Classifying Syllables in Imagined Speech using EEG Data

Barak Oshri (boshri), Nishith Khandwala (nishith), Manu Chopra (mchopra)

1. INTRODUCTION

An increasingly important advancement waiting to happen in brain-computing technologies is interfacing with speech in the mind. As fluency and dependence on technology raises the demands for faster, cleaner, and more productive interfaces, the pathways such an accomplishment would pave in the scope of our communicative abilities would lead to a revolution in our natural and digital interaction with the world.

Advancements in imagined speech technologies is suffering from evading a holistic approach to understanding how we can read and interpret language in the mind. Current research in this field is producing a limited and acute set of tools to systemize far too specific instances of a task, such as classifying "yes" and "no". Not only are these studies archetypal cases of over fitting to the details of the experiment with which they were run under, but they encourage an outlook for studying imagined speech that is myopic and fails to represent the complexity of the task, for any solution to this problem must be scalable to large groups of people and preferably multiple languages.

The shortcomings of imagined speech research, however, are not wholly unexpected. Given our severely limited neurological understanding of what conscious thoughts are and how imagined speech is carried by this mechanism, it is near impossible to make assumptions and predictions about the structure of EEG data that attempts to measure them. Also, whereas other imagined actions such as spacial movements are lateralized to one hemisphere over the other, imagined speech has equal levels of significant activity in both hemispheres.

This paper does not proclaim to overcome this challenge. What it does do is attempt to engage with and reason through an experiment that breaks through the mold of treating imagined speech studies as application oriented exercises and highlights what models and approaches may yield the best insights in the future of this promising field.

1.1 Approach

An approach to understanding imagined speech using EEG needs to be foundational and scalable. It needs to model a system from EEG data just as a natural language is understood in common use. For this reason we pursued the paradigm that phonetic qualities of words affect how their associated syntactic rudiments are stored in the brain, where it follows that language as *thought* in the mind is being produced with some correlation to how the language was learnt and understood on first basis. Given this hypothesis, the role of machine learning in imagined speech classification is to tailor models that can evaluate and predict syntactic features.

The focus of our research, then, is to discuss how multiple syllables can be classified between each other with the ambition in future research that accurate classifications of syllables will allow predictions of arbitrary strings of them (or words). It is no less than a beauty that a finite number of syllables give rise to the entire breadth of a language.

We will thus explore two machine learning approaches to do this task. The first includes using KNN and Naive Bayes to examine how we can build a model out of elementary features of the data, and the second involves the use of artificial neural networks to seek nonlinear patterns for prediction.

2. EXPERIMENT

We created our own data set by making use of Takako Fujioka's EEG lab at the Center for Computer Research in Music and Acoustics (CCRMA). We used a 10-20 system EEG with 64 channels covering the entirety of the subjects head. Three additional nodes tracked eye and upper-facial movements to assist removing blinking and face movement artifacts from the data. The EEG sampled at a rate of 500HZ.

A subject was asked to imagine speaking the pair of syllables 'ba' / 'ku' and 'im' / 'si' alternating between trials. A low and a high pitch tone were predecided before the experiment to correspond to the the pair of syllables, the lower tone corresponding to 'ba' or 'im' and the upper tone corresponding to 'ku' or 'si'. In one round of readings, 200 trials of a syllable pair, 100 of each syllable, were mixed randomly and presented to the subject as the tones. After a short break, the experiment was repeated with the other pair of syllables. We then performed the first pair again in another round and the second pair in the next. In total, 200 readings each of 'ba', 'ku', 'im', and 'si' were collected.

The length of the queuing sound lasted for 0.2 seconds, enough to perceive the pitch but not too long that response to the tone interferes with thinking. The subject was given 2.5 seconds and asked to utter the correct syllable once, after which he was asked to rest his mind until the next beep is heard.

