Classification of Auditory and Visual Stimulus using EEG Signal

Justin Faber, Sydney Feldman, Steven Emmel, Elizabeth Mills
Astronomy 278, Machine Learning Final Presentation, Winter 2021
Instructors: Tuan Do and Bernie Boscoe

Abstract

In this report, we present our findings for binary classification of minimally processed, single trial human subject EEG data. We compare several supervised, offline ML algorithms, explore options for feature selection, and specifically address the question of whether single trial EEG time-series data can be reliably classified by information from one electrode alone. Our results show that both Random Forests and Support Vector Machines achieve better accuracy than Gaussian Naive Bayes, independent of feature selection or number of channels included in the model. Additionally, holistic feature selection proved more accurate than using raw time series data points as features. Lastly, two channels achieved higher accuracy than one alone. Interestingly, two channel accuracy success depended on using channels that were in different locations on the scalp, even though adjacent channels achieved high accuracy scores on their own. Our results suggest that single channel analysis can reliably distinguish between different cognitive states, with machine learning tools.

Part 1: Big Picture

The objective of this project is to classify brain activity (measured by EEG channels from 256 locations on the head) as a reaction to visual or auditory stimuli. An additional goal is to be able to do this classification using data from only a few specific channels. It is difficult and expensive to manufacture the head caps with 256 probes that we currently use, and inconvenient to use these caps (especially outside of lab environments). If we are able to do this classification with only a few channels, the 256 probe headset could be replaced with a smaller and less expensive version for classification experiments like the one that produced this data.

We will measure our performance using accuracy score, which should be a good indication of performance since our data is evenly split between visual and auditory stimulus data. Our goal is to achieve 80% accuracy with our classification. Previous studies have achieved accuracy between 70-80% accuracy classifying stimulus present vs stimulus absent scenarios using individual EEG channels (Stewart *et al*. 2014) which was considered successful. This is the standard we hope to match/surpass by using a limited number of EEG channels. Doing so would prove that limited channels are sufficient for stimulus classification.

This project will involve supervised machine learning, as we are choosing the features, and it will be offline.

Part 2: Get the Data

Our data come from a neuroscience attention study on human subjects that extracts brain voltage signals from 256 distinct scalp locations, sampled at 250 Hz (A. Lenartowicz et al., “Neurophysiological Signals of Ignoring and Attending Are Separable and Related to Performance during Sustained Intersensory Attention,” *Journal of Cognitive Neuroscience*, 26:9, (2014) 2055-2069). The raw data are available on Hoffman Cluster, and anyone interested in working more with these data are invited to reach out to Agatha Lenartowicz at the Center for Cognitive Neuroscience at Semel Institute of UCLA. Previously, these data were analyzed by averaging EEG signal across trials and subjects, and comparing grand averages in brain signal for cases where the subject was attending, ignoring, or passively experiencing either auditory or visual stimulus. Single trial classification and analysis is novel for these data and this neuroscience subfield.

As a first step towards applying machine learning binary classification to EEG signal, our analysis focused on classification of attending auditory processing versus attending visual processing. For example, we took only auditory trials from the data whereby the subject heard a sound AND had the task of auditory attending in order to respond whether the tone was a high pitch or a low pitch. Additionally, we only took visual trials from the data whereby the subject saw an object on the screen AND had the task of visual attending in order to respond whether the object was tilted right or left on the computer screen.

While attending visual and attending auditory brain signals have been easily classified by their grand averages, this is the first time (to our knowledge) that machine learning algorithms have been applied to single trial data. We focused on single trials where the subject correctly responded to the stimulus, such that we could be confident there was successful task engagement and focused attention during that time. We omitted trials where the subject missed the stimulus (offered no response) or incorrectly responded to the stimulus.

