# Discovering Novel Neural Patterns in Visually Evoked Potentials Through CNNs

**Nelson Hidalgo**
Department of Brain and Cognitive Science and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
nelsonh@mit.edu

## Abstract

Steady State Visually Evoked Potential (SSVEP) have been widely used in creating brain-computer interfaces and studying visual attention. As convolutional neural networks (CNNs) have outperformed conventional methods of SSVEP analysis, there is an important need to unveil some of the abstractions of CNNs and allow researchers to meaningfully interpret patterns in visual processes that CNNs learn. To bridge this gap, we implemented a CNN trained to identify the target stimulus frequencies in an SSVEP study, and we evaluated the features it learned using a gradient ascent algorithm, EEGDream. Our CNN outperformed the conventional CCA analysis (70 % vs. 60%) in classifying participant independent SSVEP target frequencies, especially in conditions with poor signal quality. Our EEGDream analysis of the features learned by the CNN show strong evidence that the network learned to identify the peak spectral frequencies as well as the spatial patterns in the SSVEP signal. The EEGDream visualization also indicated that our CNN identified the presence of adjacent frequencies in the SSVEP for the classification of some of the stimulus frequencies. Thus, our research demonstrates the feasibility of using EEGDream to identify at a granular level fundamental and novel features learned by CNNs from SSVEP signals.

## 1 Introduction

### 1.1 Background

We owe much of our knowledge about the visual system to studying visually evoked potentials, which are electrical potentials in the brain produced by different visual stimuli. A widely studied visually evoked potential is the Steady State Visually Evoked Potential (SSVEP), which is a type of potential that occurs at the same frequency as a flashing light stimulus. SSVEPs are produced at frequencies between 6-90 Hz Herrmann [2001] and exhibit a distinct peak precisely at the stimulus frequency Norcia et al. [2015]. They are found mainly in the occipital area, but are widely distributed through the parietal, temporal, frontal, and prefrontal areas (Pastor et al. [2003], Di Russo et al. [2007]).

SSVEPs have been widely used in creating brain-computer interfaces (BCIs), devices that allow participants to interact with the virtual or physical world using their brain signals (Armengol-Urpi and Sarma [2018],Chen et al. [2014]). For these BCI applications, it is important to build faster, more reliable SSVEP detection algorithms. SSVEPs have also been used to study visual attention, binocular rivalry, and working memory Norcia et al. [2015].

Historically, the main tool used to analyze SSEVPs was cannonical correlation analysis (CCA), which is a statistical method that correlates SSVEP signals with template sinusoids matching the frequency of the stimuli presented (Armengol-Urpi and Sarma [2018], Nakanishi et al. [2015], Lin et al. [2006]).

Preprint. Under review.

In recent years, convolutional neural networks (CNNs) have outperformed state-of-the-art SSVEP analysis and BCI classification methods in terms of classification accuracy (Ravi et al. [2020], Kwak et al. [2017], Waytowich et al. [2018]). An important question that this success raises is whether CNNs are detecting previously discovered or new patterns in neural signals. Studies have shown that CNNs do extract some well-known properties from SSVEPs, such as the expected frequency power band (Ravi et al. [2020]) as well as amplitude and phase-related information (Kwak et al. [2017], Waytowich et al. [2018]).

While Waytowich et al. [2018] analyzed the properties of the CNN's temporal kernels, his study and others in the literature have not explored the features learned by the CNN at the level of specific neurons in the network. The development of gradient ascent methods, which consists in optimizing network inputs to produce maximal activation of neurons (Erhan et al. [2009], Mordvintsev et al. [2015]), would allow for further exploration of what features CNNs learn from neural signals.

To expand on work to interpret the features learned by CNNs in the analysis of SSVEP signals, my research focuses on visualizing and assessing the features learned by a CNN that is trained on an open-source SSVEP-based BCI data-set originally published in Nakanishi et al. [2015]. This data-set consists of electroencephalogram (EEG) recordings from participants who were tasked to look at flickering keys on a virtual keyboard. The CNN is trained to classify what specific key the participant looked at based on the participant's EEG signal.

