2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Performance Improvement of EEG-Based BCI Using Visual Feedback Based on Evaluation Scores Calculated by a Computer

Hikaru Sato[1], Aoi Yoshida[1], Takamasa Shimada[2], and Tadanori Fukami[1]

Abstract— **In the study of an electroencephalography (EEG)-based brain computer interface (BCI) using the P300, there have been many reports on computer algorithms that identify the target intended by a user from multiple candidates. However, because the P300 amplitude depends on the subject's condition and is attenuated by physical and mental factors, such as fatigue and motivation, the performance of the BCI is low. Therefore, we aim to improve performance by introducing a feedback mechanism that provides the user with an evaluation calculated by the computer during EEG measurement. In this study, we conducted an experiment in which the user input one character from four characters on the display. By changing the character size according to the evaluation score calculated by the computer, the computer's current evaluation was fed back to the user. This is expected to change the consciousness of the user to enable them to execute a task by knowing the evaluation; that is, if the evaluation is high, the user needs to maintain the current state, and if the evaluation is low, a behavioral change, such as increasing attention, is required to improve the evaluation.**

As a result of comparing 10 subjects with and without feedback, accuracy improved for seven subjects that were given feedback.

I. Introduction

The brain computer interface (BCI), which enables the input of information, such as characters, to a computer using electroencephalography (EEG), is attracting attention as a future user interface for all people, not only those who have difficulty communicating. The P300 is a typical EEG component used in EEG-based BCI. It has a transient potential that appears about 300 ms after the onset of sensory stimulus that results from imposing a task such as mental counting. Typically, the number of candidates to be input is limited, and they are randomly presented as sensory stimuli. Therefore, a target among them is specified by detecting or identifying the P300 generated by executing the task when the target stimulus is presented. Many algorithms have been proposed to input information at a high speed. Additionally, many studies have been conducted on the relationship between stimulus conditions and the P300. As an example, in the P300 speller, Ron-Angevin et al. conducted experiments by changing the parameters of character size, character spacing, and character display range[1]. Kirasirova et al. performed an experiment under visual field restriction by wearing a binocular aperture that confined their sight to the central visual field[2]. Other reports have been made on stimuli that increase user concentration[3] and the effects of matrix size and interstimulus interval on performance[4] in order to improve the performance of the P300-based BCI system.

Because the P300 is a component related to human judgment and cognition that reflects the user's intention using active task execution, its appearance depends on the user's condition. In particular, a decrease in motivation and fatigue caused by long-term use deteriorates input performance. Therefore, instead of aiming to improve identification accuracy using a computer algorithm, we attempted to develop the generation of the P300 to enable easy identification by increasing the motivation of the user. We believe that if the user in target determination is notified of the evaluation score calculated by the computer in real time, this will lead to the maintenance of the user's motivation and improvement of consciousness, and an effective P300 for identification could be obtained. Therefore, the stimulus presentation will change based on the evaluation score calculated by the computer. Some methods have been proposed for narrowing down the candidate characters (i.e., reducing the number of choices) using multiple steps[5]. If the character to be input is not included in the next stage, it is necessary to return to the previous stage. This method focuses on improving accuracy by reducing the number of candidates, but our method differs in that it focuses on changing the user's mind. Arvaneh et al. gave feedback by showing the correct letter to the user[6]. In their study, input characters are given in advance, and the classification result for the measured EEG is displayed as feedback. As a result, the subject recognizes whether the input is correct or not. However, our method differs from their study in that it does not show the output of the classifier to the user after EEG measurement, but shows the user the progress by changing the stimulus characteristics in real time.

In our previous study, as a first step, we attempted to identify two characters to ensure that the evaluation score feedback was valid. To inform the user of how the target character is being evaluated by the computer at the present time, we assigned red to the character with a higher evaluation score and blue to the other character with a lower score as feedback information. We confirmed that feedback stimulus presentation reflecting the evaluation score enhanced the P300 amplitude[7]. More color types are needed to convey more detailed score information to the user, but it is difficult to intuitively grasp the evaluation score with a slight difference in color. Therefore, in this study, we conducted an experiment by increasing the number of candidates to

[1]Hikaru Sato, Aoi Yoshida, and Tadanori Fukami are with the Graduate School of Science and Engineering, Yonezawa, Yamagata 992-8510, Japan {t212170m, t212014m}@st.yamagata-u.ac.jp, fukami@yz.yamagata-u.ac.jp

[2]Takamasa Shimada is with the School of System Design and Technology, Tokyo Denki University, Tokyo 120-8551, Japan tshimada@mail.dendai.ac.jp

four characters with feedback that changed the character size according to the evaluation score.

