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

10.3217/978-3-99161-014-4-006

INTRODUCING THE ASME-SPELLER, AUDITORY BCI SPELLER UTILIZING STREAM SEGREGATION: A PILOT STUDY

Simon Kojima[1], Shin'ichiro Kanoh[1,2]

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

E-mail: nb21106@shibaura-it.ac.jp

ABSTRACT: The auditory BCI spellers are considered the only means of communication for late-stage patients with severe neurological disorders such as amyotrophic lateral sclerosis (ALS). To date, several auditory BCI spellers have been proposed. However, they require multiple steps, visual support, or multi-channel audio systems. In this study, we proposed an ASME-speller, which stands for Auditory Stream segregation, Multiclass, ERP speller, that uses an auditory BCI paradigm based on auditory stream segregation to detect the target of the user's selective attention by presenting a QWERTY keyboard-like audio stimuli. The 64-channel electroencephalogram was measured while the six subjects carried out 15-character ASME-speller paradigms. Offline simulation using dynamic stopping showed that the ASME speller achieved an average accuracy of 0.73 and an average ITR of 3.78 bits/min. The best results were achieved with an accuracy of 0.97 and an ITR of 7.61 bits/min. These results indicate that the ASME speller can be used as a new auditory BCI speller. This study provides more users with a high-accuracy and intuitive new speller option.

INTRODUCTION

Brain-computer interfaces (BCIs) give their users communication and control channels that do not depend on the brain's normal output channels of peripheral nerves and muscles [1]. Many BCIs aimed to restore communication for locked-in patients suffering from progressive motor diseases such as amyotrophic lateral sclerosis (ALS) [2]. Many spelling protocols using visual stimuli have been proposed [3] to realize the application of BCI in communication. However, it is known that patients with late-stage ALS have unreliable gaze control [4], and the BCIs using visual stimuli are not adequate for those patients. On the other hand, auditory BCIs do not occupy their sight and can be used by visually impaired patients. Thus, it is meaningful to realize the auditory BCIs for spelling application.

Furdea et al. [5] proposed an auditory speller BCI similar to a visual P300 speller [6]. In this system, a display of a 5×5 matrix containing 25 alphabet characters and voices two number words coded with each character's position in the matrix was presented. One corresponded to the row, and one corresponded to the column. The system detected which character the users paid attention to with two steps. The target row was detected in the first step, and the target column was detected in the second step. Klobassa et al. did a similar study but with a 6×6 matrix containing all 26 alphabet characters and miscellaneous [7]. Also, they changed the human voice to environmental sounds. Schreuder et al. [8] utilized the AMUSE paradigm [9] for a spelling application. This system also detected the target character using a two-step procedure. They divided alphabet characters into six groups. The target group was detected in the first step, and the target character was detected in the second step. Each character group and character was presented from one of the six loudspeakers surrounding the subjects' heads.

Some auditory speller BCIs have been proposed; however, these studies had one of the following issues. —(1) One trial cannot determine the target character. (2) The mapping from the character to sound streams or stimuli is not intuitive and requires memorization or visual support. (3) It requires a multi-channel audio system, complicating setup and making it unavailable to patients who have hearing impairment in one ear.— Thus, we propose a novel auditory speller BCI protocol for solving these issues, the ASME-speller.

ASME paradigm: ASME (for Auditory Stream segregation, Multiclass, ERP) is the paradigm for auditory BCI based on auditory stream segregation. The auditory stream segregation is one of the auditory illusions that alternately presented sounds can be perceived as segregated multiple streams [10]. e.g., when sounds that have different frequencies (A and B) are presented alternately (ABABAB...), they can be perceived as two segregated sound streams (AAA... and BBB ...). The authors proposed an auditory BCI paradigm utilizing auditory stream segregation [11–14]. In this system, the oddball sequence was put into segregated streams and presented simultaneously to the subjects, and the subjects paid attention to the target stimuli in the target stream. The target stream was estimated by detecting ERP responses elicited by the target stimuli. To date, we tested the ASME paradigm with two streams [11, 12], three streams [13], and four streams [14].