Part 3: Explore the Data

The data were minimally preprocessed by means of a .1 Hz high pass filter. Additionally, the continuous EEG signal was separated into specific epochs, whereby each epoch had exactly 2 seconds of continuous signal that was centered (time-locked) on the onset of a particular auditory or visual stimulus. Each time that a stimulus was observed connected to a particular brain signal time-series that had its middle datapoint corresponding to the onset time of the stimulus. Specifically, each epoch time series had exactly 1000 data points, whereby the 500th datapoint corresponds to the onset of the stimulus. This meant that one could zoom into a particular region of the epoch, and measure signal at a particular time AFTER the stimulus presentation. Since our goal was to classify visual versus auditory processing, it was helpful to have stimulus onset timing exactly the same in each epoch.

In total, there were roughly 800 trials each for auditory processing and visual processing, so we had a balanced input set. All trials were from the same subject, over the course of a two hour experiment period, comprising of multiple stimulus presentation blocks.

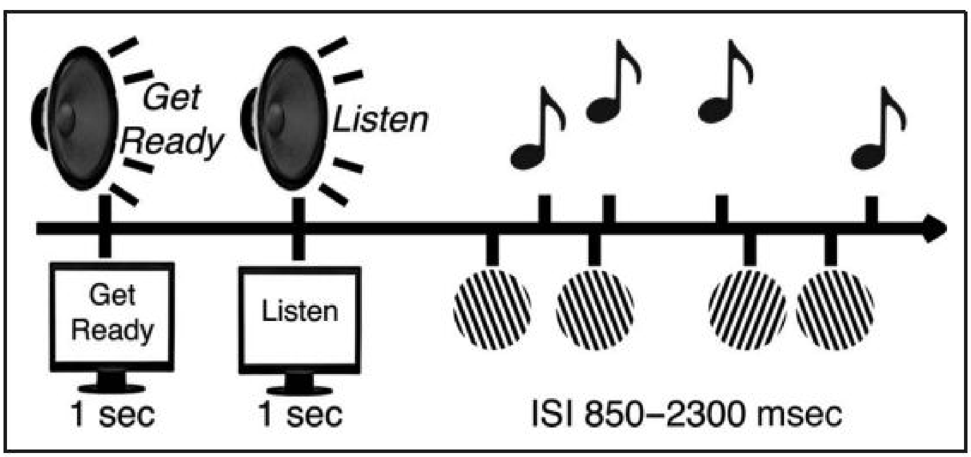

Figure 1: Schematic of experimental paradigm where subject sees and hears stimulus, and is tasked with only listening to auditory stimulus and responding to whether the tone is a high pitch or a low pitch sound. In a different block of trials, the subject was exposed to the same diversity of auditory and visual stimuli, but instead was tasked with only watching the visual stimulus and responding to whether the object on the screen was titled left or right (not shown.)

Part 4: Prepare the Data

We used 20% of the data as our test set, which was used for calculating all of the accuracy score in the histograms in the next sections. We also set aside 10% of the data as a validation set. We did not use this data for anything except for reporting the final accuracy scores in the final section of this report.

