# Observing Video Stimuli: Final Report


**Victor Ye**
Product Leader
Santa Cruz, United States
vye2@ucsc.edu

**Jacob Wynd**
Team Member
Santa Cruz, United States
jwynd@ucsc.edu

**Tunglin Lee**
Team Member
Santa Cruz, United States
tlee76@ucsc.edu



## ABSTRACT

Visual stimuli from 5 categories of videos (abstract, waterfalls, faces, Rube Goldberg mechanisms, and speed) were shown to subjects, resulting in different brain wave patterns collected by an electroencephalogram (EEG) test. This data is processed into matrices divided by video. We then train a convolutional neural network (CNN) to determine based on the EEG data, which video the subject was watching. We believe that this model could eventually be useful in determining the differences between properly functioning brains, and indications of health risks. Furthermore in learning how to identify the brain's response to certain external stimuli, it could be useful in communicating with, and improving the quality of life of people who are otherwise unable to interact with the outside world.


## MOTIVATION

Training machine recognition of EEG patterns in response to different stimuli could have many benefits, including filtering out superfluous signals when diagnosing medical conditions, and improving the ability to interpret complex thoughts from people who have become paralyzed or are otherwise unable to communicate. This project draws from the work of Bobe et al [1] [2] who collected the original dataset, and were attempting to reconstruct the entire image observed by the subject.

**Figure 1. Example videos of the 5 categories. From left to right, abstract, waterfalls, faces, Rube Goldberg mechanisms, speed**

## DATASET

Our dataset consists of 34 EDF files from 17 participants' EEG data form observing video stimuli from 5 different categories: abstract forms, waterfalls, faces, Goldberg mechanisms, and speed. *(Fig. 1)* Supplied were also CSV files per participants containing 117 entries of:

1. Label - Actual category that participant was viewing (ID of video clip category)
2. Exemplar - ID of video clip source
3. Onset-start - Start time of the video clip referring to EEG signal timeline (in seconds)
4. Onset-end - End time of the video clip referring to EEG signal timeline (in seconds)

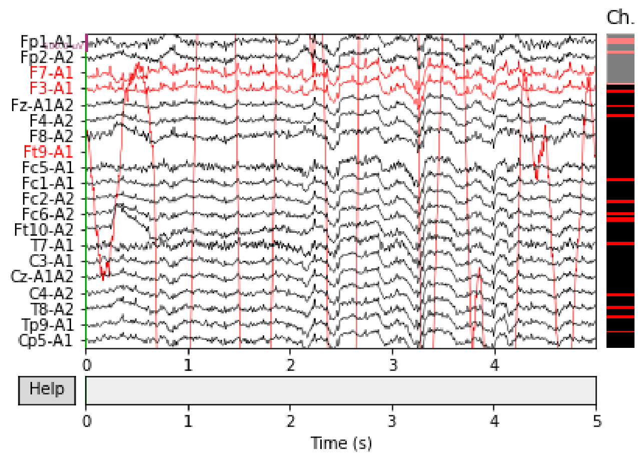


**Figure 2. Before interpolating. Bad channels F7-A1, F3-A1, Ft9-A1 have been highlighted red for easier viewing.**

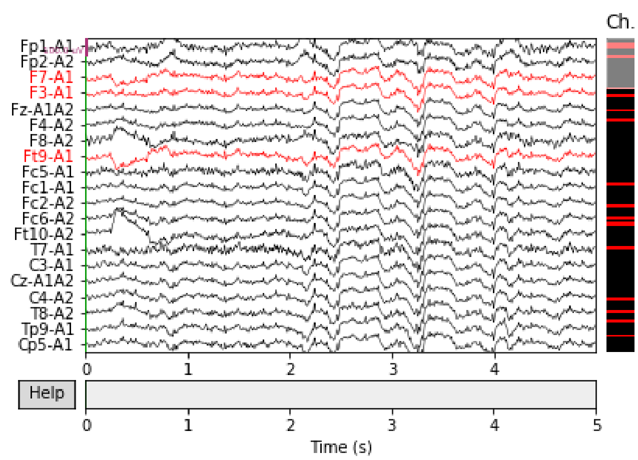


**Figure 3. After interpolating. The bad channels are now averaged out according to their neighbors.**

### Preprocessing

To provide uniform data to input to our CNN model, we made the videos the same length by truncating them to the shortest length found. While data is lost with this method, it allows us to more quickly understand if our model successfully learns useful classifications. Across the 17 subjects, there were many bad channels which were much noisier than their counterparts *(Fig. 2)*. To deal with this, we decided to interpolate the bad channels, which replaces them with the plain average of their neighbors. This allowed us to keep our data points a constant dimensionality.