The ASME speller:

The QWERTY is a keyboard layout widely used in computers and smartphones, and many personal computer

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-006

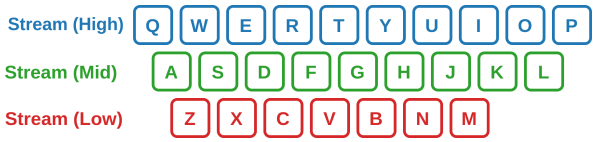

Figure 1: The conceptual diagram of the ASME-speller.

and smartphone users are expected to be familiar with the QWERTY layout. Since all 26 characters are mapped to three-row keys, the entire keyboard layout can be represented with three streams ASME paradigm. Fig. 1 shows the QWERTY layout and the corresponding tone stream on the ASME-speller. Three key rows are assigned to the sound stream, and the top, middle, and bottom rows correspond to the stream, which has high, middle, and low-frequency bands, respectively. Within each stream, each character is presented as a spoken voice. When the user is going to type "T," the user thinks of which row "T" is located in the QWERTY layout. Since the character "T" is in the top row, the user will listen to the corresponding stream (the stream with a high-frequency band) and pay attention to the "T" sound stimuli. Since one stimuli are paid attention to and the others are ignored, this sequence can be considered an oddball, and the target stimuli elicit ERPs, including P300 [11–14]. The target character can be estimated by detecting ERP responses with a machine learning approach. This study tested the ASME-speller paradigm with 15 characters as a pilot study.

MATERIALS AND METHODS

Experimental Design: Two different conditions were conducted. (1) The ASME condition: Each row of the QWERTY layout was spoken by a different person and had a different pitch, so each row could be perceived as a different sound stream. (2) The control condition: all stimuli were spoken by the same person and had the same pitch. In a session, four runs were conducted with changing conditions. In total, two ASME runs and two control runs were conducted. In a run, 15 trials were conducted. Before starting each trial, the target character was shown on the display in front of the subject, and the subject was instructed to pay attention to the target stream and the target character. All subjects were familiar with the QWERTY layout, and no visual support was provided. 225 stimuli (15 targets and 210 nontargets) were delivered in a trial, and the trial length was about 46 seconds.

Stimuli: The fifteen characters (E, R, T, I, O, A, S, D, H, L, C, V, B, N, and M) were selected for this study. Each voice stimuli were generated by Amazon Web Services (AWS) Amazon Polly. AWS Amazon Polly is a cloud service that converts text into synthesized spoken audio. The voice stimuli of characters corresponding to the top (E, R, T, I, and O), middle (A, S, D, H, and L), and bottom (C, V, B, N, and M) row on the QWERTY layout were generated with the voice ID of *Ruth* (Female), *Kevin* (Male child) and *Joey* (Male), respectively. The voice IDs were selected as the top, middle, and bottom rows could be perceived as higher, middle, and lower pitch streams.

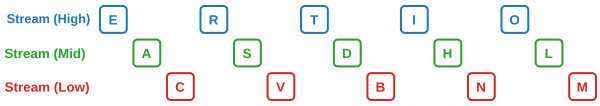

Figure 2: The time chart of presented stimuli block.

To enhance the difference between each stream, the characters corresponding to the top and bottom are shifted in pitch with $+2$ halftones and -2 halftones, respectively. Fig. 2 shows the "block" of the sequence. Each block had 15 stimuli in total. Within each stream, the order of the characters was randomized. For both ASME and control conditions, the stimuli were presented in the order of the characters corresponding to the top, middle, and bottom rows of the QWERTY layout. In a trial, 15 blocks were played. The stimulus onset asynchrony (SOA) was set to 0.2 s. All 15 characters were generated with the voice ID of *Kevin* for the control condition, and no pitch shifting was applied. All other parameters for the control condition were the same as the ASME condition. All sound stimuli were delivered by Fireface 802 (RME, Germany) with headphones (MDR-EX800ST, Sony, Japan).

Signal Acquisitions: The following 64-channel (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) electroencephalogram (EEG) were measured with Ag-AgCl passive electrodes (Easycap, Easycap GmbH, Germany). The vertical and horizontal electrooculogram (EOG) were also measured. All EEG and EOG signals were amplified and recorded with BrainAmp DC and BrainAmp MR plus (Brain Products GmbH, Germany). The reference and the ground electrodes were placed on the right and left ear mastoid, respectively. The signals were recorded at a sampling frequency of 1000 Hz. Subjects sat on a comfortable chair placed in a soundproofing electromagnetic shielded room.

