2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy. October 6-9, 2019

A Hybrid BCI Approach to Detect Brain Switch in Action Observation by Utilizing Convolution Neural Network

Xin Zhang, Guanghua Xu, *Member, IEEE*, Aravind Ravi, Sarah Pearce, and Ning Jiang, *Senior Member, IEEE*

***Abstract*—Action observation (AO) is a promising methodology for promoting cortical activations of the sensorimotor area via the mirror neuron systems. Integration of AO with brain-computer interfaces (BCIs) is appealing for neural rehabilitation applications. One of the important issues in BCI-based AO is discriminating when a user is actively engaged with the BCI, *i.e.* intentional control (IC) state versus when a user is not engaged with the BCI, *i.e.* non-intentional control (NC) state. In this study, we designed a visual gaiting stimulus, which was in the form gaiting sequence of a human, to elicit the steady-state motion visual evoked potential (SSMVEP) response in visual area. And we proposed a convolutional neural network (CNN) to discriminate IC and NC states. The results illustrated the proposed CNN can discriminate the IC and NC states. Furthermore, combining attention feature from the frontal area and SSMVEP features, *i.e.* a hybrid approach, showed significant performance improvement over only using SSMVEP features when the time window length was longer than 2s. And the accuracy achieved 91.94 ± 6.20% when utilizing the hybrid approach in IC vs. NC discrimination.**

I. Introduction

Brain computer interfaces (BCI) can enable individuals to interact with external environment, 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. One of the key features for a BCI to be practical applications is the ability to operate in an asynchronous mode, in which the BCI can detect whether or not the user intends to interact with the BCI, *i.e.* a brain switch [2][3]. To this end, the ability of any BCI to discriminate intentional control (IC) and non-intentional control (NC) states accurately based on ongoing EEG data is crucially important for its practical applications.

To improve the performance of discriminating the IC and NC states, the concept of hybrid BCI was proposed. A hybrid BCI is composed of two BCI modalities, or one BCI and another non-BCI system. For instance, a hybrid BCI based on SSVEP and P300 was presented in [4], where motor-imagery, P300, and eye blinking were combined to implement forward, backward, and stop control of a wheelchair [5]. However, different modalities of EEG or other physiological signals contain different features so that different algorithms need to be implemented and their results fused. Recently, deep learning has made impressive advances in solving real-world problems. Convolutional neural networks (CNN) is one supervised learning approach for deep leaning. It has the property of automatic feature extraction. While it is still unknown whether CNN is suitable for the classification in the hybrid BCI.

In addition, action observation (AO) is a promising methodology for promoting motor cortical activation in the neural rehabilitation [6]. 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 [7] 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. In another more recent study, a BCI-based AO rehabilitation game (a flickering action video) was used to activate the mirror-neuron system [8]. SSVEP was used to classify whether or not the participant was watching the stimulus video of a human movement, *i.e.* engaged in interacting with the BCI, resulting in a two-class scenario: NC and IC. 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 participants stared 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 and design periodic action motion to induce steady-state motion visual evoked potential (SSMVEP) [9].

Furthermore, the attention-level could be measured using the EEG signals. When a person pays attention to any stimulus, some changes should appear in the EEG signal. Thut et al. showed that the EEG activity in the alpha rhythm (8–13 Hz) was modulated by sustained voluntary attention [10].

Thus, to improve the performance of BCI integrated with AO, we proposed a new hybrid BCI, *i.e.* combining SSMVEP feature and attentional feature . A gaiting stimulus was designed to induce SSMVEP. And the attention level, when participants gazed at the stimulus, was detected. We performed CNN method, which utilizing EEG data from visual area and frontal area as the input (VF-CNN method), to

This work was supported by National Natural Science Foundation of China (No. 51775415, and No.51505363), China Scholarship Council, and an NSERC-ENGAGE grant (No. 401261605). (Corresponding authors: Guanghua Xu, Ning Jiang.).
Xin Zhang and Guanghua Xu are with school of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China, email: zx2929108zx@stu.xjtu.edu.cn and ghxu@mail.xjtu.edu.cn). Guanghua Xu is also with State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China.
Ning Jiang and Aravind Ravi are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, N2L3G1, Canada (e-mail: ning.jiang@uwaterloo.ca, aravind.ravi@uwaterloo.ca).
Sarah Pearce is with Cognixion, Toronto, ON, M5H 1K5, Canada (e-mail: sarah@cognixion.com).

discriminate IC and NC states of BCI in AO. And the results were compared with SSMVEP BCI approach, which only utilized EEG data from visual area (V-CNN method).

