Robustness analysis of decoding SSVEPs in humans with head movements using a moving visual flicker

Suguru Kanoga,* Masaki Nakanishi, Akihiko Murai, Mitsunori Tada and Atsunori Kanemura

1 Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 135-0064, Japan
2 Swartz Center for Computational Neuroscience, University of California San Diego, La Jolla, CA 92093, United States of America
3 Human Augmentation Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Chiba 277-0882, Japan
4 LeapMind Inc., Tokyo 130-0044, Japan
5 Author to whom any correspondence should be addressed.

E-mail: s.kanouga@aist.go.jp

Keywords: SSVEP, BCI, muscular artifact, EEG, spatial filter

Abstract

Objective. The emergence of mobile electroencephalogram (EEG) platforms have expanded the use cases of brain–computer interfaces (BCIs) from laboratory-oriented experiments to our daily life. In challenging situations where humans’ natural behaviors such as head movements are unrestrained, various artifacts could deteriorate the performance of BCI applications. This paper explored the effect of muscular artifacts generated by participants’ head movements on the signal characteristics and classification performance of steady-state visual evoked potentials (SSVEPs).

Approach. A moving visual flicker was employed to induce not only SSVEPs but also horizontal and vertical head movements at controlled speeds, leading to acquiring EEG signals with intensity-manipulated muscular artifacts. To properly induce neck muscular activities, a laser light was attached to participants’ heads to give visual feedback; the laser light indicates the direction of the head independently from eye movements. The visual stimulus was also modulated by four distinct frequencies (10, 11, 12, and 13 Hz). The amplitude and signal-to-noise ratio (SNR) were estimated to quantify the effects of head movements on the signal characteristics of the elicited SSVEPs. The frequency identification accuracy was also estimated by using well-established decoding algorithms including calibration-free and fully-calibrated approaches.

Main results. The amplitude and SNR of SSVEPs tended to deteriorate when the participants moved their heads, and this tendency was significantly stronger in the vertical head movements than in the horizontal movements. The frequency identification accuracy also deteriorated in proportion to the speed of head movements. Importantly, the accuracy was significantly higher than its chance-level regardless of the level of artifact contamination and algorithms. Significance. The results suggested the feasibility of decoding SSVEPs in humans freely moving their head directions, facilitating the real-world applications of mobile BCIs.

1. Introduction

An electroencephalogram (EEG)-based brain–computer interface (BCI) has been receiving a lot of attention as a next-generation communication channel between human brains and external environments [1]. Many researchers have put in significant efforts in proof-of-concept studies to increase the performance and reliability of BCIs toward various applications. Emerging technologies such as wearable sensing, mobile computing, and virtual/augmented reality (VR/AR) have shown great potentials for bridging a gap between laboratory-oriented BCI experiments and its real-world applications. In a traditional experimental setting, EEG data are acquired by using a bulky, expensive, and wired system requiring skin preparation with conductive gels. Recent studies demonstrated that non-prep dry electrodes made it possible to measure high-quality EEG signals without the laborious skin preparation [2, 3]. Several other studies have validated the feasibility of using mobile platforms (e.g. smartphones [4] and microcontrollers [5, 6]) with wireless/wearable EEG sensors to implement truly portable online BCI systems. Furthermore, a head-mounted VR display (HMD) has been successfully integrated with a mobile EEG
device, enabling to build an all-in-one closed-loop BCI system consisting of visual/auditory stimulation, data recording/processing, and feedback [7]. Such advances have significantly mitigated the restriction of conventional EEG recordings, and explosively expanded the use cases of BCI systems.

A steady-state visual evoked potential (SSVEP), which is the brain’s electrical response to repetitive visual stimulus, has been one of the most popular event-related potentials (ERPs) in implementing EEG-based BCIs [8, 9]. In the human visual cortex, the firing of neurons synchronizes to the flickering stimulation, resulting in SSVEPs characterized by sinusoidal-like waveforms at the same frequency as the stimulus and its harmonics [10, 11]. The frequency-tagging technique, which encodes multiple visual stimuli with different flickering frequencies, has been widely used in designing an SSVEP-based BCI [12]. By finding the dominant frequency component in elicited SSVEPs, the target stimulus, which a user is gazing at, can be reliably identified. In this way, the system can indirectly translate users’ intentions into commands tagged with the target stimulus for controlling external devices. Due to its robust characteristics, an SSVEP has attracted many researchers for developing a high-performance BCI system. With advances in multiple stimulus coding and target identification algorithms, the performance of SSVEP-based BCIs has been dramatically improved with a linear trend of an information transfer rate (ITR) [13, 14]. The research progresses in the past decade boosted an ITR from around 30 bits min \(^{-1}\) to over 300 bits min \(^{-1}\) [14].

Despite the considerable advances in an SSVEP-based BCI, sparse studies have explored the capability of using mobile BCI platforms in realistic environments accompanied by natural human behaviors [15–18]. For instance, Lin et al. validated the feasibility of detecting SSVEPs from participants freely walking on a treadmill using a wireless EEG headset [17], although their quality was deteriorated inversely proportional to the walking speed [18]. In another study, Yao et al. replicated the Lin et al.’s study using an HMD headset, and demonstrated an online SSVEP-based BCI system in a mobile VR environment [19]. However, little is known about the quantitative effects of other natural human behaviors on SSVEPs and the difficulty of extracting the SSVEPs from the contaminated EEG signals. For example, especially in a mobile VR environment, it is completely natural to expect that EEG signals are contaminated by muscular artifacts generated in their necks since their viewing angle is controlled by their head pointing direction. There would be numerous potential BCI applications for such mobile VR environments including hands-free controllers and communications. However, no study has explicitly allowed subjects to freely move their head in an experiment. Due to the lack of systematic and quantitative evaluation of the effect of head movements on SSVEPs, the practical and commercial BCI applications functioning in our daily life have still been hindered.