II. Method

A. EEG measurement

The EEG measurement system is shown in Fig.1. In this study, as shown in Fig.2, candidate characters are placed in four locations on the left, right, top, and bottom of the display. The subject inputs one target character specified by the experimenter in advance. The subject gazes at the character and performs the task of counting the number of times the character is highlighted. MATLAB and Psychtoolbox were used for EEG analysis and stimulus presentation.

EEGs were measured from a subject sitting in a chair 30 cm away from the display. The four candidate characters were randomly presented individually. In one cycle, all characters were presented once. Five measurement cycles were performed to input one target character. The duration of character presentation was 0.5 s and the presentation interval was 1 s.

Ten healthy males (age: 22.9±0.83 years) participated in this study. The EEG measurement was performed after informed consent from the participants and approval from the ethics committee of Yamagata University were obtained.

Using the g.tec electroencephalograph (g.MOBIlab+), electrodes (g.SAHARA) were placed at positions O_1 and O_2 using the international 10-20 method, and EEGs were measured at a sampling frequency of 256 Hz. The averaged waveform at the two electrodes was used for the analysis.

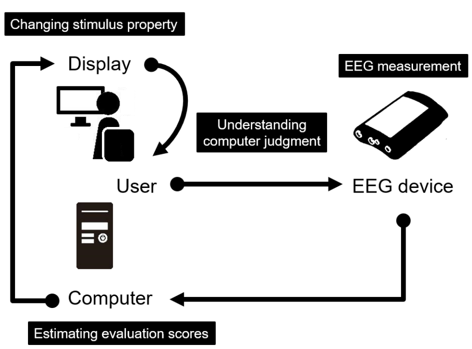

Fig. 1. EEG measurement system.

Fig. 2. Arrangement of characters

B. Feedback presentation of a visual stimulus

A bandpass filter that passes signal components between 2 and 7 Hz was applied to each response. Additionally, the P300 was detected from the averaged waveform of multiple responses. The evaluation score was defined as the peak amplitude of the P300, and the character with the maximum score was determined to be the target character. The P300 peak amplitude was simply set to the maximum value in the interval from 300 to 500 ms after stimulation for which the appearance of the P300 was expected. The evaluation score can be, for example, a feature obtained from machine learning, and higher accuracy may be obtained. However, in this study, we did not use any preprocessing tools other than the bandpass filter to evaluate only the appearance of the P300 that could be directly inspected.

We informed the subject of the evaluation using the computer by changing the character size. If the subject was able to understand the score information displayed on the computer, it was possible to change the motivation for task execution; that is, if the score of the target character that the subject intended to input was low, the subject increased his/her concentration to improve it, and conversely, if the score was high, the subject attempted to maintain that state.

An evaluation score was calculated for each candidate character and the character size was determined based on a ranking. As shown in Fig.3, the character size was set to three stages, and all characters were presented in the normal (middle) size in the first cycle because none of the characters had been evaluated. In the second cycle, the 1st place character was presented in a larger size (left), the 4th place character was presented in a smaller size (right), and the 2nd and 3rd place characters remain unchanged. Then the 1st and 4th place characters became one size larger and one size smaller, respectively, in the next cycle. One character was highlighted by the background color being set to red and the character color to white (Fig.4). The subject performed the task when the target character was highlighted. In the initial state (first cycle), all characters were of normal size, with the character arrangement shown in Fig.2. As an example, Fig.5 shows the stimulus presentation in the second cycle when character 'B' was in 1st place and 'C' was in 4th place in the first cycle, and 'B' was highlighted. The character size was updated each cycle. Finally, the 1st place character was input.

Experiments with and without feedback were performed on each subject 10 times. To eliminate the ordinal effect, in the two experiments, the first experiment and second experiment were swapped each time.

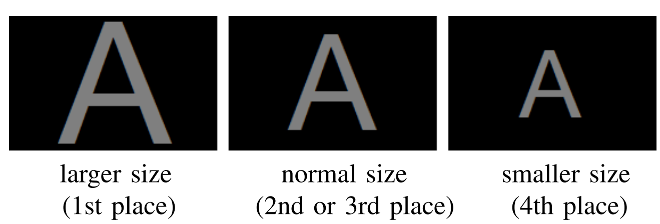

Fig. 3. Relationship between character size and ranking in the second cycle.

Fig. 4. Presentation of characters as a visual stimulus.

Fig. 5. Change of character size based on the evaluation score: the case in which 'B' is 1st place and 'C' is 4th place in the first cycle, and 'B' is presented.)

III. Results

First, Fig.6 shows an example of the waveforms used for calculating the evaluation scores. It shows the waveform for each character when the target character was 'C' for a certain subject. Because the bandpass filter was applied, the P300 peak was obtained stably. However, the latency was much more than 300 ms in many cases.

