Deep Video Canonical Correlation Analysis for Steady State motion Visual Evoked Potential Feature Extraction

1st Raika Karimi
Electrical & Computer Eng.
Concordia University
Montreal, Canada
R_arim@encs.concordia.ca

2nd Arash Mohammadi
Concordia Institute for Inf. Systems Eng.
Concordia University
Montreal, Canada
arash.mohammadi@concordia.ca

3rd Laura Rosero
Fine Arts
Concordia University
Montreal, Canada

4th Amir Asif
Electrical & Computer Eng.
Concordia University
Montreal, Canada
amir.asif@concordia.ca

***Abstract*—Recently, there has been a surge of interest in development of Brain Computer Interface (BCI) systems based on Steady-State motion-Visual Evoked Potentials (SSmVEP), where motion stimulation is utilized to address high brightness and uncomfortably issues associated with conventional light-flashing/flickering. In this paper, we propose a deep learning-based classification model that extracts features of the SSmVEPs directly from the videos of stimuli. More specifically, the proposed deep architecture, referred to as the Deep Video Canonical Correlation Analysis (DvCCA), consists of a Video Feature Extractor (VFE) layer that uses characteristics of videos utilized for SSmVEP stimulation to fit the template EEG signals of each individual, independently. The proposed VFE layer extracts features that are more correlated with the stimulation video signal as such eliminates problems, typically, associated with deep networks such as overfitting and lack of availability of sufficient training data. The proposed DvCCA is evaluated based on a real EEG dataset and the results corroborate its superiority against recently proposed state-of-the-art deep models.**

***Index Terms*—Steady State Motion Evoked Potentials, EEG Signals, Brain Computer Interfaces, Deep CCA.**

I. Introduction

Throughout history, scientists and engineers have envisioned that interactions between our brain and the outer world can be established without intervention of human's voluntary nervous system. Dedicated determination towards achieving this goal has resulted in emergence of Brain-computer-interface (BCI) systems controlling external devices directly using brain signals independently of peripheral nerve pathways [1]–[3]. BCI systems have rapidly found their path in clinical studies sparked development of rehabilitation and assistive BCI technologies [4], [5]. Providing comfortable, convenient, cost-effective, and user-friendly BCI platforms are critically challenging for rehabilitation systems [6]. To this end, nowadays, one of the main tendency of neurotechnology companies is to incorporate Augmented Reality (AR) within BCI technologies [7], which further necessitates development of advanced signal processing/learning models for BCI systems.

This work was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through the NSERC Discovery Grant RGPIN-2016-04988.

Electroencephalogram (EEG)-based BCI systems developed based of Steady-State Visual Evoked Potential (SSVEP) are considered as the main technology for potential integration with AR due to their outstanding characteristics such as high accuracy and Information Transfer Rate (ITR) [8]–[10]. Continuous utilization of SSVEPs, however, causes eye fatigue and puts excessive mental load on subjects [11] rendering its practical utilization challenging. To address this issue, Motion-Onset Visual Evoked Potentials (mVEPs) [12]–[14], which elicit $P1$, $N2$, and $P2$ components in EEG signal are introduced as attractive alternatives. Recently, there has been a surge of interest on Steady-State motion Visual Evoked Potentials (SSmVEPs) [15] benefiting from advantages of both mVEPs and SSVEPs. The SSmVEPs comprise of reversal periodic movements such as Contraction-Expansion, Rotation, Swing, and Radial-Zoom [16], [17], which are used instead of conventional flickering-based stimuli.

While SSmVEPs are posed to pave the way for advancement of AR-based BCI systems, there are still in their infancy. Although the BCI systems developed based on SSmVEPs induce less eye fatigue, their frequency detection accuracy and ITR are not comparable to that of the SSVEPs, yet. In other words, Power Spectral Density (PSD) of the EEG signals when a SSVEP are used spike more intense modulated frequencies in comparison to the case where SSmVEPs are used under similar conditions [19]. In [18], features of the brain response to different flickering images are recognized by modeling visual pathways based on artificial neural networks. SSmVEPs, however, can evoke other harmonic frequencies/features of EEG signals depending on the type of designed motion paradigm due to their complex nature. The luminance of utilized colors [20], brightness contrast ratio [21], and existence of sharp edges in the design of SSmVEPs [16], [19] are examples of the complexities that can be taken into account for design of SSmVEPs. These effects can lead to the phase shift of evoked potentials, appearance of frequency peaks in the PSD, and other extra informative event related potential components.

