

# A Comparison Study of Single- and Multiple-Target Stimulation Methods for Eliciting Steady-State Visual Evoked Potentials

Chuyang Xiao, Kuan-Jung Chiang, *Student Member, IEEE*,  
Masaki Nakanishi, *Senior Member, IEEE* and Tzyy-Ping Jung, *Fellow, IEEE*

**Abstract**—A visual stimulator plays an important role in a steady-state visual evoked potential (SSVEP)-based brain-computer interface (BCI). In conventional BCI studies, SSVEPs have been elicited by either a single stimulus whose flickering frequency varies across trials or multiple stimuli flickering at different frequencies simultaneously. It has been implicitly assumed that the SSVEPs generated by the single- and multiple-target stimulation methods are comparable. However, no study has directly compared their efficacy in eliciting SSVEPs. This study, therefore, performed a quantitative comparison of signal-to-noise ratio (SNR) and classification accuracy using 4-class SSVEPs generated by these two methods. The classification accuracy was estimated by three commonly-used target identification algorithms including calibration-free canonical correlation analysis (CCA)-based method and template-based methods with CCA- and task-related component analysis (TRCA)-based spatial filters. The results showed that the single-target stimulation method led to significantly higher SNR and classification accuracy than its multi-target counterpart.

## I. INTRODUCTION

A steady-state visually evoked potential (SSVEP) is the brain's electrical response to repetitive visual stimulation [1]. The SSVEP has been widely used in implementing an electroencephalogram (EEG)-based brain-computer interface (BCI) due to its high communication rate and little user training [2]. In an SSVEP-based BCI, users are asked to gaze at one of the multiple visual stimuli, which are modulated by different frequencies and/or phases, presented on a computer monitor. By analyzing the elicited SSVEPs, the target stimulus, which the user is gazing at, could be identified. In this way, an SSVEP-based BCI can translate users' intention into commands to control external devices. The performance of an SSVEP-based BCI has been rapidly increasing in the past decade [3], [4].

The visual stimulator plays a vital role in an SSVEP-based BCI. Visual stimuli can be presented using flashing light-emitting diodes (LEDs) or flickers on a computer monitor. Most of the recent studies employ a computer monitor as a visual stimulator, which enables us to implement and configure visual-stimulation parameters flexibly without relying on any hardware modifications [5]. With the advances

in stimulus-presentation techniques, the number of visual stimuli that can be presented on a computer monitor has been significantly increased, leading to a large number of BCI commands. For instance, frequency approximation methods made it possible to render robust SSVEPs at flexible frequencies with a high-frequency resolution [6]–[8]. More recently, hybrid frequency and phase modulation methods have been widely used in implementing multi-command SSVEP-based BCIs [3], [4], [9].

In early proof-of-concept studies, it used to be common to present a single visual stimulus at a time, which is modulated by several different frequencies across trials, to a subject and apply target identification algorithms to the sequentially-recorded SSVEPs to investigate classification accuracy [5], [7], [10]. More recently, as a computer-monitor-based stimulator becomes popular, researchers started to present multiple visual stimuli simultaneously, whose layout is designed according to their target applications, to assess its simulated online performance [3]. In the literature, it has been implicitly assumed that the two stimulation approaches – single-target and multiple-target stimulation methods – are comparable in their ability to elicit SSVEPs. However, no study has performed a systematic and quantitative comparison between the two approaches. Therefore, it remains unclear whether the results obtained in previous studies using the two approaches can be directly translated to each other.

This study aims to perform a quantitative comparison of the single-target and multiple-target stimulation approaches. To this end, we collected 4-class frequency-modulated SSVEPs elicited by the two types of stimulus presentation approaches from eight subjects. This study compared the amplitude, signal-to-noise ratio (SNR), and frequency-detection accuracy of the elicited SSVEPs. The frequency detection accuracy was assessed by three popularly-used algorithms: 1) calibration-free canonical correlation analysis (CCA)-based method, 2) template-based method with CCA-based spatial filtering, and 3) template-based method with task-related component analysis (TRCA)-based spatial filtering.