II. METHODS

A. Experiment protocol

Five healthy subjects (ages from 20 to 30, 4 males and 1 females) participated in the experiments. The experimental protocol was approved by a University of Waterloo's Office of Research Ethics (# 23152). Written Informed Consent forms were obtained from the participants before their participation in the experiments.

During the experiment, the participants seated in a comfortable chair and were briefed on the tasks to be performed. The participants were asked to watch the LCD screen on which the visual cues, stimuli, and feedback information were displayed as shown in Fig. 1. The stimulus is in the form gaiting sequence of a human as shown in Fig. 1(A). And the stimulus program was developed with MATLAB using the Psychophysics Toolbox [11]. Four gaiting targets: (F, f) = (8.57, 0.536) Hz, (12, 0.75) Hz, (10, 0.625) Hz, and (15, 0.938) Hz were displayed in the left, right, up, and down position of the screen, respectively (Fig. 1(B)). F refers to the frame rate. f refers to stride frequency.

During the experiment, a total of four experimental runs were performed with 20 trials per run, resulting in a total of 80 trails. 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 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 trail. Then the four stimuli would replace the four letters, appearing on the screen for a 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.

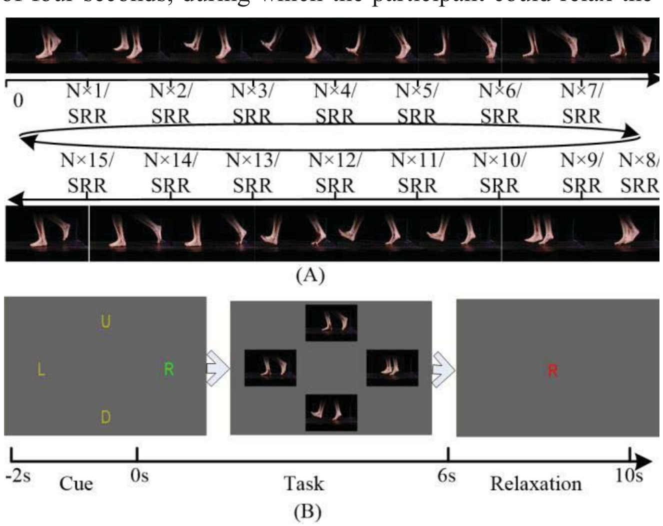

Figure 1. Generation of the gaiting stimulus and illustration of the experiment protocol. (A) Generation of the gaiting stimulus. SRR referred to screen refresh rate and N referred to the number of frames. (B) Illustration of one trial in the experimental runs, in which the task was gazing at one gaiting stimulus.

B. EEG signal measurement

EEG signals were recorded with a commercial research grade EEG system (gUSBamp and Ladybird electrodes, g.tec Guger Technologies, Austria). Eight electrodes were placed at F3, F4, PO3, POz, PO4, O1, Oz, and O2 of the international 10–20 system. Left or right 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.

C. Input data conditioning in CNN

The acquired EEG data were preprocessed by a band-pass filter from 5 Hz to 40 Hz. Each 6-second epochs of EEG signals were segmented using a T-second sliding window with an overlap of (T - 0.1) second. And four window lengths (T = 1s, 2s, 3s, and 4s) were chosen in this study. Then each T-second window data were transformed into its frequency domain representation by Fast Fourier transform. The 30 ×T frequency points between 5 Hz and 35 Hz were chosen as the input data of the network. This was performed for each of the six channels of the EEG data. For V-CNN approach, the six channels data were only from visual area (PO3, POz, PO4, O1, Oz, and O2). For the proposed VF-CNN approach, the six channels data were from visual area (PO3, POz, PO4, Oz) and frontal area (F3, F4) of the brain. The total number of samples for each participant was $4\times2\times20\times((6\text{-}T)/0.1+1)$.