Target identification is a crucial component of a SSmVEP-based BCI system where, conventionally, Canonical Correlation Analysis (CCA) is utilized [22]. The CCA tries to correlate a linear relationship between two multi-dimension variables, i.e., recorded EEG signals and template signals, which are functions of the SSmVEP frequencies. As regular CCA's performance can highly be affected by the interference of spontaneous EEG signals, its extensions, for instance, via spatial filtering are widely considered [23], [24]. The Task Related Component Analysis (TRCA) enhancing reproducibility of SSVEPs across multiple trials is another attempt to remove the unrelated background EEG activities [25]. In certain cases, CCA becomes weak in exploiting useful representatives of the underlying EEG data due to its nonlinearity. Kernel based CCA is a solution to nonlinearly project data to an embedding space, in which the linear CCA can be applied [26]. Furthermore, nonlinearity map of data can be generated from the template signals in a supervised fashion, and it was the first intuition to use the Recurrent Neural Networks (RNN) before applying CCA in SSmVEPs [29] or SSVEPs [27], [28]. Recent studies have shown that Convolutional Neural Networks (CNNs) can boost performance of BCI classifiers [30], [31]. In Reference [32], for instance, a new CNN method is applied to the complex Fast Fourier Transform (FFT) of EEG signals exploiting magnitude and phase information and outperforming CCA-based solutions. Using deep networks to find similarities between test and template signals, however, can lead to the overfitting issue due to small size of training datasets, typically, available for SSmVEPs-based training.

The paper is motivated by the desire for having a classification model that can extract all the aforementioned features of the SSmVEPs directly from the videos of stimuli. In this regard, we propose an intuitively pleasing deep learning-based classification model, referred to as the Deep Video Canonical Correlation Analysis (DvCCA), which consists of a Video Feature Extractor (VFE) layer that uses characteristics of videos utilized for SSmVEP stimulation to fit the template EEG signals of each individual, independently. The proposed VFE layer extracts features that are more correlated with the stimulation video signal as such eliminates problems, typically, associated with deep networks such as overfitting and lack of availability of sufficient training data. Moreover, in comparison to other deep networks, the proposed DvCCA architecture uses the extra information of videos, therefore, requiring less training data, which is critical for BCI systems having available few template samples of the subjects.

II. Methods and Materials

A. Stimulation Design

To achieve the objectives outlined in Section I, four novel SSmVEP based paradigms with four motion modes shown in Fig. 1 are designed. The first mode (Fig. 1(a)) is the "Reciprocal Rotation" of a green and black circle between $-45°$ and $75°$. The second motion (Fig. 1(b)) is the "Radial Zoom" in which the size of a green and black circle changes periodically. The third mode (Fig. 1(c)) is "Swing" centering around the middle of a black line. Finally, the last motion (Fig. 1(d)) is the "Sway" of a green rectangular centering

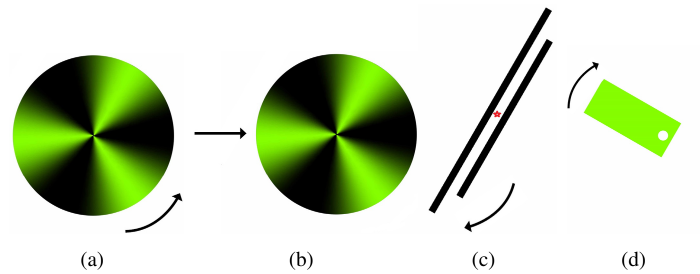


Fig. 1. The four designed SSmVEP paradigms: (a) Reciprocal-Rotation; (b) Radial-Zoom; (c) Swing; (d) Sway.

around the width. The frequency of motion direction change inside the reciprocal motion is defined as motion inversion frequency. The motion inversion frequency corresponds to the frequency of stimulation, which is equal to fundamental SSmVEP frequency in the EEG signals.