This study aimed at assessing whether it is feasible to decode SSVEPs while participants are freely moving their head direction. To this end, we proposed a novel experimental design that simultaneously induced SSVEPs as well as muscle artifacts from participants’ necks by employing moving visual flickers. In the experiment, participants were instructed to pursue the visual flickers, which moved horizontally and vertically along sinusoidal trajectories at controlled speeds, to induce intensity-manipulated muscular artifacts. The stimuli were also modulated by four different frequencies (i.e. 10 to 13 Hz with an interval of 1 Hz) for evaluating their classification performance. To quantify the effect of the muscular artifacts, the amplitude spectra and signal-to-noise ratios (SNRs) of elicited SSVEPs were compared among the moving speeds of the stimuli. In addition, the classification accuracy of the frequency-coded SSVEPs was estimated by the well-established target identification algorithms. In particular, the canonical correlation analysis (CCA)-based method [20] and the individual-template-based method incorporated with five spatial filtering techniques [21] were employed as calibration-free and fully-calibrated approaches, respectively. Importantly, this study only employed the aforementioned popularly-used algorithms and spatial filtering techniques for ease in understanding and comparing the results with previous studies. The goal of this study is (1) to quantify the signal characteristics of SSVEPs contaminated by the muscular artifacts, (2) to assess the feasibility of performing BCI tasks accompanied by head movements, and (3) to find the most reliable algorithm to detect the SSVEPs.

2. Materials and methods

2.1. Participants

This study was approved by the Institutional Review Board of National Institute of Advanced Industrial Science and Technology (HF2018-858). All experiments were performed according to relevant guidelines and regulations. Ten healthy participants (three females and seven males, aged 22–28 years, with a mean age of 24.4 years) with normal or corrected-to-normal vision took part in the experiment. The participants read and signed a written informed consent form before participating in the experiment.

2.2. Stimulus design

This study employed a moving visual flicker to simultaneously induce SSVEPs and muscular artifacts in the participants’ neck. The visual stimulus was presented on an LG 38UC99-W 37.5-inch curved liquid crystal display (LCD) screen (LG Electronics) with a resolution of 3840 × 1600 pixels and a refresh rate of 60 Hz. The visual stimulus was rendered within a 5×5 cm square. The flickering frequency \( f_s \) was selected from 10 to 13 Hz with an interval of 1 Hz. In addition, the position of the visual stimulus \( s(f_{in}) \)
on the display was continuously modulated by the following equation:

\[ s(f_m, i) = A \sin[2\pi f_m(i/f_r)], \]

where \(f_m\) indicates a moving speed, \(i\) indicates the frame index, \(f_r\) indicates the refresh rate of LCD screen, and \(A\) indicates the displacement of the sinusoidal trajectory. The \(s(f_m, i)\) indicates the relative horizontal position to the center of the display, and a vertical stimulus movement was realized by physically rotating the display 90 degrees counter-clockwise with a monitor arm (Ergotron 45-509-216). The moving speed \(f_m\) ranged from 0 to 0.3 Hz with an interval of 0.1 Hz. Note that, when the moving speed was zero (i.e. \(f_m = 0\)), the stimulus stayed still at the center of the display. The \(A\) was set to 40 cm in this study (figure 1(a)). The stimulus program was developed under MATLAB (MathWorks Inc.) using the Psychophysics toolbox extensions [22].

2.3. Experimental protocol

In the experiments, each participant was seated on a chair 60 cm away from the display to allow his/her head to rotate over 30 degrees from the initial head position. The participants were asked to chase the moving stimulus by shifting their head orientation. Importantly, chasing the moving stimulus by shifting their gaze was strictly prohibited to properly induce their head movements. To assist the participants performing the chasing tasks, low-power laser light emitted from the left side of non-prescribed glasses was used as a visual feedback (figure 1(b)). In this way, the participants could know if they were performing the tasks well or not since the trajectory of the laser light was corresponding only to their head movement and independent from their eye movement. In addition, the trajectory of the laser light was video-captured for the following behavioral analysis. An example of the trajectory of the laser light illustrated in figure 2 showed that the participant correctly pursued the stimulus moving at a fixed rate. Note that, we made sure the reflected light does not directly reach to the participants’ eyes.

The experiment for each participant was divided into two sessions corresponding to the two moving directions. Each participant performed eight blocks in a session, and completed twenty trials of the stimulus gazing task in a block. The twenty trials consisted of five trials for each of four flicking frequencies. Throughout a block, the stimulus moved continuously at a fixed speed (i.e., \(f_m\)). The moving speed was randomly selected from the four speeds for each block so that each speed was selected twice in a session. At the beginning of each trial, a moving white square was presented for 1 s, and then it started to flicker for 2 s at the frequency (i.e., \(f_s\)) randomly assigned (figure 2). To reduce ocular artifacts, participants were asked to avoid eye blinks during the stimulation period. The total number of trials obtained from a participant was 320 (i.e. 2 moving directions \(\times\) 4 moving speeds \(\times\) 4 flickering frequencies \(\times\) 5 trials \(\times\) 2 blocks/moving speed).

2.4. Data collection

EEG data were recorded with four gold cup electrodes placed over the occipital area (POz, O1, Oz, and O2). In addition, electromyogram (EMG) data were also acquired with four surface Ag/AgCl electrodes placed over bilateral upper trapezius (UT) and sternocleidomastoid (SCM) muscles. The data were recorded at a sampling rate of 250 Hz using an OpenBCI Cyton board (OpenBCI, Inc). The reference and ground electrodes were attached to the left and right earlobes, respectively. An electrolyte (Ten20 Conductive Paste, Weaver and Company, Inc) was applied to all the electrodes to reduce skin resistance. Event triggers that indicate the onset of visual stimulation were sent via the lab streaming layer (LSL) [23] to synchronize the EEG and the EMG signals and the experimental protocol.

