Proceedings of the
8th Graz Brain-Computer Interface Conference 2019 DOI: 10.3217/978-3-85125-682-6-44

EVALUATION OF AUDITORY BCI SYSTEM BASED ON STREAM SEGREGATION

S. Kanoh [1,2], S. Kojima [2]

[1] College of Engineering, Shibaura Institute of Technology, Tokyo, Japan
[2] Graduate School of Engineering, Shibaura Institute of Technology, Tokyo, Japan

E-mail: kanoh@shibaura-it.ac.jp

ABSTRACT: The auditory brain-computer interface (BCI) system based on auditory stream segregation was tested. In this system, three oddball sequences consisted of musical tones at different frequency bands were presented alternately to subject's right ear. Stimulus parameters were set that subjects could perceive these three oddball sequences as streams (stream segregation). Subjects were requested to pay attention to one of the streams during experiments. It was shown that the P300 was elicited only by the deviant stimuli in the attended target stream in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

INTRODUCTION

The brain-computer interface (BCI) system based on selective attention to specific sensory input has attracted interests. The selective attention to the sensory stimulus of user's interests is reflected by the P300, one of the event-related potential (ERP) components of human EEG. The P300 and an earlier ERP component, mismatch negativity (MMN), can be elicited when user perceives an oddball sequence, on which frequently presented stimuli (standard stimuli) are presented while some of which were randomly replaced by infrequently presented stimuli (deviant stimuli). Such a P300-based BCI has been practically applied as a virtual keyboard [1] or other types of menu selection systems [2].

On almost all of the P300-based BCIs, visual domain has been used. Visual stimulation contains rich information comparing to other sensory domains and it can be easily controlled and delivered to users. Another reason is that it is easy to pay attention to the specific visual targets. However, on P300-based BCI with visual stimuli, the visual functions are "occupied" by the BCI system and not available during operation.

The BCI system which uses auditory stimulation (auditory BCI) can be used without restricting user's visual function. And it can be applied to blind patients. However, only a few types of auditory BCI have been realized. Hill et al. proposed a P300-based auditory BCI [3]. In this system, two kinds of auditory oddball sequences are presented to user's left and right ear, respectively, and user was asked to pay attention to one of the sequences presented to left or right. Schreuder et al. proposed multi-class auditory BCI with multiple oddball tone sequences presented from different locations by speakers placed around subjects [4].

The authors proposed the novel auditory BCI based on stream segregation [5]. Stream segregation is the phenomenon that the series of tones at similar frequencies tend to be bound together and are perceived as a stream of tones, even if tones at different frequencies were inserted between segregated tones. This phenomenon is one of the auditory illusions which have been discussed in the studies of auditory scene analysis [6] of the Gestalt Psychology. The mechanism of the phenomenon can be understood that the two constraints for perceptual grouping of multiple tones, similarity of tone frequency and the factor of time continuity, compete with each other [7]. In this auditory BCI system, two individual oddball sequences at different frequency bands were presented alternately to user's single ear to realize the stream segregation. Subjects were requested to pay attention to one of the tone streams. By classifying P300 and MMN components, high detection accuracy was achieved. However, only a few EEG channel data was recorded in the previous study, and the detailed investigations were left for further study.

In this study, the auditory BCI system proposed by the authors [5] was evaluated by multichannel EEG measurements. The three streams which consisted of musical tones were used as auditory stimuli. It was shown that the P300 was elicited only by the deviant stimuli in the attended target stream (target stream) in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

MATERIALS AND METHODS

Stimulus presentation: MIDI musical tones were used for presenting stimuli. All the tones were generated by using digital auditory workstation (Cakewalk by BandLab, BandLab Technologies, Singapore) and software sound source (SampleTank3, IK Multimedia Production, Italy). Piano tones (Grand Piano 1 SE) were used to enhance the stream segregation. All the tones were presented to subjects by DSP (System3, Tucker-