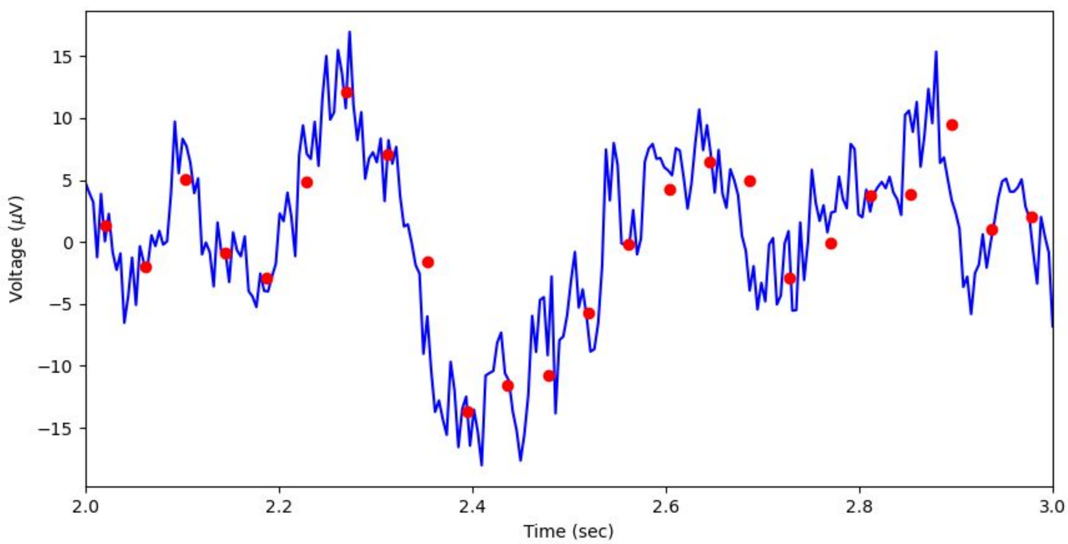

Figure 2: A single channel's voltage readouts (blue) and the average of every ten points (red)

Before performing any significant feature engineering, we averaged every ten data points over a length of time of one second (at 250 Hz), corresponding to 25 data points for each channel. This was done to reduce the number of features handled by the single channel models. Later on, more significant feature engineering will take place so as to increase the accuracy scores and involve additional channels.

Part 5: Select Model and Train

Three different models were applied to the single channel data: Gaussian Naïve Bayes, Random Forest, and SVM. Each model was fit to the training data using the default hyperparameters. The individual channels vary largely in the accuracy score obtained from fitting them. Because of this, we decided to fit each channel individually with all three models and check the resulting accuracy score on the test set. The histogram shown below summarizes our results.

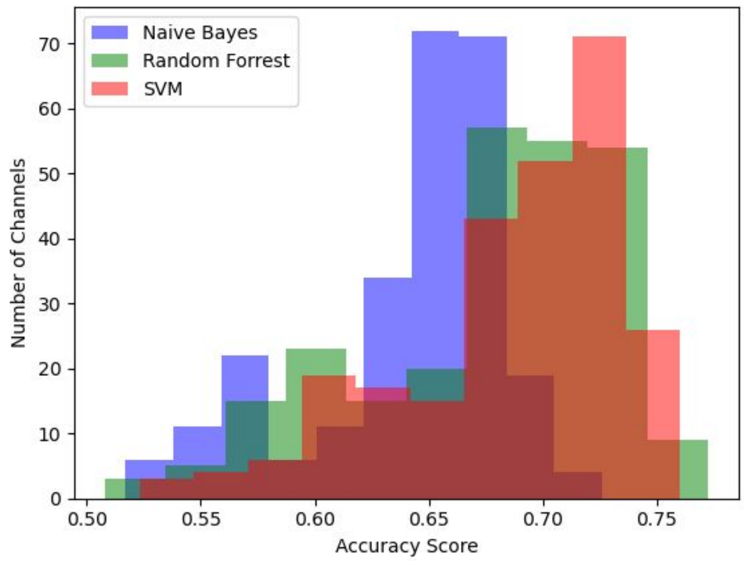

Figure 3: Histograms comparing three models across all of the channels. Classification was performed on single trials and using only one channel at a time.

The average scores for the Gaussian Naïve Bayes method was .64, for the Random Forest Classifier it was .68, and for the SVM it was .69. While the average of the SVM models was slightly higher than that for the Random Forest Classifier, the RFC channels with the highest accuracy scores were the most accurate overall. Changing the hyperparameters didn't significantly affect the distributions, although the individual channels were affected by hyperparameter tuning.

We chose to use Support Vector Machines in our primary model. This selection is based off of our comparison to other methods (see histograms). Further, SVMs have been widely used for classification of EEGs (Stewart *et al*. 2014) and SVMs have yielded the highest accuracy scores in EEG classification competitions (Blankertz et al., 2006).