2.5. Data preprocessing

Data epochs comprising four-channel EEG and four-channel EMG signals were extracted according to the event triggers and considering the latency delay in the experimental systems and the human’s visual system. The latency delay was estimated as its group delay.
formulated and successfully-used in previous studies [24–26]. The latency can be derived by measuring the phases of SSVEPs as a function of stimulus frequencies and calculating the slope of the line fitting to the phase values. In this study, 2 s EEG data epochs extracted in [0, 2] s, where the time zero indicates the stimulus onset, corresponding to the static stimulus (i.e. \( f_m = 0.0 \text{ Hz} \)) and all four stimulus frequencies were used to estimate the latency delay for each subject. As the result, the averaged estimated latency of \( 0.48 \pm 0.04 \text{ s} \) across all ten subjects was obtained. After that, data epochs containing the EEG and EMG data were extracted again in \( [0.48, 0.48 + d] \), where \( d \) indicates the data lengths used for the following analyses.

The EEG data epochs were first notch filtered at 50 Hz to remove line noise, and then band-pass filtered from 4 Hz to 90 Hz with Chebyshev Type I infinite impulse response (IIR) filter. The EMG data epochs were also notch filtered at 50 Hz, and then high-pass filtered at 20 Hz through a fifth-order Butterworth filter. Zero-phase forward filtering was implemented using the \texttt{filtfilt} function in MATLAB.

### 2.6. Signal characteristic analysis

To explore the signal characteristics of SSVEPs obtained in different conditions (i.e. moving speeds and flickering frequencies), this study first compared the amplitude spectra and SNRs of single-channel SSVEPs. The amplitude spectra \( F(f) \) was calculated by taking the absolute value of discrete Fourier transform (DFT). In this study, the SNR of SSVEPs was defined as the ratio of amplitude of the SSVEP at stimulating frequency to the mean amplitude of the background EEG activities from 1 Hz to 50 Hz as follows:

\[
\text{SNR} = 20 \log_{10} \frac{F_{\text{target}}(f)}{F_{\text{non-target}}(f)} \text{ (dB)},
\]

where \( F_{\text{target}}(f) \) indicates the amplitude value at the stimulus frequency \( f_s \), and \( F_{\text{non-target}}(f) \) indicates the mean amplitude values at frequencies within 1 Hz to 50 Hz excluding the target frequency \( f_s \). In this study, 2 s data epochs (i.e. \( d = 2 \)) and the signals obtained at Oz channel were used to calculated the signal characteristics.

The amplitude of muscular artifacts generated from the participants’ necks was also quantified by analyzing the EMG data epochs. The root-mean-squared (RMS) values of time-course EMG data were first computed at each channel [27]. After that, the RMS values at bilateral channels were averaged to quantify the activities of the UT and the SCM muscles [28, 29].

### 2.7. Target identification analysis

To evaluate the effect of head movement on the classification performance of SSVEPs, this study employed calibration-free and fully-calibrated target identification algorithms. Both approaches can be considered as a template-matching-based algorithm. Let \( \chi = (\chi)_{njkh} \in \mathbb{R}^{N_f \times N_c \times N_s \times N_t} \) and \( X \in \mathbb{R}^{N_c \times N_s} \) be training data and single-trial validation data, respectively. Here, \( n \) indicates the index of classes (i.e. stimulus frequencies), \( N_f \) is the number of classes, \( j \) indicates the channel index, \( N_c \) is the number of channels, \( k \) indicates the index of sample points, \( N_s \) is the number of sampling points, \( h \) indicates the index of training trials, and \( N_t \) is the number of training trials. The objective of the target identification is to take an input \( X \) and assign it to one of \( N_f \) classes via a template-matching-based feature extraction \( r_n = f(X, Z_n) \).

Here, \( Z_n \) is a template which characterizes an SSVEP corresponding to \( n \)th stimulus frequency. In the calibration-free approach, computer-generated
SSVEP models can be used as templates. On the other hand, in the fully-calibrated approach, individualized templates can be derived from the training data \( \chi_n \). After the feature extraction, the target class index \( \tau \) can be identified by the following rule:

\[
\tau = \arg\max_n r_n, \quad n = 1, 2, \ldots, N_f. \tag{3}
\]

### 2.7.1. Calibration-free method

In a calibration-free method, computer-generated models were employed to classify frequency-coded SSVEPs. In general, SSVEPs can be modeled by sinusoidal waveforms at fundamental and harmonic frequencies [30]. Thus, the model \( \mathbf{Y}_n \in \mathbb{R}^{N_h \times N_f} \) for \( n \)-th stimulus can be formulated as:

\[
\mathbf{Y}_n = [\sin(2\pi f_1 t), \cos(2\pi f_1 t), \ldots, \sin(2\pi f_N f_1 t), \cos(2\pi f_N f_1 t)]^T,
\]

where \( t \in \{1, \frac{2}{f_1}, \ldots, \frac{N_h}{f_1}\} \), \( f_1 \) is the sampling frequency (\( f_s = 250 \) Hz) and \( N_h \) is the number of harmonics. In this study, \( N_f \) was set to five according to the previous study [31]. To take advantage of multi-channel EEG recordings, CCA, which is a statistical way to measure the underlying correlation between two sets of multidimensional variables, has been popularly used in the studies of an SSVEP-based BCI [20, 32]. Considering the linear combinations of the validation data \( \mathbf{X}'(x) \) and the computer-generated SSVEP models \( \mathbf{Y}'(x) \), CCA finds weight coefficients \( \mathbf{w}_x(\mathbf{X}) \in \mathbb{R}^{N_h} \) and \( \mathbf{w}_y(\mathbf{Y}) \in \mathbb{R}^{N_f} \), which maximize the correlation between the two variables as:

\[
\mathbf{w}_x(\mathbf{X}), \mathbf{w}_y(\mathbf{Y}) = \arg\max_{\mathbf{w}_x, \mathbf{w}_y} \frac{E[\mathbf{w}_x(\mathbf{X})^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y(\mathbf{Y})]}{\sqrt{E[\mathbf{w}_x(\mathbf{X})^T \mathbf{X} \mathbf{w}_x(\mathbf{X})]} \sqrt{E[\mathbf{w}_y(\mathbf{Y})^T \mathbf{Y} \mathbf{w}_y(\mathbf{Y})]}},
\]

(5)

In the CCA-based method, the optimized weight coefficient \( \mathbf{w}_x(\mathbf{X}) \) and \( \mathbf{w}_y(\mathbf{Y}) \) are used as spatial filters, and the feature value is calculated as \( r_n = \text{Corr}(\mathbf{X}'(x), \mathbf{Y}'(x)) \) where \( \text{Corr}(\cdot) \) indicates the Pearson’s correlation analysis. Finally, the target frequency can be identified by the equation (3).