Subjects: Six subjects (ages 22 – 27, mean: 24.0) participated in this study. This study protocol was approved by the Review Board on Bioengineering Research Ethics of Shibaura Institute of Technology and was conducted in accordance with the Declaration of Helsinki. Before the experiment, subjects were given information orally and in writing, and written informed consent was obtained from all subjects. No subject had known neurological disorders or hearing problems.

ERP Analyses: The EOG artifacts were removed with independent components analysis (ICA). The measured signals were bandpass filtered by 2nd order Butterworth filter in the range of 1–30 Hz, and responses to each stimulus were epoched in the range of -0.1–1.0 s relative to stimulus onset. Then, all epochs were downsampled to 250 Hz. To assess the separability between the responses to the target and nontarget stimuli, signed-r^2 values [15] were obtained.

Binary Classification: The EOG artifacts were re-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-006

moved using independent components analysis (ICA). The measured signals were bandpass filtered by 2nd order Butterworth filter in the range of 0.1–8Hz, and responses to each stimulus were epoched in the range of 0–1.0 s relative to stimulus onset. Then, all epochs were downsampled to 250 Hz. The mean amplitude in the following ten intervals ([0.0, 0.1; 0.1, 0.2; 0.2, 0.3; 0.3, 0.4; 0.4, 0.5; 0.5, 0.6; 0.6, 0.7; 0.7, 0.8; 0.8, 0.9; 0.9, 1.0] seconds relative to the stimulus onset) were used as the classification feature. The dimension of the feature vector was 10intervals$\times$64channels$=$ 640. The classification accuracy (AUC: area under the receiver operating characteristic curve) between the responses to the target and nontarget stimuli was obtained by a shrinkage linear discriminant analysis (Shrinkage-LDA) [15] with 4-fold chronological cross-validation. For the binary classification, the chance level was 0.5. The information transfer rate (ITR) was calculated using the equation proposed by Wolpaw et al. [16].

BCI simulation (target character detection): In the BCI simulation, the target character of the trial was estimated. For both ASME and control conditions, 30 trials were conducted. The BCI simulation was conducted with chronological 3-fold cross-validation by training data from 20 trials and testing with data from 10 trials. From the training data, the mixing and unmixing matrices were derived using ICA to remove EOG artifacts. The mixing and unmixing matrix was applied to the training data to remove EOG artifacts, and the feature vector was obtained with the same method described in the section "*Binary Classification*", and Shrinkage-LDA was trained. The classification output $f(\mathbf{x_i}) = \mathbf{w}^T\mathbf{x_i} + b$ was defined as follows, where $\mathbf{x_i}$ is a feature vector, $\mathbf{w}$ is the weight vector obtained by LDA, and b is a bias. Each feature vector $\mathbf{x_i}$ had a corresponding class label $y_i \in \{-1, 1\}$, and assumed that class label $+1$ is the target and -1 is nontarget. The LDA was trained as $f(\mathbf{x}) \geq 0$ if $\mathbf{x_i}$ was in class $+1$ and $f(\mathbf{x}) < 0$ if $\mathbf{x_i}$ was in class -1. The mixing and unmixing matrix derived using ICA was applied for epoch data in each trial in test data, and the feature vectors were obtained. Then, the classifier output $f(\mathbf{x_i})$ for each feature was computed, and the class with the largest mean value of classifier output was estimated as the final classification result. The classification results were evaluated by accuracy. For the BCI simulation, the chance level was 0.067.

Dynamic Stopping: To optimize the trial length, the dynamic stopping strategy [17, 18] was also tested for BCI simulation. Dynamic stopping could be triggered after presenting the 75 stimuli in each trial. A one-sided Welch's t-test was applied to the classifier outputs $f(\mathbf{x})$ of the class, between which the mean value of $f(\mathbf{x})$ was the largest and second largest. If the difference was significant ($p < 0.05$), the classification procedure was stopped, and the classification result for the trial was determined with the data up to that stimuli.

RESULTS

Table 1: Binary classification results. The classification accuracy (AUC) for the ASME and control conditions are shown. The chance level was 0.5.