## II. METHODS

### A. Experiment

1) *Subjects*: Eight healthy adults with normal or corrected-to-normal vision participated in this experiment. Each participant read and signed an informed consent form approved by the Human Research Protections Program of University of California San Diego before the experiment.

\*Research supported in part by the Army Research Laboratory under Grants W911NF-10-2-0022 and W911NF2020088, and the US National Science Foundation (NSF) under Grant CBET-1935860. Chiang's and Jung's participation was funded by a contract from Microsoft Corporation. (C. Xiao and K. -J. Chiang contributed equally to this work.)

C. Xiao, M. Nakanishi, K. -J. Chiang and T. -P. Jung are with Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA 92093, USA. (e-mail: masaki@sccn.ucsd.edu).



Fig. 1. Illustrations of the two stimulus layouts: (a) single-target stimulus presentation, and (b) multi-target stimulus presentation.

2) *Stimulus Design*: The experiment consisted of two stages. In both stages, visual stimuli were presented on a virtual-reality (VR) head-mounted display (HMD) Odyssey (Samsung Electronics Co., Ltd.). In the first stage, the single-target stimulation method, in which a single visual stimulus was presented at the center of the screen, was tested (Fig. 1(a)). The participants were asked to gaze at the visual stimulus that was modulated by four frequencies (9, 10, 11, and 12 Hz) assigned randomly in every single trial. The subjects performed twelve trials (4 frequencies  $\times$  3 times) of 2-s gazing tasks with an interval of 1s in a block and completed three blocks. In the second stage, the multi-target stimulus presentation was tested. That is, four visual stimuli modulated by the four frequencies were presented simultaneously (Fig. 1(b)). The subjects were asked to gaze at a target stimulus indicated by the stimulus program in a random order for 2s. At the beginning of each trial, a pink square marker appeared for 1s at the position of the target stimulus. As is the case in the single-target stimulation method, the subjects completed three blocks consisting of twelve trials. After each block, there was a short break lasting for several minutes. Therefore, each subject completed 36 trials (4 frequencies  $\times$  9 trials) for each stimulation method. The stimulus program was developed using a Unity game engine (Unity Technologies).

3) *EEG Recording*: EEG data were recorded with the ten Ag/AgCl electrodes covering the occipital area (Pz, PPO1h, PPO2h, PO3, POz, PO4, POO1, POO2, and Oz) and Cz using a BioSemi ActiveTwo system (Biosemi, Inc.). EEG data were digitized at a sampling rate of 2,048 Hz. The VR headset was put over the EEG cap. The event triggers that indicated the onsets of visual stimulation were sent from a parallel port of a computer to the EEG amplifier and recorded on an event channel synchronized to the EEG data.

## B. Target Identification Algorithms

1) *Preprocessing*: The EEG data were epoched with a 1-s window starting from 0.15 s after the stimulus onset of each trial. The nine Occipital channels mentioned in II-A.3 were re-referenced to the channel Cz and down-sampled to 256 Hz. Finally, a band-pass filter from 4 Hz to 120 Hz was applied.

2) *CCA-based method*: CCA has been widely used to detect the frequency of SSVEPs [11], [12]. In an SSVEP-based

BCI, CCA measures the underlying correlation between multi-channel EEG recordings  $\mathbf{X} \in \mathbb{R}^{N_c \times N_s}$  and reference signals for  $n$ -th stimulus  $\mathbf{Y}_n \in \mathbb{R}^{N_h \times N_s}$  by finding the weight vectors  $\mathbf{w}_x$  and  $\mathbf{w}_y$ , which maximize the correlation between those projections as follows:

$$r_n = \max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\mathbf{w}_x^T \mathbf{X} \mathbf{Y}_n^T \mathbf{w}_y]}{\sqrt{E[\mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x] E[\mathbf{w}_y^T \mathbf{Y}_n \mathbf{Y}_n^T \mathbf{w}_y]}}. \quad (1)$$