### 2.7.2. Fully-calibrated method

In fully-calibrated methods, individualized templates \( \bar{\mathbf{X}}_n \in \mathbb{R}^{N_h \times N_f} \) obtained by averaging multiple training trials as \( \bar{\mathbf{X}}_n = \frac{1}{N_t} \sum_{t=1}^{N_t} \chi_n \) are employed to better characterize SSVEPs than the computer-generated models [33–35]. It has also been proven that spatial filtering techniques are effective to enhance the performance of the fully-calibrated methods [21]. Assuming the stationarity in EEG signals, spatial filters derived from training data can be applied in the following validation stage. Let \( \mathbf{w}_n \in \mathbb{R}^{N_h \times N_f} \) be a spatial filter for \( n \)-th stimulus frequency, where \( N_h \) is the number of spatial filter (\( N_k \leq N_f \)), an ensemble spatial filter \( \mathbf{W} \in \mathbb{R}^{N_h \times (N_f \times N_f)} \) can be constructed as follows:

\[
\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N_f}].
\]

The correlation-based feature extraction in the fully-calibrated methods can be formulated as \( r_n = \text{Corr}(\mathbf{X}'(\mathbf{X}), \bar{\chi}_n^T \mathbf{W}) \). Since more complicated spatial filters can be derived from training trials in the fully-calibrated methods than in the calibration-free method, several kinds of algorithms have been proposed in previous studies [21]. The following four promising algorithms were employed in this study.

- Minimum energy combination (MEC): The MEC-based spatial filtering technique was originally proposed by Friman [36]. The MEC finds a weight coefficient that maximizes the energy of nuisances signals in EEG signals. To extract the nuisance signals \( \bar{\mathbf{X}}_n \in \mathbb{R}^{N_h \times N_f} \), potential SSVEP components are first removed from the individualized templates using computer-generated SSVEP models \( \mathbf{Y}_n \) formulated in equation (4) as:

\[
\bar{\mathbf{X}}_n = \mathbf{X}_n - \mathbf{Y}_n^T \mathbf{Y}_n \mathbf{X}_n^T 
\]

Then, spatial filters can be derived by minimizing the energy of nuisance signals as follows:

\[
\mathbf{w}_n = \arg\min_w \| \mathbf{X}_n^T \mathbf{w} \|^2 = \arg\min_w \| \mathbf{w} \|^2 \mathbf{X}_n \mathbf{X}_n^T \mathbf{w}.
\]

(8)

The solution of the minimizing problem is given by the eigenvalue decomposition. Note that, not only the eigenvector corresponding to the smallest eigenvalue but also some of the other eigenvectors might contain nuisance signals. Therefore, \( N_k \) eigenvectors were included in the spatial filters so as to discard 90% of the energy in the nuisance signals according to its original study [36].

- Canonical correlation analysis (CCA): The CCA-based spatial filters can be derived by computing CCA between the individualized templates \( \bar{\mathbf{X}}_n \) and the computer-generated SSVEP models \( \mathbf{Y}_n \). That is, it can be done by replacing the validation data \( \mathbf{X} \) by the individual templates \( \bar{\mathbf{X}}_n \) in the equation (5), and the weight coefficients \( \mathbf{w}_n \in \mathbb{R}^{N_h} \) obtained by \( \bar{\mathbf{X}}_n \) and \( \mathbf{Y}_n \) can be used as a spatial filter \( \mathbf{w}_n \) for \( n \)-th stimulus frequency.

- Independent component analysis (ICA): ICA decomposes multidimensional signals into statistically independent components which is widely used in EEG analysis [37–39]. In this approach, training trials for \( n \)-th stimulus frequency are first transformed from a three-way tensor \( \mathbf{X}_n \in \mathbb{R}^{N_h \times N_f \times N_t} \) into a matrix \( \mathbf{X}_n \in \mathbb{R}^{N_h \times (N_f \times N_t)} \). By applying infomax ICA algorithm implemented in EEGLAB [38] to the transformed data, a demixing matrix \( \mathbf{A} \in \mathbb{R}^{N_h \times N_f} \) and independent components \( \mathbf{U} = \mathbf{A}^T \mathbf{X}_n \in \mathbb{R}^{N_h \times (N_f \times N_t)} \) can be obtained. However, the correspondence between the independent components and source activities are unknown. To find the components which might contain SSVEP components, canonical correlation coefficients between independent
components and computer-generated SSVEP models are known. In this study, \( N_s \) independent components whose canonical correlations exceed a threshold were considered as components related to SSVEPs, and their corresponding demixing vectors were extracted as spatial filters \( w_n \), for \( n \)th stimulus frequency. The selection threshold of canonical correlation coefficients was set to 0.4 [21].