Accuracy comparisons were made with and without feedback. However, not all subjects showed the same tendency: seven people achieved better results with feedback, but three people achieved better results without feedback. Therefore, we divided the 10 subjects into two groups, calculated the accuracy, and considered those results. We show the relationship between the number of cycles and the mean accuracy for the seven subjects who were better in the experiment with feedback in Fig.7, and for the three subjects who were better without feedback in Fig.8. Because feedback was not applied in the first cycle, we show the results for two or more cycles. Here, when paired t-test was performed with the accuracy with and without feedback in 10 subjects, no significant difference appeared. However, significant differences were observed in the 5th cycle by dividing into seven subjects and three subjects.

Next, we show a typical example for each group of the relationship between the number of cycles and the evaluation score of the target character. Figures 9 and 10 show the results of two subjects who achieved better results with feedback and without feedback, respectively. In both cases, the graph shows the average value for 10 trials.

IV. Discussions

We divided the 10 subjects into a group with higher accuracy with feedback and a group with higher accuracy without feedback. Because the two groups showed completely different trends, as shown in Figs.7 and 8, we considered the results separately. In the group for which the accuracy was higher with feedback, the accuracy increased as the number

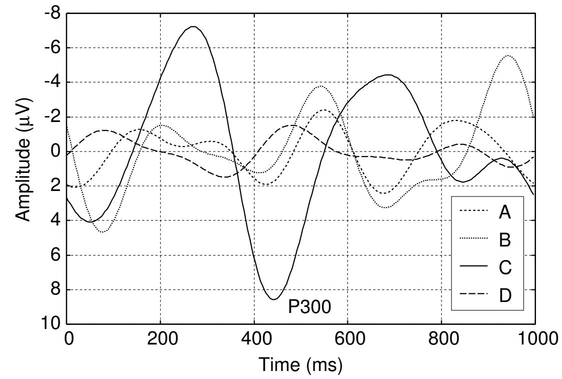

Fig. 6. Example of the waveform for each character when the target character is 'C' for a certain subject.

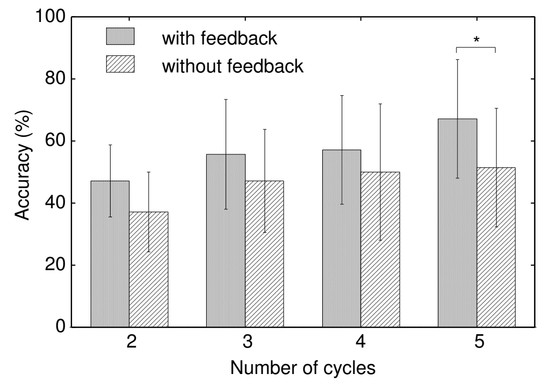

Fig. 7. Relationship between the number of cycles and the mean accuracy for the seven subjects who were better in the experiment with feedback. (*:$p < 0.05$)

of cycles increased when there was feedback, but the increase in accuracy stagnated when the number of cycles was three or more without feedback. The mean accuracies after the third cycle without feedback were 47.1%, 50.2%, and 51.4%. A post-experiment questionnaire was used to obtain various opinions from subjects. One subject commented as follows: "If the target character size was the largest, I was able to keep my motivation high and succeeded in inputting that correctly." Another subject commented about the feedback: "Even if the target character was the smallest at an early stage, I thought I would do my best in the next cycle to return to the initial size." Similar opinions were obtained from the other subjects who achieved better accuracy with feedback. These positive opinions were expected. We considered that the reason for the accuracy improvement when feedback was used was that the subject's decrease in motivation caused by anxiety about whether the input was successful could be prevented by the subject obtaining information about the evaluation scores during the measurement. Figure 9 shows a typical example, and the gradient of the evaluation score with respect to the number of cycles was very large when feedback was used.

By contrast, in the group with higher accuracy without feedback, the increase was particularly remarkable from two to three cycles, as shown in Fig.8; however, after that, it

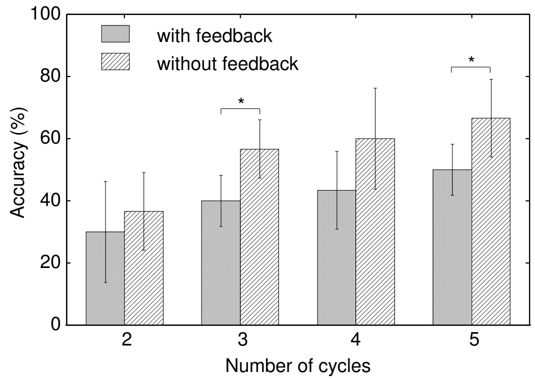

Fig. 8. Relationship between the number of cycles and the mean accuracy for the three subjects who were better in the experiment without feedback. (*:p <0.05)

