# Classifying Syllables in Imagined Speech using EEG Data

Stanford University

Barak Oshri, Manu Chopra, Nishith Khandwala

## Problem

Imagined speech refers to the process in which a subject imagines speaking a given word without moving any muscle or sound. The ability to understand imagined speech will fundamentally change the way we interact with our devices. We'd like to classify the syllables "ba", "ku" ,"im" and "si" from imagined speech EEG signals. These syllables were selected since they contain no semantic meaning so that classification would be performed on the imagined speech instead of the semantic contribution to imagined speech production. Our model is successfully able to classify syllable pairs from the EEG data with over 90 percent accuracy.

## Features

We implemented two feature extractors.

**1. Mean Feature Extractor**

For every channel, we average the data for each syllable from all the trials. The signal for each channel is divided into 8 parts. Each part is averaged and hence, every channel corresponds to a set of 8 values. The graph below shows a sample wave from a channel which is divided into 8 parts and then averaged. This allows us to discount any unexpected increase/decrease in the data.

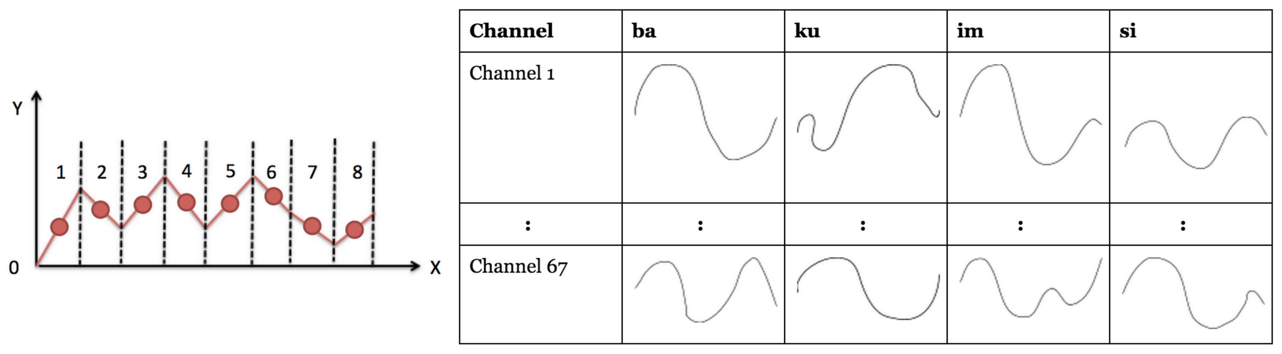


**Figure 3.** A graph showing the division and averaging of a signal

Let $ch_{(i,j)}$ represent the value of the $j^{th}$ part of the $i^{th}$ channel. That is for the $n^{th}$ channel, the first part can be written as $ch_{(n,1)}$. At the end of this operation, our feature set looks like this:

$$[ch_{(1,1)}, ch_{(1,2)}, ch_{(1,3)} \ldots ch_{(1,8)}, ch_{(67,1)}, ch_{(67,2)} \ldots ch_{(67,8)}]$$
$$[y \in [1,2,3,4]]$$

Here, we use the notation y = 1 to donate "ba", y = 2 to donate "ku" and so on.

**2. Discrete Wavelet Transform.**

We also extracted the wavelet coefficients by decomposing the EEG signal using a Discrete Wavelet Transform (DWT). DWT has proven useful in characterizing the signals of EEG data because it uses non-stationary time series analysis and leads to good time-frequency localization by using longer time windows at low frequencies and vice versa. Since the transform leads to an excessively large coefficient space, we performed Principal Component Analysis independently on each of the approximation matrix, first level horizontal and vertical images of the transform to produce a smaller four dimensional space. Each point was the coefficients of a channel wave with number of dimensions equal to the length of the transformed signal. The projection then represents the space of coefficients that best characterize the transformed signal.