In order to analyze the features learned by the network, we designed a gradient ascent method, which we term EEGDream as it is inspired by DeepDream (Mordvintsev et al. [2015]). EEGDream reconstructs simulated EEG data, which consists purely of Gaussian noise, into a signal that most optimally activates the output neurons. Our hypothesis is that our model will not only learn fundamental features in SSVEP, but also identify novel spectral or temporal properties of SSVEP signals in the brain that conventional CCA was not able to detect.

### 1.2 Previous work

The studies that utilize CNNs to analyze SSVEP use 2D CNNs and often rely on the frequency power spectrum of EEG signals rather than the original signal (Ravi et al. [2020], Kwak et al. [2017], Waytowich et al. [2018]). One limitation of this method is that only power spectrum information is being used to analyze the signals, omitting time-series information in the signal. The CNN we implemented overcomes this issue by using a 1D CNN that allows us to analyze EEG signals directly, which could allow us to detect novel features in the data. Our architecture draws on some key ideas: (1) implementing a convolution across time with linear activation at the first layer to automatically extract spectral information in the signal (Khok et al. [2020], Waytowich et al. [2018]), (2) using depth-wise and separable convolutions to minimize the number of trainable parameters in the network (Waytowich et al. [2018]).

Current studies visualize the features used by CNNs using t-Distributed Stochastic Neighbor Embedding (t-SNE), a method for visualizing high-dimensional features in different clusters (Kwak et al. [2017], Waytowich et al. [2018]). However, this method only allows us to visualize clusters that may confirm or not that the network learned to differentiate between different frequencies, but it does not allows us to visualize the features learned by CNNs in terms of more granular features in the SSVEP signal. The development of the DeepDream gradient ascent algorithm (Mordvintsev et al. [2015]), which has been previously used to visualize the features that networks trained on images learn, motivated my design of EEGDream to investigate novel neural patterns detected by CNNs from EEG signals.

## 2 Methods

### 2.1 Data acquisition and training methods

The SSVEP experiment was conducted by Nakanishi et al. [2015] and made publicly available at `https://github.com/nel-hidalgo/EEGDream`. All participants were asked to provide informed consent approved by The Human Research Protections Program of the University of California San Diego, and there was no personally identifiable inforation in the dataset. As illustrated in Figure 1.A, data was collected from the scalp of 10 participants using the BioSemi ActiveTwo EEG system (Biosemi, Inc.), an external Ag/AgCl EEG system, as they looked at one of 12 flickering keys on a 27

inch LCD screen. The EEG data was recorded from channels PO7, PO3, POz, PO4, PO8, O1, Oz, and O2. Each key on the screen was flickering at frequencies between 9.75-14.75 Hz and phases ranging from 0-1.5 radians as specified in 1.A.i. The participants performed 15 blocks of the experiment, where each block consisted in 12 4-seconds trials in which they focused on one of the keys indicated at random by the program while all the keys flickered on the display simultaneously. The data made available was sampled at 256 Hz. In my experiment, each 4-seconds trial was subdivided into 1-second windows in order to increase the number of training/testing examples, making for a total of 60 samples per class for each participant. In addition, we filtered the data between 9 and 30 Hz using a 2nd order Butterworth filter in order to eliminate noise and preserve only the first and second harmonics of the SSVEP potential. We also standardized the data between 0 and 1.

For the purpose of investigating SSVEP patterns that are participant-independent, we use a leave-one-subject-out cross validation method. Each model was trained on data from 9 participants and tested on the data from one participant, who was left out from training. The proposed CNN algorithm is tested against CCA, which is one of the most widely used SSVEP analysis methods (Armengol-Urpi and Sarma [2018]). The analysis methods discussed in Nakanishi et al. [2015], such as Individual Template CCA or Combined CCA, which achieved highest accuracy on this dataset originally are excluded from this analysis because they are participant-dependent methods that performed poorly in a participant independent study by Waytowich et al. [2018].