- Task-related component analysis (TRCA): TRCA is a method for finding a linear coefficient vector \( \mathbf{w} \in \mathbb{R}^{N_s} \) to maximize inter-trial correlation [40] when there are \( h \)th trial (\( h = 1, \ldots, N_s \)) in the training data for \( n \)th stimulus \( \mathbf{x}^{(h)}_n \in \mathbb{R}^{N_s \times N_t} \) and task-related components \( y_n^{(h)} = \mathbf{w}^T \mathbf{x}^{(h)}_n \in \mathbb{R}^{N_s} \). All possible combination for the covariance \( C_{h_1,h_2} \) of different training trials can be expressed as:

\[
C_{h_1,h_2} = \text{Cov}(y_n^{(h_1)}, y_n^{(h_2)})
\]

where the matrix \( \mathbf{S}_n \in \mathbb{R}^{N_s \times N_t} \) is defined as:

\[
(S_n)_{h_1,h_2} = \sum_{h_1,h_2=1}^{N_s} \text{Cov}(\mathbf{x}^{(h_1)}_n, \mathbf{x}^{(h_2)}_n),
\]

To normalize the cumulative covariance, the variance of \( y_n \) is constrained by concatenated training trials \( \mathbf{x}_n \):

\[
\text{Var}(y_n) = \sum_{h=1}^{N_s} \mathbf{w}_j^T \mathbf{w}_j \text{Cov}(\mathbf{x}_{n,j}, \mathbf{x}_{n,h}) = \mathbf{w}^T \mathbf{Q}_n \mathbf{w} = 1.
\]

Based on these equations, the optimal linear coefficient vector can be found as:

\[
\mathbf{w}_n = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_n \mathbf{w}}{\mathbf{w}^T \mathbf{Q}_n \mathbf{w}}.
\]

The largest eigenvalue of the matrix \( \mathbf{Q}_n^{-1} \mathbf{S}_n \) is selected as the spatial filter \( \mathbf{w}_n \).

2.8. Performance evaluation

The behavioral performance was evaluated by analyzing the video captures of the trajectory of the laser light. The behavioral accuracy was defined to be the percentage of the time period in which the trajectory of the light stayed within the moving flicker (5 x 5 cm square) in the total stimulation period at each moving direction and speed. Since there were some segments of time in which the reflection of the light was not able to be reliably detected because of unexpected fluctuations in room brightness, those segments were excluded from the behavioral analysis.

The classification accuracy was estimated using a leave-one-block-out cross validation (CV) for each moving direction and speed. Here, a block contains four trial data of different frequencies (4-class SSVEPs). From each experimental condition, we obtained 10 block data (5 trials x 2 times 4 classes); thus, 9 blocks and 1 block were used as training and validation data, respectively, in each validation process. In the case of participants 1, 3, 4, and 9, the first block was discarded because some of the first samples did not contain data due to slow startup of wireless communication; thus, the number of training blocks was 8 (i.e., \( N_s = 8 \)). The classification accuracy was calculated with the aforementioned five algorithms and different data lengths (i.e., \( d \)) from 0.2 to 2.0 s with an interval of 0.2 s. Here, the calibration-free and fully-calibrated methods are called as the standard CCA-based and the template-based methods, respectively, in the following sections. Three-way repeated measures analysis of variances (ANOVAs) were applied to the amplitude and SNR of SSVEPs to investigate the effects of the two moving directions, four moving speeds, and four stimulus frequencies. In addition, a four-way repeated measures ANOVA was applied to the classification accuracy to explore the effect of the movement directions, four moving speeds, ten data lengths, and five algorithms.

3. Results

3.1. Behavior and muscular artifacts

Table 1 lists the behavioral performance in pursuing the moving stimulus at each moving direction and speed for each participant. The length of tracking period varied across participants. The trajectory of the light was not able to be detected for a participant (s1) at all, so the pursuit accuracy was not estimated for the participant. For most of the others, the trajectory was reliably detected during around 80% of the total experimental time. In addition, the results showed that the participants precisely pursued the moving stimulus with the accuracy ranged from 80.8% to 99.4% under all the conditions.

Figure 3 shows the averaged RMS values of the EMG data measured over the UT and SCM muscles across participants while they were pursuing the horizontally and vertically moving visual stimulus. When \( f_m = 0.0 \), there was no directionality since the stimulus was not moving. The figure revealed that both UT and SCM were activated by both horizontal and vertical head movements. It also shows that the vertical head movements tended to induce stronger activations than the horizontal ones. Figure 3(b) revealed that the activities (i.e., RMS values) at SCM increased as the moving speed increased, showing linear relationships.

3.2. Signal characteristics of elicited SSVEPs

Figure 4 represents averaged amplitude spectra of elicited SSVEPs across all participants while they
are staring at the moving visual stimuli flickering at 10 Hz. The signals at the Oz electrode were used for the frequency analysis. Regardless of the moving directions and speeds, the amplitude spectra showed distinct peaks at the fundamental frequency (i.e. 10 Hz). However, as the moving speeds increased, the amplitudes at the stimulus frequency decreased and the background activities increased. Figure 5 shows the amplitude values and SNRs of SSVEPs for the four stimulus frequencies with different moving directions and speeds.

A three-way repeated measures ANOVA revealed that there were significant main effect of stimulus frequencies on the amplitude of SSVEPs ($F(3,225) = 3.94, p = 0.019$). However, there were no significant main effects of moving direc-

Table 1. Accuracy (%) of pursuing the moving stimulus for each participant. The accuracy was calculated by data while the trajectory of the laser light was properly detected from video captures (i.e. tracking period).

| Subject | $f_m = 0.1$ | $f_m = 0.2$ | $f_m = 0.3$ | Tracking Period (%) |
|---------|-------------|-------------|-------------|--------------------|
| s1      | N/A         | N/A         | N/A         | 0.0                |
| s2      | 96.0        | 97.4        | 96.8        | 55.4               |
| s3      | 92.0        | 92.9        | 91.1        | 80.8               |
| s4      | 98.7        | 94.3        | 97.8        | 72.6               |
| s5      | 97.4        | 98.1        | 97.9        | 85.9               |
| s6      | 96.7        | 94.1        | 96.6        | 88.9               |
| s7      | 98.7        | 97.2        | 96.6        | 100.0              |
| s8      | 99.3        | 98.2        | 99.0        | 0.0                |
| s9      | 98.5        | 98.7        | 89.9        | 88.7               |
| s10     | 99.1        | 98.0        | 98.3        | 88.3               |

Figure 3. Averaged RMS values in response to the stimuli horizontally and vertically moving at different rate across participants. (a) Upper trapezius (UT), (b) Sternocleidomastoid (CSM).