To elaborate on the motion choices utilized in our designs, we note that as mentioned in Reference [17], any paradigm with periodic motion can be used as stimuli of SSmVEPs. Each designed SSmVEP paradigm has a focal point where the paradigm oscillates/moves around this fixed center-point. Other factors that play crucial roles in our designs are brightness and luminance-contrast of the colors used in the paradigm. To make the proposed designs as efficient as possible, we capitalize on previous works [21] performed to investigate effects of different SSmVEP patterns with different brightness on accuracy and fatigue.

The refresh rate of the monitor showing the SSmVEP stimuli is a limiting factor restricting the frequencies that can be designed when an equal number of frames are used during consecutive cycles (one motion direction change). To implement flexible target frequencies, we need to have designs with variable number of frames per cycle; as such, we used the technique in Reference [34] and verified in [35] to implement our designs. Given the refresh rate of the monitor, the first task is to design binary stimulus sequences (i.e., the number of frames per each half-cycle, typically, asymmetric) with the goal of allocating frames based on the specified target frequencies. Following Reference [34], the number of frames per half-cycle of our paradigms are constructed as follows

$$S(f,i) = \text{square}\big[2\pi f(\frac{i}{R_r})\big], \tag{1}$$

where i indicates the frame index; f denotes the target frequency; R_r represents the monitor's refresh rate, and; $S(f,i)$ denotes the stimulus sequences associated with target frequency f. Note that, each half-cycle corresponds to one contraction or expansion or half of the reciprocal motion. Consequently, a motion oscillation at a target SSmVEP frequency up to $f \leq R_r/k$ can be generated as a stimulus. In other words, for one cycle of an SSmVEP paradigm to be understandable, k minimum number of frames per half-cycle is required for the subject to realize the motion cycle comfortably. The proposed SSmVEP paradigms are displayed as stimulations for EEG data collection. The first step is pre-processing of the EEG signals as they are, typically, exposed to artifacts and high/low frequency noises. To extract the SSmVEP from the EEG signals, applying spatial and time domain filters are, therefore, necessary. In time domain filtering, first, zero-phase Chebyshev Type I band-pass filter (2-40 Hz) is applied to smooth the data and remove high-frequency artifacts.

B. Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is a statistical method utilized to study the linear relationship between two groups of multi-dimensional variables. For two sets of signals arranged in matrices denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}$, the goal is to find two linear projection vectors $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$, such that the linear combination of the two groups of signals $\boldsymbol{w}_x^T\boldsymbol{X}$ and $\boldsymbol{w}_y^T\boldsymbol{Y}$ has the largest correlation coefficient, i.e.,

$$\rho = \max \frac{E(\boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_y)}{\sqrt{E(\boldsymbol{w}_x^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_x)E(\boldsymbol{w}_y^T\boldsymbol{X}\boldsymbol{Y}^T\boldsymbol{w}_y)}}. \quad (2)$$

Conventionally, the reference signals are constructed at the stimulation frequency f_i as (*in contrary, the proposed DvCCA uses a deep architecture to construct reference signals*)

$$\boldsymbol{y}_i = [\cos 2\pi f_i t, \sin 2\pi f_i t, \ldots, \cos 2\pi N_h f_i t, \sin 2\pi N_h f_i t]^T,$$

where $t = \frac{1}{f_s}, \ldots, \frac{m}{f_s}$, the f_s is the sampling rate, m is sample points, and N_h is the number of harmonics, which is dependent on the paradigm, and is obtained experimentally from the Welch Power Spectrum of signals.