### 2.2 Custom CNN model

The architecture shown in 1.B, which is fully implemented together with EEGDream at `https://github.mit.edu/nelsonh/SSVEPNet`, was used to analyze the SSVEP signals. Its first layer is a convolutional layer with a filter of dimension 128. According to Waytowich et al. [2018] and Khok et al. [2020], the first layer acts as a type of band pass filter that extracts spectral information from the signal across time. This first layer is a depth-wise convolution, which mean that they convolve channels across time without mixing information across channels. The following layers are two depth-wise separable convolutions with exponential linear unit (ELU) acitavtions. ELU activations have been shown to perform better in neural networks for EEG classification (Lawhern et al. [2018]). Depth-wise separable convolutions perform a convolution along the time domain in one step and mix that information with another set of filters in a second step. This property means that there is less trainable parameters, thus making this CNN more appropriate for EEG applications in which data is often scarce. Another important characteristic of these layers is that they implement large filters of dimension 128 and 64, as those have been shown to be more appropriate for the analysis of time-series data with CNNs (Khok et al. [2020]). The hyper-parameters in this network were

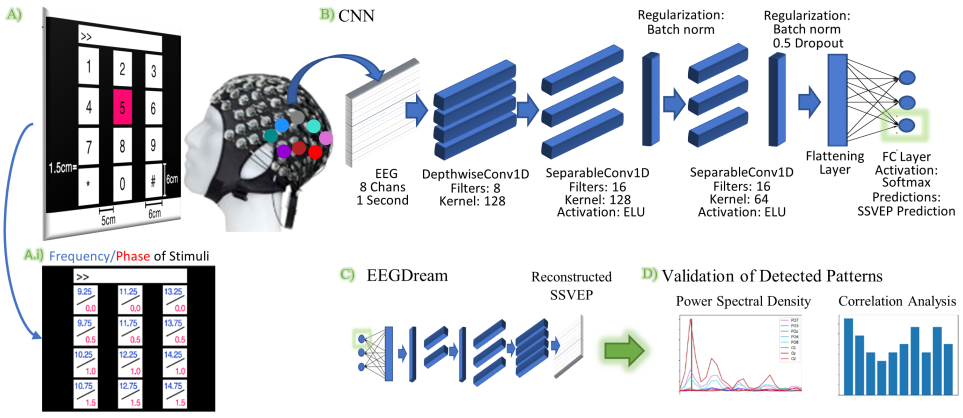


Figure 1: Diagram of experimental methods: (A) the participant looked at flickering keys on a screen while brain response in the occipital region is recorded with EEG, (B) EEG recordings are processed through our CNN to predict the key the participant looks at, (C) EEGDream gradient ascent produces EEG signals based on features learned by CNN, and (D) the synthesized eeg signals are validated via spectral analysis and CCA analysis.

determined based on existing literature, including Waytowich et al. [2018] and Khok et al. [2020], and further hyper-parameter search will be implemented in future work.

### 2.3 CCA Analysis

We compare our model with the well-established CCA method. CCA is a statistical analysis tool that compares two sets of data $X$, which is the EEG data shaped as channels by samples, and $Y_f$, which is a simulated sinusoid of frequency $f$ with the same number of samples as $X$. There is a $Y_f$ matrix per stimulus frequency $f$, and the maximum correlation value $\rho_f$ is computed between $X$ and each $Y_f$ according to Equation 1. Then, the algorithm predicts that the participant was looking at frequency $f$, which is the frequency with maximum correlation $\rho_f$.

$$\rho_f = \max_{W_x, W_y} \frac{E\left[W_x X^T Y_f W_y^T\right]}{\sqrt{E\left[W_x X^T X W_x^T\right] E\left[W_y Y_f^T Y_f W_y^T\right]}} \tag{1}$$

### 2.4 EEGDream and validation

Our EEGDream algorithm adapted the implementation of DeepDream (Mordvintsev et al. [2015]), a method of gradient ascent that reconstructs an input image by modifying it so as to produce maximal activation, meaning maximum mean difference, in a chosen layer.

