Proceedings of the
9th Graz Brain-Computer Interface Conference 2024

10.3217/978-3-99161-014-4-007

A NEW AUDITORY BRAIN-COMPUTER INTERFACE BASED ON STREAM SEGREGATION UTILIZING ASSR

Shin'ichiro Kanoh[1], Naoki Mizukami[1], Simon Kojima[1]

[1]Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo, Japan

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

ABSTRACT: The authors have proposed an auditory brain-computer interface (BCI) based on stream segregation which detects users' selective attention to one the multiple segregated streams. In this system, several oddball sequences with different frequency bands were presented to users. To detect the target stream, this system needed to wait for the arrival of deviant stimuli in oddball sequences. In this study, auditory steady-state response (ASSR) was utilized to achieve a higher information transfer rate (ITR) system. Two streams consisting of sinusoidally amplitude-modulated (SAM) tones were presented to subjects, and they were requested to attend to one of the two streams. From the results of the electroencephalogram (EEG) measurement experiment, it was found that the user's selective attention enhanced ASSR corresponding to the modulation frequency of SAM sounds in the target tone stream, and the target stream which subject paid attention to was detected at an average accuracy of 0.77. The best accuracy was 0.92. It was concluded that an auditory BCI based on stream segregation utilizing ASSR is feasible, and it provides high performance and practical BCI options to users.

INTRODUCTION

Brain-computer interfaces (BCIs) provide their users communication and control channels without using brain's normal output channels of peripheral nerves and muscles [1]. Since Vidal et al. [2] proposed electroencephalogram (EEG) based noninvasive BCI, it has attracted interests by a number of researchers.

Among BCIs, reactive BCI detects a stereotypical brain response elicited by sensory stimuli presented to users [3]. To date, a number of studies on visual-based BCI have been conducted since it is easy to present visual stimulus to users with a precise time resolution. As the visual modality is the most important sense in daily life, users can not do anything else when they use such visual-based BCIs. On the other hand, as auditory BCIs do not rely on visual modality, they might not disturb users' daily activities during operations.

The authors have proposed an auditory BCI based on stream segregation. Stream segregation is one of the illusory phenomena in auditory perception, on which multiple tone sequence with different frequencies which are alternately presented in time tend to be perceived as multiple segregated sound streams. For example, when two kinds of sounds that have different frequencies (A and B) are presented alternately in time (ABABAB...), such a sound sequence can be perceived as two segregated sound streams (AAA... and BBB...) (Fig. 1) .

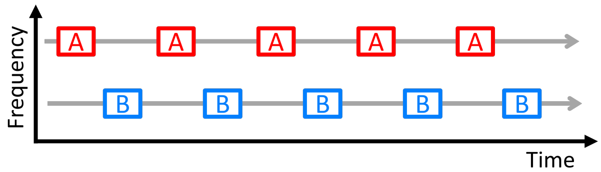

Figure 1: Schematic diagram of auditory stream segregation

The authors have proposed an auditory brain-computer interface (BCI) based on stream segregation which detects users' selective attention to one the multiple segregated streams [4, 5]. In these studies, two different oddball sequences perceived as segregated streams were presented to one ear of subjects, and they were requested to pay attention to one of the streams. It was shown that the target stream of selective attention could be estimated by detecting ERPs (event-related potentials) including P300 component which was elicited by the deviant stimuli embedded in the target stream. The authors proposed 2-class [4, 5], 3-class [6], and 4-class [7] BCI systems based on auditory stream segregation.

The feature of the auditory BCI based on segregation is that it operates in the frequency domain of the incoming monaural sound. As human auditory system have a rich capability to perceive pitch and melody of the sounds as well as location and movement in three dimensional auditory space, it is expected that such a system could enhance conventional auditory BCIs.

However, since such an auditory BCI detects P300 responses to the deviant stimuli in the oddball sequence perceived as stream, the system needs to wait for the deviant stimuli to detect the target of selective attention. As the frequency of the presentation of the deviant stimuli is low in oddball sequences, the average response time for P300 detection is long and it causes lower information transfer rate (ITR).