A time line for a single trial for the syllable pair ('ba', 'ku') is shown below:

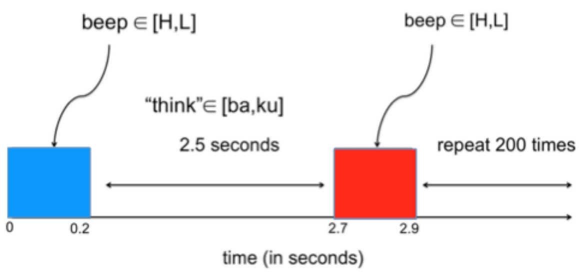


Figure 1. A timeline of the experiment

The obtained EEG data contained the presence of artifacts from blinking and facial muscle contractions. The EEG data was preprocessed to remove these artifacts in EEGLAB. The electromyographic artifacts were removed and the signals from electrodes closest to the ear / neck discarded. Data processing identified that 134 of the 800 trials recorded were too noisy, mostly in the pair of trials 'ba' / 'ku'. The results for 'im' / 'si' were accordingly given greater significance.

After processing out the artifacts, we trimmed out each trial to the length of the relevant data. Given that it takes a regular subject approximately 0.5 seconds to imagine speaking the syllable, we trebled that time frame to include the decision process, imagination, and decline of the thought signal, so we reduced each trial to 0.2 to 1.7 of the original signal starting at 0.0.

Note that by deciding to include the part of the signal where the subject is simultaneously reacting to the sound and deciding which syllable correctly corresponds to its pitch, we have inadvertently created a separating criterion that allows a machine learning model to classify the trials based on the brain response to the pitch. This is a nontrivial complication that is not accounted for in most syllable studies, but for which we have considered by making multiclass classifications that include classifications between syllables that were imagined with the same pitch.

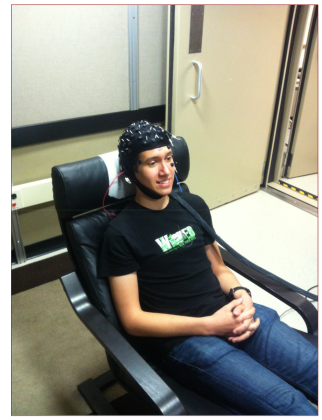


Figure 2. A subject during the experiment at CCRMA.

3. FEATURES

3.1 Mean Feature Extractor

Perhaps the most natural extractor is to approximate each wave with a series of points that are averages of a segment at the point's position. This is principally useful because the feature space considering every recording of the wave for all 64 channels would amount to a feature dimension that is too large and would overfit the data. We also use this feature extractor because it is neurologically relevant and can be used to identify Event-Related Potentials (ERPs) and other characteristic motifs by identifying points and times that are especially informative of the class.

The mean feature extractor also provides an intuitive way of producing canonical syllable waves with higher signal-to-noise ratio by averaging all trials of a syllable together. This will be useful in making measurements of deviation of the individual trials from a representative. The figure below shows a wave from a channel which is divided into 8 parts and averaged with all trials of the class.

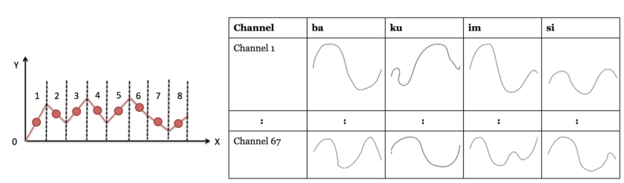


Figure 3. A graph showing the division and averaging of a signal

So if $ch_{(\mathrm{i,j})}$ represents the value of point j in channel i, then the feature vector is

$$[ch_{(1,1)}, ..., ch_{(1,8)}, ch_{(2,1)}, ..., ch_{(64,8)}][y \in [1, 2, 3, 4]]$$

which has $64 \times 8 = 512$ dimensions. We use the notation y = 1 to donate ba, y = 2 to donate ku and so on.

3.2 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.

4. CLASSIFICATION

4.1 K Nearest Neighbors

Using K Nearest Neighbors (KNN) allows us to make inferences about how distinct the syllables are from each other. It is useful as a measure of how successful our feature space is in drawing out salient linear features of the classes. If KNN classifies accurately, than the syllables partition the feature space into Voronoi cells with boundaries that have neurological significance for why they distinguish syllables. If KNN is less predictive then we know that nonlinear patterns are needed for more effect.