EEGDream, as shown in Figure 1.C, takes as input a simulated 8-channels EEG signal made of Gaussian noise. Then, the input signal undergoes gradient ascent following Equation 2,where $\eta$ is the step size and $X$ is the input signal. We use the loss function in Equation 3, where $y_i$ is the target frequency label and $CNN(X^{(i)})$ is the output prediction of the CNN. These equations effectively perform gradient ascent by maximizing the activation of the output neuron that predicts a frequency.

$$X^{(i)} = X^{(i-1)} - \eta \nabla_X L(X^{(i)}) \tag{2}$$

$$L(X^{(i)}) = -\sum_{n=1}^{12} y_i log(CNN(X^{(i)})) \tag{3}$$

A reconstructed EEG signal will be obtained for each target stimulus frequency, making for 12 reconstructed EEG signals for each of the 10 leave-one-subject-out CNN networks. We later perform an averaged spectral analysis across all 10 networks, from which we will obtain 12 averaged spectra, one per target stimulus frequency. As shown in Figure 1.D, analysis of the reconstructed signal via this technique and CCA will help us assess how well the network learned to reproduce SSVEP signals with appropriate peak frequencies. We also compute a loss score by training the network with one EEG channel turned off at a time to understand what channels contained most important features.

## 3 Results

### 3.1 CNN vs. CCA performance

The performance of the CNN and CCA were evaluated on 1 second long EEG data from each participant with 60 trials per stimulus frequency. As shown in Figure 2, our CNN network outperformed CCA across most participants. On average, our CNN performed ten percent better than the CCA network. The CCA performance is a proxy to the level of noise in the signal because, theoretically, if we had high signal to noise ratio in the signal, CCA would have perfect performance. In particular, the CNN and CCA performed similarly on participant s1, which had significant amount of noise in the data, making it difficult to identify SSVEP signals in it. However, even in a noisy environment such as that observed in participant s0, the CNN was better able to classify SSVEP patterns. In the cases where CCA performance was higher than 50 percent, which indicates that there was a cleaner signal, the CNN achieved significantly higher performance (participants s2, s3, s4, s6, s7, and s9).

In addition, the extent to which our CNN predicted frequencies accurately is presented in Figure 3.On average, the CNN had 70 percent accuracy in identifying different frequencies, meaning that it

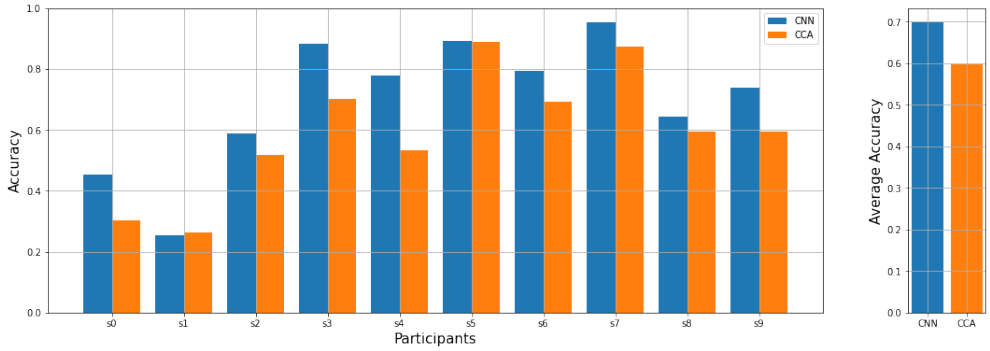


Figure 2: Accuracy of our CNN model and CCA on 60 samples of 1-second EEG data per stimulus frequency from each participant. While CCA did not require prior training, our CNN was trained based on leave-one-subject-out cross validation and tested on the subject which was left out.