Therefore, in this pilot study, an auditory BCI system based on stream segregation without using oddball sequences was proposed and tested. In this proposed system, instead of P300 components, auditory steady-state

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-007

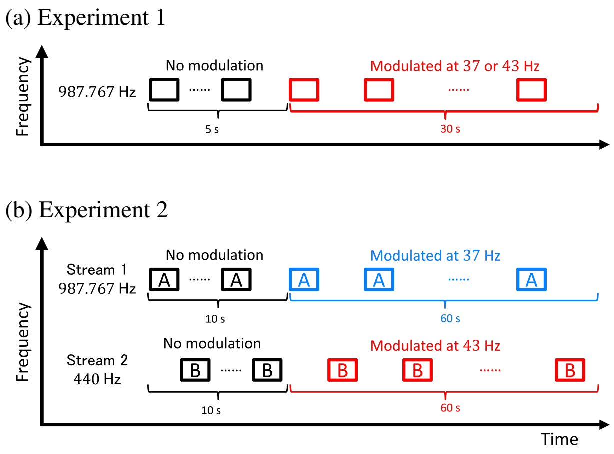

Figure 2: Time chart of presented stimuli used in Experiment 1 (a) and Experiment 2 (b)

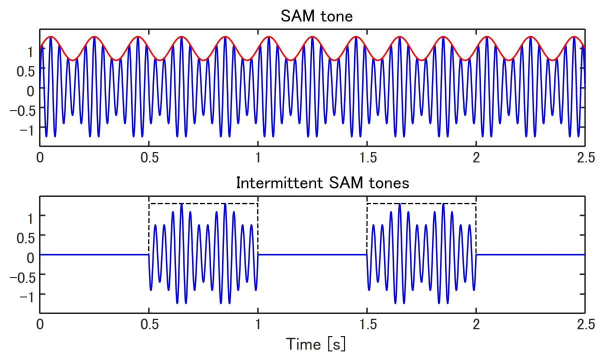

Figure 3: Schematic diagram of SAM tone and intermittent SAM tones. For viaualization, the following parameters are used: tone (carrier) frequency 25 Hz, modulation frequency 5 Hz, duration of intermittent SAM tone 0.5 s. Red line shown in the upper figure is an envelope of the amplitude modulation.

response (ASSR) was used to detect user's selective attention to the segregated stream. ASSRs are elicited by temporally modulated auditory stimulation, such as an amplitude-modulated (AM) tone [8].

After starting the presentation of a sustained AM tone, EEG or magnetoencephalogram (MEG) rapidly entrains to the modulation frequency and phase of the stimulus [8]. Lopez et al. [9] reported that ASSR is modulated by selective attention and can be utilized as a BCI paradigm. If ASSR is elicited by segregated AM tone streams and is modulated by selective attention, P300 component can be replaced by ASSR as a new identifier of the target of selective attention on auditory BCI based on selective attention. And it is expected that the target of the users' attention can potentially be detected in a shorter time period than our proposed system [4, 5].

In this paper, the results of two experiments are shown. Experiment 1 is a preliminary experiment to confirm ASSRs elicited by the intermittent tones can be observed and detected. And in Experiment 2, the feasibility of the auditory BCI based on stream segregation utilizing ASSR was investigated.

EXPERIMENT 1: ASSR TO AMPLITUDE-MODULATED INTERMITTENT TONE SEQUENCE

Objective: In the previous studies [4–7], sequence of intermittent short sounds perceived as segregated streams were presented to subjects. To apply ASSR to such an auditory BCI, it is required that ASSR is elicited by a sequence of intermittent short sound. However, it is general to measure ASSR by presenting long lasting AM tones to subjects, and whether ASSRs are evoked using intermittent short sound stimuli has not been evaluated. Hence, in Experiment 1, the ASSRs to amplitude-modulated intermittent sound stimuli were evaluated as a preliminary experiment.