III. Deep Video Canonical Correlation Analysis (DvCCA)

A. Initial Weight Extraction from Videos

The proposed DvCCA model is developed based on the following facts: (1) Changing color of each pixel in the stimulation video can evoke EEG signals; (ii) Harmonic frequencies of the signal of each pixel across time can appear in the EEG data depending on the distance of the pixel from the focal point of the target, and; (iii) The final visual evoked potential consists of the impact of the time signal of all the pixels. Therefore, the visual field incorporating the distance between the subject and the monitor and location of two pixels in the image is a suitable criterion to measure the mentioned distance in this context. The visual field (in degree) between the focal point and any arbitrary pixel is called eccentricity [13]. The more eccentricity of each pixel is from the focal point, the more trivial the trace of the time signal of the corresponding pixel will be in the VEP. Therefore, the first layer of the one side of our network tries to translate, via a dense layer, time series of different pixels to come up with a new representation for SSmVEP signal. The eccentricity of each pixel will be used for weight assignment in this layer.

B. The DvCCA Network Architecture

Architecture of the proposed DvCCA is shown in Fig. 2. The network has two segments consisting of 3 layers of dense neural networks. The objective is to project two datasets $\boldsymbol{X}$, and $\boldsymbol{Y}$ to a new space where the linear CCA is compatible to the types of the two dataset. Matrix $\boldsymbol{X} \in \mathbb{R}^{K\times m}$ represents the EEG time samples from K channels and is the input of the first segment. Matrix $\boldsymbol{Y} \in \mathbb{R}^{3f\times m}$, where m denotes number of time sample, and $3f$ denotes the output size of the Video Feature Extractor (VFE) layer, is the input of the second segment and output of the VEP layer. The propagation rule of each segment with input $\boldsymbol{X} \in \mathbb{R}^n$ is defined as follows

$$\boldsymbol{H}^{(i+1)} = \sigma(\boldsymbol{W}^X\boldsymbol{H}^{(i)} + \boldsymbol{b}^X), \quad (3)$$

where $\boldsymbol{W}^X \in \mathbb{R}^{n_i\times n_{i+1}}$ is a matrix of weights, $\boldsymbol{b}^X \in \mathbb{R}^{n_{i+1}}$, is a bias vector, H_i is the representation of data, and n_i is the number of nodes at i^{th} layer, and σ is Rectified Linear Unit (ReLU). The ultimate loss function is the linear CCA between output layers of the two segments. Adam optimizer is utilized for performing the optimization task.

Video Feature Extractor (VFE) Layer, is a dense layer in which the input is the video of stimulus with frame size $a \times b$ after applying sinc interpolation in time domain, and the output is time features associated with the video for each of the three RGB color channels. The time signals corresponding to pixels with identical distance to the focal point are averaged to generate a new tensor $\boldsymbol{Z} \in \mathbb{R}^{c\times m\times 3}$, where c is the number of groups of pixels. Each of the three RGB color channels, is provided separately to a different dense network and the outputs are concatenated. The input of the network is $\boldsymbol{Z}_i \in \mathbb{R}^c$ at the i^{th} time sample, and output of each dense network is

$$output = \sigma\big(\boldsymbol{W}^{\boldsymbol{Z}}\boldsymbol{Z}_i + \boldsymbol{b}^{\boldsymbol{Z}}\big), \quad (4)$$

where $\boldsymbol{W}^{\boldsymbol{Z}} \in \mathbb{R}^{c\times f}$ is a matrix of weights, and $\boldsymbol{b}^X \in \mathbb{R}^f$, is a bias vector. The learnable variables of the network are measured via back propagation of CCA loss function.

User Specific Training Procedure: For each subject, the method is fitted individually and the network is validated on the template data of the same participant. Moreover, we perform 8-Fold cross-validation to test the proposed DvCCA classifier. Each trial is divided into 8 equal segments and the algorithm is performed for different window lengths obtained from these segments.