performed significantly higher than chance, which was 8.3 % in this case. We observe in Figure 3 that the frequencies which the network misclassified most were frequencies adjacent to the target frequency. For example, in the 12.25 Hz frequency label, the network mistakes that frequency most often with 11.75 and 12.75 Hz, which are the light blue classes along the 12.25 Hz row. This is expected as those are the closest frequencies to the target stimulus.

### 3.2 Spectral and CCA Analysis of Learned Features

Following the methodology described in the last paragraph of Section 2.4, we optained the averaged power spectra in Figure 4. We obtained a power spectrum that characterizes the features learned by CNNs independent of which participants it was trained on. For example, the psd plot labeled *(*Freq 9.25) is the average power spectrum across reconstructed SSVEP signals from 10 training instances of the CCN network corresponding to an SSVEP at 9.25 Hz. The results in Figure 4 show that a spectrum peak was present at precisely the expected frequencies, indicated as a black vertical line, which show that all reconstructed EEG signals match real SSVEP potentials recorded with EEG in terms of power spectrum. Higher peaks, such as frequencies 9.25, 11.25, and 10.25, indicate SSVEP signals that were best learned by the network.

In addition to spectral analysis, we performed CCA analysis on the reconstructed signal as a way to assess how closely it resembles real SSVEP signals in the time domain. The results in Figure 5.C

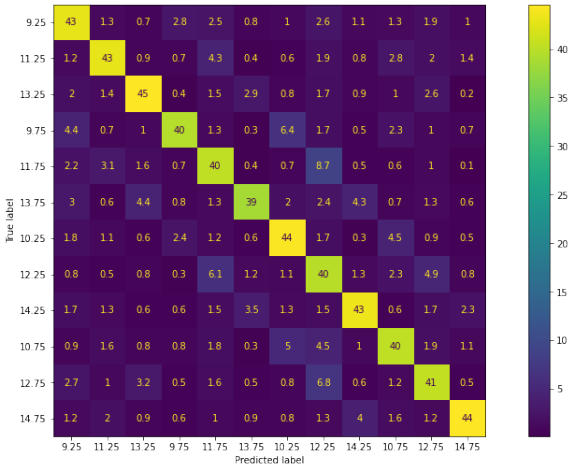


Figure 3: Confusion matrix of CNN SSVEP prediction out of 60 trials per frequency averaged across participants. On the y-axis are the true labels and on the x-axis the predicated labels.