Figure 4. Comparisons of the averaged amplitude spectra of SSVEPs in response to the 10 Hz stimuli (a) horizontally and (b) vertically moving at different rate across participants. The dashed lines indicate the fundamental and its harmonic frequencies.
sions \((F(1225) = 0.02, p = 0.882)\) and speeds \((F(3225) = 1.17, p = 0.340)\) on the amplitude of SSVEPs. On the other hand, another three-way repeated measures ANOVA revealed that there were significant main effects of the moving directions \((F(1225) = 6.79, p = 0.029)\), speeds \((F(3225) = 16.9, p < 0.001)\), and stimulus frequencies \((F(3225) = 4.08, p = 0.016)\) on the SNR of SSVEPs.

### 3.3. Target identification accuracy

Figure 6 shows the averaged accuracy across participants with different data lengths from 0.2 s to 2.0 s with an interval of 0.2 s. It represents the following three trends: (i) the classification accuracy increased as the data length increased; (ii) the classification accuracy decreased as the moving speed increased; and (iii) the fully-calibrated approaches (i.e. template-based methods) outperformed the calibration-free approach (i.e. standard CCA-based method) especially when using short data lengths (e.g. 0.2–1.2 s). Figure 7 redraws the subset of averaged accuracy across participants shown in figure 6 as functions of the moving speeds at different data lengths from 0.4 s to 2.0 s with an interval of 0.4 s. It suggests that the deterioration of the accuracy is linearly associated with the moving speeds at each data length.

A four-way repeated measures ANOVA applied to the results illustrated in figure 6 showed significant main effects of the moving directions \((F(13279) = 6.54, p = 0.031)\), speeds \((F(33279) = 16.05, p < 0.001)\), algorithms \((F(43279) = 39.86, p < 0.001)\), and data lengths \((F(93279) = 216.20, p < 0.001)\) on the classification accuracy of SSVEPs. All pairs using speed factor have significant interactions (to moving directions \((F(33279) = 208.54, p < 0.001)\), algorithms \((F(123279) = 11.77, p < 0.001)\), and data lengths \((F(273279) = 6.04, p < 0.001)\). In addition, there were significant interactions between moving directions and algorithms \((F(43279) = 3.69, p < 0.005)\).

### 4. Discussion

#### 4.1. Validity of the proposed experimental design
The experimental design proposed in this study aimed to induce SSVEPs contaminated by intensity-manipulated muscular artifacts corresponding to...
participants’ head movements. The novelties of our experimental design were twofold: (1) employing the moving visual stimulus to induce participants’ head movements, and (2) using the laser light to track them. Note that, an eye tracking was not suitable for tracking head movements since it cannot distinguish between eye and head movements. The visual feedback based on the laser light indeed helped the subjects properly pursue the moving stimulus as shown in figure 2 and table 1. It was also beneficial to regulating the subjects’ saccadic eye movements because they needed to pay overt attention to the laser light. Although microsaccade could still affect the signal characteristics of SSVEPs, previous studies have not taken it into consideration for their experimental condition. We assume that the effect of microsaccade on SSVEPs is negligible since the experimental condition in the present study is fairly consistent with that in previous studies as long as the subjects properly pursued the stimulus.

The level of artifact contamination was manipulated by adjusting the moving speeds of visual stimulus since rapid movements require strong muscular activation. Figure 3 showed the level of muscular activation indeed increased as the moving speed increased. Interestingly, the vertical head movements tended to induce stronger artifacts than the horizontal ones. Its possible explanation is that vertical head movements require strong activation in related muscles such as the SCM muscle because of the time-varying pitch moment induced by gravity, whilst horizontal head movements do not require much muscular activity because yaw rotations are independent from the effect of gravity. This tendency was consistent with the SNR of SSVEPs. Figures 4 and 5 showed that the amplitudes at the stimulus and non-stimulus frequencies were deteriorated and increased, respectively, as the moving speeds increased, resulting in degraded SNR of SSVEPs. The results confirmed the validity of the proposed experimental design for inducing intensity-manipulated muscular artifacts as well as eliciting SSVEPs in EEG recordings.

Figure 7. Averaged target identification accuracy across participants as functions of the moving speeds \( f_m \) at five different data lengths for each algorithm with linear regression lines. The horizontal dashed lines indicate the chance-level accuracy in the four-class classification (i.e. 25%).

Figure 8. Comparison of the target identification algorithms in its robustness. The baseline accuracy and sensitivity were defined as the intercept and slope of linear regression line fitting to the obtained classification accuracy at each data length with different moving speeds. The size of markers is proportional to the standard deviation of sensitivity across participants.
4.2. Feasibility of detecting SSVEPs and its robustness

As shown in figure 6, all the algorithms achieved higher accuracy than its chance-level (i.e. 25%) regardless of the moving direction and speed of the visual stimulus. The results verified the feasibility of detecting SSVEPs from the EEG signals contaminated by muscular artifacts generated by head movements. In most cases, the template-based methods outperformed the standard CCA-based method. This result seems to be counter-intuitive since not only the validation data but also the templates (i.e. training data) for constructing classification models must be affected by the muscular artifacts. However, the spatial filtering techniques worked well to enhance the SNR of SSVEPs by reducing the effect of the artifacts. When \( f_m = 0.0 \), the template-based methods with the CCA- and the TRCA-based spatial filtering outperformed the other methods. This result is consistent with a previous study [21] which compared the performance of different algorithms on a benchmark SSVEP dataset [26]. However, the accuracy of the template-based method with the TRCA-based spatial filtering significantly dropped when \( f_m = 0.3 \). This might be because the TRCA-based spatial filtering is completely a data-driven approach. In fact, the CCA-based spatial filtering, which is a model-based approach, showed robust performance regardless of the intensity of muscular artifacts.