We run a modified KNN where each trial is assigned to the averaged wave form of the four syllables it is closest to. This is so that we can measure how noisy an individual trial is with respect to an ideal of its class. For a given trial w_i, we assign to w_i

$$\text{argmin}_{s \in [\text{ba}_{\text{avg}}, \text{ku}_{\text{avg}}, \text{im}_{\text{avg}}, \text{si}_{\text{avg}}]} D(\phi(w_i), \phi(s))$$

where $D(\phi(w)_i, s)$ is the Euclidean distance metric between the feature extraction ϕ of w_i and a class-average syllable.

	ba	ku	im	si
Accuracy	0.2875	0.4932	0.5000	0.2384

Table 1: Multiclass KNN

The results for KNN are promising but not ideal. 'ku' and 'im' are being classified at rates significantly better than random (25%), but the prediction rates for 'ba' and 'si' are insignificant. Since one syllable in the two trials performs well and the other doesn't, it leads suspect that biases between conjoining the two experiments are leading to this symmetrical result.

4.2 Naive Bayes

Naive Bayes can overcome some of the biases in KNN by generating a model for each class independently. Essentially, the Naive Bayes assumption in this case means that a wave is characterized by its amplitudes and that syllables are matched to sample waves that have the right registers.

The parameters for each class are given as $l = 1, 2, 3, 4$, so

$$\phi_{k|y=l} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n_i} 1\{x_j^{(i)} = k \wedge y^{(i)} = l\} + 1}{\sum_{i=1}^{m} 1\{y^{(i)} = l\} n_i + |V|}$$

where V is the size of the feature extracted signal and

$$\phi_{y=l} = \frac{\sum_{i=1}^{m} 1 y^{(i)} = l}{m}$$

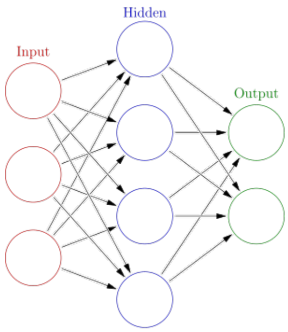
After the parameters and probabilities are trained, we make estimates for test cases using max-likelihood statistics. We then use 5-fold cross validation to get test predictions.

	ba	ku	im	si
Accuracy	0.6966	0.4420	0.6666	0.6855

Table 2: Multiclass Naive Bayes

4.3 Neural Networks

In artificial neural networks, the burden of making assumptions about the structure of the data is transferred to the training of hidden layers that solve "subproblems" of the given input. This allows us to find nonlinear relationships of features along and between channels that can offer greater predictive power than the earlier models discussed.



The number of inputs is the size of the feature vector used. The hidden layers are linear weighted subproblems over a sigmoid activation function. If w_i is a signal, for each hidden unit h_j,

$$h_j = \sigma(\mathrm{v}_j \cdot \phi(w_i))$$

with v_j a learned weight and logistic activation function

$$\sigma(z) = (1 + e^{-z})^{-1}$$

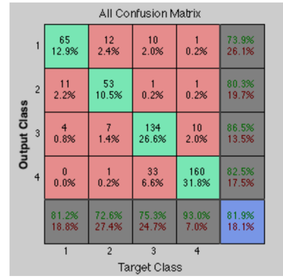
We trained a feedforward network using a scaled conjugate gradient backpropagation to update the weights and measured performance using cross entropy. We evaluated the performance on different size hidden layers and found that a hidden layer of size 6 maximizes the training and test performance. 85% of the trials were used for training and 15% of the trials were held out for testing.

	ba	ku	im	si
Accuracy	0.8125	0.7260	0.7528	0.9302

Table 3: Multiclass Neural Networks

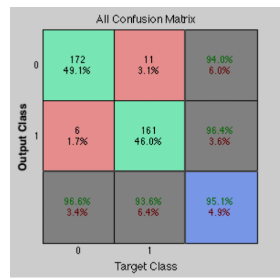
5. ANALYSIS

An 81.9% overall classification rate for neural networks is stunning and far exceeds results of similar studies in the field, which generally have prediction accuracies between 60% and 70% and often times on binary classifiers. Below is a confusion matrix for the four class classification of syllables 'ba', 'ku', 'im', and 'si' in that order:



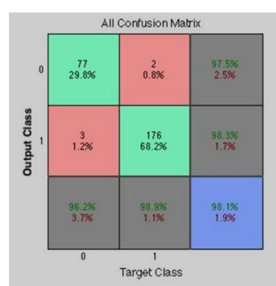
A classification of 81.9% on the entire data set with 87.8% training accuracy and 66.7% testing accuracy is suffers from mild overfitting but not to excess. The success of neural networks suggest that the most predictive features occur in patterned combinations that are not immediately identifiable with single feature expansions.

The neural network with hidden layer of size 6 also worked remarkably well in binary classification of the syllable pair 'im' and 'si'.

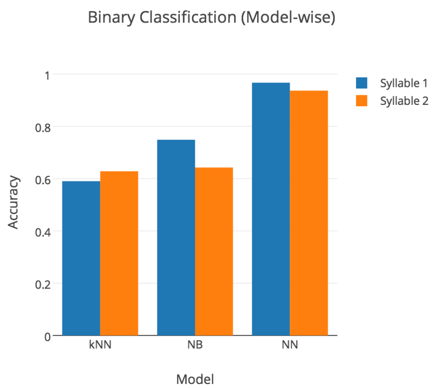


This is not surprising given the success of the multiclass classifier. Note, however, that the network scaled to the multiclass case because the number of correct predictions in the binary and multiclass classifier are not too distinct.

The neural network also classifies the pair of syllables 'ba' and 'im' that were recorded in response to the same pitch with equal accuracy, which is especially interesting because now the nature of the experimental data has changed (gathered from independent rounds of measurement).



Unlike the Naive Bayes and KNN classifiers, the neural networks performed consistently well on the variety of cases tested on, whereas Naive Bayes and KNN did not scale as well when they classified four syllables instead of two. This is noticeable comparing the charts of binary and multiple classifications of the models studied.



6. FUTURE WORK

Neural networks should be further studied for their potential to uncover neural patterns that we do not know from existing neurological sources. We note informally that a hidden layer of size 6 was the optimal size for all cases studied with neural networks in this experiment, and that this warrants special attention for it reflects on a level of patterns implicit in the data and in the functional sites of the brain it originated from.

Given the success of multiple classification of four syllables, we further propose that neural networks be tested against more a wider range of syllables and that resting state EEG data is used for control experiments. We think that the accuracy reported in this paper is beyond that expected for given past results in the field, and that further experiment should be conducted on a larger data set including multiple subjects.

An ideal approach to imagined speech and general BCI applications can encompass new functionality, and many of the approaches traditionally used, especially support vector machines, do not have the representational mechanics to pursue general imagined speech understanding and expansive BCI needs. Neural networks are versatile tools for modelling networks of sizes that grow, alter and expand, and we believe that they show promising hopes for elucidating brain data.

7. ACKNOWLEDGEMENTS

Special thanks to CS229 TA Dave Deriso for his incredible support and assistance in this project. We could not have collected our own data set without Takako Fujioka's patience for teaching us how to use and run the EEG in her lab. And thanks to Andrew Ng without whom we wouldn't be introduced to this wonderful material.

8. REFERENCES

1. Bhagavatula V.,Advanced Signal Processing and Machine Learning Approaches for EEG Analysis, Jun 2010. `http://www.researchgate.net/publication/235054600_Advanced_Signal_Processing_and_Machine_Learning_Approaches_for_EEG_Analysis`

2. Mingjun Zhong et al.,Classifying EEG for Brain Computer Interfaces using Gaussian Process, 2008 `http://people.rennes.inria.fr/Anatole.Lecuyer/prl08_zhong.pdf`

3. Bekir Karlk and Sengl Bayrak Hayta, Comparison Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG, Aug 2014 `http://iwbbio.ugr.es/2014/papers/IWBBIO_2014_paper_1.pdf`

4. Michael DZmura et al.,Toward EEG Sensing of Imagined Speech ,2009 `http://link.springer.com/chapter/10.1007/978-3-642-02574-7_5#page-2`