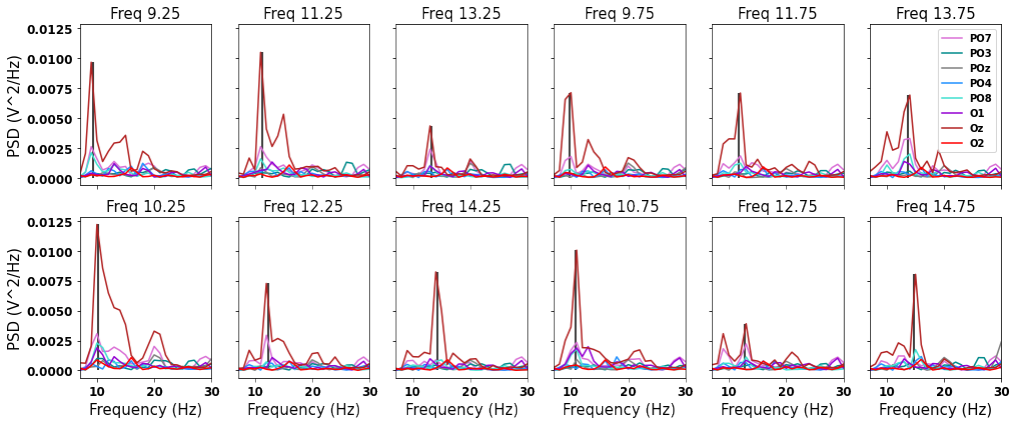


Figure 4: Averaged power spectrum across all 10 network initializations ranging from 9.25 to 14.5 SSVEP frequency. The dark vertical line indicates the frequency of the target stimulus. Peaks in the spectrum indicate frequencies that are strongly present in the reconstructed SSVEP signal.

show the results of this analysis on the reconstructed EEG signals, which resulted in higher than 80 percent accuracy for 9 out of ten of the reconstructed signals.

The loss score, shown as a headplot in Figure 5.B was computed to assess what channels were most important in the network's predictions. Compared to the headplot of average power spectrum from all EEG channels across trials in Figure 5.C, which represents what channels contain most information in the orginal signal, our network presents a similar distribution of most important channels.

## 4 Discussion

In our research, we evaluated the performance of our CNN architecture against CCA. Our results provide sound evidence that our CNN outperforms the standard CCA method. Our network was better able to identify SSVEP patterns, especially in situations where the signal quality may not have been optimal in which CCA performed with lower than 60 % accuracy. These results indicate that our CNN model may be applied to SSVEP-based BCI applications to improve (1) information transfer rate, as our model only requires 1 second of data to detect SSVEP, and (2) accuracy that is independent of the participant. However, a limitation of this study is that it did not deploy the model in real-time, so it must be tested further on a real-time BCI study.

Given the improved performance of our model, we went on to test what features it had learned from the SSVEP signal. Our hypothesis that the CNN model would be able to learn fundamental features in the SSVEP, such as the peak in spectra at the same frequency as the target stimulus, was confirmed

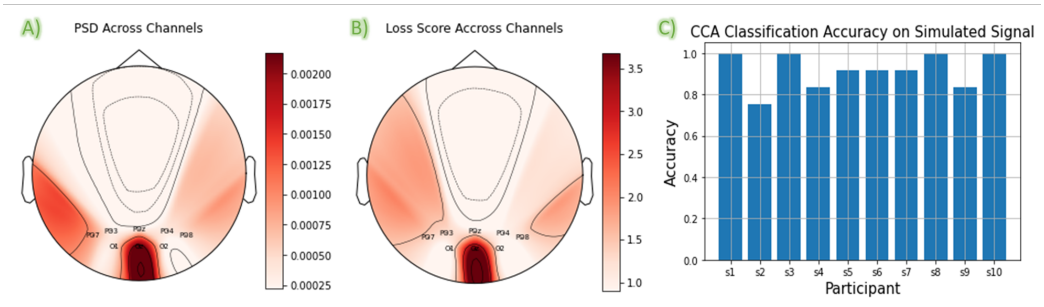


Figure 5: Headplot (A) shows darker shades of red for channels that have strongest activation for frequencies between 7 and 20 Hz in the original SSVEP signal averaged across participants. Headplot (B) shows darker shades of red for channels that cause greatest loss in performance when eliminated. The barplot (C) shows the accuracy of CCA ran on the reconstructed SSVEP signal.

based on our analysis in spectral and CCA analysis of the EEGDream reconstructed signals in Section 3.2. In addition, we learned that our CNN identified spatial features that closely match the features of real SSVEP signals. As a result, we can conclude that our model did not over fit to noise or artifacts particular to specific participant data, but rather it learned well-defined features in EEG. These results are consistent with previous findings that used hand-crafted EEG features rather than time-domain signals in their CNN implementations (Kwak et al. [2017], Ravi et al. [2020]). This establishes the feassibility of using our EEGDream and loss-score based spatial analysis to assess the features learned by CNN models from frequency-coded EEG data. A limitation of this result, and a topic of future research, is that a more compact metric should be developed which allows us to use our EEGDream-based methods as a tool to compare how well different CNN models stack against one another in terms of identifying relevant EEG features.