Methods: Three males participated in this experiment. The subjects sat on a comfortable chair in a soundproofing electromagnetically shielded room, and 9-channel EEG (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4) were recorded with passive Ag-AgCl electrodes (Easycap, Easycap GmbH, Germany). Reference and ground electrodes were placed on the right and left ear mastoid, respectively. EEG signal was amplified and recorded with an biosignal amplifier (BrainAmp MR plus, Brain Products, Germany) at a sampling frequency of 1000 Hz. Before sampling, bandpass filter (0.1–100 Hz) was applied to the recorded data.

Fig. 2(a) shows the time chart of the tone sequence presented to subjects in Experiment 1. Sequence of tones (duration 180 ms) was presented to subjects, and they were requested to listen to the presented tones. The stimulus onset asynchrony (SOA) was set to 400 ms.

The length of a trial was 35 s. Frequency of each tone was set to 987.767 Hz (musical pitch B5). The first 25 tones (10 s) were pure tone without amplitude modulation, and the following 75 tones (30 s) were sinusoidally amplitude-modulated (SAM) tones (schematic diagram is shown in Fig. 2). Two kinds of modulation frequency (37 Hz or 43 Hz) were tested in different trials.

All stimuli were generated by MATLAB (Mathworks, USA). Tone stimuli were presented to subjects' left ear by an audio interface (Fireface 802, RME, Germany) and headphones (HDA200, Sennheiser, Germany).

MATLAB (Mathworks, USA) and EEGLAB [10] were used for analysis. The recorded EEG signal was bandpass filtered at 20–150 Hz.

The data of 30 s during which SAM tones were presented (colored red in Fig. 2(a)) was analyzed by the following way. Additionally, EEG data with removing gaps (EEG data during presentation of each SAM tone was extracted and concatenated, total duration 13.5 s) was also analyzed and results were compared.

EEG signal was segmented into 5 s intervals with an overlap of 4.5 s. Then, canonical correlation analysis (CCA) was applied to the segmented data. The reference signal $\mathbf{Y}(f_t)$ was set as follows.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-007

