IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 6, 2019, accepted June 16, 2019, date of publication June 21, 2019, date of current version July 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924185

Fusing Frontal and Occipital EEG Features to Detect "Brain Switch" by Utilizing Convolutional Neural Network

XIN ZHANG[1], GUANGHUA XU[1,2], (Member, IEEE), ARAVIND RAVI[3], WENQIANG YAN[1], AND NING JIANG[3], (Senior Member, IEEE)

[1]School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[2]State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada

Corresponding authors: Guanghua Xu (ghxu@mail.xjtu.edu.cn) and Ning Jiang (ning.jiang@uwaterloo.ca)

This work was supported in part by the National Natural Science Foundation of China under Grant 51775415 and Grant 51505363, in part by the China Scholarship Council, and in part by the Early Researcher Award from the Ministry of Research Innovation and Science of the Province of Ontario under Grant ER17-13-183.

ABSTRACT Flicker is the most widely used steady-state visual evoked potential (SSVEP) stimulus. In addition, checkerboard can induce steady-state motion visual evoked potential (SSMVEP) in the occipital area. More recently, the action video is proposed to simultaneously elicit SSMVEP and induce sensorimotor area activations via the mirror neuron systems through Action Observation (AO). Integration of AO with brain–computer interface (BCI) is appealing for neural rehabilitation applications. In order to make such a BCI paradigm more feasible in neural rehabilitation, it is essential to discriminate whether a user is actively engaged with the BCI, *i.e.* intentional control (IC) state, or not engaged, *i.e.* non-intentional control (NC) state. In this study, the EEG responses to these three types of visual stimuli were compared for the first time and a convolutional neural network (CNN) was proposed to discriminate IC and NC states. A visual gaiting stimulus was designed to realize BCI-based AO. The results showed that the power of alpha rhythms from frontal area decreased more when the participants engaged at the gaiting stimuli than when the participants engaged at the other two types of stimuli. In addition, the correlation coefficient between the EEG from the occipital area and the template signals increased when the participants engaged at the stimuli. The results also clearly demonstrated the proposed CNN method can discriminate the IC and NC states. In addition, the combination of the attention feature from the frontal area and the SSVEP/SSMVEP feature from the occipital area showed significant performance improvement for the gaiting stimulus, but not for the other two types of stimuli.

INDEX TERMS Brain-computer interface, steady-state motion visual evoked potential, convolutional neural network.

I. INTRODUCTION

Brain-computer interfaces (BCI) can enable individuals to interact with external environments, such as a computer or other equipment without using regular output pathways of peripheral nerves and muscles [1]. Electroencephalographic (EEG) signals are mainly used in non-invasive BCIs, by exploiting neural information in the EEG such as motor imagery, movement-related cortical potential, P300, steady-state visual evoked potential (SSVEP) and so on. Particularly, SSVEP is evoked by periodic visual stimulation with a stationary distinct spectrum with characteristic SSVEP peaks in EEG recording over occipital cortex [2]. SSVEP-based BCIs have advantages of high information transfer rate (ITR) and no need for subject training [3].

Flicker, checkerboard motion stimulus, and recently proposed action video are three different types of visual stimuli. Flicker is the most widely used SSVEP stimulus [3]. The checkerboard motion stimulus, *i.e.* the periodic motion stimulus, is able to induce steady-state motion visual evoked potential (SSMVEP) in the occipital area in the brain [4]–[6]. And the SSMVEP paradigms show low-adaptation characteristic

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Zhang.

and less visual discomfort for BCI applications [6]. The third type of stimulus, the action video is mainly used in action observation (AO), which is a promising methodology for promoting motor cortical activation in neural rehabilitation applications [7]. And studies have shown AO can activate the motor neurons as those responsible for producing the observed action via the brain's mirror neuron system (MNS). One recent study [8] suggested that AO could be a good option for patients with stroke who have difficulty using motor imaginary to effectively stimulate cortical-peripheral motor pathways. And the period action motion can also active the occipital cortex. While few researches had tried to compare the EEG response to these three types of visual stimuli.

Furthermore, researches have shown that patients' engagement in rehabilitation could significantly improve its outcome which is critical for rehabilitation [9], [10]. And the attention level, which is measured using the EEG signals, was used to evaluate patients' engagement [11]. Thut showed that the EEG activity in the alpha rhythm (8–13 Hz) was modulated by sustained voluntary attention [12]. And the alpha activity was known to increase when the user was disengaged [13]. However, the changes in the attention level during and after the participants observing the visual stimulus were rarely reported. Thus, the current study explored the differences and changes of the attention level and the responses in the occipital area to these three types' stimuli.

To our best knowledge, only one BCI-based AO was reported [14]. A flickering action video (a BCI-based AO rehabilitation game) was used to activate the mirror-neuron system. And SSVEP was used to classify whether or not the participant was watching the stimulus video of a human movement, *i.e.* a "brain switch" [15], [16], resulting in a two-class scenario: intentional control (IC) and non-intentional control (NC). And the results showed that when the user was engaged in BCI interactions, significantly stronger mirror-neuron system was activated than the condition when the user was not engaged with BCI interaction. However, it was not possible to identify whether the participant was visually engaged at the foreground of the video, or simply the background in the video. In the current study, we used the action video without flickering as the stimulus (a gaiting stimulus) and design periodic action motion to induce steady-state motion visual evoked potential (SSMVEP) [4].

One of the key features of a BCI that is essential for practical applications is the ability to operate in an asynchronous mode, *i.e.* the BCI can detect whether or not the user intends to interact with the BCI, *i.e.* "brain switch". To this end, the ability of a BCI to discriminate between IC and NC states accurately from continuous EEG data is crucially important for its practical applications. For asynchronous SSVEP-based BCIs, the EEG data from the occipital area are widely used for this purpose. Different algorithms were proposed to improve the classification performance, *e.g.* a novel pseudo-key-based approach [17], k-nearest neighbors (KNN) [18], support vector machine [19], etc. The current study tried an alternative approach to improve the performance of discriminating the IC and NC states, *i.e.* a hybrid BCI.