 DOI: 10.3217/978-3-85125-682-6-44

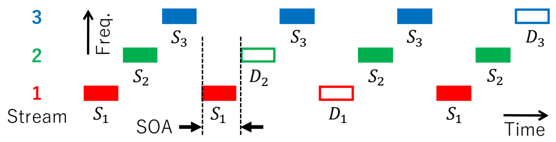

Figure 1: Schematic diagram of tone sequences

Table 1: Frequencies of the tones used in this study

Stream	Standard tone	Deviant tone
1	$\boldsymbol{S_1}$: C3 (131 Hz)	$\boldsymbol{D_1}$: G3 (196 Hz)
2	$\boldsymbol{S_2}$: D5 (587 Hz)	$\boldsymbol{D_2}$: A5 (880 Hz)
3	$\boldsymbol{S_3}$: E7 (2637 Hz)	$\boldsymbol{D_3}$: B7 (3951 Hz)

Davis Technologies, USA) and headphones (HDA200, Sennheiser, Germany). Timing and order of presented tones were controlled by Arduino UNO.

Stimulus setting: Stimulation tone sequences consisted of three streams with different frequency bands. Fig.1 shows a schematic diagram of stimulation tone sequence. Stream n $(n = 1, 2, 3)$ consisted of standard tone S_n $(p = 0.9)$ and deviant tone D_n $(p = 0.1)$. The frequencies of standard and deviant tones in the three streams were shown in Tab. 1. The duration and rise/fall time of each tone was 150 ms and 10 ms, respectively. SOA was set to 180 ms.

Experiments: Three volunteers (one female, 21 ~ 22 years old) with normal hearing functions took part in the experiments as subjects. Subject sat on a comfortable armchair in a sound-proof electromagnetically-shielded room (Music Cabin, Takahashi Kensetsu Co. Ltd., Japan). During experiment, subject was requested to pay attention to one of the three streams (indicated by experimenter before starting experiments) and to count the number of deviant tones in the target stream. The experiment to each subject consisted by two sessions, which were performed in a day. In each session, three sets of experiment were executed. In each set in a session, the object of selective attention was stream 1, 2 and 3, respectively.

EEG measurement: 64-channel EEG (Fp1, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT9, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, FT10, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP9, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, TP10, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2) was measured during experiments by Ag-AgCl electrodes (EasyCap, EasyCap GmbH, Germany). Reference and ground electrodes were placed at right and left earlobe, respectively. Measured signal was amplified at 0.1 Hz to 100 Hz, and acquired to PC with sampling frequency of 1000 Hz (BrainAmp MRplus/DC/ExG, Brain Products GmbH, Germany).

Data analysis: Acquired data was analyzed by MATLAB. The data was band-pass filtered (2nd order Butterworth filter, 1~40 Hz), and the responses to the tone were extracted (−100~500 ms from the stimulus onset). The extracted data was referred to the mean