D. Architecture of the convolutional neural network

The CNN consisted of five sequential layers in this study as shown in Fig. 2. The input data were processed as described in the previous section. L1 and L2 were both 2-D convolutional layer with batch normalization. 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 × 6 with a stride of 1. Similar, a total of 30 convolutional kernels were

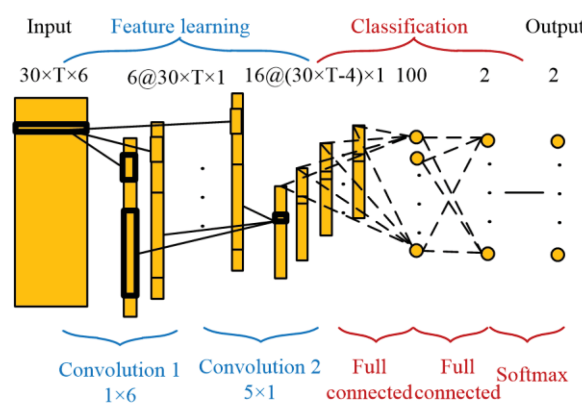

Figure 2. Illustration of the convolutional neural network architecture.

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 last layer, $L5$, used the softmax function. And the loss function for classification was cross entropy for two mutually exclusive classes.

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.

E. Attention-level detection

The power spectral density (PSD) estimation is performed through the periodogram technique to detect the attention-level of the participant [12]. Let S(*f*) be the value of the periodogram at frequency *f* (in Hz):

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

where S(*f*) is the value of the periodogram at frequency *f*, *x* is the EEG signal (average of the EEG data from F3, F4 channels) of n samples, and *N* is the total number of samples of the signal. EEG data with 2-second window length were analyzed and the window was moved in steps of 0.1s. The power of alpha rhythms (8-13 Hz) was extracted below.

$$P = \sum_{f=8}^{13} S(f) \qquad (2)$$

F. Canonical correlation analysis

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. 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 were 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. The components of the template signal were designed below.

$$\begin{bmatrix} F_1 & 2\times F_1 & F_1 + 2\times f_1 \\ F_2 & F_2 - 2\times f_2 & F_2 + 2\times f_2 \\ F_3 & F_3 - 2\times f_3 & 2\times F_3 \\ F_4 & F_4 - 2\times f_4 & F_4 + 2\times f_4 \end{bmatrix} \qquad (4)$$

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.

G. Statistical analysis

The mixed effect model of Analysis of variance (ANOVA) was used for statistical analysis. Methods (V-CNN and VF-CNN) and time window lengths 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.

III. Results

A. The EEG responses in the frontal area and visual area

Fig. 3 showed the average of the power (*P* in Eqn. (2)) and the average of ρ values (Eqn. (3)) across all the trails in each target in participant 5. The power was in alpha rhythms of the frontal channels (F3 and F4). And the ρ values were the CCA correlation coefficient between EEG signals and the temple signals at the target stimulation frequency. When the participant gazing at the gaiting stimulus (started at 0s), the correlation coefficient values increased and the power of alpha rhythms decreased in all the four targets. Then these values maintained relatively stable during the task period (time < 6s). During the gaze relax period (6 to 10s), the correlation coefficient values decreased and the power of alpha rhythms increased. Therefore, as the participant gazed at the gaiting stimulus, the attention-level changed and reflected in the power of alpha rhythms from the frontal area of the brain. Furthermore, the changes of the correlation coefficient values implied that the designed gaiting stimuli could induce the corresponding stimulation frequencies as shown in equation (4). Thus, both the EEG features from frontal area and visual area could be selected to do classification for "brain switch".

B. Comparison of the accuracies using the two CNN-based methods

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 4s with a step of 1s) were calculated and showed in Fig. 4. We observed that the accuracy increased with longer time window for both methods. And the accuracies, using VF-CNN method, achieved 79.07 ± 3.95%, 86.69 ± 5.18 %, 89.79 ± 5.78 %, and 91.94 ± 6.20% when the time window lengths were 1s, 2s, 3s, and 4s, respectively.

To further quantify the difference between these two methods, we used mixed effect model of ANOVA on the accuracies. The fixed factors methods and time window lengths both had significant effects on the accuracies ((F = 41.07, p< 0.001), (F = 61.60, p< 0.001)). Post-hoc comparison revealed that the accuracy using VF-CNN method were

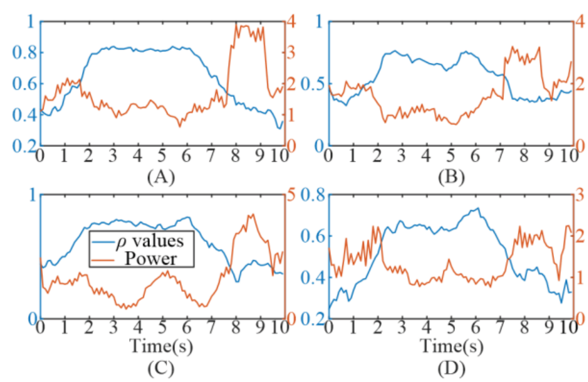

Figure 3. The average of the power of alpha rhythms and the average of the correlation coefficient values across all the trails in each target in participant 5. (A) left target (B) right target (C) up target (D) down target. Time 0s indicates beginning of gazing the stimulus and time 6s indicates the beginning of gaze relax.

