Is an auditory P300-based Brain-Computer Musical Interface feasible?

Zacharias Vamvakousis[1] and Rafael Ramirez[1]

Music Technology Group, Pompeu Fabra University
`zacharias.vamvakousis@upf.edu`

Abstract. One of the most robust ways of building a voluntarily controlled BCI that wouldnt require almost any training on behalf of the user, is through the P300 potential. In this study we implemented and evaluated a simple auditory P300-based musical interface. The interface consists of a dynamic 5-class auditory oddball paradigm. Each stimulus is tuned to a specific note, forming a part of a chord. At the same time each stimulus is mapped to a different chord. The user can change the harmony of the auditory oddball paradigm by mentally counting the occurrences of the desired stimulus. The interface was evaluated on 8 healthy subjects on two scenarios: (i) the audiovisual and (ii) the auditory. The Inter-Onset Interval was set to 220ms and within a trial each stimulus appeared 20 times. As a result a user is able to change the harmony of the interface every 22 seconds. Both scenarios were evaluated on 8 healthy subjects. In the audiovisual scenario the average online accuracy was 80% and in the auditory 44%. According to these results an auditory only P300 based Brain-Computer Musical interface is hardly feasible.

Keywords: P300, auditory oddball, BCMI

1 Introduction

A Brain-Computer Interface (BCI) works by capturing the users brain activity and converting it to meaningful information in order to control a computer. Most BCIs are built using the the electroencephalogram (EEG). An EEG device captures the electromagnetic activity of the brains cortex, using electrodes in touch with the skin of the users scalp. The target group that could benefit the most from the development of BCIs is this of people with severe physical disabilities, such as patients with locked-in syndrome.

Using existing BCI applications someone can perform various tasks, such as controlling a wheel chair, writing, drawing, browsing the internet, playing computer games or controlling musical parameters of an interface [1]. Several Brain-Computer Interfaces for controlling musical parameters have been proposed in previous research. The first proposed musical Interface was Music for Solo Performer [2], in the 1960s by Alvin Lucier. The amplified EEG signals were driven to loudspeakers. The vibrations caused were triggering sounds though a set of percussive instruments attached in the loudspeakers. This case though, is better described as a sonification of the brain activity, rather than a BCI. An attention-based BCI was first proposed by David Rosenboom [3]. In this interface EEG components, related with the shifts in the selective attention of the user, were introduced as parameters in a generative music system. It is uncertain though whether the features used were indeed related to the users selective attention. Many approaches propose the direct mapping of certain EEG bands to musical parameters [4, 5]. In these approaches though, the amount of control the user has over the interface is questionable. It would require extensive training for a user to be able to manipulate his own brains activity. The limits between a biofeedback interface and an interface where the user voluntarily controls its functions are not always clear.

Probably the most robust way of building a voluntarily controlled BCI that wouldnt require almost any training on behalf of the user, is through the P300 potential. The P300 potential is a positive deflection of the captured electromagnetic activity, 300ms after a rare or unexpected event is perceived, centred around the vertex of the cortex and spread all over the cortex. In a multi-class P300-based BCI, a number of stimuli are presented to the user in a random order and the user draws his attention to a specific stimulus (usually by mentally counting its occurrences). After a number of repetitions of each stimulus, the system is able to predict on which stimulus the user was focusing on. The nature of the stimulus might be visual, auditory, tactile or combination of these. By altering his attention to different stimulus the user is able to perform different actions.

The most well-known P300-based multi-class BCI is the P300 speller proposed in 1988 by Farwell and Donchin [6]. In the typical P300-speller paradigm the user stares at a screen where the characters are placed on a grid. As the characters are flushing in a random order, the user focuses on the character he/she wants to spell. Every time the attended character flashes, a P300 potential is generated. After a number of repetitions, the character that causes the stronger P300 peaks is classified by the system as the attended character. Implementations of the P300-speller have also been proposed using auditory instead of visual stimuli [7, 8].

Apart from typing, a big variety of P300-based BCIs - targeted mainly for locked-in patients- has been proposed, such as controlling the mouse cursor [9], controlling an internet browser [10], controlling a wheelchair [11], painting [12], or controlling musical interfaces [13, 14]. In ICMC 2008 [13], it was presented a P300-based BCI where the user selects the midi-note number placed on a grid, in a similar way a user spells letters in the P300 speller. The maximum speed achieved among 5 subjects was one note every 7 seconds.

Another P300 based BCI proposed [14], integrates the idea of the P300 speller in a music 8x8 step sequencer. The notes of the sequencer are flashing in a random order, and the user selects them as he/she would select letter in the speller. At the same time the melody produced by the sequencer is played back.

