Does Real-Time Feedback Improve User Performance in SSVEP-based Brain-Computer Interfaces?

Jing Mu, Po-Chen Liu, David B. Grayden, Ying Tan, and Denny Oetomo

Abstract— **Offline and online experiments are both widely used in SSVEP-based BCI research and development for different purposes. One of the major differences between offline and online experiments is the existence of real-time feedback to the user while they are using the interface. However, the role of feedback in SSVEP-based BCIs has not yet been well studied. This work focuses on understanding the effect of feedback in SSVEP-based BCIs and if there exists any relationship between offline and online BCI performance. An experiment was designed to compare directly the accuracies of the BCI with and without feedback for participants. Results showed that feedback can improve performance in a complex task, but no clear improvement was observed in a simple task.**

I. Introduction

Brain-computer interfaces (BCIs) are alternative communication pathways between human and computer to replace or complement physical commands. The steady-state visual evoked potential (SSVEP) is one of the most commonly used brain signals in non-invasive BCIs. SSVEPs are reactive brain activities that respond to a periodically flickering visual stimulation, and is considered to have minimum user training requirement and higher decoding accuracy compared to other modalities such as motor imagery [1].

In SSVEP-based BCI research and development, two types of experiments have been used: offline experiments, where no feedback is provided to the user during the experiment, and online experiments, where users have access to real-time feedback (real-time feedback refers to instant feedback after each decoding cycle here) on their performance. Usually, offline experiments are performed to understand SSVEP responses from selected stimulation setups and online experiments are conducted to test and validate the performance of the designed interface.

Both offline and online experiments are used widely in SSVEP-based BCI research and sometimes both are included in a study. However, the role of feedback in SSVEP-based BCIs has not yet been carefully studied to understand how feedback to users during usage would affect SSVEP-based BCI performance and if offline experiment results can meaningfully imply online experiment results. Previous studies [2]–[4] presented results for both experiments without feedback and with feedback (offline and online experiments); however, how feedback affected each participant's result was not investigated. Researchers have also studied and compared the effectiveness of different visual feedback methods in SSVEP-based BCIs [5]. Even though a significant improvement in speed was observed when feedback was provided compared to no feedback conditions, it was a comparison between each subject's best performing feedback modality and the no feedback case. Moreover, the best performing feedback method varied from person to person [5].

This paper describes an experiment designed to compare directly the accuracies of the BCI when feedback was removed vs. provided to the participants. In order to further understand if the performance from the two cases and the difference between the two changes with the complexity of the interface, small (8 targets) and large (24 targets) interface setups were tested in the experiment. The results from the experiment were then analysed to understand the effect of feedback on performance and whether or not the performance from offline SSVEP-based BCI experiments implies online performance.

II. Methods

A. Experimental Setup

In this work, Unity (Unity Technologies, USA) was used to design the SSVEP interface and Simulink (The MathWorks, Inc., USA) was used to record the data in synchrony with the stimulation.

Participants sat on a chair 70 cm away from the screen (Alienware AW2518HF; 24.5 inch, 1920×1080) as shown in Fig. 1 with the centre of screen in the sagittal plane and the height of the chair adjusted to their comfort. The experiment was carried out in a normal engineering dry lab environment.

Stimulation consisted of either a large (4×6) or a small (2×4) matrix of square flashing targets that are of size 108 $\times$ 108 pixels each and targets were placed 108 pixels apart to avoid potential attenuation on the stimulation frequency [6]. These two layouts are referred to as large grid or small grid, respectively, in the rest of this paper. The square flashing targets were white in colour and patterned by a square wave with zero phase. Stimuli were presented at a 120 Hz refresh rate on the screen. For the large grid, the 24 frequencies ranged from 8 Hz to 12.6 Hz (inclusive) in steps of 0.2 Hz. For the small grid, the 8 frequencies ranged from 8 Hz to 9.4 Hz (inclusive) in steps of 0.2 Hz. In both cases, the frequencies were randomly shuffled between the targets.

A g.USBamp amplifier (g.tec medical engineering GmbH, Austria) at a sampling rate of 512 Hz was used to record

This work was supported by the Valma Angliss Trust.