After interpolating the bad channels *(Fig. 3)*, we split the EEG data into 6 second segments according to the time frame of the video stimulus observed, and computed for their Fast Fourier Transform (FFT) matrices. Then we selected frequencies up to 40 hertz (since typical brain activity occurs within that range). We also experimented with frequencies up to 128 hertz so that we had a uniform square matrix. Ultimately, these were the matrices passed to our model.

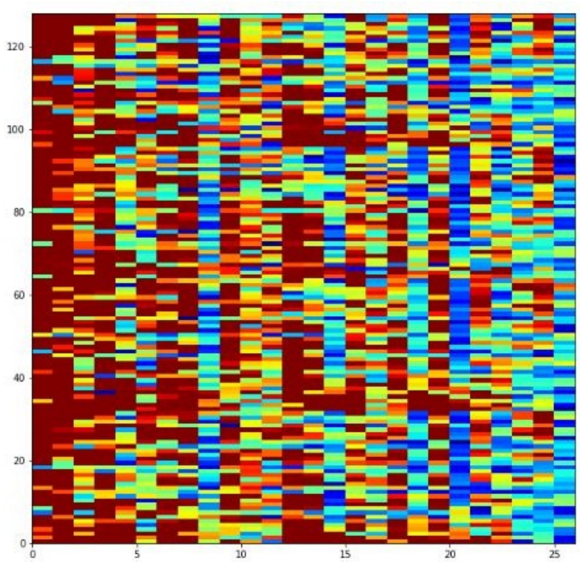

**Figure 4. 128 channel by 40 frequencies**

### MODELS AND ALGORITHMS

We created and tested several convolutional neural networks (CNN) to determine which one would produce the best results. Each attempted network informed the changes of future networks. All models used sparse categorical crossentropy (a version of categorical crossentropy that saves memory by allowing labels to be values rather than vectors) to calculate loss, and used a five neuron output layer with softmax activation to determine the label.

The goal behind using a CNN is to leverage its image recognition capabilities to identify the different waveforms produced by observing different visual stimuli. We have several convolution layers in the network in order to identify these waveforms across different scales and increase the likelihood

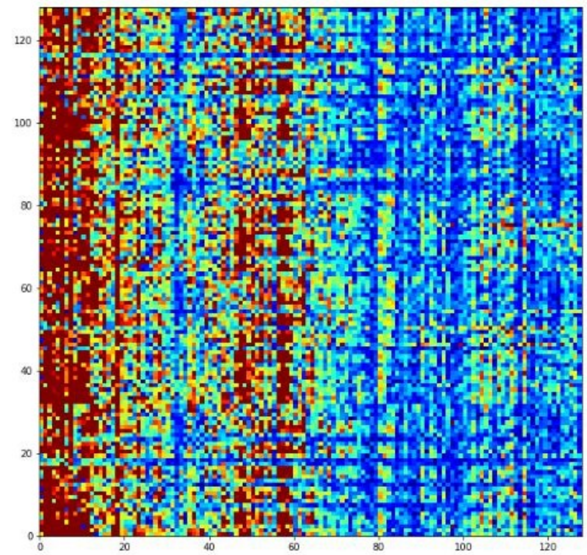

**Figure 5. 128 channel by 128 frequencies**

of finding useful features. Our output layer is based on softmax so that we can get a definitive result with five possible labels.

#### Model 1

Our first CNN had 3 convolution layers, each of which was followed by a max-pooling layer, this was then flattened and fed into one hidden densely connected layer before being set to the output layer. This model could not produce results with accuracy above 25%, and was prone to increasing loss on the test set as training continued.

#### Model 2

Next we trained a model that had 5 convolution layers, each of these layers had more neurons than those in the previous model, and were followed by max-pooling layers. We then flattened the output, and added a single dense layer with 8,192 nodes. This was then fed to our output layer. This model produced better results, achieving an accuracy of 36% with loss of 1.5.

#### Model 3

Our final model had the same beginning structure of convolution and max-pooling layers, however it deepened the feedforward section after the flatten layer. This CNN had 5 densely connected layers after the network was flattened, the first 3 of which were followed by dropout layers to prevent overfitting. This model achieved the best results, with accuracy of 45% *(Fig.7)* and loss of 1.3*(Fig.8)*.