Performance Evaluation: Classification accuracy and ITR evaluate the performance of the paradigms. The ITR is a valid criterion to assess the speed of BCI systems and is computed as

$$\text{ITR} = \frac{60}{T}\left[\log_2 K + \sigma\log_2\sigma + (1-\sigma)\log_2(\frac{1-\sigma}{K-1})\right], \quad (5)$$

where T is the sum of time of each trial and the resting state time between two trials, K is the number of stimuli, and σ is the recognition accuracy. Additionally, the accuracy-difference between trials of two consecutive sessions can be used as an effective assessment to measure the fatigue associated with a designed paradigm. Evaluation of SSmVEP performance via the one-way analysis of variance (ANOVA) is a renowned metric, therefore, ANOVA is run to guarantee that the subject's response to the SSmVEP-stimuli are statistically meaningful. The statistical significance was defined as $p < 0.05$.

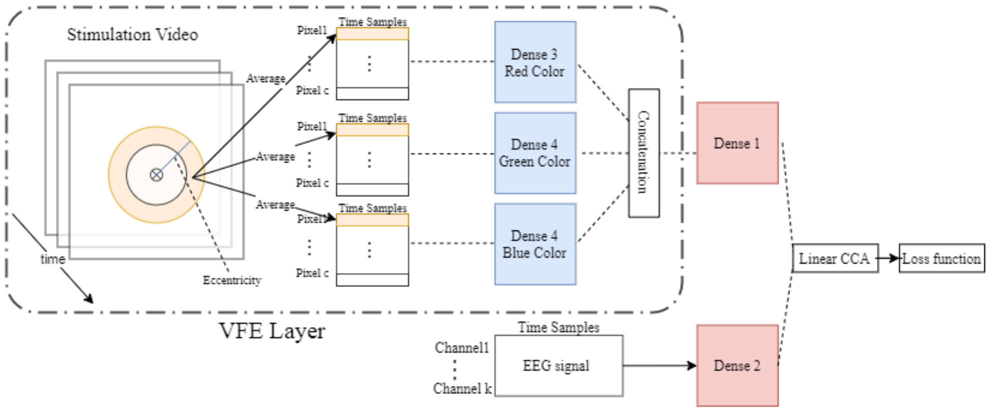


Fig. 2. Architecture of the proposed Deep Video Canonical Correlation Analysis (DvCCA).

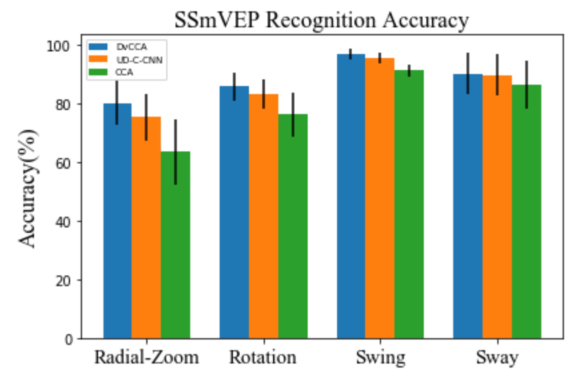
IV. Experimental Results

Experimental Setup: In this section, we evaluate the proposed DvCCA based on a real dataset consisting of 10 individuals, 5 women and 5 men between age of 20 to 27. Participants have no evidence of visual or color-recognition ailments. Five subjects had experience of BCI experiments. The EEG signals were collected using a portable and wireless bio-signal acquisition system (32 bipolar active wet electrodes with sampling rate 500Hz), g.Nautilus from g.tech Medical Engineering. The reference and the ground electrodes of the headset were placed at the earlobe and frontal position (Fpz), respectively. The electrodes P_z, P_{o7}, P_{o3}, P_{o4}, P_{o8}, and O_z from the parietal-occipital region were chosen to collect EEG signals. Stimuli with a white background were displayed on a 21.5-inch LED screen at 60Hz refresh rate. The resolution of the screen was 1920×1080, and the viewing distance was 70cm. The data were collected with the policy certification number 3007997 of Ethical acceptability for research involving human subjects approved by Concordia University.