These last two proposed interfaces use visual stimuli for controlling the musical interface. In a previous study [18] we implemented an audiovisual P300-based Brain-Computer Musical Interface (BCMI) and did a preliminary evaluation. The average selection accuracy was 83%. In the current paper we implemented a similar interface and evaluated on more subjects in both the audiovisual and auditory modality.

2 Materials and Methods

2.1 Materials

The Enobio 8 wireless electrophysiology sensor system was used for recording the brain signals. The signal processing and classification process were performed using OpenVibe software [15]. Using the VRPN server object, stimulations are sent from OpenVibe to a c++ application implemented in openframeworks toolkit. The openframeworks application was used to visualize the interface and send midi messages through LoopBe 3 virtual midi port to propellerhead Reason 5.0 for sonifying a synthesizer. The system was tried on a laptop with a 2.53GHz i5 460M processor with 4GB of RAM running windows 7 OS, using the laptops internal Realtek ALC269 sound card. The resulting latency of the sound stimuli was 46ms and the standard deviation 4.38 ms. A pair of Sennheiser CX150 In-Ear Headphones was used for auditory feedback.

2.2 The Interface

The interface consists of a dynamically changing 5-class P300 oddball paradigm. Two different scenarios were implemented: the audiovisual and the auditory. Table 2.2 summarizes the timbre of each stimulus (which is the musical instrument that generates its sound), along with its stereo panning and the pitch. The pitch of each stimulus varies, depending on which is the currently selected chord.

	Harp/-90°	Harpsichord/-45°	Harp/0°	Harpsichord/45°	Harp/90°
F	a3	c4	f4	a4	c5
C	g3	c4	e4	g4	c5
G	g3	b3	d4	g4	b4
Dm	a3	d4	f4	a4	d5
Am	a3	c4	e4	a4	c5

Table 1. The notes corresponding to each chord. The name of the note is followed by its octave. For example a3 corresponds to the note 'a' at the 3rd octave. The musical instrument used and the stereo panning of each stimulus are also shown.

During a trial all stimuli appear in a random order, 20 times each. The Inter Onset Interval (IOI) (time interval between the onset of two consecutive stimuli) was set to 220ms and the duration of each stimulus is 80ms. At least one stimulus interferes between two occurrences of the same stimulus. Each stimulus is mapped to a different chord. In an order from left to right in the stereo panning position, the stimulus are mapped to the following chords: F major, C major , G major , D minor and A minor. If during a trial the user wants to select one of the chords, he/she has to switch his attention to the corresponding stimulus and mentally count its occurrences. At the end of each trial the system detects which was the attended stimulus and tunes all stimuli to the corresponding chord. There is no time interval between two trials. Instead, after switching to the selected chord, each stimulus appears three time, with an order from left to right. This helps the user to spot the next stimulus he/she wishes to attend.

In front of the user there is a computer screen showing the available chords to be selected. The chords are placed in the same order as the corresponding stimuli (see figure 1). The currently selected chord is marked in red. Just in the case of the audiovisual scenario, the chord names are flashing when the corresponding stimulus sounds.

In the stimuli design there are 3 discriminating cues: timbre, pitch and stereo spatialization. From left to right, the sound of a harp alternates with a sound of a harpsichord in a way that two neighboring sounds do not share the same timbre. Both of the selected instruments are stringed and have similar timbre. This was preferred to preserve the musicality of the interface.

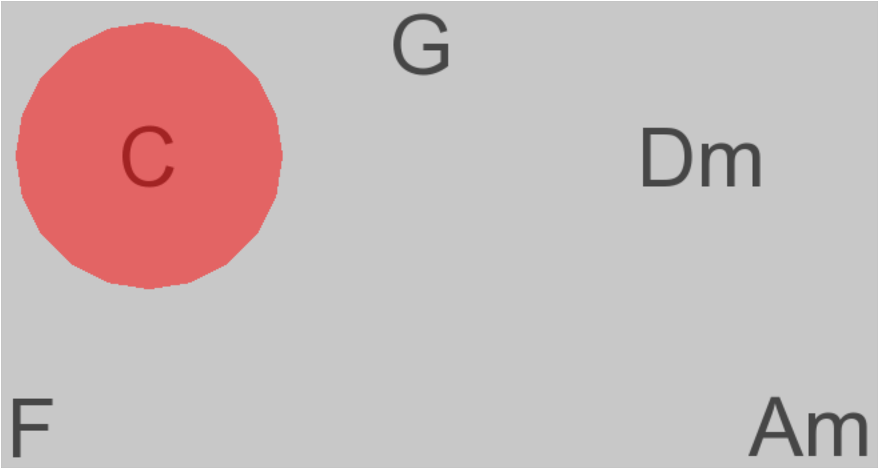

Fig. 1. The visual feedback. In the case of the audiovisual scenario the chord names flash when the corresponding stimulus sounds. The red circle indicates which is the currently selected chord.