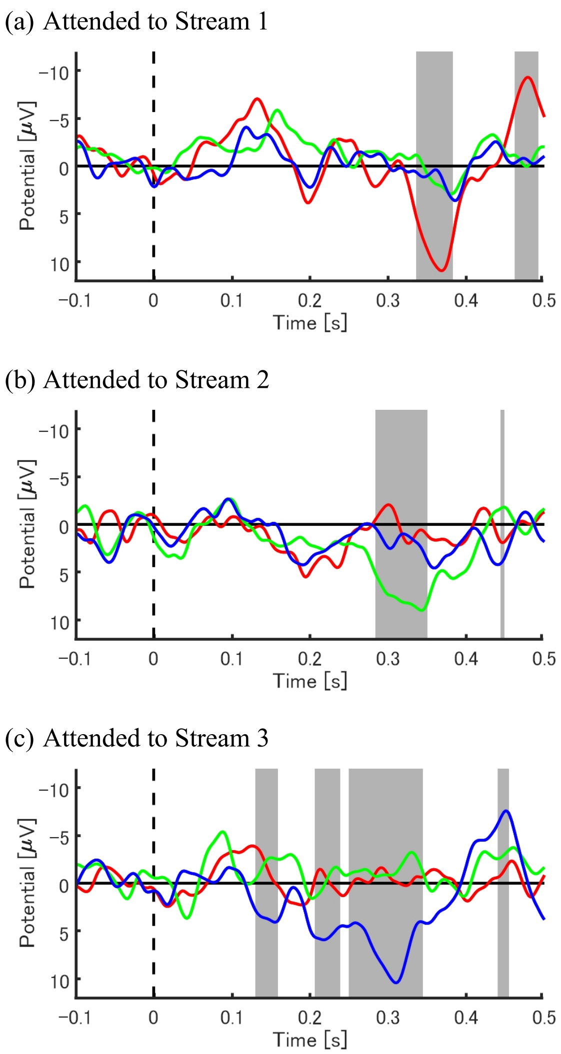

Figure 2: Averaged responses to deviant tones in Stream 1 (D_1: red), Stream 2 (D_2: green) and Stream 3 (D_3: blue) on Subject A (electrode Pz). Figures (a), (b) and (c) show responses when the subject paid attention to Stream 1, 2 and 3, respectively. Grey box denotes the significant difference between responses to target (attended) and non-target deviant tones.

amplitude in the pre-stimulus baseline (−50~0 ms). EEG epochs with amplitudes exceeded ±100 μV were excluded from further analysis. The pre-processed epoch data was averaged over trials, and the difference between the responses to standard and deviant tones was statistically tested.

Pattern classification: The ability to detect the selective attention to the target stream was evaluated by pattern classification. Three linear discriminant analysis (LDA) classifiers [8] to detect event-related potentials (P300) elicited by the deviance on stream 1, 2 and 3 were designed.

The feature vectors were calculated in the following way. The pre-processed data (shown above) at electrodes F3, Fz, F4, Cz, P5, Pz and P6 was used for pattern

 DOI: 10.3217/978-3-85125-682-6-44

Table 2: Classification accuracy to detect the target stream for selective attention. Number of the data set for sample and test data, and the accuracy to detect Streams 1, 2 and 3 are shown.

Subject	Data Set		Accuracy [%]		
	Sample	Test	Stream 1	Stream 2	Stream 3
A	1	2	93.3	93.3	87.1
	2	1	93.5	96.7	96.7
B	1	2	100.0	90.0	93.5
	2	1	100.0	80.0	96.8
C	1	2	93.5	73.3	60.0
	2	1	77.4	96.7	90.0

classification. Data taken from each channel was down-sampled to 500 Hz. The resulting data was concatenated ($300\ \text{samples} \times 7 = 2100$ dimensions). The principal component analysis (PCA) was applied to sample data which has dimensions of $2100 \times \mathrm{N}$ (N: number of samples) to reduce dimensions and to determine a linear transformation matrix by limiting the variance to 99%. Dimensions of test data was reduced by the linear transformation matrix determined above.

A set of the data taken in a session was used as a sample data set, and that in another session was used as a test data set. For evaluating detection rate, all the responses to deviant tones in 10 s time window were classified and a classification accuracy was calculated in each stream [5].

RESULTS AND DISCUSSION

Fig. 2 shows the examples of averaged responses to deviant tones in three streams. Grey box denotes the significant difference between responses to target (attended) and non-target deviant tones (Student's t-test, $p < 0.01$). On Figs. 2(a), 2(b) and 2(c), the responses to the deviant tones when this subject paid attention to Stream 1, 2 and 3 are shown, respectively. In each of the graph, the responses to the deviant tones in Stream 1 (D_1: red), Stream 2 (D_2: green) and Stream 3 (D_3: blue) are displayed. From these figures, it was found that only the corresponding deviants to the target stream for selective attention elicited P300 at the latencies around 300 to 400 ms (D_1 in Fig. 2(a), D_2 in Fig. 2(b) and D_3 in Fig. 2(c)). This result was observed clearly on Subjects A and B.