Part 6: Fine Tune

We found that fine tuning the hyper parameters of the model only minimally increases our accuracy score. Instead, we saw the largest improvement when adjusting which features we were using and how we were extracting them. In the previous sections, we showed results from using values of the averaged windows of the signal as features. In this section, we use statistics of the full signal as features.

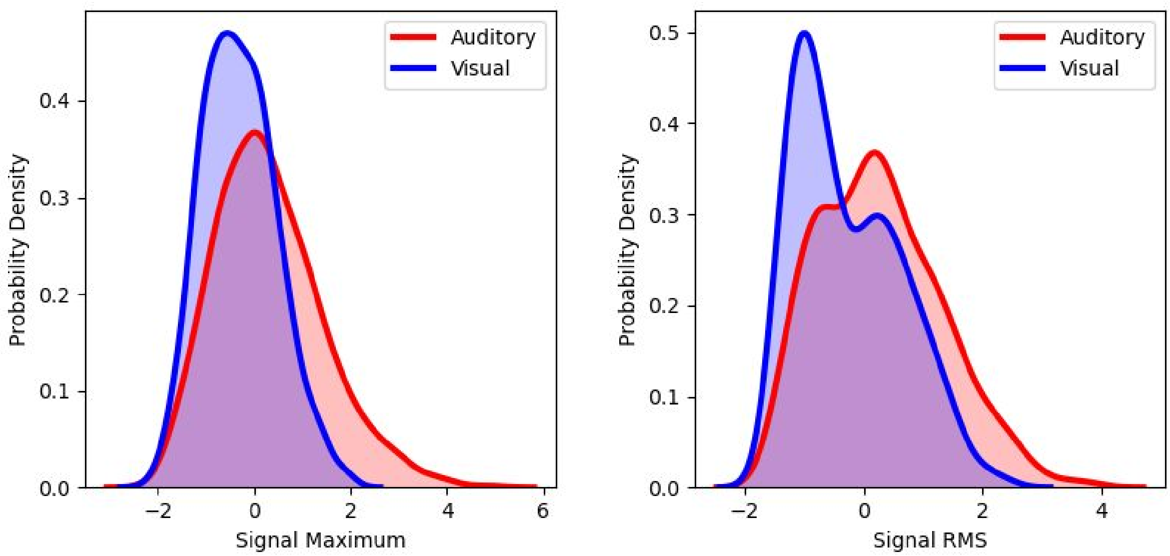

Figure 4: Histograms of the maximum value of the signal and the signal RMS for visual and auditory stimulus. These two features were helpful for improving classification.

Additionally, we used the frequequency domain to engineer more features. Several of the Fourier components showed significant differences between the two types of stimulus. Including all of the Fourier components did not improve our accuracy score, so we instead chose to include only a few at the dominant frequencies of the signal.