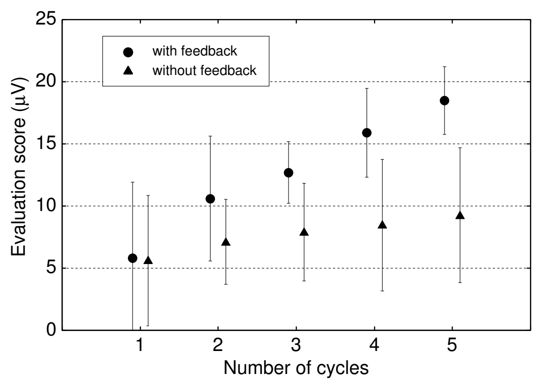

Fig. 9. Relationship between the number of cycles and the evaluation score for the one subject who was better in the experiment with feedback.

was almost the same as the increase with feedback. The subjects who belonged to this group commented on the feedback: "The target character became the largest size at least once, but then I was relieved" and "I unknowingly checked other characters." Because the amplitude of the P300 fluctuated depending on the subject's concentration on the task, we speculate that such a decrease in concentration may have led to a decrease in accuracy. Figure 10 shows an example. At an early stage, there was no substantial difference depending on the presence or absence of feedback, but as the number of cycles increased, the evaluation score with feedback gradually decreased.

In this study, the electrodes were placed on O_1 and O_2. This is because P300 with a relatively high amplitude appeared and was easy to fix. However, since it has been reported that the putamen, supplementary motor area, and superior temporal gyrus are related to attention to stimulation[8], we would like to examine the electrode position in the future.

V. Conclusions

In this paper, we attempted to improve input performance by changing the size of characters and feeding back evaluation scores calculated by a computer to enable subjects to change their consciousness. When accuracy was compared

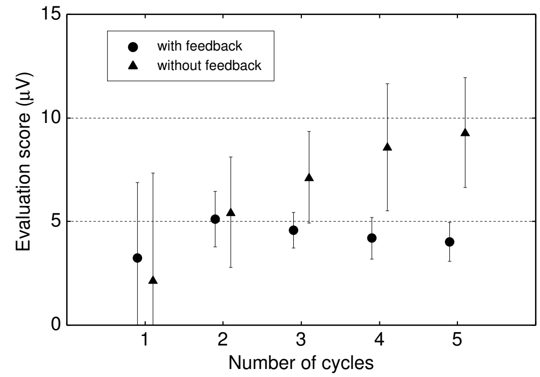

Fig. 10. Relationship between the number of cycles and the evaluation score for the one subject who was better in the experiment without feedback.

with and without feedback, 7 out of 10 subjects achieved better results with feedback. In the opinion of the subjects who achieved better results without feedback, some stated that that were distracted by changing the size of characters next to his/her target character, and that they lost their motivation because of the low evaluation of the target character. This result means that the feedback of evaluation scores is effective for many users, though not all.

In future research, we will increase the number of data, examine the effectiveness of this method in detail from a statistical point of view, and determine effective feedback for as many people as possible.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19K12077. We thank Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript.

References

[1] R. Ron-Angevin, L. Garcia, Á. Fernández-Rodríguez, J. Saracco, J. M. André and V. Lespinet-Najib, "Impact of speller size on a visual P300 braincomputer interface (BCI) system under two conditions of constraint for eye movement," Comput. Intell. Neurosci., 7876248, 2019.

[2] L. Kirasirova, V. Bulanov, A. Ossadtchi, A. Kolsanov, V. Pyatin and M. Lebedev, "A P300 brain-computer interface with a reduced visual field," Front Neurosci., vol.14, 604629, 2020.

[3] J. Jin, H. Zhang, I. Daly, X. Wang, A. Cichocki, "An improved P300 pattern in BCI to catch user's attention," J. Neural Eng., vol.14, no.3, 036001, 2017.

[4] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, J. R. Wolpaw, "A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance," Biol. Psychol. vol.73, no.3, pp.242–252, 2006.

[5] S. Ikegami, K. Takano, K. Kondo, N. Saeki and K. Kansaku, "A region-based two-step P300-based brain-computer interface for patients with amyotrophic lateral sclerosis," Clin. Neurophysiol., vol.125, pp.2305–2312, 2014.

[6] M. Arvaneh, I. H. Robertson, T. E. Ward, "A P300-based brain-computer interface for improving attention," Front Hum. Neurosci., vol.12, 524, 2019.

[7] T. Fukami and T. Shimada, "Enhancement of P300 amplitude by feedback stimulus presentation reflecting evaluation value in BCI," Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., WePoS–24.10, 2018.

[8] S. Halder, T. Leinfelder, S. M. Schulz, A. Kübler, "Neural mechanisms of training an auditory event-related potential task in a brain-computer interface context," Hum. Brain Mapp. vol.40, no.8, pp.2399-2412, 2019.