2.3 Classification Process

Before using the interface in the online session, a training session is performed in order to acquire data to train a two class LDA classifier. During both the training and the online session the user is comfortably seated in a chair, in front of a screen. He/she is asked to remain still and avoid as much as possible swallowing or moving any facial muscles. During the training session an arrow is pointing at the stimulus the user should attend. The training session consists of 10 trials, in which each stimulus is set as a target twice.

An epoch consists of the 8-channel recording of the time interval 150 to 750ms after the presentation of a stimulus. The signal is downsampled to 50Hz and band-pass filtered to 1-12Hz (Butterworth 4rd order). Using the xDAWN [16] Spatial Filter Trainer in Openvibe, a 8 to 3 channels spatial filter is acquired. The resulting features per epoch are then used to train a two-class Linear Discriminate Analysis Classifier (LDA) to distinguish target from non-target epochs. Once the spatial filter and the LDA classifier parameters are acquired, the user might start using the interface.

During the on-line session the signal is processed as in the training session. Then, for each stimulus a voting classifier computes the sum of the hyperplane distances -given by the LDA classifiers-, and outputs as the attended stimulus the one with the lowest sum.

2.4 Evaluation and Results

The proposed interface was evaluated on 8 healthy subjects (5 male, mean age 29 years, standard deviation 5.46). All subjects gave oral consent to participate in the study. On each user it was first evaluated the audiovisual and then the auditory scenario. For each scenario a 10-trials training session was followed by a 15-trials online session. In order to evaluate the accuracy of the system, the users were asked to select all the chords with order from left to right as shown in figure 1. As the training session consisted of 10 trials, each stimulus was set as target in 2 trials in that case, while it was set as target 3 times in the online session. An arrow was always pointing to the stimulus to be attended. The first two repetitions of each stimulus were not taken into account during the classification process. The reason for that is that presumably it required some time for the users to spot the desired stimulus.

Table 2.4 summarizes the selection accuracy and 10-cross fold validation of all users. The average selection accuracy for the audiovisual scenario is 80% while for the auditory scenario it is 44%.

3 Discussion

In this study we implemented and evaluated an P300-based BCMI and evaluated it in two scenarios: the audiovisual and the auditory. In our paradigm the user is able to change the harmony of the sound stimuli by switching his attention to

User	Audiovisual		Auditory	
	10-fold Cross Validation	Accuracy out of 15	10-fold Cross Validation	Accuracy out of 15
M32	65.90%	8	63.10%	12
M27	71.00%	14	64.70%	5
M34	66.90%	14	62.70%	7
M38	65.30%	8	59.80%	5
F20	70.50%	12	58.30%	8
F27a	66.20%	11	61.10%	7
M28	77.50%	15	60.90%	3
F27b	70.10%	14	57.30%	6
AVG	69.18%	12.00 (80%)	60.99%	6.63(44%)

Table 2. The 10-fold Cross Validation and selection accuracy for all users in both the audiovisual and auditory scenario.

each one of them. This is a special case of an auditory oddball paradigm, where the properties of the stimuli change according to the selections of the user. This probably makes the oddball task even more difficult. In the audiovisual scenario, where the users also make use of the visual cue, the selection accuracy is 80%. When the visual cue is removed, the selection accuracy falls to 44%.

In the audiovisual scenario all users reported that they counted all 20 occurrences of the attended stimulus in every trial. On the contrary in the case of the auditory scenario several subject reported that in some of the trials they were not able to count all occurrences of the attended stimulus. Subject F27a reported that she was confusing the 3rd with the 4th stimulus (as they are placed in the stereo panning from left to right). The same was reported by subject M28. Subject F27b reported that was counting about 18 occurrences instead of 20 in all the trials. Subject M27 could not count correctly the 3rd stimulus in one of the trials. The rest of the subjects reported that they were able to count 20 trials of the target stimulus in all trials. These subjects happen to have musical training. The auditory scenario was more difficult for all subjects, apart from the subjects with musical training (M32, M38, M34 and F20). The difficulty of the subjects with no musical training to spot the stimulus in the auditory scenario is also reflected in their selection accuracy. “Musicians” average 50% while “non-musicians” 33.3%. Subject M32 was the only subject that had previous experience with auditory P300-based BCIs. He was the only one that achieved a high selection accuracy. This indicates that training might be crucial in such interfaces.