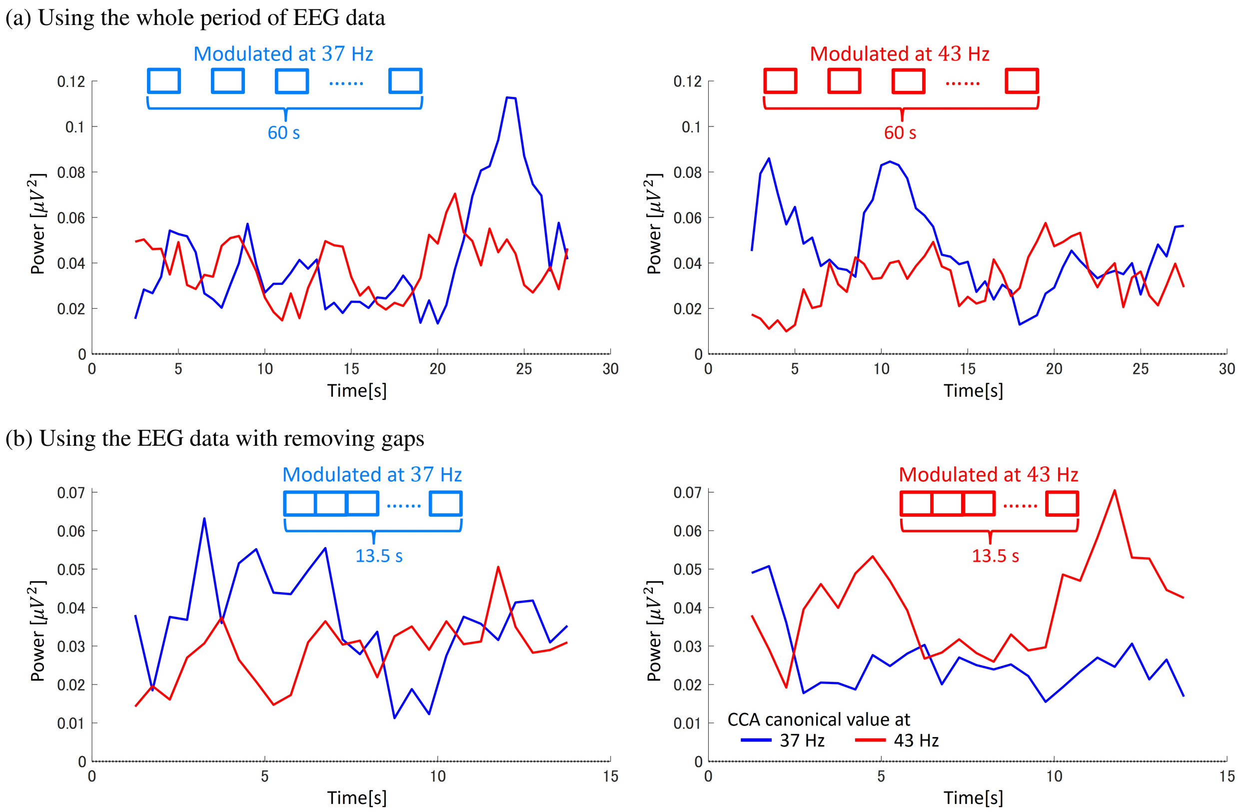

Figure 4: An example of the time courses of the power spectrum value of CCA canonical variables at 43 (red) and 37 Hz (blue) (Subject A). Top row: results of analysis using the whole period of EEG data (30 s) . Bottom row: results of analysis of EEG data with removing gaps (13.5 s). The results when this subject was presented SAM tones modulated at 37 Hz (left) and 43 Hz (right) are shown.

$$\mathbf{Y}(f_t) = \begin{bmatrix} \sin(2\pi f_t t) \\ \cos(2\pi f_t t) \\ \sin(4\pi f_t t) \\ \cos(4\pi f_t t) \\ \sin(6\pi f_t t) \\ \cos(6\pi f_t t) \end{bmatrix} \quad t = \frac{1}{f_s}, \frac{2}{f_s}, \cdots, \frac{N_s}{f_s} \qquad (1)$$

where, f_s is the sampling frequency, N_s is the number of samples of the 5 s segmented data, and $f_t \in \{37, 43\}$ Hz is the target frequency to be detected by CCA, since SAM tones were modulated at 37 or 43 Hz. Fast Fourier transform (FFT) was applied to canonical variable obtained by CCA, and the time course of the power spectrum value at each target frequency f_t was obtained. If the target frequency to be detected is f_d, corresponding power spectrum value if FFT power spectrum of canonical variable using reference signal $\mathbf{Y}(f_t)$ at frequency f_t.

Results: Fig. 4 shows the time courses of the power spectrum value of CCA canonical variables at 37 and 43 Hz of Subject A.

In Fig. 4 (a), the results of analysis using the whole period of data (30 s) are shown. And the responses during the time range when subjects were presented intermittent sounds were discarded are shown in Fig. 4 (b).

When the data from the whole trial was analyzed, there was no clear difference between the power spectrum values of CCA components at two modulation frequencies (Fig. 4 (a)). However, the CCA power values at corresponding modulation frequency were increased if the segments when the sounds were delivered to this subject were concatenated and analyzed (Fig. 4 (b)). The same tendencies were also observed on the other two subjects, however, the time course of the power spectrum values of CCA components were varied across subjects.

As a result, despite the short length of each SAM tone (180 ms), ASSRs were evoked while the stimuli were played. Thus, it can be concluded that intermittent short tone can also evoke ASSR. It was also confirmed that the power of ASSR corresponding to the modulation frequency of presented tones was increased, and it could be observed clearly if the data was analyzed when the data during the presentation of each intermittent sound was extracted and concatenated.

EXPERIMENT 2: DETECTION OF SELECTIVE ATTENTION TO SEGREGATED STREAM OF SAM TONES

Objective: In this experiment, tone sequence of SAM tones, which were perceived as two segregated tone streams, were presented to subjects, and pattern classification to detect subjects' target of selective attention was estimated offline. This experiment aimed to evaluate the feasibility of the auditory BCI paradigm based on stream segregation utilizing ASSR.