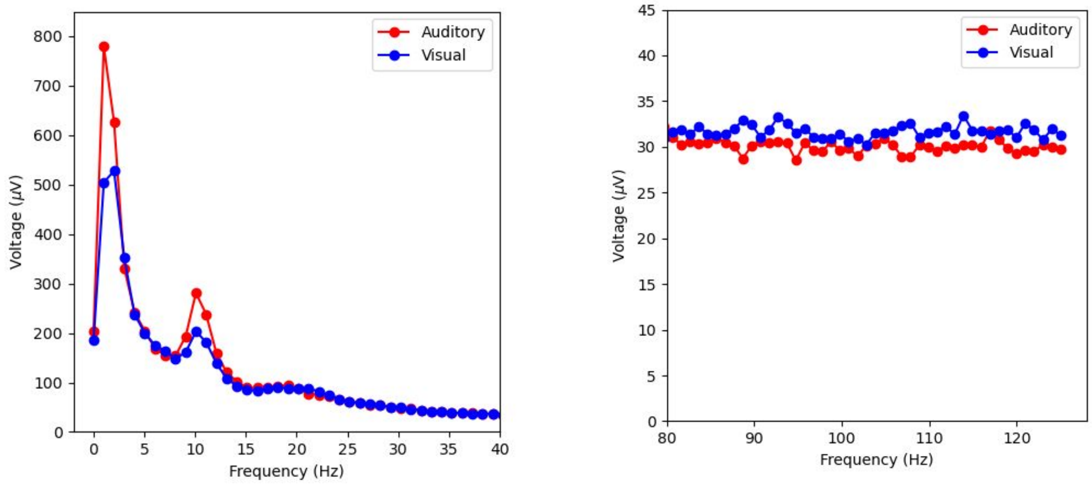

Figure 5: Fourier transform of the signal averaged across all 801 trials for each type of stimulus. We used 3 Fourier components from at the peaks as features. We also took the average of the high frequency components (80-125 Hz) and used that average as a feature.

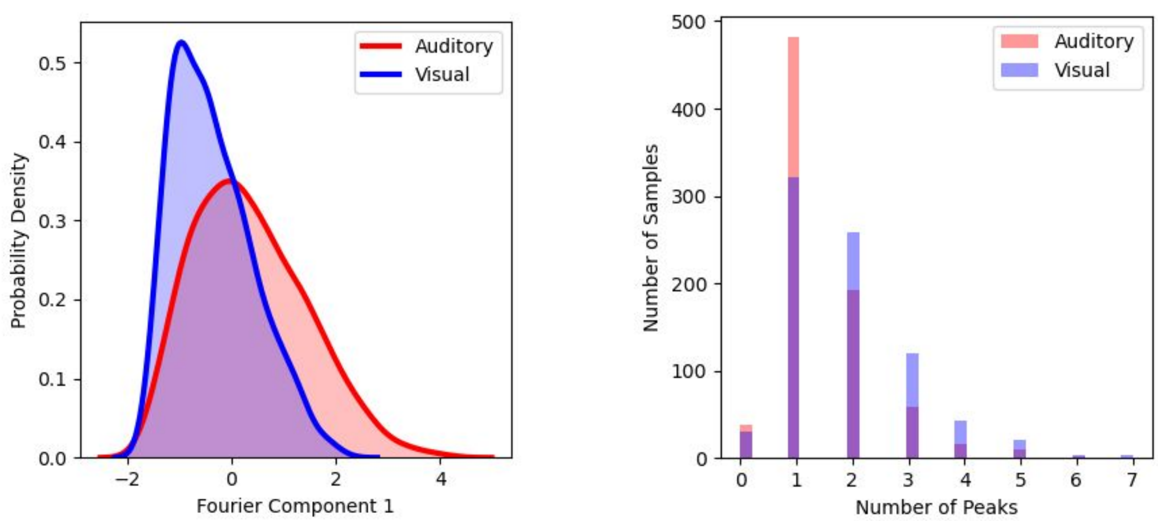

Figure 6: Histograms of the Fourier component at the first peak (~2 Hz) and the number of peaks in a smoothed signal in the time domain.