Here,  $N_c$  is the number of channels,  $N_s$  is the number of sampling points per epoch, and  $N_h$  is the number of harmonics ( $N_h = 3$  in this study) to be considered in the analysis. The reference signals are constructed by sine-cosine waveforms as:

$$\mathbf{Y}_n = \begin{bmatrix} \sin(2\pi f_n t) \\ \cos(2\pi f_n t) \\ \vdots \\ \sin(2\pi N_h f_n t) \\ \cos(2\pi N_h f_n t) \end{bmatrix}, t = \frac{1}{F_s}, \frac{2}{F_s}, \dots, \frac{N_s}{F_s}, \quad (2)$$

where  $F_s$  refers to the sampling rate and  $f_n$  is the stimulation frequency. The frequency of the reference signals which maximize the correlation  $r_n$  was selected as the frequency of SSVEPs.

3) *Template-based method*: The template-based target identification algorithm was first proposed in our previous study [9]. When using the method, users first need to collect individual training data for each visual stimulus to construct spatial filters and templates. The individual training data for  $n$ -th stimulus are denoted by a three-way tensor  $\chi_n = (\chi_n)_{jkh} \in \mathbb{R}^{N_c \times N_s \times N_t}$ . Here,  $j$  indicates the channel index,  $k$  indicates the index of sample points,  $h$  indicates the index of training trials, and  $N_t$  is the number of training trials. For each visual stimulus, individual templates  $\bar{\chi}_n \in \mathbb{R}^{N_c \times N_s}$  can be obtained by averaging multiple training trials as  $(\bar{\chi}_n)_{jk} = \frac{1}{N_t} \sum_{h=1}^{N_t} (\chi_n)_{jkh}$ . Spatial filters  $\mathbf{w}_n \in \mathbb{R}^{N_c}$  can also be derived from the training data for each stimulus as described later, and then they are applied to the individual templates and input EEG data  $\mathbf{X}$  to enhance the SNR of SSVEPs. After spatial filtering, the correlations  $r_n$  between the input data and templates are calculated as feature values as follows:

$$r_n = \rho(\mathbf{X}^T \mathbf{w}_n, \bar{\chi}_n^T \mathbf{w}_y), \quad (3)$$

where  $\rho(a, b)$  indicates the Pearson's correlation analysis between two signals  $a$  and  $b$ . The template that maximizes the correlation  $r_n$  is selected as the one corresponding to the target stimulus.

In an SSVEP-based BCI, the effectiveness of the template-based method can be further improved by applying spatial filters. In recent studies, the CCA- and TRCA-based spatial filters have been widely used [4], [9], [13]. The CCA-based spatial filters can be derived as  $\mathbf{w}_x$  which is obtained when inputting the individual templates  $\bar{\chi}_n$  into the equation (1) instead of  $\mathbf{X}$ . TRCA is the method that extracts spatial filters that maximize the reproducibility during the task period [4], [14]. Considering an observed multi-channel EEG signal  $\mathbf{x}(t) \in \mathbb{R}^{N_c}$ , the TRCA finds a linear coefficient vector

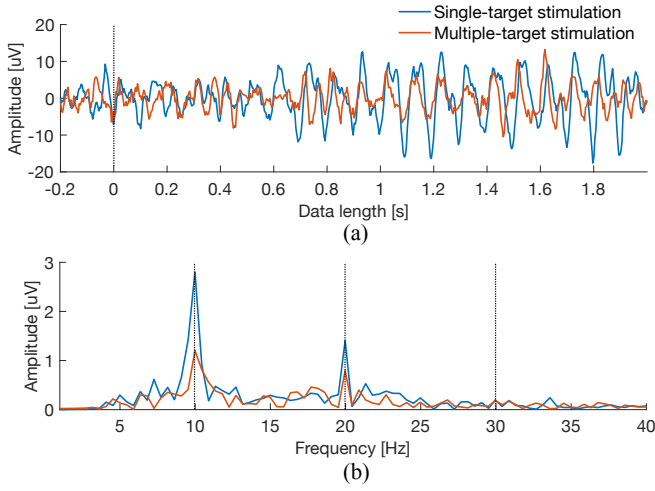


Fig. 2. (a) Time series and (b) amplitude spectra of SSVEPs elicited by a visual flicker at 10 Hz presented using the two stimulation methods from a subject.