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-007

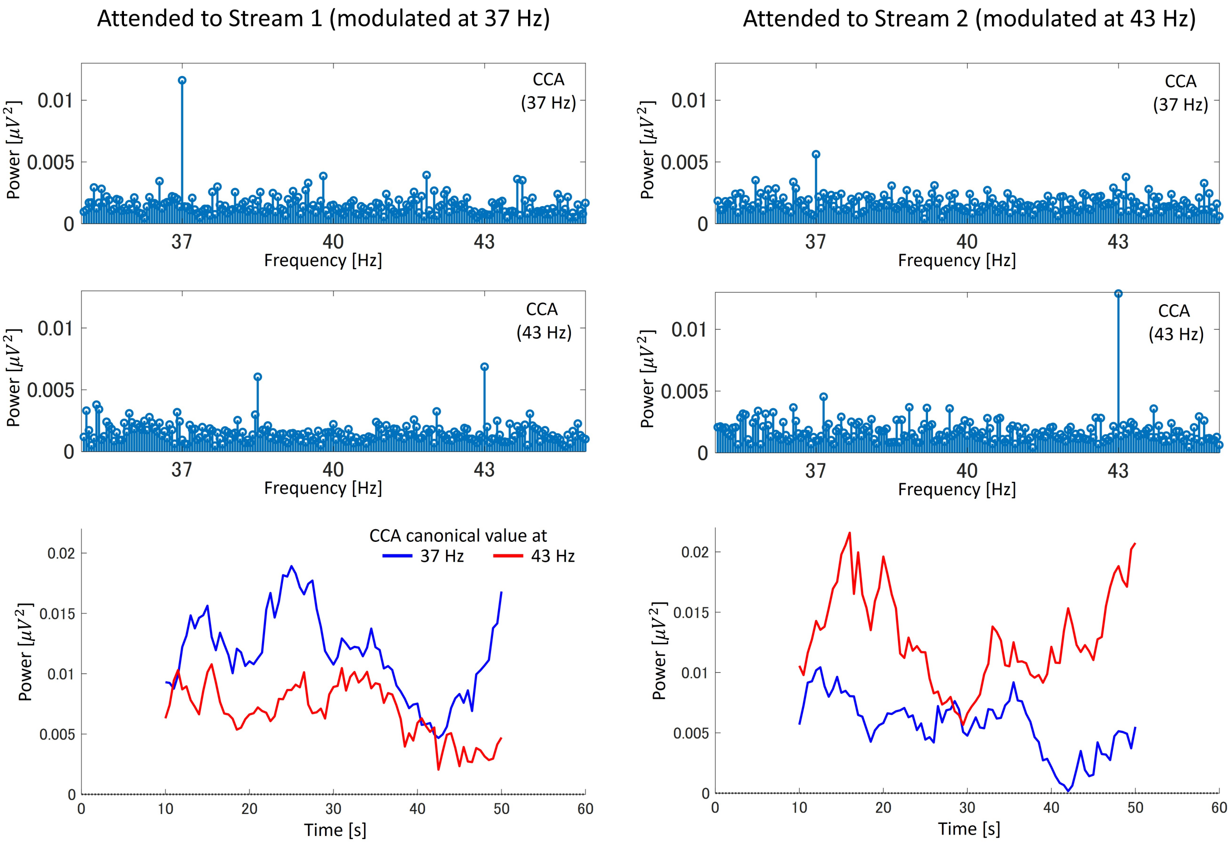

Figure 5: An example of the power spectrum of the CCA canonical variable on 37 and 43 Hz, and time courses of the power spectrum value at 37 (blue) and 43 Hz (red) which a time windows of 20 s (Subject B). Results when the subject pays attention to and Stream 1 (modulation frequency 37 Hz) and Stream 2 (43 Hz) are shown in left and right column, respectively.

Methods: Eight male subjects (aged between 21–23) participated in this experiment. The subjects sat on a comfortable chair in a soundproofing electromagnetically shielded room, and 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) were recorded with passive Ag-AgCl electrodes (Easycap, Easycap GmbH, Germany) . Other conditions for measurement were the same as in Experiment 1.