Hybrid BCIs usually compose of two BCI modalities [20], [21], or one BCI and another non-BCI system [22], [23]. For instance, a hybrid BCI based on SSVEP and P300 was presented in [24], where motor-imagery, P300, and eye blinking were combined to implement forward, backward, and stop control of a wheelchair [25]. For a hybrid BCI, two systems can be combined sequentially or simultaneously [26]. In a simultaneous hybrid BCI, both modalities are processed in parallel. And input signals can be two different EEG signals. In sequential hybrid BCIs, the output of one system is used as the input of the next system. This approach is mostly used when the first system task is to indicate whether the user intends to communicate, *i.e.* "brain switch". The current study was designed to detect "brain switch" by fusing two different EEG signals simultaneously.

However, different modalities of EEG signals contain different features so that different algorithms need to be implemented and their results integrated. Recently, deep learning has made impressive advances in solving real-world problems. Convolutional neural network (CNN) is one supervised learning approach for deep leaning. It has the property of automatic feature discovery and extraction. Here, we proposed a CNN to perform information fusion of two different EEG signals to realize the discrimination of NC state and IC state.

In this study, we compared the EEG responses to three types of BCI based visual stimuli (flicker, checkerboard, gaiting). And the EEG data from the occipital and frontal area were used as inputs to a CNN (VF-CNN method), which was designed to discriminate IC and NC states. The results were compared with the SSVEP BCI approach, where only EEG data from the occipital area were used (V-CNN method).

II. METHODS

A. EXPERIMENT PROTOCOL

Ten healthy subjects (ages from 20 to 30, 8 males and 2 females) participated in the experiments. The experimental protocol was approved by a University of Waterloo's Office of Research Ethics (ORE # 23152). Written Informed Consent forms were obtained from the participants before their participation in the experiments.

In the current study, there were three types of stimuli, *i.e.* flicker, checkerboard, and gaiting, as shown in Fig. 1. The first two types of stimulus, as described in the Introduction, are widely reported in the literature [5], [19]. The gaiting stimulus, which was a new proposed action observation stimulus, was in the form of gaiting sequence of a human.

The stimuli were the frame-based stimulation pattern. All the stimuli programs were developed with MATLAB using the Psychophysics Toolbox [26].

During the experiment, the participants were seated in a comfortable chair and were briefed on the tasks to be performed. The participants were asked to watch the LCD screen

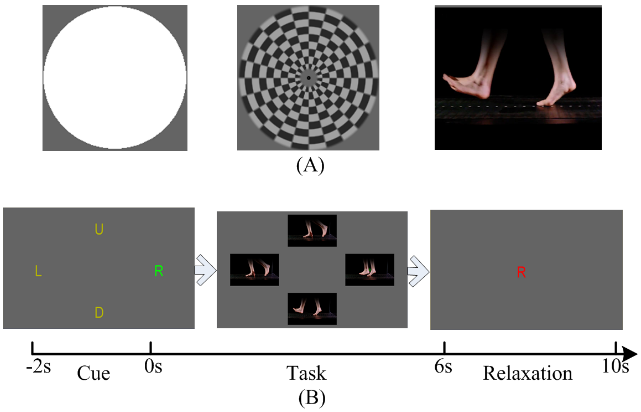


FIGURE 1. Illustration of the experiment protocol. (A) Three types of visual stimuli (flicker, checkerboard, and gaiting). (B) Illustration of one trial in the experimental runs, in which the task was gazing at one gaiting stimulus.

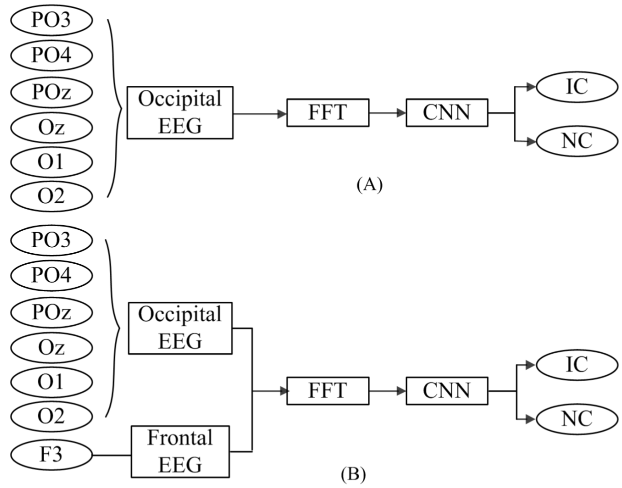


FIGURE 2. The schematic diagram of the two CNN pipelines. (A) The V-CNN pipeline. (B) The VF-CNN pipeline.

on which the visual cues, stimuli, and feedback information were displayed as shown in Fig. 1(B). The refresh rate of the LCD screen was 60Hz and the distance between the screen and the participant was 50cm. During the task period, four targets with different frequencies were displayed in the left, right, up, and down position of the screen. For the flicker stimulus and checkerboard stimulus, the frequencies of the four targets were 8.57 Hz, 12 Hz, 10 Hz, and 15 Hz. The frequencies of the four gaiting targets (F, f) were (8.57, 0.536) Hz, (12, 0.75) Hz, (10, 0.625) Hz, and (15, 0.938) Hz (Fig. 1(B)). F refers to the frame rate. f refers to stride frequency.