$\mathbf{w} \in \mathbb{R}^{N_c}$  to maximize inter-trial correlation of its projections  $t(t) = \mathbf{w}^T \mathbf{x}(t)$ , which is called a task-related component. The  $h$ -th trial in the observed EEG signals and task-related components are described as  $\mathbf{x}^{(h)} \in \mathbb{R}^{N_c \times N_s}$  and  $\mathbf{y}^{(h)} \in \mathbb{R}^{N_s}$ , respectively. The covariance  $C_{h_1, h_2}$  between  $h_1$ -th and  $h_2$ -th trials of  $\mathbf{y}^{(h)}$  is described as:

$$\begin{aligned} C_{h_1, h_2} &= \text{Cov}(\mathbf{y}^{(h_1)}, \mathbf{y}^{(h_2)}) \\ &= \sum_{j_1, j_2=1}^{N_c} w_{j_1} w_{j_2} \text{Cov}(\mathbf{x}_{j_1}^{(h_1)}, \mathbf{x}_{j_2}^{(h_2)}). \end{aligned} \quad (4)$$

All possible combination of  $N_t$  trials are summed as:

$$\begin{aligned} \sum_{\substack{h_1, h_2=1 \\ h_1 \neq h_2}}^{N_t} C_{h_1, h_2} &= \sum_{\substack{h_1, h_2=1 \\ h_1 \neq h_2}}^{N_t} \sum_{j_1, j_2=1}^{N_c} w_{j_1} w_{j_2} \text{Cov}(\mathbf{x}_{j_1}^{(h_1)}, \mathbf{x}_{j_2}^{(h_2)}) \\ &= \mathbf{w}^T \mathbf{S} \mathbf{w}. \end{aligned} \quad (5)$$

Here, the matrix  $\mathbf{S} \in \mathbb{R}^{N_c \times N_c}$  is defined as:

$$(S)_{j_1, j_2} = \sum_{\substack{h_1, h_2=1 \\ h_1 \neq h_2}}^{N_t} \text{Cov}(\mathbf{x}_{j_1}^{(h_1)}, \mathbf{x}_{j_2}^{(h_2)}). \quad (6)$$

To obtain a finite solution, the variance of  $y(t)$  is constrained as:

$$\begin{aligned} \text{Var}(y(t)) &= \sum_{j_1, j_2=1}^{N_c} w_{j_1} w_{j_2} \text{Cov}(\mathbf{x}_{j_1}(t), \mathbf{x}_{j_2}(t)) \\ &= \mathbf{w}^T \mathbf{Q} \mathbf{w} = 1. \end{aligned} \quad (7)$$

The constrained optimization problem can be solved as:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmax}} \frac{\mathbf{w}^T \mathbf{S} \mathbf{w}}{\mathbf{w}^T \mathbf{Q} \mathbf{w}} \quad (8)$$

The optimal coefficient vector is obtained as the eigenvector of the matrix  $\mathbf{Q}^{-1} \mathbf{S}$ . The eigenvector corresponding to the largest eigenvalue is selected as the spatial filter to extract task-related components.

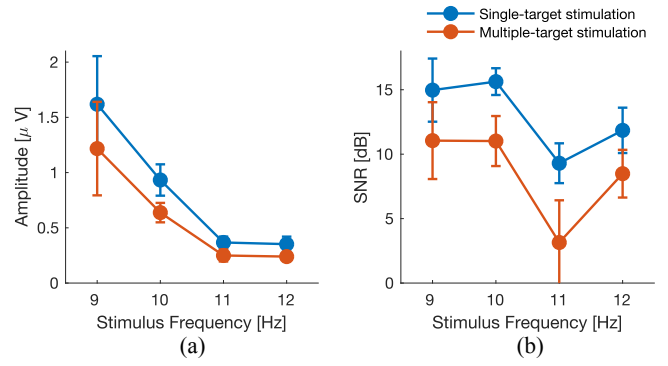


Fig. 3. Averaged (a) amplitude and (b) SNR of SSVEPs across subjects elicited by the two stimulation methods.

