledgements

Supported by NIH (EB006356).

References

Hill NJ, Schölkopf B. An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. *J Neural Eng,* 9(2):026011, 2012.

Hill NJ, Moinuddin A, Kienzle S, Häuser A-K, Schalk G. Communication and control by listening: toward optimal design of a two-class auditory streaming brain-computer interface. *Front Neurosci,* 6:181, 2012.

Höhne J, Krenzlin K, Dähne S, Tangermann M. Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *J Neural Eng*, 9(4):045003, 2012.