In one experimental session, a total of four experimental runs were performed with 20 trials per run, resulting in a total of 80 trials for each type of stimulus. At the beginning of each trial, four letters ('L', 'R', 'U', 'D') would appear at the screen for 2 seconds, at the left, right, up, and down positions of the monitor, respectively. And one of the four letters would be green while the other three were yellow. The green letter indicated the target stimulus for the trial at which participant would then engage his or her gaze for the remainder of the trial. Then the four stimuli would replace the four letters, appearing on the screen for the duration of six seconds, during which the stimuli were modulated at the four frequencies stated above. The participants were asked to gaze at the target appearing at the same position of the green letter (shown between -2 and 0 seconds) for the entire six seconds duration of the trial. This was followed by a relaxation period of four seconds, during which the participant could relax the gaze. And the online classification result using canonical correlation analysis (CCA) would be displayed in the middle of the screen. Then the next trial would begin. And each target was repeated for five times in one run. The participants were asked to avoid moving their heads and avoid performing any sudden jerking movements during the experimental trials.

All EEG data and event time stamps (the beginning and end of each trial) were recorded by PC.

B. DATA PREPARATION

EEG signals were recorded with a commercial research-grade EEG system (gUSBamp and Ladybird electrodes, g.tec Guger Technologies, Austria). Seven electrodes were placed at F3, PO3, POz, PO4, O1, Oz, and O2 of the international 10–20 system. Left earlobe was used as the reference and Fpz was used as ground. All electrodes' impedances were kept below 5 kΩ following the guideline of the manufacturer. The sampling frequency was 1200 Hz. The signals were band-pass filtered between 0.1 and 100 Hz and a notch filter from 58 Hz to 62 Hz was used to eliminate the power line interface.

The acquired EEG data were preprocessed by a band-pass filter from 5 Hz to 40 Hz. According to the time stamps recorded in the experiments, the EEG data were segmented into IC groups and NC groups. IC groups responded to the task period (0s −6s). NC groups responded to the cue and relaxation period (−2s − 0s, 6s − 10s). Each 6-second epochs of EEG signals were segmented using a T-second sliding window with an overlap of (T − 0.1) second. And three window lengths (T = 1s, 2s, and 3s) were chosen in this study. Then each T-second window data were transformed into its frequency domain representation by Fast Fourier transform as shown in Fig. 2. And T-second window data were padded with trailing zeros to 4096. The 105 frequency points between 5 Hz and 35 Hz were chosen as the input data of the network. This was performed for each of the seven channels of the EEG data. Thus, the size of the input data was 105 × channel.

For V-CNN approach, the EEG data from the occipital area (PO3, POz, PO4, O1, Oz, and O2) were chosen as the input of the network as shown in Fig. 2(A). The channel was equal to 6. For the proposed VF-CNN approach, the EEG data from the occipital area (PO3, POz, PO4, O1, Oz, and O2) and frontal area (F3) in the brain were chosen as the input of the network as shown in Fig. 2(B). The channel was equal to 7. And the total number of samples for each participant was $4\times2\times20\times((6\text{-}T)/0.1+1)$.

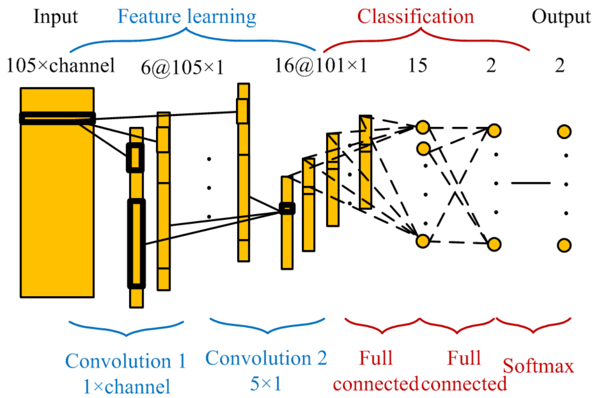


FIGURE 3. Illustration of the convolutional neural network architecture.

C. ALGORITHM MODEL FOR CLASSIFICATION

The proposed CNN consisted of five sequential layers in this study as shown in Fig. 3. The input data were processed as described in the previous section. L1 and L2 were both 2-D convolutional layer with batch normalization. Batch normalization reduced the internal covariance within the input samples [27]. And the rectified linear unit (ReLU) was used as the activation function in these two layers. A total of six convolutional kernels were used in L1 and the kernels had a size of $1 \times$ channel with a stride of 1 (channel = 6 in V-CNN approach, channel = 7 in VF-CNN approach). Similar, a total of 30 convolutional kernels were used in L2 and the kernels had a size of 5×1 with a stride of 1. These two layers effectively performed data-driven feature exaction from the data. The subsequent layers of the CNN shifted to performance classification. Layers L3 and L4 were fully-connected layers with dropout. The numbers of the units in the output of these two layers were 15 and 2. The last layer, L5, used the *softmax* function. And the loss function for classification was cross entropy for two mutually exclusive classes.

The network weights were learned based on the stochastic gradient descent learning algorithm which used the standard error back propagation to optimize network weights. The cross-entropy function was used as the loss function. The learning rate was set at 0.01. The number of training epoch was set as 30, and the size of mini-batch for stochastic gradient descent was set to 32.

A 5-fold cross-validation scheme was performed for each participant's data. All the four runs' data were partitioned into five equal-sized subsamples sequentially in time. Of the five parts, a single part was retained as the validation data for testing the model, and the remaining four parts were used as training data. The cross-validation process was then repeated five times, with each of the five parts used exactly once as the validation data.