To address our hypothesis that our CNN model learned novel features in the SSVEP data, we relied on the aforementioned EEGDream-based visualizations. Our results in Figure 4 indicate that our CNN model did not only detect the peak corresponding to the stimulus frequency in making predictions, but it also detected frequencies that the participants were not directly focusing on. At the 9.25 , 11.25, 10.25, and 9.75 Hz target frequency, for example, frequencies around 14.75 Hz show a small peak that seems to be important in the network prediction. Nonetheless, we were not able to identify a consistent pattern of smaller SSVEP frequencies that indicate shifts in the participants' attention across participants. Therefore, the main conclusion we can draw is that the network classified based on frequencies other than the main stimulus frequency, but it is not clear if those patterns reflect a relationship between different frequency stimulations in the brain. One limitation we face is not being able to establish patterns that exist accross different SSVEP-based BCI studies given that the scope of this study only included one dataset.

## 5 Conclusion

Our research has implemented a CNN that may be applied to SSVEP-based BCIs to obtain higher classification accuracy. We have also implemented a set of validation techqniques, including EEG-Dream, to assess what types of features the CNN learned from SSVEP. Our results contribute to work on debunking some of the abstractions about the features learned by CNNs in the classification of neural signals. In addition, our work constitutes an important initial step towards learning what novel features CNNs learn from neural signals. On that regard, further work is needed to expand on this research effort, especially through research that involves large-scale analysis of several state-of-the-art CNN architectures on well-constrained SSVEP experiments.

## References


Christoph S Herrmann. Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental brain research*, 137 (3):346–353, 2001.

Anthony M Norcia, L Gregory Appelbaum, Justin M Ales, Benoit R Cottereau, and Bruno Rossion. The steady-state visual evoked potential in vision research: A review. *Journal of vision*, 15(6):4–4, 2015.

Maria A Pastor, Julio Artieda, Javier Arbizu, Miguel Valencia, and Jose C Masdeu. Human cerebral activation during steady-state visual-evoked responses. *Journal of neuroscience*, 23(37):11621–11627, 2003.

Francesco Di Russo, Sabrina Pitzalis, Teresa Aprile, Grazia Spitoni, Fabiana Patria, Alessandra Stella, Donatella Spinelli, and Steven A Hillyard. Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. *Human brain mapping*, 28(4):323–334, 2007.

Alexandre Armengol-Urpi and Sanjay E Sarma. Sublime: a hands-free virtual reality menu navigation system using a high-frequency ssvep-based brain-computer interface. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–8, 2018.

Xiaogang Chen, Zhikai Chen, Shangkai Gao, and Xiaorong Gao. A high-itr ssvep-based bci speller. *Brain-Computer Interfaces*, 1(3-4):181–191, 2014.

Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PloS one*, 10(10):e0140703, 2015.

Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE transactions on biomedical engineering*, 53(12):2610–2614, 2006.

Aravind Ravi, Nargess Heydari Beni, Jacob Manuel, and Ning Jiang. Comparing user-dependent and user-independent training of cnn for ssvep bci. *Journal of neural engineering*, 17(2):026028, 2020.

No-Sang Kwak, Klaus-Robert Müller, and Seong-Whan Lee. A convolutional neural network for steady state visual evoked potential classification under ambulatory environment. *PloS one*, 12(2): e0172578, 2017.

Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of neural engineering*, 15(6):066031, 2018.

Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer features of a deep network. *University of Montreal*, 1341(3):1, 2009.

Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Inceptionism: Going deeper into neural networks. 2015.

Hong Jing Khok, Victor Teck Chang Koh, and Cuntai Guan. Deep multi-task learning for ssvep detection and visual response mapping. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1280–1285. IEEE, 2020.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.