### RESULT AND ANALYSIS

Our best convolutional neural network achieved a maximum accuracy of 45% on test data. while this model is not

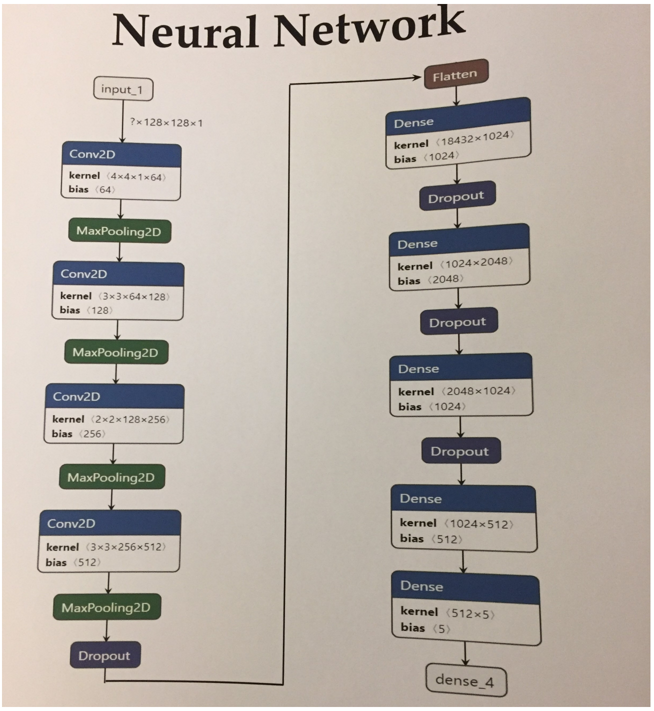


**Figure 6. The structure of Model 3.**

accurate enough to be reliable, the model is correct twice as often when identifying the type of video being watched based on the EEG data than random chance would suggest. In an attempt to bring more accuracy to the neural network, more dense layers were added after the first, which produced more accurate results. It is possible that with longer training or more samples, such a network could become more effective than the current model.

In addition it is worth noting that while it took longer to train, using the 128 by 128 version of the data yielded significantly better results, as compared to the 128 by 40 version. With larger dimensionality in our FFTs, our matrices held more data points and thus were able to hold more distinguishable features. This allowed our CNN model to more effectively identify these features and provide a better result than those of the 128 by 40 version.

## CHALLENGES

EEG data is prone to noise, therefore one of our primary challenges was to find a method of normalizing the data in such a way as to filter bad data, such as those produced by bad channels, and false stimuli caused by eye blinks and other factors. In addition, the data provided was for videos of varying length, but our neural network required input of fixed size. To resolve this, we truncated the data down to the size of the shortest video (about 6 seconds), because it was a simple yet effective solution to implement that would provide the CNN with longest uninterrupted EEG scan examples.

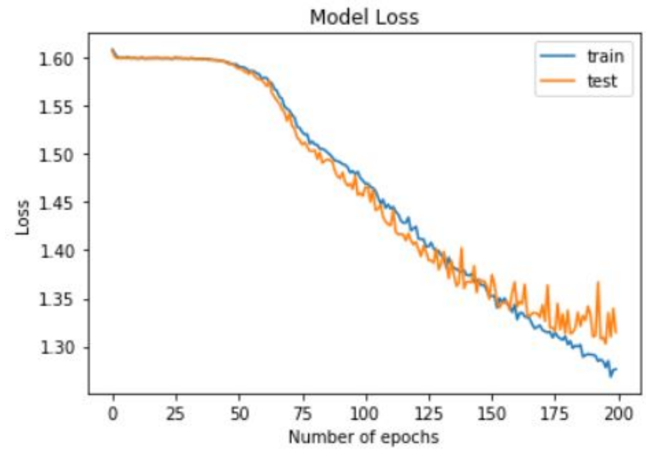


**Figure 7. Model 3 achieved 45% accuracy.**

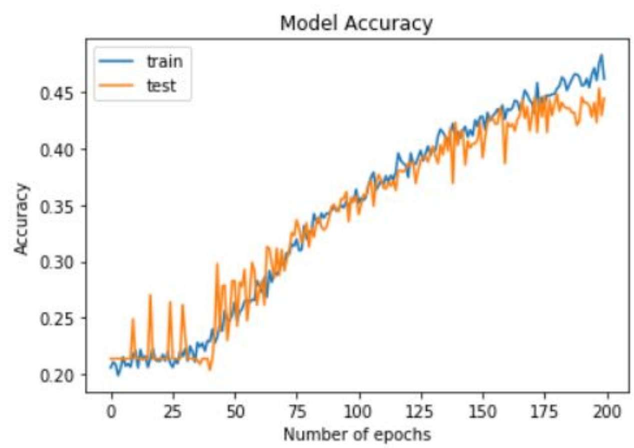


**Figure 8. Model 3 achieved loss of 1.3.**

Converting the EEG data from the .edf files to a format usable by the CNN proved challenging as few resources were provided with the initial dataset. Furthermore, the documentation for spectrograms, FFT, and other EEG concepts were extremely dense and often required lots of prior knowledge. Though their implementation syntax was not entirely difficult, we found it challenging to verify correctness.