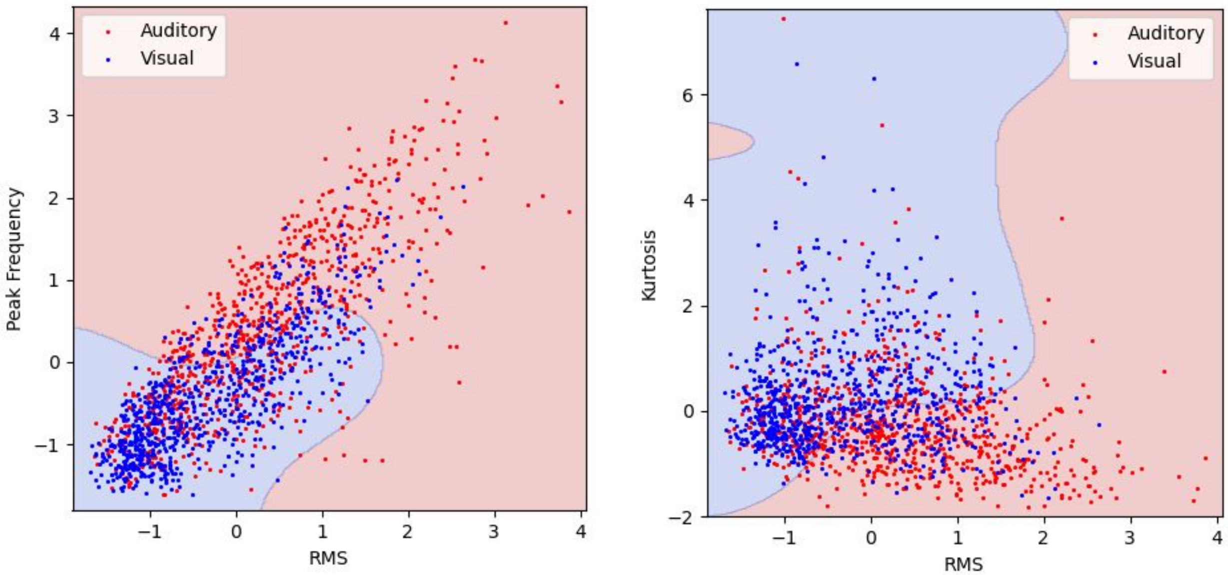

Figure 7: Decision boundaries for SVM classification, using only two features at a time. The RMS-Peak Frequency model yields an accuracy score of 0.68, while the RMS-Kurtosis model yields an accuracy score of 0.65.

The SVMs appear to be classifying the data as we would expect for two-feature models. We now combine all of the features discussed in this section to create an 8-feature model. These 8 features are:

- 3 key Fourier components (2 Hz, 3 Hz, and 10 Hz)
- The average of the high frequency components (80-125 Hz)
- The maximum of the signal in the time domain
- The RMS of the full signal
- The kurtosis of the full signal
- The number of peaks in the time domain of the signal

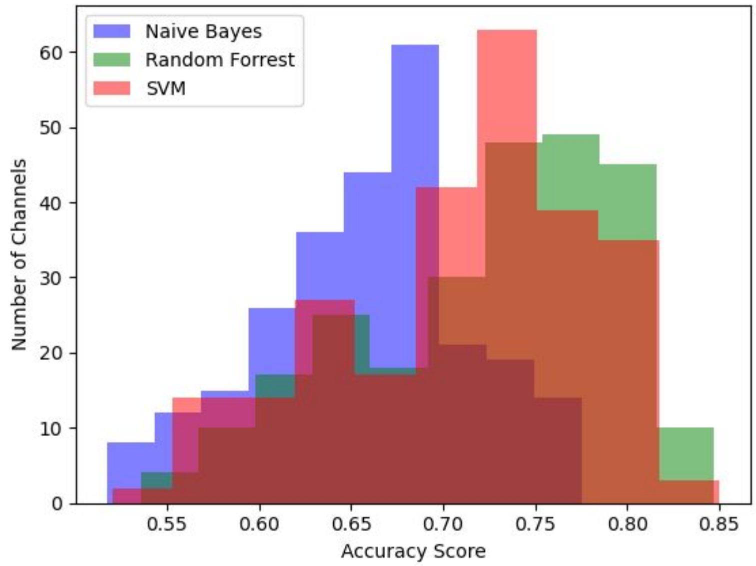

Figure 8: Histograms comparing the three models. Random Forests and SVMs gave similar accuracy scores, with a mean of about 0.75. Many of the channels gave an accuracy score above 0.8. This is an improvement from the previous method of using averaged windows of the signal as features.

Part 7: Present the Solution

Using the 8 feature model, we were able to achieve up to 84% accuracy using a single channel of single trial EEG time series data. The channels with the best individual performance are highlighted below in yellow.