As is the case with the SNR (i.e. figure 5(b)), the target identification accuracy decreased as the moving speed increased, and its effect was larger in vertical head movements than in horizontal head movements. Figure 7 further indicated that the deterioration was linearly associated with the speed of head movements regardless of the moving directions. Therefore, we employed the slope and the intercept of linear regression lines to evaluate the robustness of the target identification algorithms. For instance, the steeper the slope of an algorithm is, the more the algorithm is sensitive to artifacts. In addition, the intercept can be considered as a baseline accuracy of the algorithm when \( f_m = 0.0 \). Figure 8 illustrates the relationship between the sensitivity (i.e. \(-1 \times \text{slope}\)) and the baseline accuracy (i.e. intercept) for each algorithm at different data lengths. The projections onto the baseline accuracy and data length plane revealed that the template-based method with the CCA-based spatial filtering had the highest baseline accuracy among all the methods with both horizontal and vertical head movements.

The projections onto the sensitivity and baseline accuracy plane indicated that both sensitivity and baseline accuracy increased as the data length increased regardless of algorithms. It also clearly showed that all the target identification algorithms tested in this study were more sensitive to the vertical head movements than the horizontal ones. The projections onto the sensitivity and data length plane revealed that the template-based method with the TRCA-based spatial filtering had the highest sensitivity among all the methods. It is worth emphasizing that the template-based method with the CCA-based spatial filtering outperformed that with the TRCA-based spatial filtering in terms of both baseline accuracy and sensitivity.

In this study, the CCA-based spatial filtering achieved the highest performance in five algorithms. However, the best algorithm might be different for each application. Since the performance generally depends on the number of visual stimuli, stimulus frequencies/phases, etc, target identification algorithms should be optimized for each target application. Note that, since the best data length is also different depending on such parameters [13, 14], it should be optimized for each target application and algorithm so as to maximize its ITR.

4.3. Limitations

A limitation of this study is that only four electrodes were used to measure the EEG signals. A previous study, which compared different SSVEP detection algorithms, demonstrated that fewer electrodes located within the occipital area achieved better performance than more electrodes covering the whole scalp for most of the algorithms [21]. The classification accuracy might be able to be further improved by densely placing a large number of electrodes over the occipital area. However, even with a small number of electrodes, the present results successfully proved the feasibility of detecting SSVEPs from the EEG signals contaminated with the muscular artifacts. In practice, the number of electrodes should be reduced as much as possible to improve the usability of BCI applications.

Another limitation is that this study only tested the simplified head movements. Therefore, there has still been a gap between the simplified head movements and more complicated ones occurring in practical environments. Since most complicated head movements can be considered as the combination of simplified vertical and horizontal movements, the present results would give insight into the effect of realistic head movements on SSVEPs. In addition, roll movements were excluded from the study because they rarely happen while experiencing an HMD-based VR environment and different experimental designs were required to induce them. It would be of interest to further investigate the effect of roll head movements on SSVEPs toward providing more comprehensive guideline in designing practical BCI applications. Nonetheless, the results obtained by analyzing the highly-manipulated muscular artifacts would be suggestive to evaluate the applicability of BCI systems to real-world situations.

4.4. Future visions

Our future works will focus on how to improve the performance of detecting SSVEPs in natural environments from two viewpoints. Previous studies have proposed a variety of methods for reducing artifacts from EEG signals [37, 41–46]. For instance,
ICA [37, 41], CCA [42], and artifact subspace reconstruction (ASR) [43, 44] have traditionally been employed to reduce certain types of artifacts from EEG signals. In addition, the combination method of multivariate empirical mode decomposition (MEMD) and multiset CCA (MCCA) showed better performance in artifact removal than other existing methods [45, 46]. Although the spatial filtering techniques used in this study somewhat have ability to improve the SNR of SSVEPs, combining artifact reducing techniques could further enhance the detectability of SSVEPs. Our previous semi-simulation study, in which a benchmark SSVEP data [26] and electromyogram (EMG) measured from participants’ neck were artificially mixed to simulate artifact-contaminated SSVEP signals, showed that the existing artifact removal methods could improve the target identification accuracy [47]. However, its effectiveness on real data has still remained unclear; thus, it could be addressed by using the dataset recorded in this study.

5. Conclusions

This study assessed the effect of horizontal and vertical head movements on the signal characteristics of SSVEPs and its classification performance. We proposed a novel experimental design that simultaneously induced SSVEPs as well as muscle artifacts from participants’ neck by using the moving visual flicker. For both horizontal and vertical head movements, the amplitude and SNR of SSVEPs significantly degraded as the moving speed increased.

The vertical head movement tended to generate stronger muscular artifacts than the horizontal movement. Even with the severe contamination, the popularly-used target identification methods could reliably detect SSVEPs. The model-based approaches including the standard CCA-based method and the individual-template-based method with CCA-based spatial filtering showed more robust classification performance than the data-driven approaches such as TRCA-based spatial filtering. The results suggest the feasibility of performing an SSVEP-based BCI in practical and real-world situations, where users can freely behave, including VR environments.

Acknowledgments

This work was supported in part by the New Energy and Industrial Technology Development Organization (NEDO).

ORCID iDs

Suguru Kanoga  https://orcid.org/0000-0002-2214-3622
Masaki Nakanishi  https://orcid.org/0000-0003-0415-072X