### C. Data analysis

This study first compared the signal characteristics of SSVEPs such as their amplitudes and SNRs between the two stimulus presentation methods. The SNR was computed as the ratio of the amplitude of SSVEPs to the mean amplitude of neighboring frequencies with a range of 2 Hz [5]. In addition, this study used the aforementioned three algorithms (i.e., the CCA-based method, the template-based method with CCA- and TRCA-based spatial filtering) with a filter bank analysis (the number of sub-bands = 5) [15] to identify target stimuli. The classification accuracy was estimated by a leave-one-out cross validation (LOOCV) except for the calibration-free CCA-based method. This study also employed factorial repeated measures analysis of variances (ANOVAs) to investigate the statistical difference in classification accuracy between the two visual stimulation methods.

## III. RESULTS

Fig. 2 depicts an example of time series and amplitude spectra of SSVEPs elicited by the single-target and multiple-target stimulation methods, and Fig. 3 shows averaged amplitude and SNRs of SSVEPs across subjects. In general, the single-target stimulation method led to higher amplitude and SNR than the multiple-target stimulation method. Two (stimulation approaches)  $\times$  four (stimulus frequencies)  $\times$  four (harmonics) three-way ANOVAs showed significant main effect of stimulation approaches in amplitude ( $F(1, 233) = 77.77, p < 0.001$ ) and in SNR ( $F(1, 233) = 13.04, p < 0.001$ ).

Fig. 4 shows the averaged classification accuracy across subjects using the three methods as functions of data length for the two stimulus-presentation methods. Regardless of the target identification algorithms, the single-target stimulation method achieved higher classification accuracy than the multiple-target stimulation method. Indeed, a two (stimulation approaches)  $\times$  three (target identification algorithms)  $\times$  ten (data lengths) three-way ANOVA showed significant main effect of stimulation approaches ( $F(1, 438) = 31.82, p < 0.001$ ).

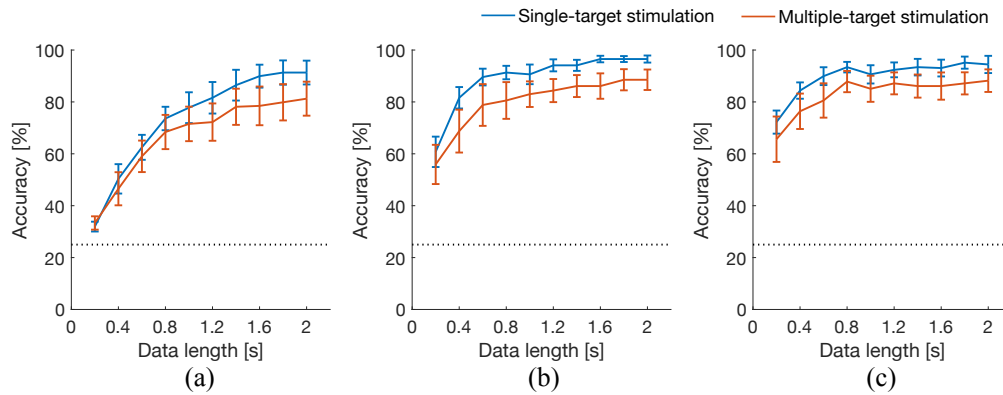


Fig. 4. The averaged classification accuracy of SSVEPs generated by the single- and multiple-target stimulus presentation across subjects estimated by (a) the CCA-based method, (b) the template-based method with CCA-based spatial filtering, and (c) the template-based method with TRCA-spatial filtering. The error bars indicate the standard errors. The dashed lines indicate the chance-level accuracy (i.e., 25%).