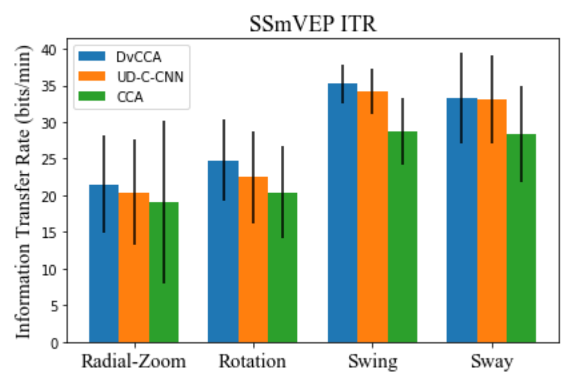
The size of the VFE layer is set to 10 in our experiments. The Learning Rate, Epoch Number, and Mini batch size are set to 10^{-3}, 50, and 100, respectively. The regularization parameter is set to 10^{-5} to avoid gradient exploding.

Experimental Protocol: Four paradigms, i.e., (i) Rotation; (ii) Radial zoom, and; (iii) Swing; (v) Sway, are experimented in separate runs. The target frequencies were 5, 6, 7, 8, and 9Hz and subjects are asked to focus on the focal point. Each run of a video included two consecutive sessions. Subjects were required to stare at a target using a pointer. In each session, each existing target is pointed in four trials. Each trial lasted 3.5 seconds with a 2.5 seconds break between consecutive trials and 10 minutes break between two runs.

Result: The proposed DvCCA is applied to the data of each subject separately. A recently proposed technique, the UD-C-CNN with the configuration introduced in Reference [32], is applied to our dataset for comparison purposes. More specifically, comparisons are performed between UD-C-CNN method, regular CCA, and the proposed DvCCA method. The target detection is done for eight time windows and the best



(a)



(b)

Fig. 3. Comparison of accuracy and Information Transfer Rate (ITR) across four different SSmVEP paradigms: (a) Mean and standard deviation of SSmVEP recognition accuracy comparisons. (b) ITR of different target identification methods across the 10 subjects.

results among time windows are represented in the diagrams.

1) Accuracy Comparisons: Fig. 3(a) evaluates performance in terms of the mean and standard accuracy across subjects. It can be observed that the proposed DvCCA outperforms its counterparts across all the tested paradigms. The DvCCA results in the accuracy of 80.1 ± 7.5, 85.8 ± 4.7, 96.8 ± 1.7, and 90.1 ± 7.1% for the Radial-Zoom, Reciprocal Rotation, Swing, and Sway paradigms, respectively. The swing paradigm has the best accuracy with one-way ANOVA showing statistical significance ($F = 5.72$, $p = 0.015$).

2) ITR Comparisons: Fig. 3(b) represents mean and standard deviation of the ITRs associated with different SSmVEP detectors across the 10 subjects. The DvCCA outperforms baselines in terms of speed of detection resulting in ITR of 21.5 ± 6.7, 24.8 ± 5.5, 35.2 ± 2.7, and 33.3 ± 6.2 for the Radial-Zoom, Reciprocal Rotation, Swing and Sway paradigms, respectively. The Swing paradigm with DvCCA classifier has the best ITR performance. Through our comparisons, the proposed DvCCA shows superiority for classifying SSmVEP paradigms.

V. Conclusion

Given recent surge of interest on SSmVEP-based BCI systems, the paper proposes a new deep learning based classifier, referred to as the Deep Video Canonical Correlation Analysis (DvCCA). The proposed DvCCA model consists of a Video Feature Extractor (VFE) layer that uses characteristics of videos to fit to the template EEG signals of each individual independently, which results in the extracted features to be more correlated with the stimulation video signal. Th proposed VFE uses characteristics of videos eliminating problems associated with more complicated networks such as overfitting and/or lack of enough training data. The proposed DvCCA is evaluated based on a real EEG dataset outperforming its state-of-the-art counterpart and achieving accuracy of 80.1 ± 7.5, 85.8 ± 4.7, 96.8 ± 1.7, 90.1 ± 7.1 over the four SSmVEPs.