D. PERFORMANCE ANALYSIS

The power spectral density (PSD) estimation was performed through the periodogram technique to detect the attention-level of the participant [28]. Let S(f) be the value of the periodogram at frequency f (in Hz):

$$S(f) = \frac{T_s}{N}\left|\sum_{n=1}^{N} x(n)e^{-j2\pi fnT_s}\right|^2 \qquad (1)$$

where $S(f)$ is the value of the periodogram at frequency f, x is the EEG signal from electrode F3 of n samples, and N is the total number of samples of the signal. EEG data with 2-second window length was analyzed and the window was moved in steps of 0.1s in the current study. The average power of alpha rhythms (8-13 Hz) is extracted below.

$$P = \frac{1}{N_\alpha} \times \sum_{f=8}^{13} S(f) \qquad (2)$$

The CCA algorithm is widely used in SSVEP processing, where it is used to calculate the correlations between template signals and multi-channel EEG data [29]. The formula of CCA is:

$$\rho = \max \frac{E[w_x^T XY^T w_y]}{\sqrt{E[w_x^T XX^T w_x]E[w_y^T YY^T w_y]}} \qquad (3)$$

where ρ is the correlation coefficient, X is the EEG data and Y is the template signals.

In this study, the EEG data X was composed of the EEG signals from PO3, POz, PO4, O1, Oz, and O2 electrodes. The template signals Y were composed of several groups of sine and cosine signals. For the flicker and checkerboard stimuli, the template signal was shown below.

$$Y = \begin{Bmatrix} \sin(2 \times \pi \times F_i \times t) \\ \cos(2 \times \pi \times F_i \times t) \\ \sin(2 \times \pi \times 2 \times F_i \times t) \\ \cos(2 \times \pi \times 2 \times F_i \times t) \end{Bmatrix}, \quad i = 1, 2, 3, 4 \qquad (4)$$

where F_i is the frequency of each stimulus target.

Besides, the spectrum of the SSMVEP induced by the gaiting stimuli was more complex than the other two types of stimuli. Thus we designed the combination of frequency components of the template signal which was shown below.

$$Cb = \begin{bmatrix} F_1 & 2 \times F_1 & F_1 + 2 \times f_1 \\ F_2 & F_2 - 2 \times f_2 & F_2 + 2 \times f_2 \\ F_3 & F_3 - 2 \times f_3 & 2 \times F_3 \\ F_4 & F_4 - 2 \times f_4 & F_4 + 2 \times f_4 \end{bmatrix} \qquad (5)$$

where $F_1 = 8.57$ Hz, $f_1 = 0.536$ Hz, $F_2 = 12$ Hz, $f_2 = 0.75$ Hz, $F_3 = 10$ Hz, $f_3 = 0.625$ Hz, $F_4 = 15$ Hz, $f_4 = 0.938$ Hz.

Thus the templates Y in CCA for the gaiting stimulus were shown below.

$$Y = \begin{Bmatrix} \sin(2 \times \pi \times Cb_{i,1} \times t) \\ \cos(2 \times \pi \times Cb_{i,1} \times t) \\ \sin(2 \times \pi \times Cb_{i,2} \times t) \\ \cos(2 \times \pi \times Cb_{i,2} \times t) \\ \sin(2 \times \pi \times Cb_{i,3} \times t) \\ \cos(2 \times \pi \times Cb_{i,3} \times t) \end{Bmatrix}, \quad i = 1, 2, 3, 4 \qquad (6)$$

where $Cb_{i,j}$ is the value in row i and in the column j of Cb in Eqn. (5), j= 1, 2, 3.

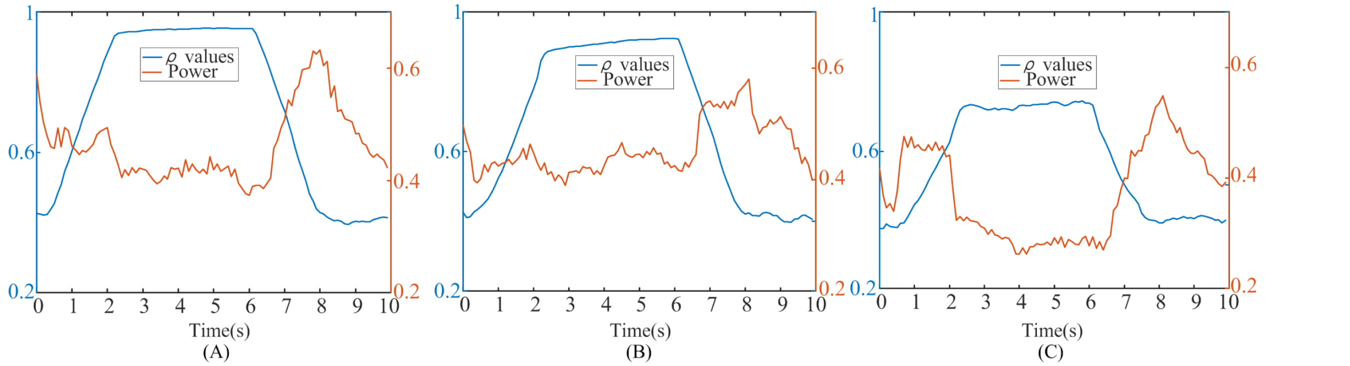


FIGURE 4. The average of the power of alpha rhythms and the average of the correlation coefficient values during the experiments across all the participants in each type of stimuli. (A) Using the flicker stimulus. (B) Using the checkerboard stimulus. (C) Using the gaiting stimulus. Time 0s indicates beginning of gazing the stimulus and time 6s indicates the beginning of gaze relaxation.

E. STATISTICAL ANALYSIS

The mixed effect model of analysis of variance (ANOVA) was used for statistical analysis of the differences in the power of alpha rhythms. Stimuli (flicker, checkerboard, and gaiting) and states (task and relaxation) were used as fixed factors and participant was used as the random factor. The power of alpha rhythms was the response variable. Similar, mixed effect ANOVA was used for the analysis of the classification accuracies performed by the two CNN-based pipelines. Methods (V-CNN and VF-CNN) and time window lengths (1s, 2s, and 3s) were used as fixed factors and participant was used as the random factor. Accuracy was the response variable. The Bonferroni post hoc analysis was used to assess significance. The statistical significance level was 0.05 for all tests.