J. Mu, P. Liu, Y. Tan, and D. Oetomo are with the Department of Mechanical Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.

D. B. Grayden is with the Department of Biomedical Engineering, The University of Melbourne, Parkville, VIC 3010, Australia.

j.mu@student.unimelb.edu.au, liupochen@gmail.com, {grayden, yingt, doetomo}@unimelb.edu.au

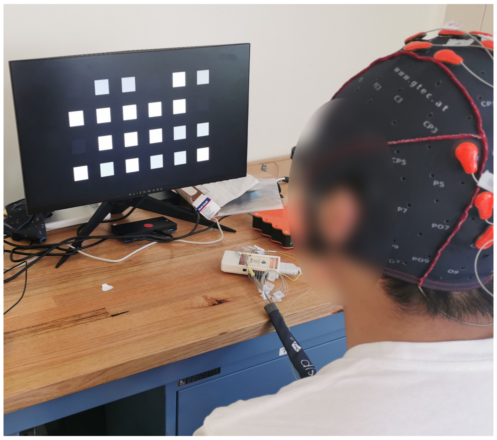


Fig. 1: Experimental setup. Participant sits 70 cm away from the screen where SSVEP targets are displayed. In this figure, the large grid (4 $\times$ 6) is being presented.

EEG data from 16 g.SAHARA dry electrodes positioned at Fz, FCz, FC1, FC2, Cz, C1, C2, Pz, P3, P4, PO3, POz, PO4, O1, Oz and O2 according to the international 10-10 system. The EEG signal was notch filtered at 50 Hz and 0.5 Hz - 100 Hz bandpass filtered on all channels during data recording.

B. Experimental Protocol

The experiment included four sessions with each session containing four tests, as shown in Fig. 3. For a fair comparison of performance under with and without feedback conditions, the four sessions were arranged using an ABBA protocol such that sessions 1 and 4 were without feedback and sessions 2 and 3 were with feedback. Each session contained two tests with large grid and two tests with small grid. The tests within the same session is also arranged in ABBA pattern and alternated with BAAB in different sessions. Breaks were provided after each test with varying durations as shown in Fig. 3.

Two setups were used in the tests: a small grid with 8 targets or a large grid with 24 targets. In each test, participants were asked to follow a cue to go through all targets on the screen one-by-one in a fixed sequence (from left to right, top to bottom), 1 trial per target. Therefore, each test contained 8 or 24 trials for the small grid or large grid test, respectively.

For the feedback sessions, each trial consisted of a 1 second cue, 3 seconds of stimulation, 1 second of rest and 1 second of feedback as depicted in Fig. 2. During the feedback presentation, the intended target turned to solid green or red to represent correct (green) or incorrect (red) decoding from the classifier. Trials in no feedback sessions had a 2 seconds rest after the stimulation instead. In the feedback sessions, a score was displayed on the screen by the end of the test with the score being the number of correctly decoded trials in the test.

C. Participants

Eight participants (2 females and 6 males) of age 20 - 26 years (22.5 $\pm$ 2.35) participated in this experiment. All

No feedback	1s Cue	3s Stimulation	1s Rest	
With feedback	1s Cue	3s Stimulation	1s Rest	1s Feedback

Fig. 2: Trial structure in no feedback (top row) and with feedback (bottom row) sessions.

Session 1: No feedback							
Small grid	*1min rest*	Large grid	*2min rest*	Large grid	*1min rest*	Small grid	*3min break*
Session 2: Feedback							
Large grid	*1min rest*	Small grid	*2min rest*	Small grid	*1min rest*	Large grid	*5min break*
Session 3: Feedback							
Small grid	*1min rest*	Large grid	*2min rest*	Large grid	*1min rest*	Small grid	*3min break*
Session 4: No feedback							
Large grid	*1min rest*	Small grid	*2min rest*	Small grid	*1min rest*	Large grid	End

Fig. 3: Experimental structure. Sessions 1 and 4 are no feedback sessions. Sessions 2 and 3 are feedback sessions. Each session contains 2 small grid and 2 large grid tests.

participants were right-handed with no known neurological diseases. Two participants (P7 and P8) have had previous BCI experience. This experiment was approved by the University of Melbourne Human Research Ethics Committee (Ethics ID 1851283). Written consent was collected from all participants prior to the experiment.