Subject	ASME	control
A	0.72	0.53
B	0.71	0.52
C	0.86	0.63
D	0.76	0.58
E	0.63	0.56
F	0.74	0.61
Average	0.74	0.57

Fig. 3 shows grand averaged ERP responses to the target and nontarget stimuli. In the time range from 0.2 to 0.4 seconds, N2 was observed in the ASME condition. Furthermore, in the time range from 0.4 to 0.8 seconds, P300 was observed. The amplitude of N2 and P300 were larger for the target stimuli than for nontarget stimuli, and the absolute value of the signed-r^2 was also larger, which implies it was informative for the machine learning model for classification. In contrast, a clear difference between the responses to the target and the nontarget stimuli was not observed in the control condition. The absolute value of the signed-r^2 was small compared to that for the ASME condition; the separability between the response to the target and nontarget stimuli was small compared to that for the ASME condition. Tab. 1 shows the binary classification accuracy (AUC) for the ASME and control conditions. The accuracy for the ASME condition was significantly larger than that for the control condition ($p = 0.031$, two-sided Wilcoxon signed-rank test). Fig. 4 shows the result of the BCI simulation (detecting the target character of the trial) without dynamic stopping. The average accuracy was 0.72 (ASME) and 0.31 (control), and the accuracy of the ASME condition was significantly larger ($p = 0.031$, two-sided Wilcoxon signed-rank test). The average ITR was 2.79 bits/min (ASME) and 0.726 bits/min (control), and the ITR of the ASME condition was significantly larger ($p = 0.031$, two-sided Wilcoxon signed-rank test). Fig. 5 shows the result of the BCI simulation using dynamic stopping. The average accuracy was 0.73 (ASME) and 0.31 (control), and the accuracy of the ASME condition was significantly larger ($p = 0.031$, two-sided Wilcoxon signed-rank test). The average ITR was 3.78 bits/min (ASME) and 0.775 bits/min (control), and the ITR of the ASME condition was significantly larger ($p = 0.031$, two-sided Wilcoxon signed-rank test). After applying dynamic stopping, the ITR (Information Transfer Rate) was improved without any drop in accuracy. The best ITR was reached at 7.61 bits/min with an accuracy of 0.97 (subject C). The worst ITR was 0.902 bits/min with an accuracy of 0.40 (subject E). Fig. 6 shows the results of BCI simulation using dynamic stopping in the confusion matrix.

DISCUSSION

The letters in each row of the QWERTY keyboard layout, mapped to three sound streams, were presented as voice