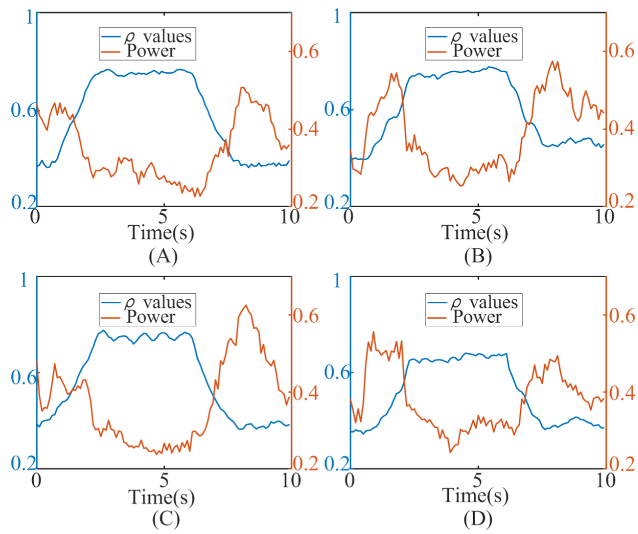


FIGURE 5. The average of the power of alpha rhythms and the average of the correlation coefficient values using the gaiting stimulus across all the participants and all the trials in each target. (A)The left target. (B)The right target. (C)The up target. (D)The bottom target.

III. RESULTS

A. THE EEG RESPONSES IN THE FRONTAL AREA AND OCCIPITAL AREA

Fig. 4 showed the average of the power (P in Eqn. (2)) and the average of ρ values (Eqn. (3)) across all the participants in each type of stimuli. The power was in alpha rhythms of EEG signals from the frontal channels (F3). And the ρ values were the maximum CCA correlation coefficient between EEG signals from the occipital area and the temple signals at the target stimulation frequency. The EEG signals were averaged across all the trials in each participant. When the participant gazing at the stimulus (started at 0s), the correlation coefficient values increased and the power of alpha rhythms decreased in all types of stimuli (Please note that the length of time window was 2s. Thus the results from 0s to 2s in Fig. 4 were obtained from the EEG data contained the cue period as shown in Fig. 1). Then these values maintained relatively stable during the task period (time $<$ 6s). During the gaze relaxation period (6 to 10s), the correlation coefficient values decreased and the power of alpha rhythms increased. Therefore, as the participant gazed at all types of the SSVEP/SSMVEP stimulus, the attention-level changed and reflected in the power of alpha rhythms from the frontal area in the brain. Thus, both the EEG features from the frontal area and occipital area could be selected to do classification for "brain switch". Furthermore, the changes in the correlation coefficient values in Fig. 4(C) implied that the designed gaiting stimuli could induce the corresponding stimulation frequencies as shown in equation (5).

To further compare these three types of stimuli's influence on EEG, we compared the CCA correlation coefficient and the power of alpha rhythms separately. Table 1 showed the average CCA correlation coefficient values during the task period in each participant. The ρ values (Eqn. (3)) using the gaiting stimulus was the lowest. And ρ values using the checkerboard were slightly lower than the values using the flicker stimulus. It indicated that the performance of classification in the gaiting stimulus would be poorer than that in flicker and checkerboard stimuli if CCA was used.

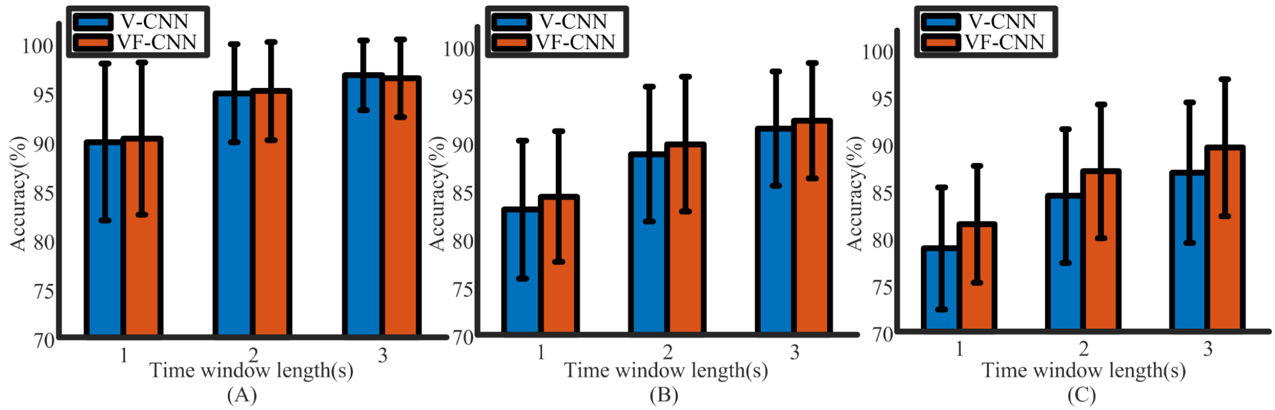


FIGURE 6. The average classification accuracies with different time window lengths in each type of stimuli. (A) Using the flicker stimulus. (B) Using the checkerboard stimulus. (C) Using the gaiting stimulus.

TABLE 1. The average correlation coefficient values during the task period in each participant.

Subjects	Flicker	Checkerboard	Gaiting
S1	0.97	0.94	0.77
S2	0.98	0.92	0.85
S3	0.94	0.96	0.74
S4	0.98	0.92	0.62
S5	0.92	0.92	0.70
S6	0.89	0.75	0.55
S7	0.96	0.93	0.76
S8	0.95	0.90	0.69
S9	0.95	0.90	0.71
S10	0.92	0.89	0.62
Mean ±SD	0.95 ± 0.03	0.90 ± 0.06	0.70± 0.09

For the power of alpha rhythms, the mixed effect ANOVA was applied to quantify the differences. The factors stimuli and states both had a significant influence on the power of alpha rhythms ($p < 0.001$, $p < 0.001$). The post-hoc comparison revealed that there was no significant difference on the power during the relaxation period between flicker and checkerboard ($p > 0.1$), flicker and gaiting ($p > 0.1$), checkerboard and gaiting ($p > 0.1$). While the power during the task period using gating stimulus was significantly lower than that using flicker stimulus ($p = 0.009$) and using checkerboard stimulus ($p = 0.001$). And there was no significant difference on the power during the task period using flicker and checkerboard ($p > 0.1$). It meant the gaiting stimulus attracted more attention of the participants than flicker and checkerboard stimuli.