## Data Collection

We created our own data set by making use of Professor Takako Fujioka's EEG lab at Center for Computer Research in Music and Acoustics (CCRMA). In our experiment, the subject imagined speaking the four syllables 'ba / ku' and 'im / si' based on pre decided audio cues while their electrical brain wave activity was being recorded by EEG. The audio cue corresponded to either a high beep or low beep. The beeps were randomized and each beep type corresponded to a particular syllable. A timeline for a single trial for the syllable pair "ba, ku" is shown below:

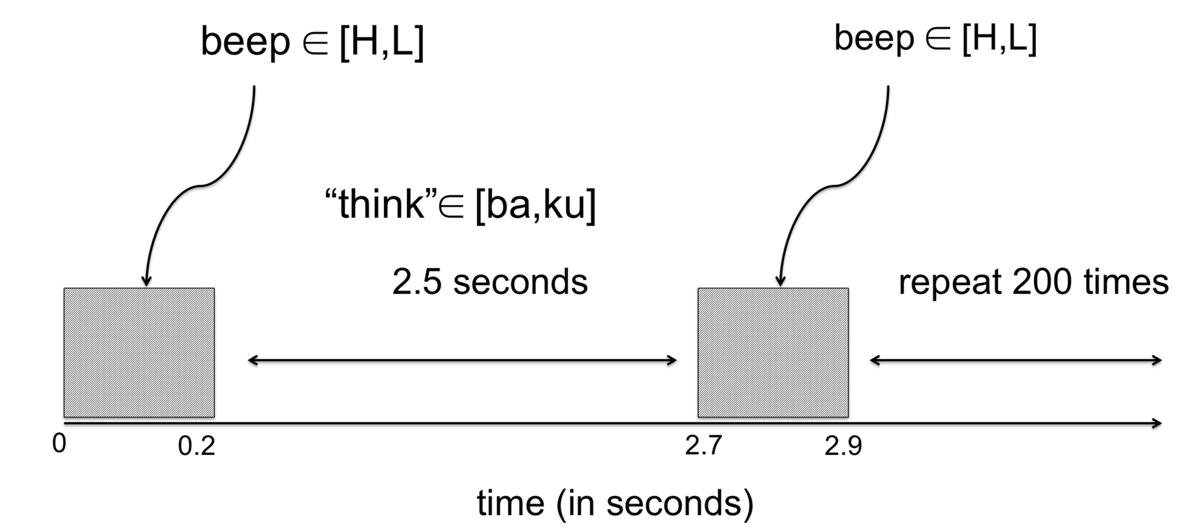


**Figure 1.** A timeline of the experiment

## Results

We used several models to classify our data including k-nearest neighbors, Naïve Bayes, Support Vector Machines, Discrete Wavelet Transform and Neural Networks. After some extensive experiments, we realized that our Neural Networks model with a mean feature extractor [highlighted below] gave us the best result.

| Feature | Model | Results |
|---|---|---|
| Mean Feature Extractor | k-nearest neighbors | {'ba' = 0.287 'ku' = 0.493 'im' = 0.500 'si' = 0.238} {'ba' = 0.625 'ku' = 0.547} {'im' = 0.589 'si' = 0.627} |
| Mean Feature Extractor | Neural Networks | {'ba' = 0.975 'ku' = 0.835} {'im' = 0.96 'si' = 0.93} |
| Mean Feature Extractor | Naïve Bayes | binary class = {0.748, 0.641} |
| Discrete Wavelet Transform | k-nearest neighbors | multiclass = {'ba' = 0.325 'ku' = 0.5068 'im' = 0.5337 'si' = 0.325} |

Results comparing 'Ba' with 'Im' have similar scores, which eliminates the fear that the model is characterizing response to pitch. The use of Artificial Neural Networks in classification of imagined speech is relatively undocumented in the sphere of BCI. Given the promising results of neural networks in this experiment, we feel that there is an uncalled for confidence in the use of feature methods in the research community and that more attention needs to be paid to nonlinear analysis of thoughts in EEG data.