The lower performance of auditory P300-based BCIs compared to the visual ones is known by previous studies. In the current study though we observe an even higher difference. This could be explained by the fact that the sound stimulus are dynamically changing according to the selections of the user. Another explanation is that the musicality of the interface was taken into account in stimuli design. The musical interval between to neighbor stimulus was a 3rd or 4th, while all stimuli are always harmonic with each other. Dissonant notes might have been easier to distinguish. Another parameter is the selected timbre of the stimuli. Instead of using a different musical instrument for each stimulus, we only selected two different instruments with similar timbre in order to preserve the musicality of the interface.

A way to improve the system's accuracy could by increasing the number of repetitions of the oddball paradigm. This would make the interaction slower. Twenty-two seconds per selection is already a big time interval. Given the results of the current study, implementing an auditory P300-based BCMI is hardly feasible.

Acknowledgments This work has been partly sponsored by the Spanish TIN project TIMUL (TIN2013-48152-C2-2-R).

References

1. Blankertz, B., Tangermann, M., & Mller, K. R. (2012). 23— BCI APPLICATIONS FOR THE GENERAL POPULATION. Brain-Computer Interfaces: Principles and Practice, 363.
2. Lucier, A. (1976). Statement on: music for solo performer. Biofeedback and the Arts, Results of Early Experiments. Vancouver: Aesthetic Research Center of Canada Publications, 60-61.
3. Rosenboom, D. (1990). The performing brain. Computer Music Journal, 48-66.
4. Hinterberger, T., & Baier, G. (2005). Poser: Parametric orchestral sonification of eeg in real-time for the self- of brain states. IEEE Trans. Multimedia, 12, 70.
5. Mealla, S., Vljame, A., Bosi, M., & Jord, S. (2011). Listening to your brain: Implicit interaction in collaborative music performances. In The International Conference on New Interfaces for Musical Expression (NIME).
6. Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and clinical Neurophysiology, 70(6), 510-523.
7. Schreuder, M., Rost, T., & Tangermann, M. (2011). Listen, you are writing! Speeding up online spelling with a dynamic auditory BCI. Frontiers in neuroscience, 5.
8. Kthner, I., Ruf, C. A., Pasqualotto, E., Braun, C., Birbaumer, N., & Halder, S. (2013). A portable auditory P300 braincomputer interface with directional cues. Clinical neurophysiology, 124(2), 327-338.
9. Citi, L., Poli, R., Cinel, C., & Sepulveda, F. (2008). P300-based BCI mouse with genetically-optimized analogue control. Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 16(1), 51-61.
10. Mugler, E. M., Ruf, C. A., Halder, S., Bensch, M., & Kubler, A. (2010). Design and implementation of a P300-based brain-computer interface for controlling an internet browser. Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 18(6), 599-609.
11. Rebsamen, B., Teo, C. L., Zeng, Q., Ang, V. M. H., Burdet, E., Guan, C., ... & Laugier, C. (2007). Controlling a wheelchair indoors using thought. Intelligent Systems, IEEE, 22(2), 18-24.

12. Van de Laar, B., Brugman, I., Nijboer, F., Poel, M., & Nijholt, A. (2013, February). BrainBrush, a multimodal application for creative expressivity. In ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions (pp. 62-67).
13. Grierson, M. (2008, July). Composing with brainwaves: minimal trial P300b recognition as an indication of subjective preference for the control of a musical instrument. In Proceedings of International Cryogenic Materials Conference (ICMC08).
14. Chew, Y. C. D., & Caspary, E. (2011, November). MusEEGk: Design of a BCMI. In Proceedings of the 8th ACM conference on Creativity and cognition (pp. 325-326). ACM.
15. Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., ... & Lcuyer, A. (2010). OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. Presence: teleoperators and virtual environments, 19(1), 35-53.
16. Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to braincomputer interface. Biomedical Engineering, IEEE Transactions on, 56(8), 2035-2043.
17. Vamvakousis, Z., & Ramirez, R. A High-Throughput Auditory P300 Interface for Everyone. 12th European AAATE Conference, IOS Press. Portugal, Volamoura: IOS Press, 19/09/2013 2013, pp. 478482.
18. Vamvakousis, Z., & Ramirez, R. (2014). P300 Harmonies: A Brain-Computer Musical Interface. ICMC— SMC— 2014.