As the gaiting stimulus showed significant difference with the flicker and checkerboard stimuli during the task period, we further explored the stimulation frequency's effect on the power of alpha rhythms using the gaiting stimulus. Fig. 5 showed the average of the power (P in Eqn. (2)) and the average of ρ values (Eqn. (3)) using the gaiting stimulus across all the participants and all the trials in each target. The power of alpha rhythms (mean $\pm$ SD) during the whole task period in each target were $0.29 \pm 0.02\ \mu V^2$, $0.31 \pm 0.05\ \mu V^2$, $0.28 \pm 0.04\ \mu V^2$, and $0.32 \pm 0.05\ \mu V^2$, respectively. Even though the stride frequencies were different in the four targets, the power of alpha rhythms showed similar trends and scope. Thus different stride frequencies showed little influence on the power of alpha rhythm.

B. COMPARISON OF THE ACCURACIES USING THE TWO CNN-BASED APPROACHES

To compare the classification performance of the VF-CNN and V-CNN approaches, the average classification accuracies with different time window lengths (from 1s to 3s with a step of 1s) were calculated and showed in Fig. 6. We observed that the accuracy increased with longer time window for both methods among all these three types of stimuli. For the flicker stimulus, the accuracies, using VF-CNN method, was $90.3 \pm 7.74\%$, 95.1 ± 4.00 %, and $96.4 \pm 3.93\%$ when the time window lengths were 1s, 2s, and 3s, respectively. According to the Bonferroni post hoc analysis, there was no significant difference on the accuracy using the V-CNN method and VF-CNN method when the time window lengths were 1s, 2s, and 3s (all $p > 0.1$). And the average accuracy using VF-CNN method was slightly lower than the accuracy using V-CNN method. For the checkerboard stimulus, the accuracies, using VF-CNN method, achieved $84.4 \pm 6.77\%$, $89.8 \pm 7.02\%$, and $92.2 \pm 6.01\%$ when the time window lengths were 1s, 2s, and 3s, respectively. The average accuracies using VF-CNN method were slightly higher the accuracies using V-CNN method when the time window lengths were 1s, 2s, and 3s. However, according to the Bonferroni post hoc analysis, there was no significant difference on the accuracy using the V-CNN method and VF-CNN method when the time window lengths were 1s, 2s, and 3s (all $p > 0.1$). For the gaiting stimulus, the accuracies, using VF-CNN method, achieved $81.4 \pm 6.22\%$, 87.0 ± 7.11 %, and $89.5 \pm 7.27\%$ when the time window lengths were 1s, 2s, and 3s, respectively. According to the mixed effect ANOVA, the fixed factors methods and time window lengths both had significant effects on the accuracies ((F = 30.22, $p < 0.001$),

($F = 107.85$, $p < 0.001$)). The post-hoc comparison revealed that the accuracy using VF-CNN method was significantly higher than the accuracies using V-CNN method when the time window lengths were1s, 2s, and 3s ($p = 0.031$, $p = 0.024$, and $p = 0.021$, respectively).

Thus the VF-CNN method obtained the ability to do classification in the hybrid BCI and this method showed significant performance improvement than the V-CNN method mainly for the gaiting stimulus.

IV. DISCUSSION AND CONCLUSION

In this study, we investigated the EEG responses to three different types of visual stimuli (flicker, checkerboard, gaiting). The results showed that the power of alpha rhythms from frontal area decreased and ρ values (correlation coefficient between the EEG from occipital area and template signals) increased when the participants gazed at the stimuli. And we demonstrated that the gaiting stimulus attracted more attention of the participants than flicker and checkerboard stimuli. Furthermore, CNN was utilized to detect NC and IC states in a hybrid SSMVEP-based BCI. And the results showed that fusing the frontal features and occipital features could significantly improve the performance of "brain switch" for the gaiting stimulus.

To our best knowledge, this is the first study exploring the attention level difference in three types of SSVEP/SSMVEP stimuli. One existing study used EEG signals from frontal lobe as the medium to observe students' attentiveness during learning [30]. Another study used alpha-band modulations from the occipital area as measures of covert attention to bilaterally located visual targets [12]. Yaomanee et al. involved three experiments (*i.e.* reading a book, locating 3D figures, and answering questionnaires) for determining whether the subjects were attentive and the results showed that alpha rhythm was slightly higher when the subjects were in a relaxed state [31]. The current study illustrated that the power of the alpha rhythms decreased when observing the stimuli and the power increased in the relaxed state. More importantly, the results showed that the gaiting stimulus attracted more attention than the flicker and checkerboard stimuli.

Besides, the EEG signals from the frontal area were selected to evaluate the attention level in the current study. Because a person's mental state and attentiveness were governed by various parts of the brain in the forehead region. So observing the EEG signals from this area was a viable method for determining whether the participants were attentive [30]. Another reason was that the frequencies of the stimuli in this study overlapped with alpha rhythms.

Furthermore, the results demonstrated that CNN was feasible for the classification in a hybrid BCI. And the proposed hybrid BCI approach achieved performance improvement compared with the SSMVEP BCI approach in discriminating IC and NC states. The performance improvement mainly occurred in the gaiting stimulus. For the flicker and checkerboard stimuli, adding one more channel's data as the input of the CNN could not improve the classification performance. The reason might be that the changes in the attention level were not strong enough in the flicker and checkerboard stimuli as shown in Fig. 4. Thus, only adding useful information to the input of the proposed CNN could improve the classification performance.