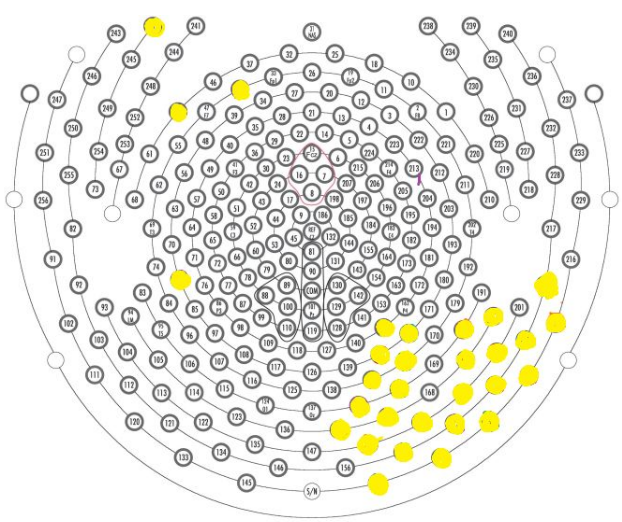

Figure 9: Diagram of EEG headset used in data collection, with

best performing individual channels highlighted in yellow

The location of these channels is somewhat unsurprising, as visual processing is known to occur at the back of the head while auditory processing is known to occur at the front and sides. We can hypothesize that channels physically close to each other are likely picking up similar information from the brain signals, and combining well-performing channels that are spatially separated may provide additional distinguishing information useful for classifying auditory and visual tasks. To test this, we used the model to fit data from 8 features, from 2 different channels (for a total of 16 features per auditory or visual trial). The combinations of channels were created by choosing the 5 best performing channels on the back of the head and the 3 well-performing channels on the front of the head and iterating over all possible pairs (15 total).

Using this method, we increased our accuracy score up to 90% on the validation set using the best performing pair of channels, which we are extremely satisfied with. It should be noted that trials of three channel combinations were performed with decreased accuracy, and combinations of two high performing pairs that were NOT spatially separated also resulted in decreased accuracy.

Another science question we had was whether or not EEG headsets can or should be pared down to be smaller and less expensive. One potential “solution” to the inconvenience of EEG headsets is the Muse Headband, shown below.

Figure 10: The Muse EEG headband. This is only one example of a potential alternative to EEG headsets.

To analyze the performance of visual and auditory classification tasks within the confines of a small headband (with limited channels), we selected 6 channels on either side of the head (without considering their individual performance in the original model. The selections were made to mimic the positions of a headband such as the one in Figure 2, and these channels are highlighted in yellow below.

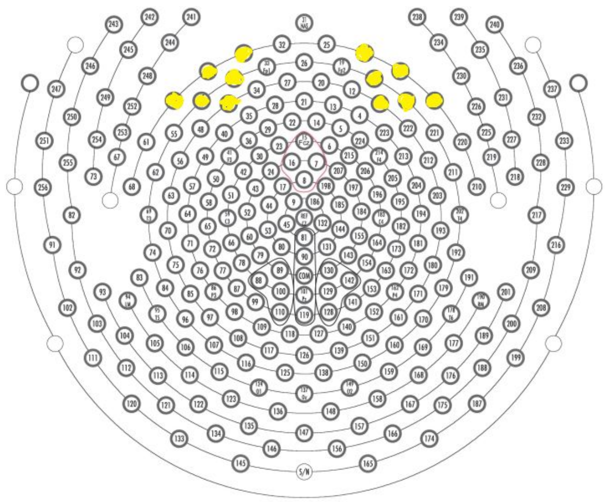

Figure 11: Channels selected to mimic nodes on an EEG headband

Iterating over pairs of left forehead and right forehead channels resulted in very well performing models. We attained between 81-89% accuracy using these specific channel combinations, which demonstrates that EEG headbands with limited channels are suitable for these specific classification tasks, and indicates that they may be a useful replacement for the 256 channel headset for other, similar tasks.