Attempts to change the neural network to improve accuracy proved difficult as resource limitations on the systems that we were using prevented us from exploring all possible variations. Finally, due in part to initial optimism regarding how much we could accomplish in the time available, we were unable to implement different types of neural networks.

## CONTRIBUTION

As a group, we did not have any experience with neuroscience. As such, a fair amount of time was spent conducting individual research on EEG topics.

### Victor Ye

Victor worked on preprocessing the EEG data so that we had a cleaner and less noisy dataset. He also worked on developing, testing, and optimizing the CNN model to improve the model accuracy and loss. He collected everyone's recorded audio for the script and created the video. For documentation, he has contributed to the proposal, progress report, post, and video script.

### Jacob Wynd

Jacob assisted Victor to do the work of preprocessing the EEG data. He also helped with developing, testing and optimizing the CNN models, with a particular focus on Model 2. Jacob also attended office hours with Tunglin to consult for future endeavors. In addition he wrote the rough draft of the script for the video, many portions of both the progress report, and this paper.

### Tunglin Lee

Tunglin created the first versions of the CNN model to analyze the EEG data. After Victor and Jacob finished preprocessing the EEG data, Tunglin helped with building the CNN model. Tunglin also worked on testing and improving the accuracy of the CNN model. For documentation, he has assisted with proposal, progress report, post video script. He also created rough drafts of presentation slides for the video presentation.

## FUTURE WORK

As the EEG data is presented as a time series, it is possible that more accurate classification could be achieved using a recurrent neural network (RNN). In addition, it would likely be beneficial to train on more varied data, while there were several thousand individual segments of EEG data to analyze, there were only 17 subjects as part of the initial study. More data would likely allow the neural networks to more precisely distinguish the useful data from the noise. Finally, as you may be able to see from the volatile spikes in the model accuracy *(Fig.7)* and loss *(Fig.8)* charts, our data was not thoroughly cleansed. More time could be prioritized in cleaning the dataset. More specifically, our dataset was tainted by human blinks, which could cause false stimuli. In future endeavors, we could incorporate independent component analysis to unmix signals for a cleaner dataset. With cleaner data, results would be more definite and much more accurate.

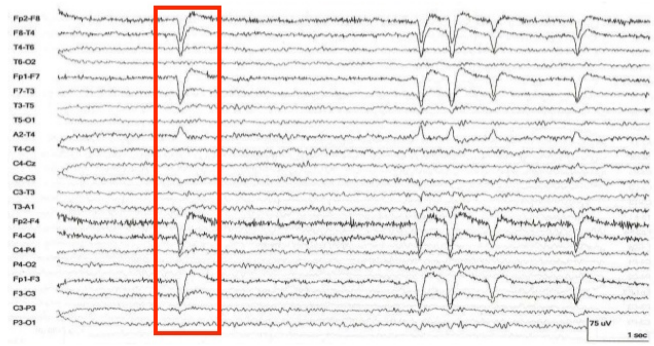

**Figure 9. An example of an eye blink in EEG data. [3]**


## ACKNOWLEDGEMENTS

We would like to thank our professor Dr. Narges Norouzi, and our TA Jonathan Scott for their support during this project.

## REFERENCES


[1] MIPT Contribution: General research guidance Dmitry Fastovets Maria Komarova Grigory Rashkov Andrey Alekseev Anatoly Bobe, Anatoly Bobe Unspecified Neurorobotics Lab. 2019a. EEG data for observing the video stimuli. (November 2019). https://data.mendeley.com/datasets/s2dxrv45fr/1

[2] Maria Komarova Grigory Rashkov Andrey Alekseev Anatoly Bobe, Dmitry Fastovets. 2019b. *Natural image reconstruction from brain waves: a novel visual BCI system with native feedback*. Neuroscience. bioRxiv. https://www.youtube.com/watch?v=9bZkp7q19f0.

[3] Stepwards. 2016. EEG Findings: Eye Blinks. (2016). https://www.stepwards.com/wp-content/uploads/2016/08/eeg-artifacts-75-638.jpg