D. Data Processing and Analysis

In all sessions, data was processed and decoded in real-time with the entire 3 seconds of SSVEP recording. However, in no feedback sessions, decoding results were not shown to the participants. Canonical correlation analysis (CCA) was used as the decoding algorithm with up to the third harmonic [7]. Channels Pz, P3, P4, PO3, POz, PO4, O1, Oz and O2 were used in SSVEP processing and decoding.

III. Results

The results are presented in terms of score and accuracy in this paper. Scores were calculated as the number of trials correctly decoded in each test; therefore, the maximum score was 8 for small tests and 24 for large tests. Accuracies were calculated as the percentage of correct identifications in the total number of trials of a test. Information transfer rate (ITR) is not reported as the trial durations were consistent throughout the experiment.

A boxplot of the scores (number of correct trials in the test) from all tests and all participants is shown in Fig. 4. The top and bottom whiskers represent the maximum and minimum values, respectively. The tops and bottoms of the boxes represent the third and first quartiles of the data, respectively. The orange lines represent the median values. The

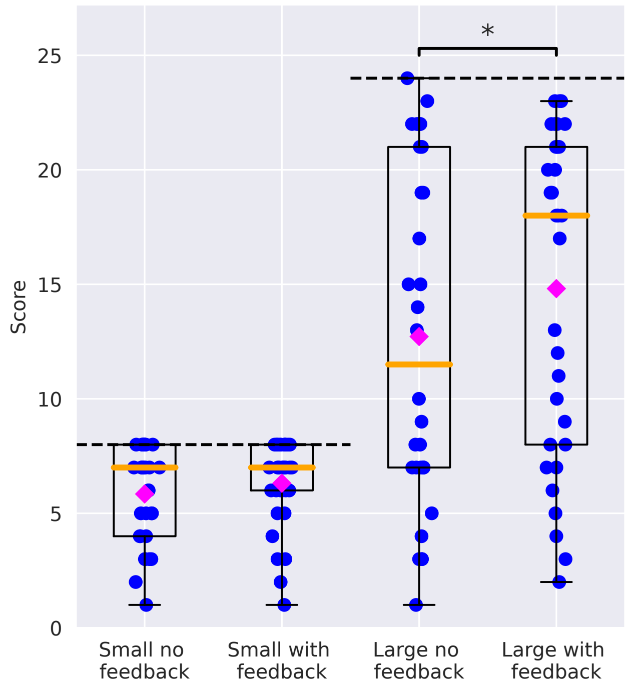


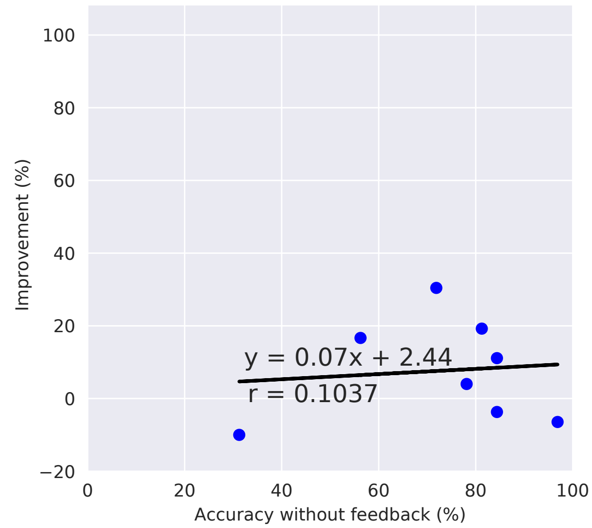
Fig. 4: Boxplots of scores for all participants and all tests. The orange line in each box represents the median, the magenta diamond represents the mean, and the black dashed lines indicate maximum possible scores (8 for the small grid and 24 for the large grid). The asterisk labels the significant improvement ($p = 0.0175$) found between large no feedback and large with feedback using two-way ANOVA.

magenta diamonds indicate mean values. Maximum scores achievable in small (8) and large (24) tests are indicated with black dashed lines.