References

[1] S. Shahtalebi and A. Mohammadi, "Bayesian Optimized Spectral Filters Coupled With Ternary ECOC for Single-Trial EEG Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2249-2259, 2018.

[2] M. Mirgholami, S. Shahtalebi, W. Cui, R. Karimi, A. Asif, and A. Mohammadi, "Adaptive Subject-specific Bayesian Spectral Filtering for Single Trial EEG Classification," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1-5, 2019.

[3] G. Schalk, *et al.*, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System," *IEEE Trans. Biomedical Eng.*, vol. 51, no. 6, pp. 1034-1043, 2004.

[4] L.I. Jovanovic, N. Kapadia, L. Lo, V. Zivanovic, M.R. Popovic, C. Marquez-Chin "Restoration of Upper Limb Function After Chronic Severe Hemiplegia: A Case Report on the Feasibility of a Brain-Computer Interface-Triggered Functional Electrical Stimulation Therapy," *American J. Physical Medicine & Rehabilitation*, vol. 99, 2020.

[5] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-Computer Interfaces for Communication and Rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, 2016.

[6] Q. Jiang, F. Shao, W. Lin, and G. Jiang, "On Predicting Visual Comfort of Stereoscopic Images: A Learning to Rank based Approach," *IEEE Signal Process. Lett.*, vol. 23, vol. 2, pp. 302-306, 2016.

[7] H. Si-Mohammed, *et al.*, "Towards BCI-based Interfaces for Augmented Reality: Feasibility, Design and Evaluation," *IEEE Trans. Vis. Comput. Graphics*, 2018.

[8] E.K. Kalunga, S. Chevallier, and Q. Barthélemy, "Transfer Learning for SSVEP-based BCI using Riemannian Similarities Between Users," *European Signal Process. Con. (EUSIPCO)*, pp. 1685-1689, 2018.

[9] E. Kinney-Lang, A. Ebied, and J. Escudero, "Building a Tensor Framework for the Analysis and Classification of Steady-state Visual Evoked Potentials in Children," *European Signal Processing Conference (EUSIPCO)*, pp. 296-300, 2018.

[10] P. Stawicki, *et al.*, "Investigating Spatial Awareness within an SSVEP-based BCI in Virtual Reality," *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 615-618, 2018.

[11] J. Xie, *et al.*, "Effects of Mental Load and Fatigue on Steady-state Evoked Potential based Brain Computer Interface Tasks: A Comparison of Periodic Flickering and Motion-reversal based Visual Attention," *PloS one*, vol. 11, no. 9, 2016.

[12] R. Beveridge, S. Wilson, M. Callaghan, and D. Coyle, "Neurogaming with Motion-onset Visual Evoked Potentials (MVEPs): Adults versus Teenagers," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 572-581, 2019.

[13] J. Chen, *et al.*, "A Single-stimulus, Multitarget BCI based on Retinotopic Mapping of Motion-onset VEPs," *IEEE Trans. Biomedical Eng.*, vol. 66, no. 2, pp. 464-470, 2018.

[14] T. Ma, *et al.*, "The Extraction of Motion-onset VEP BCI Features based on Deep Learning and Compressed Sensing," *J. Neuroscience Methods*, vol. 275, pp. 80-92, 2017.

[15] C. Han, G. Xu, J. Xie, C. Chen, and S. Zhang, "Highly Interactive Brain–Computer Interface based on Flicker-Free Steady-state Motion Visual Evoked Potential," *Scientific Reports*, vol. 8, no. 1, 2018.

[16] X. Chai, Z. Zhang, K. Guan, G. Liu, and H. Niu, "A Radial Zoom Motion-based Paradigm for Steady State Motion Visual Evoked Potentials," *Frontiers in Human Neuroscience*, vol. 13, no. 127, 2019.

[17] 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. Biomedical Eng.*, vol. 65, no. 8, pp. 1696-1704, 2017.

[18] F. Sobreira, C. Tremmel, and D. J. Krusienski, "Modeling the Visual Pathway for Stimulus Optimization in Brain-Computer Interfaces," *European Signal Processing Conference (EUSIPCO)*, pp. 1672-1675, 2018.