Fig. 3 shows the topographies of the averaged responses to deviant tones in Stream 1, 2 and 3 (D_1, D_2 and D_3 respectively) on the same subject if Fig. 2 (Subject A). On Figs 3(a), 3(b) and 3(c), time courses of the topographies on three conditions when the subject paid attention to Stream 1, 2 and 3 are shown. From this figure, it can be found that the P300 component was only observed in the response to the deviant tones in the target (attended) stream. Like the result shown in Fig. 2, such a phenomenon could be observed clearly on Subject A and B. On Subject C, although no such a clear phenomenon was observed, the unique and corresponding responses were shown on the ERP responses to the deviant tones according to the target stream to pay attention.

Tab. 2 shows the classification accuracy to detect the object stream for subject's selective attention. There are two sessions that were executed for each subject, and the data sets taken from each of the two sessions were used as sample and test data for pattern classification. The classification accuracies on Subjects A and B were high, and the result on Subject C was much higher than chance level.

On the results shown above, it was shown that if the subject paid attention to one of the three tone streams, the deviant tone in the attended target tone stream elicited P300 activities and the target tone stream could be detected by pattern classification.

CONCLUSION

In this study, the auditory brain-computer interface (BCI) system based on auditory stream segregation was tested. This BCI system was the extension of the BCI proposed by the authors, on which two kinds of streams were used [5]. In this study, three oddball sequences consisted of musical tones at different frequency bands were presented alternately to subject's right ear, and stimulus parameters were set so the subjects could perceive these tone sequences as three segregated tone streams. Subjects were requested to pay attention to one of the three streams during experiments.

It was shown that the P300 was elicited only by target stream in two out of three subjects, and the target stream could be detected by pattern classification with high accuracies on all subjects.

The results in this study indicate that such an auditory stimuli based on auditory scene analysis [6] or auditory illusion can extend the auditory sound space for stimulus presentation and improve the current auditory BCI system. The evaluation of the present system by many subjects, and the detailed investigation on elicited ERP components and their pattern classification are left for further study.

 DOI: 10.3217/978-3-85125-682-6-44

REFERENCES

[1] Donchin E, Spencer KM, Wijesinghe R. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehab. Eng., 2000; 8(2): 174-179.

[2] Acqualagna L, Blankertz B. Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). Clin. Neurophysiol., 2013; 124(5); 901-908.

[3] Hill NJ, Lal N, Bierig K, Birbaumer N, Scholkof B. An auditory paradigm for brain-computer interfaces. Advances in Neural Information Processing Systems, 2005; 17: 569-576.

[4] Schreuder M., Blankertz B., and Tangermann M. A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. PLoS ONE, 2010; 5(4): e9813.

[5] Kanoh S, Miyamoto K, Yoshinobu T. A brain-computer interface (BCI) system based on auditory stream segregation. Journal of Biomechanical Science and Engineering, 2010; 5(1): 32-40.

[6] Bregman AS, Auditory scene analysis: The perceptual organization of sound. The MIT Press, 1990.

[7] Kanoh S, Futami R, Hoshimiya N. Sequential grouping of tone sequence as reflected by the mismatch negativity. Biological Cybernetics, 2004; 91(6): 388-395.

[8] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd Edition, Wiley Interscience Publication, 2000.