Two kinds of SAM tones, 987.767 Hz (musical pitch B5) pure tones amplitude-modulated at 37 Hz and 440 Hz (musical pitch A4) of pure tones amplitude-modulated at 43 Hz were used. The duration of the tones was 180 ms. Fig. 2(b) shows the time chart of the sequence used in this experiment. The stimulus onset asynchrony (SOA) was set to 200 ms. All stimuli were presented to subjects' left ear by an audio interface (Fireface 802, RME, Germany) with headphones (HDA200, Sennheiser, Germany). The length of a trial was 70 s, and the subjects were requested to pay attention to the instructed stream. The unmodulated tones were presented for the first 10 s.

MATLAB (Mathworks, USA) and EEGLAB [10] were used for analyses. The recorded EEG signals were band-pass filtered in a range of 20–150 Hz. The data of 60 s during which SAM tones were presented was used. Based on the result of Experiment 1, EEG data during gap period was removed and concatenated before analysis. To find the optimal length of the window for analysis, four window lengths (5, 10, 15, and 20 s with the overlap of 4.5 s, 9.5 s, 14.5 s, and 19.5 s, respectively) were used for segmentation. Then, CCA was applied to the segmented data. The same reference signals as Experiment 1 were used.

The subjects' target of selective attention was estimated with linear discriminant analysis (LDA). Two features (the power spectrum of CCA canonical variables at 37 and 43 Hz) of each data segment were used for classification. The classification accuracy was evaluated by 5-fold cross-validation.

Results: Fig. 5 shows an example of the result from Subject B. The power spectrum of the CCA canonical variable on $f_t = 37$ and 43 Hz, and time courses of the power spectrum value at 37 (blue) and 43 Hz (red) which a time windows of 20 s are shown. It was shown that the CCA canonical variable at the modulation frequency of the attended stream was larger than that of the unattended stream. On 5 subjects out of 8, the same tendency was observed. However, like in Experiment 1, the time course

Published by
Verlag der Technischen Universität Graz

 10.3217/978-3-99161-014-4-007

Table 1: Classification accuracy for each length of time window.

Subject	5 s	10 s	15 s	20 s
A	0.65	0.64	0.72	0.73
B	0.69	0.72	0.82	0.92
C	0.66	0.70	0.73	0.84
D	0.56	0.65	0.77	0.83
E	0.56	0.63	0.67	0.80
F	0.64	0.62	0.70	0.68
G	0.55	0.58	0.68	0.72
H	0.57	0.54	0.57	0.62
Average	0.61	0.64	0.71	0.77

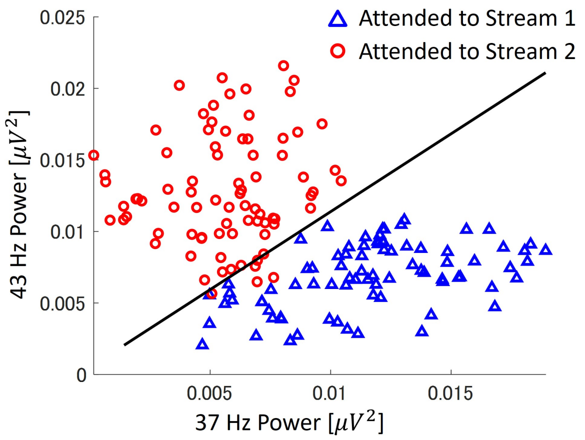

Figure 6: An example of the feature distributions with 20s of the time window (Subject B)

of the power spectrum values of CCA components were varied across subjects.

Table 1 shows the classification accuracy for each length of time window. When 20 s time window was used, the average accuracy reached 0.77. Fig. 6 shows the feature distribution of the data from subject B using 20 s time window. Feature vectors of the two classes were distributed differently in feature space and could be discriminated by linear function.