[19] R. Karimi, L. Rosero, M. Nirgholami, A. Asif, and A. Mohammadi, "Study on Novel designs with Reduced Fatigue for Steady State Motion Visual Evoked Potentials," *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp 1-5, 2019.

[20] W. Yan, , *et al.*, Steady-state Motion Visual Evoked Potential (SSmVEP) based on Equal Luminance Colored Enhancement," *PloS one*, vol. 12, no. 1, 2017.

[21] W. Yan, *et al.*, "Study on the Effects of Brightness Contrast on Steady-state Motion Visual Evoked Potential," *IEEE Int. Con. Eng. Medicine and Biology Society (EMBC)*, pp 2263-2266, 2017.

[22] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency Recognition based on Canonical Correlation Analysis for SSVEP-based BCIs," *IEEE Trans. Biomedical Eng.*, vol. 53, no. 12, pp. 2610-2614, 2006.

[23] Y. Zhang, *et al.*, "Hierarchical Feature Fusion Framework for Frequency Recognition in SSVEP-based BCIs," *Neural Networks*, vol. 119, 2019.

[24] Y. Zhang, *et al.*, "Two-stage Frequency Recognition Method based on Correlated Component Analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1314-1323, 2018.

[25] M. Nakanishi, , *et al.*, "Enhancing Detection of SSVEPs for a High-speed Brain Speller using Task-related Component Analysis," *IEEE Trans. Biomedical Eng.*, vol. 65, no. 1, pp. 104-112, 2017.

[26] Z..M Zhang and Z.D. Deng, "A Kernel Canonical Correlation Analysis based Idle-state Detection Method for SSVEP-based Brain-Computer Interfaces," *Advanced Materials Research*, vol. 341, pp. 634-640, 2012.

[27] H. Vu, B. Koo, and S. Choi, "Frequency Detection for SSVEP-based BCI using Deep Canonical Correlation Analysis," *IEEE Int. Con. Systems, Man, & Cybernetics (SMC)*, pp. 1983-1987, 2016.

[28] Q. Liu, *et al.*, "Efficient Representations of EEG Signals for SSVEP Frequency Recognition based on Deep Multiset CCA," *Neurocomputing*, vol. 378, pp. 36-44, 2020.

[29] Z.K. Gao, *et al.*, Multivariate Weighted Recurrent Network for Analyzing SSmVEP Signals from EEG Literate and Illiterate," *Europhysics Letters*, vol. 127, no. 4, 2019.

[30] N. Khadijah, *et al.*, "On the Classification of SSVEP-based Dry-EEG Signals via Convolutional Neural Networks," *IEEE Int. Con. Systems, Man, & Cybernetics (SMC)*, pp. 3726-3731, 2018.

[31] T.H. Nguyen and W.Y. Chung, "A Single-Channel SSVEP-based BCI Speller using Deep Learning," *IEEE Access*, vol. 7, pp. 1752-1763, 2018.

[32] A. Ravi, N. Heydari Beni, J. Manuel, and N. Jiang, "Comparing User-dependent and User-independent Training of CNN for SSVEP BCI," *J. Neural Engineering*, 2020.

[33] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep Canonical Correlation Analysis," *International Conference on Machine Learning*, pp. 1247-1255, 2013.

[34] M. Nakanishi, Y. Wang, Y.T. Wang, Y. Mitsukura, and T.P. Jung, "An Approximation Approach for Rendering Visual Flickers in SSVEP-based BCI using Monitor Refresh Rate," *IEEE Int. Con. Eng. Medicine and Biology Society (EMBC)*, pp. 2176-2179, 2013.

[35] M. Nakanishi, Y. Wang, Y.T. Wang, Y. Mitsukura, and T.P. Jung, "Generating Visual Flickers for Eliciting Robust Steady-state Visual Evoked Potentials at Flexible Frequencies using Monitor Refresh Rate," *PloS one*, vol. 9, no. 6, 2014.