DOI: 10.3217/978-3-85125-682-6-44

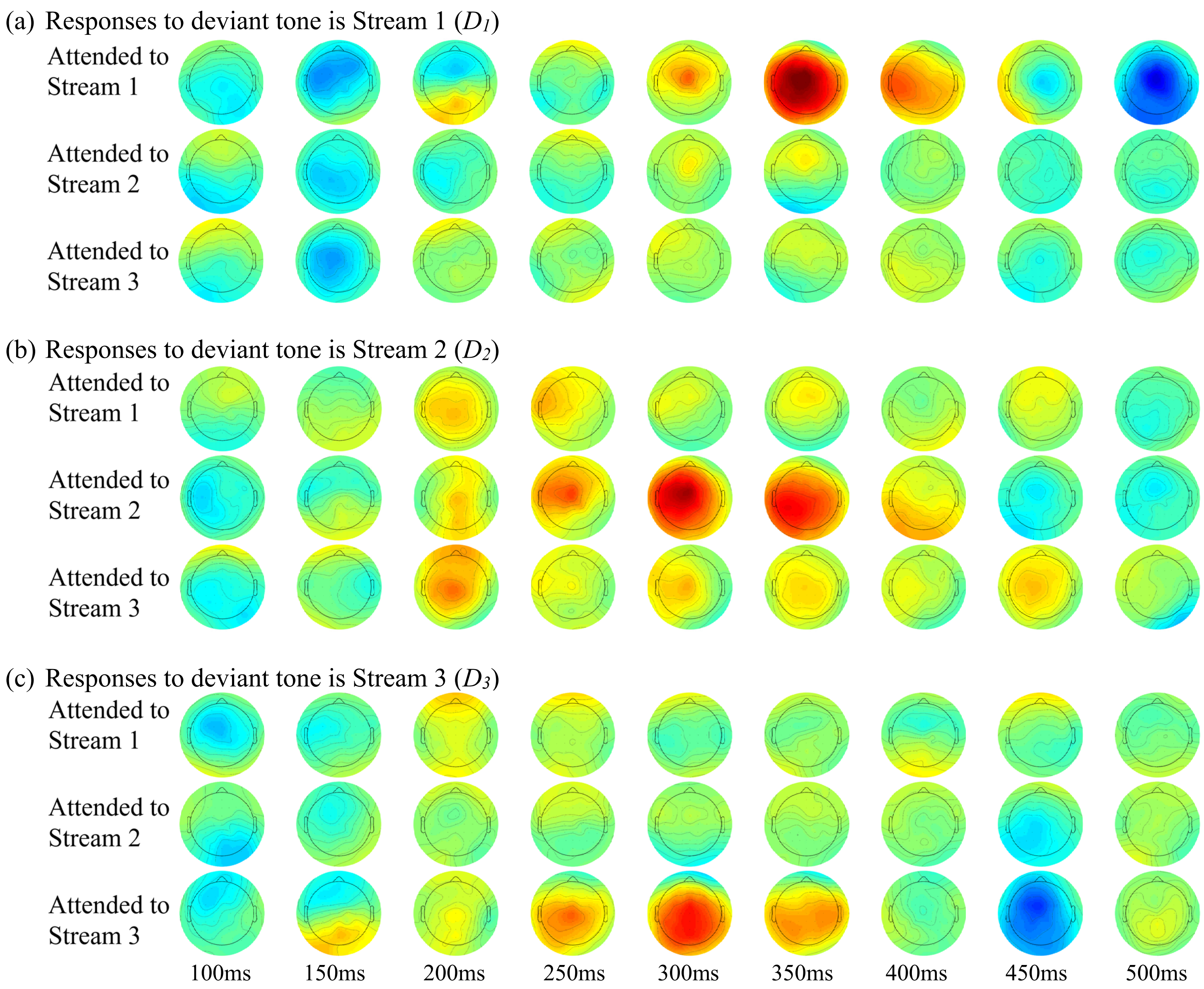

Figure 3: Topography of the averaged responses to deviant tones in Stream 1 (D_1: a), Stream 2 (D_2: b) and Stream 3 (D_3: c) of Subject A. In Figures (a), (b) and (c), time courses of the responses to corresponding deviant tones when the subject paid attention to Stream 1, 2 and 3 are shown.