It was observed that the accuracy using the gaiting stimulus was lower than the accuracies using flicker and checkerboard stimuli. However, the gaiting stimulus, which is an action observation stimulus, had the ability to activate the MNS [14]. Thus the gaiting stimulus had potential in neurorehabilitation applications. And the classification performance for the gaiting stimulus needed to be improved. The current study demonstrated that the proposed VF-CNN method could significantly improve the classification accuracy for the gaiting stimulus.

Furthermore, only one recent study reported detecting IC and NC in AO in the context of BCI [14]. However, they used the flickering action video as the stimulus to induce SSVEP and supposedly produce MNS activation. And the SSVEP response was used to classify whether the stimuli were being attended to. However, if the participant stared at the background of the video (flickered white and black), they would still get the SSVEP response. For the gaiting stimulus in the current study, the background was always black and refreshed at the screen refresh rate (60Hz). Thus, the SSMVEP response could only be induced when the participants gazed the action in the video.

Note that SSMVEPs are just several specific frequency values and the attention level is obtained from a wider frequency range which is completely different from SSMVEP features. Thus, the CCA, which is widely used in SSVEP processing, is not suitable for the detection the attention level. Since CNN has the ability of automatic feature extraction, different data processing and fusion procedures, which were needed in other hybrid BCIs [24], [32], [33], [34], were not necessary for this study. We can perform CNN and simply add one channel' data from the frontal area as the input data of CNN. And the proposed approach did not require the participants performing additional mental tasks, *e.g.* motor imagery. The requirement for the participants was exactly the same as the requirement in the traditional SSVEP experiments.

This study compared three types of SSVEP/SSMVEP stimuli and the results showed that the gaiting stimulus attracted more attention than the flicker and checkerboard stimuli. Moreover, fusing the frontal features and occipital features, i.e. a hybrid BCI method, by utilizing CNN could improve the classification performance for the gaiting stimulus. However, the current study only focused on offline classification for "brain switch". So the future work will be developing a real-time BCI-based AO neurorehabilitation system using the proposed VF-CNN method.

ACKNOWLEDGMENT

We want to thank the participants for participating in these experiments and anonymous reviewers for their helpful comments.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.

[2] 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, pp. 418–438, Apr. 2010.

[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. USA*, vol. 112, no. 44, pp. E6058–E6067, Nov. 2015.

[4] X. Zhang, G. Xu, J. Xie, and X. Zhang, "Brain response to luminance-based and motion-based stimulation using inter-modulation frequencies," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188073.

[5] W. Yan, G. Xu, J. Xie, M. Li, and Z. Dan, "Four novel motion paradigms based on steady-state motion visual evoked potential," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1696–1704, Aug. 2018.

[6] J. Xie, G. Xu, J. Wang, F. Zhang, and Y. Zhang, "Steady-state motion visual evoked potentials produced by oscillating Newton's rings: Implications for brain-computer interfaces," *PLoS ONE*, vol. 7, no. 6, Jun. 2012, Art. no. e39707.

[7] G. Rizzolatti, L. Cattaneo, M. Fabbri-Destro, and S. Rozzi, "Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding," *Physiol. Rev.*, vol. 94, no. 2, pp. 655–706, Apr. 2014.

[8] M. Tani, Y. Ono, M. Matsubara, S. Ohmatsu, Y. Yukawa, M. Kohno, and T. Tominaga, "Action observation facilitates motor cortical activity in patients with stroke and hemiplegia," *Neurosci. Res.*, vol. 133, pp. 7–14, Aug. 2018.

[9] E. Tunik, S. Saleh, and S. V. Adamovich, "Visuomotor discordance during visually-guided hand movement in virtual reality modulates sensorimotor cortical activity in healthy and hemiparetic subjects," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 198–207, Mar. 2013.

[10] M. Steinisch, M. G. Tana, and S. Comani, "A post-stroke rehabilitation system integrating robotics, VR and high-resolution EEG imaging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 5, pp. 849–859, Sep. 2013.

[11] B.-S. Lin, H.-C. Hsu, G. E. Jan, and J.-L. Chen, "An interactive upper-limb post-stroke rehabilitation system integrating BCI-based attention monitoring and virtual reality feedback," in *Proc. 3rd Int. Conf. Comput. Meas. Control Sensor Netw. (CMCSN)*, May 2016, pp. 44–47.

[12] G. Thut, A. Nietzel, S. A. Brandt, and A. Pascual-Leone, "α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection," *J. Neurosci.*, vol. 26, no. 37, pp. 9494–9502, Sep. 2006.

[13] D. L. Schomer, "The normal EEG in an adult," in *The Clinical Neurophysiology Primer*, A. S. Blum S. B. Rutkove, Eds. Totowa, New York, NY, USA: Humana Press, pp. 57–71, 2007.

[14] H. Lim and J. Ku, "A brain–computer interface-based action observation game that enhances mu suppression," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2290–2296, Dec. 2018.

[15] S. He, R. Zhang, Q. Wang, Y. Chen, T. Yang, Z. Feng, Y. Zhang, M. Shao, and Y. Li, "A P300-based threshold-free brain switch and its application in wheelchair control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 715–725, Jun. 2017.

[16] R. V. Sastry, R. Aravind, G. Panwar, S. Indrapriyadarsini, and A. G. Ramakrishnan, "Automatic user customized brain switch," in *Proc. IEEE Annu. India Conf. (INDICON)*, Dec. 2016, pp. 1–5.

[17] J. Pan, Y. Li, R. Zhang, Z. Gu, and F. Li, "Discrimination between control and idle states in asynchronous SSVEP-based brain switches: A pseudo-key-based approach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 3, pp. 435–443, May 2013.