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-006

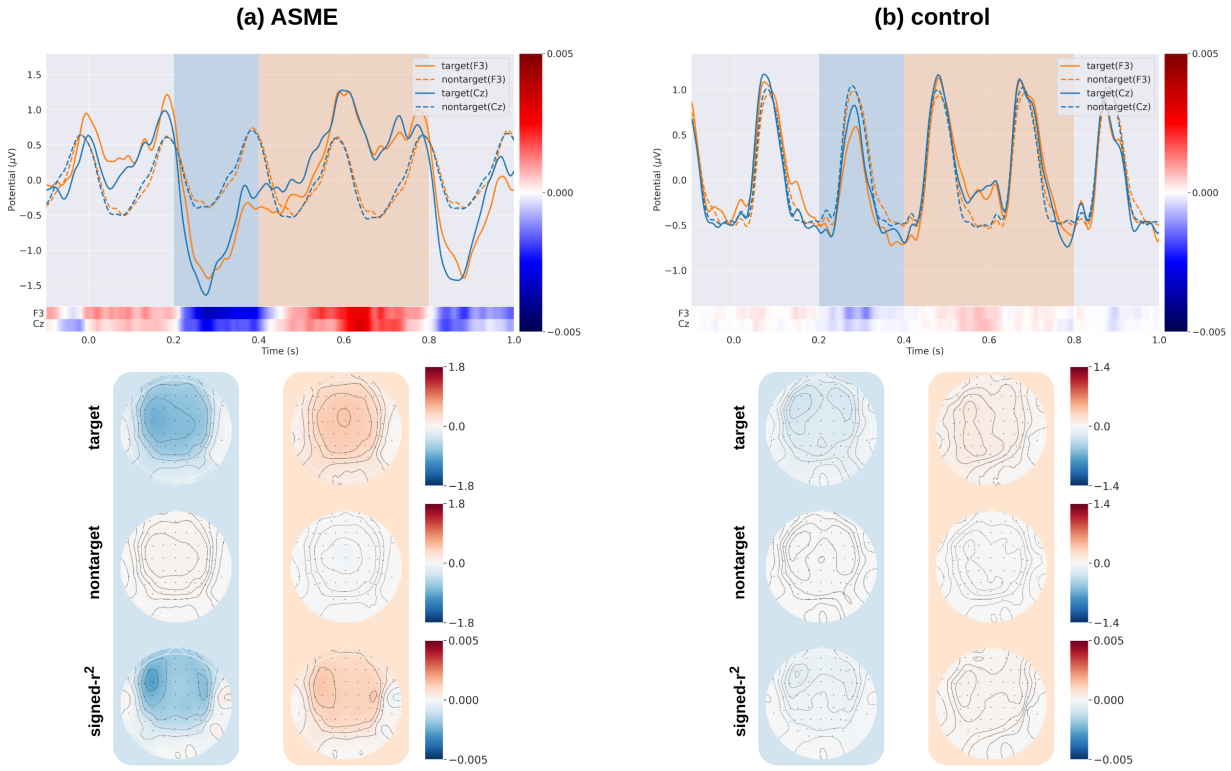

Figure 3: The grand averaged ERP responses to the target (solid line) and nontarget (dashed line) stimuli on electrode Cz (Orange) and F3 (Blue). (a) ASME condition. (b) control condition. Timepoint 0 is the stimulus onset. Stimuli were delivered with SOA of 0.2 s. The colormap below each ERP plot shows the signed-r^2 values on each electrode. The topography map shows the responses to the target, nontarget stimuli, and the signed-r^2 values from the top in each time range denoted as blue and orange mesh in the ERP plot.

stimuli, and it was shown that it is possible to pay selective attention to a single target letter stimulus. In addition, ERPs such as P300 and N2 were elicited only to the target stimuli by paying attention to them. Furthermore, the target letter could be detected with a machine learning approach with high accuracy. It can be concluded that the ASME-speller can be realized. However, the number of characters in this study was limited to 15. Thus, the speller with 26 letters needs to be tested. By applying the dynamic stopping procedure, the average Information Transfer Rate (ITR) was found to be 3.78 bits/min. Tab. 2 shows ITRs achieved in previous studies. The ITR of this study is superior to other studies except for the work by Schreuder et al. [8]. However, the best ITR was higher than theirs (7.61 bits/min v.s. 7.55 bits/min). ASME-speller has the capability to achieve higher or competitive ITR (Information Transfer Rate) and deliver high performance.

Necessity of stream segregation: The ASME-speller was achieved by dividing the sound stimuli into three groups, corresponding to each row of the QWERTY keyboard layout. However, these stimuli can also be delivered with a single stream, and it was not clear whether the sound stimuli needed to be delivered with segregated groups. Therefore, as a control, the condition of delivering all stimuli with a single stream was also tested in this study. Compared to the ASME condition, the amplitude of ERPs was smaller in the control condition, resulting

Table 2: ITRs achieved in previous studies. The ITR of [19] was read from a figure.

Average ITR (bits/min)	Authors
1.54	Furdea et al. [5]
2.0	Klobassa et al. [7]
3.4	Höhne et al. [20]
5.26	Schreuder et al. [8]
about 1.3	Höhne et al. [19]
1.11	Kleih et al. [21]
2.38	Markovinović et al. [22]
3.78	Kojima et al. (this paper)

in low classification results. In this study, the SOA was set to 0.2 s; however, in ASME condition, SOA within the stream was 0.6 s. It is expected that this slower SOA within the stream made the subject find the target stimuli easier and feel less overlap between stimuli. It can be concluded that utilizing the ASME paradigm makes SOA within the stream slower and making easier to find the target stimuli from the sequence.

CONCLUSION

In this study, the ASME-speller, which detects the users' target letter from 15 characters mapped to three sound streams corresponding to the QWERTY keyboard layout, can be realized as an auditory speller BCI. The achieved ITR was faster than most of the proposed auditory BCI spellers. This study also proved to provide stimuli di-