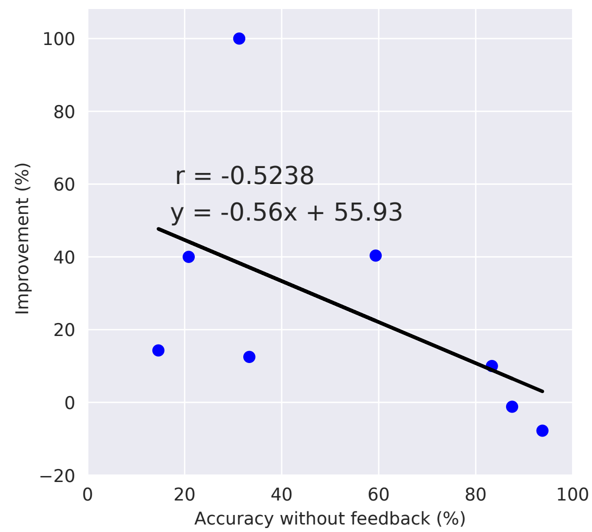
In small tests the median and mean scores were similar with and without feedback, however, in large tests the median score was higher with feedback. Statistically significant improvement was found in the large test with feedback compared to without feedback ($p = 0.0175$) using two-way ANOVA with setups and participants as factors. However, no significance was found in the small test ($p = 0.5906$).

Percentage of improvement was calculated as the difference between with feedback and no feedback accuracies divided by the accuracy without feedback. Fig. 5 shows the percentage of improvement plotted against the overall accuracy of tests without feedback for each participant. As shown in Fig. 5a, there was, in general, a positive improvement when feedback was introduced compared to without feedback in the small grid test, but the amount of improvement did not vary with accuracy without feedback (correlation coefficient $r = 0.1037$, $p = 0.8070$). In the large grid test, as shown in Fig. 5b, a general positive improvement is also observed; moreover, it was found that the percentage of improvement with feedback is negatively correlated (correlation coefficient $r = -0.5238$, $p = 0.1827$) to the accuracy without feedback.

Accuracies from all participants in all setups are summarised in Fig. 6. Heights of the bars represent the average



(a) Small grid.



(b) Large grid.

Fig. 5: Average accuracy of the participants for tests without feedback versus their respective improvement in percentage given tests with feedback for small (a) and large (b) grids.

accuracy from each participant in different setups. Error bars indicate standard error. The asterisks indicate statistical significance ($p < 0.05$) with Wilcoxon rank-sum test between no feedback and with feedback tests in each of the setups.

IV. Discussion

The results showed that there was a statistically significant improvement between the population performance of the large grid SSVEP-based BCI when feedback was hidden or provided to the participants. However, the performance varied between the individuals: 2 out of 8 participants showed significant improvement in large grid tests with feedback compared to without feedback; in the 6 with no significance, 2 showed some decrease and 4 showed some increase in accuracy. No significant differences were observed in small grid tests with or without feedback. But, in general, participants tended to perform better (positive improvement) when