#### IV. DISCUSSIONS AND CONCLUSIONS

To the best of our knowledge, this preliminary work is the first study to quantitatively investigate the effect of using single-target or multi-target stimulus presentation. The study results (Fig. 4) revealed that the single-target stimulation method can elicit SSVEPs with significantly higher SNR than the multiple-target stimulation method, leading to higher classification performance in all the three algorithms. This difference in SNR might be because the subjects were less distracted by the surrounding non-target stimuli under the single-target condition. The time series of SSVEPs in Fig. 2(a) clearly showed higher amplitudes of rhythmic signals (i.e., SSVEPs) under the single-target condition. The spectral amplitude of the SSVEPs in Fig. 2(b) also showed that the spectra had greater peaks at both fundamental and second harmonic frequencies (i.e., 10 and 20 Hz) under the single-target condition than under the multi-target one. Note that, in practical applications, multiple visual stimuli need to be presented simultaneously to provide users multiple BCI commands. Therefore, the SNR and classification accuracy of SSVEPs elicited by the single-target stimulation method cannot be an indicator of those in real-world applications. Nonetheless, since the single-target stimulation method can elicit SSVEPs with high SNR, this approach could be employed to acquire better training data for training machine-learning models than the multiple-target stimulation method. In our future work, we will investigate whether such machine-learning models could enhance the classification accuracy in an online BCI experiment.

The main limitation of this study is that the order of the two stages in the experiment was fixed. That is, the stage of the single-target stimulation was always used in the first half of the experiment, and the multi-target one was always at the later half. Therefore, the difference in the SNR of SSVEPs between the methods might be caused by the change of fatigue level, motivation, etc. Our future work will randomize the order of the two conditions. We also plan to explore the effect of the size and layout of visual stimuli on the SNR of SSVEPs and BCI performance.

#### REFERENCES

- [1] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: focus on essential paradigms and future perspectives," *Prog. Neurobiol.*, vol. 90, no. 4, pp. 418–438, Apr. 2010.
- [2] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, 2008.
- [3] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain-computer interface," *Proc. Nat. Acad. Sci. U. S. A.*, vol. 112, no. 44, pp. E6058–E6067, 2015.
- [4] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, 2018.
- [5] X. Chen, Y. Wang, S. Zhang, S. Xu, and X. Gao, "Effects of stimulation frequency and stimulation waveform on steady-state visual evoked potentials using a computer monitor," *J. Neural Eng.*, vol. 16, no. 6, p. 066007, Oct. 2019.
- [6] Y. Wang, Y.-T. Wang, and T.-P. Jung, "Visual stimulus design for high-rate SSVEP BCI," *Electron. Lett.*, vol. 46, no. 15, pp. 1057–1058, 2010.
- [7] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, "Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate," *PLoS One*, vol. 9, no. 6, p. e99235, 2014.
- [8] X. Chen, Z. Chen, S. Gao, and X. Gao, "A high-ITR SSVEP-based BCI speller," *Brain-Comp. Interf.*, vol. 1, no. 3–4, pp. 181–191, 2014.
- [9] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 06, p. 1450019, 2014.
- [10] H. Bakardjian, T. Tanaka, and A. Cichocki, "Emotional faces boost up steady-state visual responses for brain-computer interface," *Neurorep.*, vol. 22, no. 3, pp. 121–125, 2011.
- [11] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for ssvep-based bcis," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, 2006.
- [12] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method," *J. Neural Eng.*, vol. 6, no. 4, p. 046002, 2009.
- [13] M. Nakanishi, Y. Wang, and T. P. Jung, "Spatial filtering techniques for improving individual template-based SSVEP detection," in *Signal Processing and Machine Learning for Brain-Machine Interfaces*, T. Tanaka and M. Arvaneh, Eds. The Institution of Engineering and Technology, 2018, pp. 219–242.
- [14] H. Tanaka, T. Katura, and H. Sato, "Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data," *NeuroImage*, vol. 64, pp. 308–327, 2013.
- [15] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, 2015.