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-006

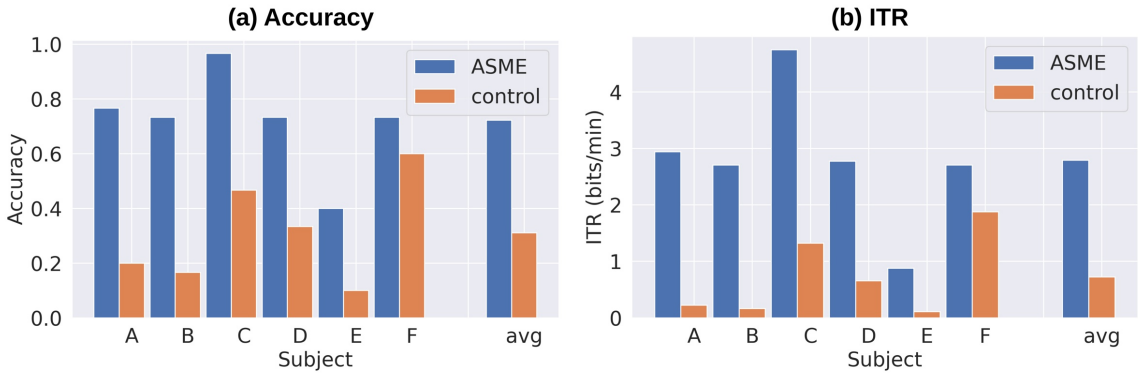

Figure 4: The results of the BCI simulation (estimating the target character of the trial) without dynamic stopping. (a) Accuracy and (b) ITR.

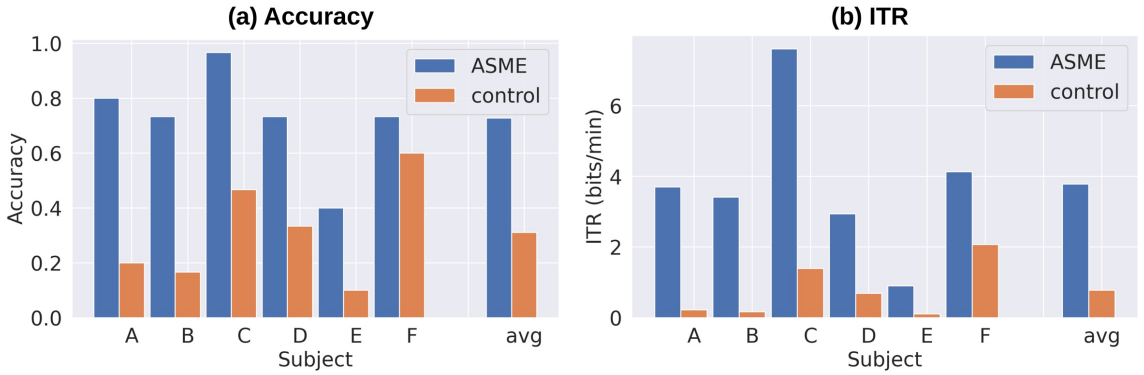

Figure 5: The results of the BCI simulation (estimating the target character of the trial) using dynamic stopping. (a) Accuracy and (b) ITR.