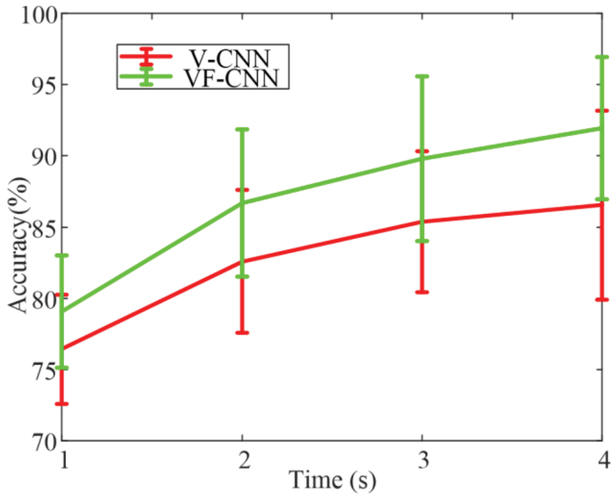

Figure 4. The classification accuracies using V-CNN method and VF-CNN method with different time window lengths.

significant higher than the accuracies using V-CNN method when the time window lengths were 2s, 3s, and 4s (p = 0.049, p = 0.021, and p = 0.001, respectively). While there was no significant difference on the accuracies using VF-CNN and V-CNN with 1s window length (p> 0.1). Thus the VF-CNN method showed significant performance improvement than the V-CNN method when the time window length was longer than 2s.

IV. Discussion and Conclusion

In this study, we utilized CNN to detect NC and IC states in a hybrid SSMVEP-based BCI integrated with AO. We illustrated the attentional level difference between participants engaged and disengaged with the gaiting visual stimuli. The results showed that the power of alpha rhythms from frontal area decreased and the correlation coefficient values increased when the participants gazed at the gaiting stimuli. We further demonstrated that combining the attentional feature and SSMVEP features could improve the performance of brain switch.

To our knowledge, only one recent study reported detecting IC and NC in AO in the context of BCI [8]. 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. In the current study, the background was always black and refreshed at 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. Since CNN has the ability of automatic feature extraction, different data processing and fusion procedures, which were needed in other hybrid BCIs [4], were not necessary in this study. We can perform CNN and simply replace two channels' data from visual area with the data from frontal area as the input data of CNN.

The proposed hybrid BCI approach achieved performance improvement comparing with the SSMVEP BCI approach in discriminating IC and NC. Furthermore, 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. However, current study only focused on offline classification for brain switch. So the future work will be developing a real-time brain switch system using 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, doi:10.1109/TRE.2000.847807.

[2] 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, doi:10.1109/TNSRE.2016.2591012.

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

[4] Yuanqing Li, Jiahui Pan, Fei Wang, and Zhuliang 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, doi:10.1109/TBME.2013.2270283.

[5] 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, doi:10.1007/s11571-014-9296-y.

[6] 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, doi:10.1152/physrev.00009.2013.

[7] 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, 2018, doi:10.1016/j.neures.2017.10.002.

[8] 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, doi:10.1109/TNSRE.2018.2878249.

[9] X. Zhang, G. Xu, J. Xie, and X. Zhang, "Brain response to luminance-based and motion-based stimulation using intermodulation frequencies," *PLoS One*, vol. 12, no. 11, p. e0188073, 2017, doi:10.1371/journal.pone.0188073.

[10] G. Thut, "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, doi:10.1523/jneurosci.0875-06.2006.

[11] D. H. Brainard, "The Psychophysics Toolbox," *Spat. Vis.*, vol. 10, no. 4, pp. 433–436, 1997, doi:10.1163/156856897X00357.

[12] 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, p. 56007, 2015, doi:10.1088/1741-2560/12/5/056007.