[18] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A lower limb exoskeleton control system based on steady state visual evoked potentials," *J. Neural Eng.*, vol. 12, no. 5, Aug. 2015, Art. no. 56009.

[19] K. Suefusa and T. Tanaka, "Asynchronous brain–computer interfacing based on mixed-coded visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 2119–2129, Sep. 2018.

[20] C. Brunner, B. Z. Allison, C. Altstätter, and C. Neuper, "A comparison of three brain–computer interfaces based on event-related desynchronization, steady state visual evoked potentials, or a hybrid approach using both signals," *J. Neural Eng.*, vol. 8, no. 2, Mar. 2011, Art. no. 25010.

[21] B. Rebsamen, E. Burdet, Q. Zeng, H. Zhang, M. Ang, C. L. Teo, C. Guan, and C. Laugier, "Hybrid P300 and Mu-Beta brain computer interface to operate a brain controlled wheelchair," in *Proc. 2nd Int. Conv. Rehabil. Eng. Assistive Technol.*, May 2008, pp. 51–55.

[22] X. Yong, M. Fatourechi, R. K. Ward, and G. E. Birch, "The design of a point-and-click system by integrating a self-paced brain–computer interface with an eye-tracker," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 4, pp. 590–602, Dec. 2011.

[23] Y. Punsawad, Y. Wongsawat, and M. Parnichkun, "Hybrid EEG-EOG brain-computer interface system for practical machine control," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug./Sep. 2010, pp. 1360–1363.

[24] Y. Li, J. Pan, F. Wang, and Z. Yu, "A hybrid BCI system combining P300 and SSVEP and its application to wheelchair control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 3156–3166, Nov. 2013.

[25] H. Wang, Y. Li, J. Long, T. Yu, and Z. Gu, "An asynchronous wheelchair control by hybrid EEG–EOG brain–computer interface," *Cogn. Neurodyn.*, vol. 8, no. 5, pp. 399–409, Oct. 2014.

[26] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[27] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," Mar. 2018, *arXiv: 1803.01164*, [Online]. Available: https://arxiv.org/abs/1803.01164

[28] P. F. Diez, A. G. Correa, L. Orosco, E. Laciar, and V. Mut, "Attention-level transitory response: A novel hybrid BCI approach," *J. Neural Eng.*, vol. 12, no. 5, Aug. 2015, Art. no. 056007.

[29] 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, Dec. 2006.

[30] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors," *Sensors*, vol. 13, no. 8, pp. 10273–10286, Aug. 2013.

[31] K. Yaomanee, S. Pan-ngum, and P. I. N. Ayuthaya, "Brain signal detection methodology for attention training using minimal EEG channels," in *Proc. 10th Int. Conf. ICT Knowl. Eng.*, Nov. 2012, pp. 84–89.

[32] R. C. Panicker, S. Puthusserypady, and Y. Sun, "An asynchronous P300 BCI with SSVEP-based control state detection," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1781–1788, Jun. 2011.

[33] R. Leeb, H. Sagha, R. Chavarriaga, and J. R. del Millán, "A hybrid brain–computer interface based on the fusion of electroencephalographic and electromyographic activities," *J. Neural Eng.*, vol. 8, no. 2, Apr. 2011, Art. no. 025011.

[34] Y. Su, Y. Qi, J.-X. Luo, B. Wu, F. Yang, Y. Li, Y.-T. Zhuang, X.-X. Zheng, and W.-D. Chen, "A hybrid brain-computer interface control strategy in a virtual environment," *J. Zhejiang Univ. Sci. C*, vol. 12, no. 5, p. 351, May 2011.

XIN ZHANG received the B.S. and M.S. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree in mechanical engineering.

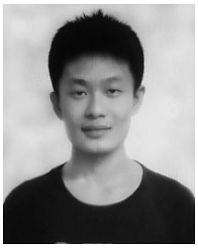
His current research interests include brain–computer interface, biomedical signal processing, and rehabilitation robot.

GUANGHUA XU received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, in 1986, 1989, and 1995, respectively, where he is currently a Professor with the State Key Laboratory for Manufacturing Systems Engineering.

His current research interests include biomedical signal processing, brain computer interface, rehabilitation robot, and condition monitoring and fault diagnosis.

ARAVIND RAVI received the B.S. degree in electronics and communication from PESIT Bangalore South Campus, VTU, India, in 2016. He is currently pursuing the master's degree with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada.

His research interests include pattern recognition, machine learning, biomedical signal processing, and brain–computer interfaces (BCI).



WENQIANG YAN received the B.S. degree in mechanical engineering from the Dalian University of Technology, Dalian, China, in 2015. He is currently pursuing the Ph.D. degree in mechanical engineering with Xi'an Jiaotong University, Xi'an, China.

In 2013, he was an Intern with Dalian Machine Tool Plant. His current research interests include biomedical signal processing and brain–computer interface.

NING JIANG received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1998, and the M.S. and Ph.D. degrees in biomedical engineering from the University of New Brunswick, Canada, in 2004 and 2009, respectively. He is currently an Assistant Professor with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada.

His research interests include signal processing methods and artificial intelligence algorithms for biological signals, such as electroencephalogram (EEG) and electromyography (EMG), for neurorehabilitation engineering applications.

He also Co-Chaired the Neurorehabilitation and Prosthetics Symposium at the IEEE International Neural Engineering Conference, in 2017 and the Track Chair of the Myoelectric Control Symposium, in 2017. He was the Co-Chair of the Workshop on BCI in Aging Research at the IEEE International Conference on Systems, Man and Cybernetics, in 2018. His service to the scientific and engineering community has been extensive, including as an Associate Editor for the IEEE Journal in Biomedical and Health Informatics, *Frontiers in Neuroscience*, and the *Brain-Computer Interfaces*. He was an Associate Editor of the IEEE International Conference of Robotics and Automation, in 2016.