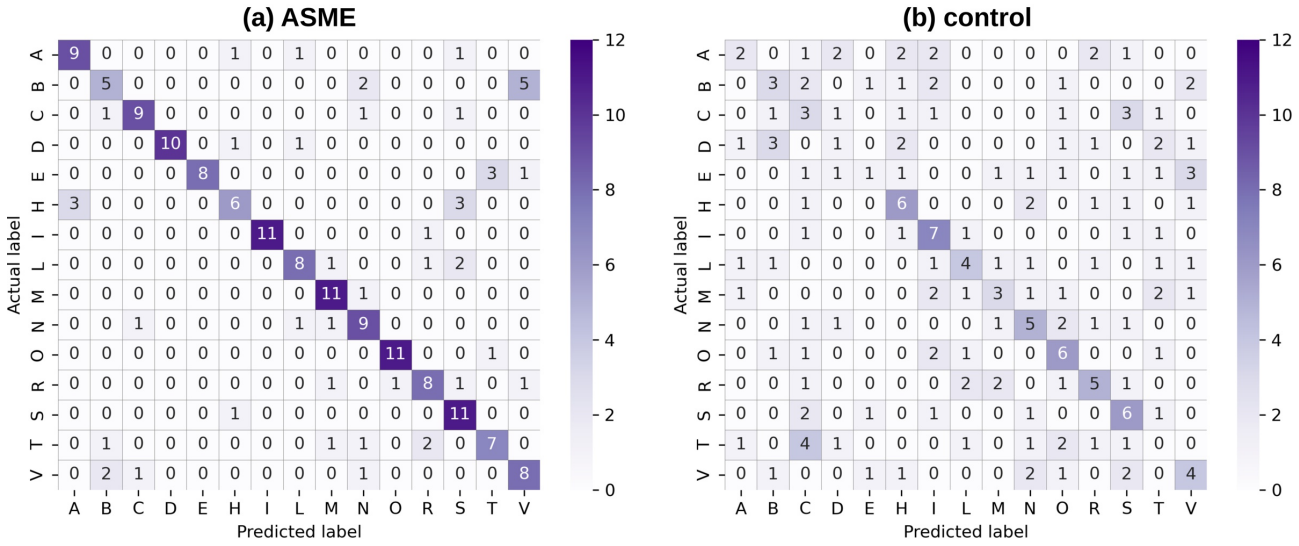

Figure 6: The confusion matrix for conditions (a) ASME and (b) control in the BCI simulation using dynamic stopping.

viding into groups by using auditory stream segregation, drastically improving the ASME-speller's performance. Furthermore, the target letter can be determined with a single trial, and no visual support is required if the users are familiar with the QWERTY layout. Additionally, All sound stimuli can be delivered with a monaural audio channel, solving the issues proposed by the auditory BCI spellers. This system has the potential to be used by patients who have severe motor impairment or hearing impairment in one ear with high ITR, and it provides users with more choices of auditory BCI spellers.

ACKNOWLEDGEMENT

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-006

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.
[2] Rao RPN. Brain-computer interfacing: An introduction. Cambridge University Press: Cambridge New York Melbourne New Delhi Singapore (2019).
[3] Kundu S, Ari S. Brain-Computer Interface Speller System for Alternative Communication: A Review. IRBM. 2022;43(4):317–324.
[4] Choi YJ, Kwon OS, Kim SP. Design of auditory P300-based brain-computer interfaces with a single auditory channel and no visual support. Cognitive Neurodynamics. 2023;17(6):1401–1416.
[5] Furdea A *et al.* An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology. 2009;46(3):617–625.
[6] Farwell L, 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.
[7] Klobassa DS *et al.* Toward a high-throughput auditory P300-based brain–computer interface. Clinical Neurophysiology. 2009;120(7):1252–1261.
[8] Schreuder M, Rost T, Tangermann M. Listen, You are Writing! Speeding up Online Spelling with a Dynamic Auditory BCI. Frontiers in Neuroscience. 2011;5.
[9] Schreuder M, Blankertz B, Tangermann M. A New Auditory Multi-Class Brain-Computer Interface Paradigm: Spatial Hearing as an Informative Cue. PLOS ONE. 2010;5(4):e9813.
[10] Bregman AS. Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press (1990).
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components — A tutorial. NeuroImage. 2011;56(2):814–825.
[16] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clinical Neurophysiology. 2002;113(6):767–791.
[17] Verschore H, Kindermans PJ, Verstraeten D, Schrauwen B. Dynamic Stopping Improves the Speed and Accuracy of a P300 Speller. In: Artificial Neural Networks and Machine Learning – ICANN 2012. Springer: Berlin, Heidelberg, 2012, 661–668.
[18] Schreuder M, Höhne J, Blankertz B, Haufe S, Dickhaus T, Tangermann M. Optimizing event-related potential based brain–computer interfaces: A systematic evaluation of dynamic stopping methods. Journal of Neural Engineering. 2013;10(3):036025.
[19] Höhne J, Tangermann M. Towards User-Friendly Spelling with an Auditory Brain-Computer Interface: The CharStreamer Paradigm. PLOS ONE. 2014;9(6):e98322.
[20] Höhne J, Schreuder M, Blankertz B, Tangermann M. A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Frontiers in Neuroscience. 2011;5.
[21] Kleih SC, Herweg A, Kaufmann T, Staiger-Sälzer P, Gerstner N, Kübler A. The WIN-speller: A new intuitive auditory brain-computer interface spelling application. Frontiers in Neuroscience. 2015;9.
[22] Markovinović I, Vrankić M, Vlahinić S, Šverko Z. Design considerations for the auditory brain computer interface speller. Biomedical Signal Processing and Control. 2022;75:103546.

Published by
Verlag der Technischen Universität Graz