References

[1] Wolpaw J R, Birbaumer N, Mcfarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control Clin. Neurophysiol. 113 767–91
[2] Chi Y M, Wang Y T, Wang Y, Maior C, Jung T P and Cauwenberghs G 2012 Dry and non-contact EEG sensors for mobile brain–computer interfaces IEEE Trans. Neural Rehabil. Syst. Eng. 20 1–7
[3] Xing X, Wang Y, Pei W, Guo X, Liu Z, Wang F, Ming G, Zhao H, Gui Q and Chen H 2018 A high-speed SSVEP-based BCI using dry EEG electrodes Sci. Rep. 8 14708
[4] Wang Y T, Wang Y and Jung T P 2011 A cell-phone-based brain–computer interface for communication in daily life J. Neural Eng. 8 025018
[5] Güneysu A and Akin H L 2013 An SSVEP-based BCI to control a humanoid robot by using portable EEG device 35th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society pp 6905–8
[6] McCormin C M, Fu J L, Wang M, Lopes L S, Wang P T, Karimi-Bidhendi A, Liu C Y, Heydari P, Nenadic Z and Do A H 2017 Performance assessment of a custom, portable, and low-cost brain–computer interface platform IEEE Trans. Biomed. Eng. 64 2313–20
[7] Nakanishi M, Wang Y T, Jung T P, Zao J K, Chien Y Y, Diniz-Filho A, Daga F B, Lin Y P, Wang Y and Medeiros F A 2017 Detecting glaucoma with a portable brain–computer interface for objective assessment of visual function loss JAMA Ophthalmol. 135 550–7
[8] Vialatte F B, Maurice M, Dauwels J and Cichocki A 2010 Steady-state visually evoked potentials: focus on essential paradigms and future perspectives Prog. Neurobiol. 90 415–38
[9] Gao S, Wang Y, Gao X and Hong B 2014 Visual and auditory brain–computer interfaces IEEE Trans. Biomed. Eng. 61 1436–47
[10] Regan D 1966 Some characteristics of average steady-state and transient responses evoked by modulated light Electroencephalogr. Clin. Neurophysiol. 20 238–48
[11] Regan D 1990 Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine Br. J. Ophthalmol. 74 255
[12] Wang Y, Gao X, Hong B, Jia C and Gao S 2008 Brain–computer interfaces based on visual evoked potentials IEEE Eng. Med. Biol. Mag. 27 64–71
[13] Chen X, Wang Y, Nakanishi M, Gao X, Jung T P and Gao S 2013 High-speed spelling with a noninvasive brain–computer interface Proc. Natl Acad. Sci. 112 E6508–67
[14] Nakanishi M, Wang Y, Chen X, Wang Y T, Gao X and Jung T P 2018 Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis IEEE Trans. Biomed. Eng. 65 104–12
[15] Gramann K, Gwin J T, Bigdely-Shamlo N, Ferris D and Makeig S 2010 Visual evoked responses during standing and walking Frontiers Hum. Neurosci. 4 202
[16] Debener S, Minow F, Emkes R, Gandras K and de Vos M 2012 How about taking a low-cost, small, and wireless EEG for a walk? Psychophysiology 49 1617–21
[17] Lin Y P, Wang Y and Jung T P 2014 Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset J. Neuroeng. Rehabil. 11 919
[18] Lin Y P, Wang Y, Wei C S and Jung T P 2014 Assessing the quality of steady-state visual evoked potentials for moving humans using a mobile electroencephalogram headset Frontiers Hum. Neurosci. 8 182

Akihiko Murai  https://orcid.org/0000-0002-2035-4346
Mitsunori Tada  https://orcid.org/0000-0001-6784-2016
Atsunori Kanemura  https://orcid.org/0000-0001-7425-3949
[19] Yao Z, Wang Y, Yang C, Pei W, Gao X and Chen H 2018 An online brain–computer interface in mobile virtual reality environments Int. J. Comput. Aided Eng. 26 345–60

[20] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs IEEE Trans. Biomed. Eng. 53 2610–4

[21] Nakanishi M, Wang Y and Jung T P 2018 Spatial filtering techniques for improving individual template-based canonical correlation analysis detection Signal Process. Mach. Learn. Brain–Mach. Interfaces (London: IET) pp 219–239

[22] Brainard D H and Vision S 1997 The psychophysics toolbox Spat. Vis. 10 433–6

[23] Kotha C A and Makeig S 2013 BCILAB: a platform for brain–computer interface development J. Neural Eng. 10 056014

[24] Pan J, Gao X, Duan F, Yan Z and Gao S 2011 Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis J. Neural Eng. 8 036027

[25] Nakanishi M, Wang Y, Wang Y T, Mitsukura Y and Jung T P 2018 Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate PLoS One 9 e99235

[26] Wang Y, Chen X, Gao X and Gao S 2017 A benchmark dataset for SSVEP-based brain–computer interfaces IEEE Trans. Neural Syst. Rehabil. Eng. 25 1746–52

[27] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis J. Neurosci. Methods 134 9–21

[28] Zerafa R, Camilleri T, Falzon O and Camilleri K P 2018 To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs J. Neural Eng. 15 051001

[29] Friman O, Volosyak I and Graser A 2007 Multiple channel detection of steady-state visual evoked potentials for brain–computer interfaces IEEE Trans. Biomed. Eng. 54 742–50

[30] Jung T P, Makeig S, Humphries C, Lee T W, McKeown M J, Iragui V and Sejnowski T J 2000 Removing electroencephalographic artifacts by blind source separation Psychophysiol. 37 163–78

[31] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis J. Neurosci. Methods 134 9–21

[32] Van Huffel S 2006 Canonical correlation analysis applied to electroencephalographic data IEEE Trans. Biomed. Eng. 53 2583–7

[33] Mullen T R, Kothe C A, Chi Y M, Ojeda A, Kerth T, Makeig S, Jung T P and Cauwenberghs G 2013 A real-time wearable EEG system with using wearable dry EEG IEEE Trans. Biomed. Eng. 62 2553–63

[34] Chang C Y, Hsu S H, Pion-Tonachini L and Jung T P 2019 Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings IEEE Trans. Biomed. Eng. accepted (https://doi.org/10.1109/TBME.2019.2930186)

[35] Chen X, Lu A, McKeown M J and Wang Z J 2017 The use of multivariate EMD and CCA for denoising muscle artifacts from channel EEG recordings IEEE Trans. Instrum. Meas. 67 359–70

[36] Chen X, Liu Q, Tao W, Li L, Lee S, Liua A, Chenq P, Cheng J, McKeown M J and Wang Z J 2019 ReMAE: a user-friendly toolbox for removing muscle artifacts from EEG IEEE Trans. Instrum. Meas. accepted (https://doi.org/10.1109/TIM.2019.2920186)

[37] Kanoga S, Nakanishi M, Murali A, Tada M and Kanemura A 2018 Semi-simulation experiments for quantifying the performance of SSVEP-based BCI after reducing artifacts from trapezius muscles 40th Annual Int. Conf. of the IEEE Eng. Med. Biol. Soc. (EMBC) (IEEE) pp 4824–7