## Denoising and Data Cleaning

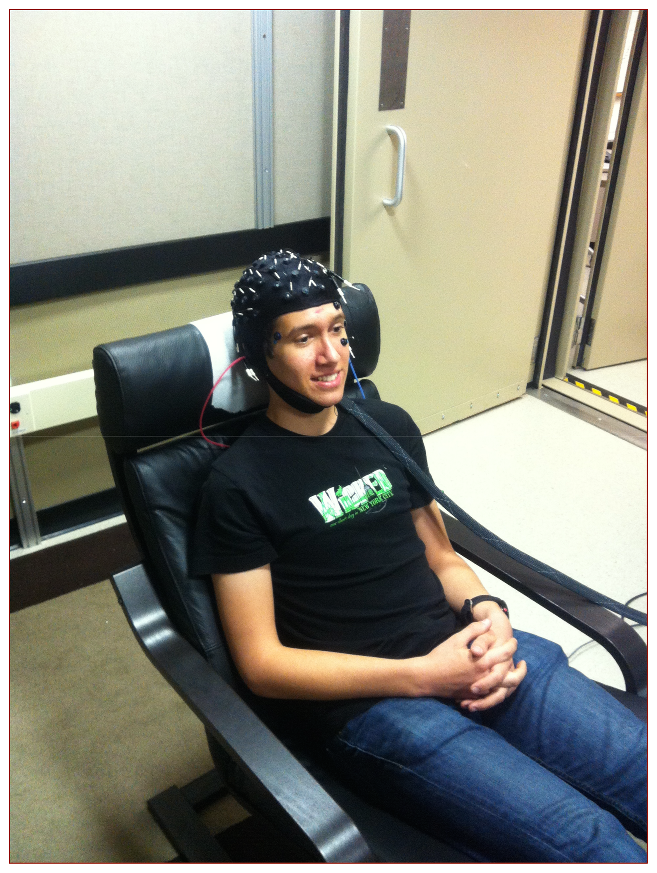

**Figure 2.** A subject during the experiment at CCRMA.

The obtained EEG data contained some presence of artifacts (i.e., changes in EEG amplitudes that don't correspond to brainwave activity but eye movements or muscle movements instead). These artifacts contribute to signal noise and hence, the EEG data was preprocessed to remove these artifacts. The electromyographic (EMG) artifacts (i.e., muscle artifacts) are removed using Neurosky and a Matlab toolbox and the signals from electrodes closest to the ear /neck are discarded to improve the signal to noise ratio. We also attempted manual channel selection.

## Results (cont…)

This is perhaps most relevant in light of the fact that there is little to no understanding of the neural mechanisms behind thought formation including imagined speech, which makes neural networks the best direction forward until we can more aptly characterize the features (such as action or event-related potentials) that characterize syllables and words. The confusion matrices for the neural network classification model of "ba/ ku" and "im/ si" are shown below.

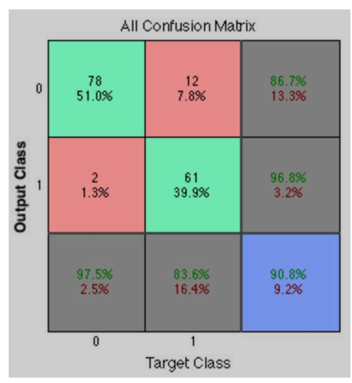

**Figure 4.** "ba/ku" NN model

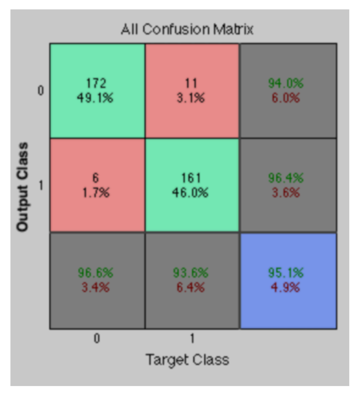

**Figure 5.** "im/si" NN model

## Future Work

Our results prove that it is possible to successfully classify certain syllables from imagined speech EEG data. In the future, we would like to classify more syllables, combining them to make words and finally, sentences. A breakthrough in classifying sentences in imagined speech would revolutionize our digital interaction. We would also like to test our models on more subjects and generalize our system to multiple subjects to ensure a more robust Brain Computer Interface (BCI) system.