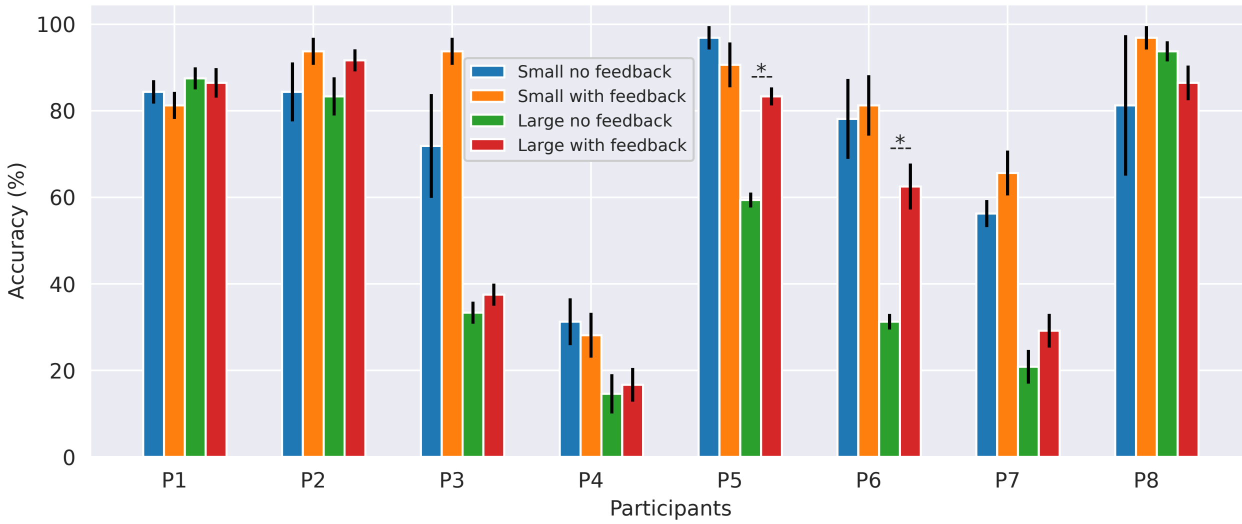


Fig. 6: Bar charts for each participant comparing feedback vs no feedback tests. Heights of the bars represent the average accuracy for each participant in different setups. Error bars label standard error. The asterisks indicate the significance with Wilcoxon rank-sum test (*: $p < 0.05$).

feedback was provided during the experiments (Fig. 5).

It was also observed from the large tests that the percentage of improvement in performance when feedback was provided was negatively correlated to the participant's performance (accuracy) without feedback.

Overall, these results suggest that offline (without feedback) experimental results can well represent online (with feedback) experiment results when the task is relatively simple (small grid). In a more complex task (large grid), significant difference was found between the performance between offline and online experiments, even though the mean and spread of offline experimental results does not differ much from online results when focusing on population performance (Fig. 4).

A. Differences Between Small Grid and Large Grid Tests

Figs. 5a and 5b show different trends in percentage of improvement in performance with feedback vs. performance without feedback in small and large grid tests, respectively. A moderate negative correlation was found from the large grid tests, but not from the small grid tests. This means that, in large grid tests, participants who performed well without feedback tended to have a smaller improvement in accuracy with feedback compared to those who did not perform well without feedback. Except for the fact that the high-performers have lesser room for improvement, we believe that feedback helped participants to focus on the task, which resulted in more relative improvement for the low-performing participants. The reason for the small grid test not showing a similar trend is potentially because the small grid test only had 8 targets, which made the task easier and shorter in time, such that the focus level of the participants did not change as much during the test as in the large grid test. Note that with a sample size of 8, it is difficult to get significance statistically and only the general trend observed from this 8 samples is discussed here.

B. Future Work

Due to COVID-19 restrictions, only 8 participants were recruited in this study to show a preliminary result. A larger scale experiment would be desirable to show more comprehensive results.

V. Conclusion

In this paper, experiments were done under the conditions where real-time feedback was hidden or presented to the user to investigate the role of feedback in SSVEP-based BCIs and the relationship between the performance of offline and online experiments. Results show that providing feedback to the user can improve performance in non-trivial tasks and offline performance can well represent online performance when focusing on population results. However, each individual responds differently to the feedback.

References

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[2] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proceedings of the National Academy of Sciences*, vol. 112, no. 44, pp. E6058–E6067, 2015.

[3] S. Ge, Y. Jiang, M. Zhang, R. Wang, K. Iramina, P. Lin, Y. Leng, H. Wang, and W. Zheng, "SSVEP-based brain-computer interface with a limited number of frequencies based on dual-frequency biased coding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 760–769, 2021.

[4] G. Ming, W. Pei, H. Chen, X. Gao, and Y. Wang, "Optimizing spatial properties of a new checkerboard-like visual stimulus for user-friendly SSVEP-based BCIs," *Journal of Neural Engineering*, vol. 18, no. 5, p. 056046, 2021.

[5] M. Benda and I. Volosyak, "Comparison of different visual feedback methods for SSVEP-based BCIs," *Brain Sciences*, vol. 10, no. 4, p. 240, 2020.

[6] J. Mu, D. B. Grayden, Y. Tan, and D. Oetomo, "Spatial resolution of visual stimuli in SSVEP-based brain-computer interface," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2019, pp. 928–932.

[7] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2610–2614, 2006.