DISCUSSION

When subjects paid attention to one of the streams consisting of SAM tones, ASSR at modulation frequency was increased, and the subjects' target of selective attention to the stream could be detected with high accuracy by the machine-learning approach. It was proved that the auditory BCI based on stream segregation utilizing ASSR is feasible. Compared to the previous study which detects P300 component by using oddball sequence as presented tone stimuli, the ASSR-based paradigm doesn't need to wait for the deviant stimuli to be presented. Furthermore, it does not require the precise timing of stimulus onset, while the oddball-based paradigm requires accurate timing of stimulus onset. Therefore, it is expected that an ASSR-based BCI using stream segregation can realize higher ITR with a simple setup.

In this pilot study, classification accuracy was below 0.7 when the length of time window to observe power spectrum of CCA canonical variable was 10 s (Table 1). Because it was needed to observe EEG data for the width of time window, power spectrum values could not be calculated at the beginning of the data (on offline analysis in this study, power spectrum values were lacked for half width of the time window at the beginning and at the end of the EEG data. See Figs. 4 and 5).

Two different target frequencies f_t could be discriminated by relative power spectrum values of ASSR at corresponding the the modulation frequencies (37 or 43 Hz) of SAM tomes (Fig. 6). It was shown that the values of the feature vector (i.e. power spectrum value of CCA canonical variable at two modulation frequencies) changed over time during attending to the target segregated stream (Figs. 4 and 5), and it was different between subjects. Such phenomenon might reflect the temporal change of the degree of attention on each individual subject.

However, to improve both classification accuracy and ITR, optimization of feature extraction of EEG using CCA is required. This is left for the further study.

CONCLUSION

An auditory BCI system based on stream segregation of SAM tone sequence was proposed and tested. In this proposed system, instead of P300 components[4–7], ASSR was used to detect user's selective attention to the segregated stream. It was found that the power spectrum of ASSR at the modulation frequency of the target SAM tone stream is increased by selective attention, which is perceived as segregated streams by stream segregation. This result shows that an auditory BCI based on stream segregation utilizing ASSR is feasible.

With this approach, the target stream can be estimated without waiting for special stimuli (e.g., deviant stimuli in oddball sequence), and it can achieve high ITR in principle. Furthermore, as opposed to an ERP-based system, precise timing of stimulus presentation is not required, and it is beneficial for making the system simpler. In principle, it will lead the auditory BCI based on stream segregation to high-performance and more practical and offers users more options for auditory BCIs.

Further investigations on feature extraction and optimization of presenting auditory tone stimuli are needed to improve the performance of the present auditory BCI system.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number JP23K11811.

REFERENCES

[1] Wolpaw J *et al.* Brain-computer interface technology: A review of the first international meeting. IEEE Transactions on Rehabilitation Engineering. 2000;8(2):164–173.

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-007

[2] Vidal JJ. Toward Direct Brain-Computer Communication. Annual Review of Biophysics and Bioengineering. 1973;2(1):157–180.
[3] Rao RPN. Brain-Computer Interfacing: An Introduction. Cambridge University Press: Cambridge (2013).
[4] Kanoh S, Miyamoto Ki, Yoshinobu T. A brain-computer interface (BCI) system based on auditory stream segregation. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Aug. 2008, 642–645.
[5] Kanoh S, Miyamoto Ki, 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] Kanoh S, Kojima S. Evaluation of auditory BCI system based on stream segregation. In: Proceedings of the 8th Graz Brain-Computer Interface Conference 2019. Graz, 2019.
[7] Kojima S, Kanoh S. Towards realizing multi-class auditory brain-computer interface paradigm based on stream segregation: A preliminary study. In: 2023 15th Biomedical Engineering International Conference (BMEiCON). IEEE: Tokyo, Japan, Oct. 2023, 1–5.
[8] O'Donnell BF, Vohs JL, Krishnan GP, Rass O, Hetrick WP, Morzorati SL. Chapter 6 - The auditory steady-state response (ASSR): A translational biomarker for schizophrenia. In: Supplements to Clinical Neurophysiology. Elsevier, Jan. 2013, 101–112.
[9] Lopez MA, 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 brain–computer interfaces. Neurocomputing. 2009;72(16):3617–3623.
[10] Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 2004;134(1):9–21.

Published by
Verlag der Technischen Universität Graz