A Novel Approach to Classify and Detect Thoughts using Electroencephalography

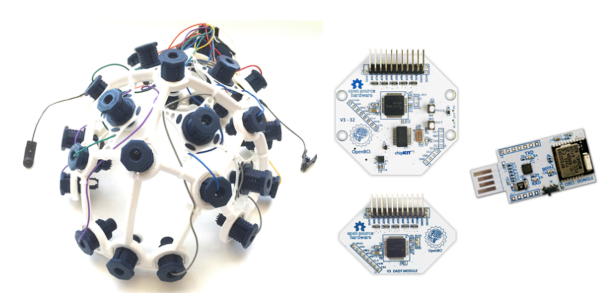
Viraat Das

Senior at Morgantown High School, Morgantown WV

1. Introduction

Electroencephalography (EEG) headsets, being highly portable, present a convenient, noninvasive method of recording the brain's electrical activity. This project seeks to develop a method for the classification of thoughts, memories, and stimuli from EEG data, for potential use in brain-computer interface (BCI), and the treatment of Aphasia and disabilities that affect verbal communication. The two headsets used were the Emotiv EEG headset and OpenBCI containing 14 electrodes and 16 electrodes, manifesting data as 14 dimensional and 16 dimensional time series. For analysis of multidimensional data machine learning techniques were adopted to train models in the classification The model can classify events given new EEG data, and stochastic neighbor embedding allows visualization of the data in 2-D. Technology enabling nonverbal communication is exciting not only for its futurist appeal, but its potential to help millions suffering from Aphasia and other speech impairments. While methods such as MRI capture data with extreme resolution, EEG headsets would provide a noninvasive, inexpensive and accessible means of translating brain activity. **This research aims to contribute to the future development of an EEG based BCI for therapeutic and rehabilitative purposes.**

Emotiv Headset



OpenBCI

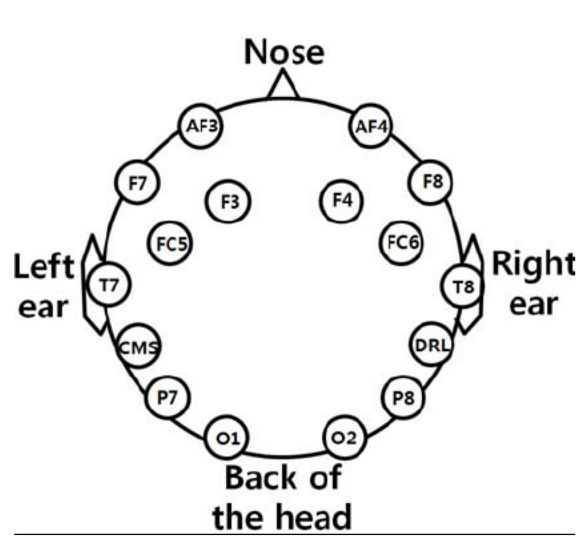
2. Research Goals

The research consisted of three basic goals

- If "thoughts" in the form of EEG Data can, in fact, be classified
- A systematic way to predict the classification of new thoughts based on old data for a particular individual
- Make sure the results were scalable with a high data and feature count size
- Create a model that scales with each individuals data

3.1 Implementation

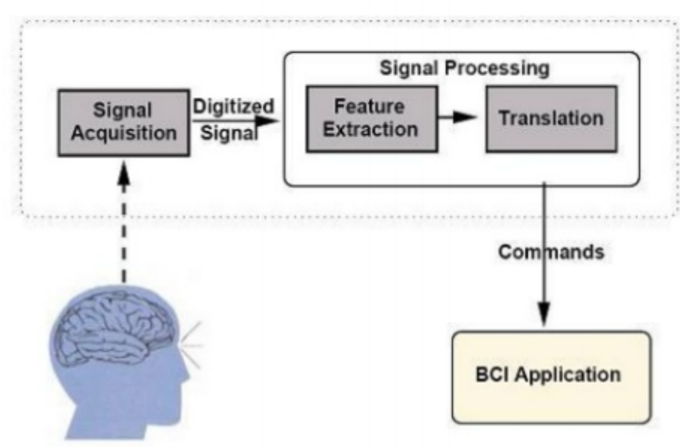
Figuring out the best set of algorithms to implement required having a visual sense of the data. Every electrode mapped out its own time series.

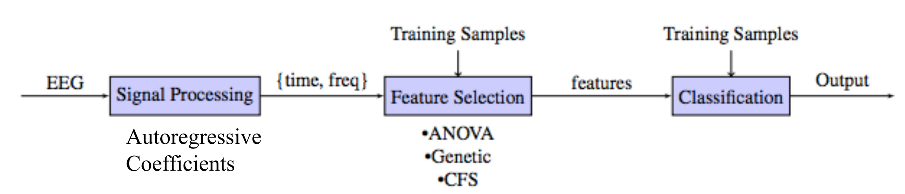


General electrode configuration of an EEG headset

Visualizing 14-D and 16-D data required

3.1 Implementation *(continued)*





3.3 Frequency Processing

Inputs from Headset

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

Yule-Walker Equations

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_p \end{bmatrix}$$

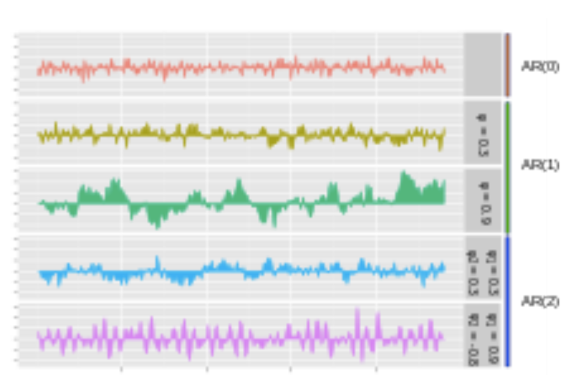
Based on desired data configurations

$$\rho(\tau) = \sum_{k=1}^{p} \varphi_k \rho(k-\tau)$$

Recursive calculation of the AR Parameters

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

These values can be fed into classifiers

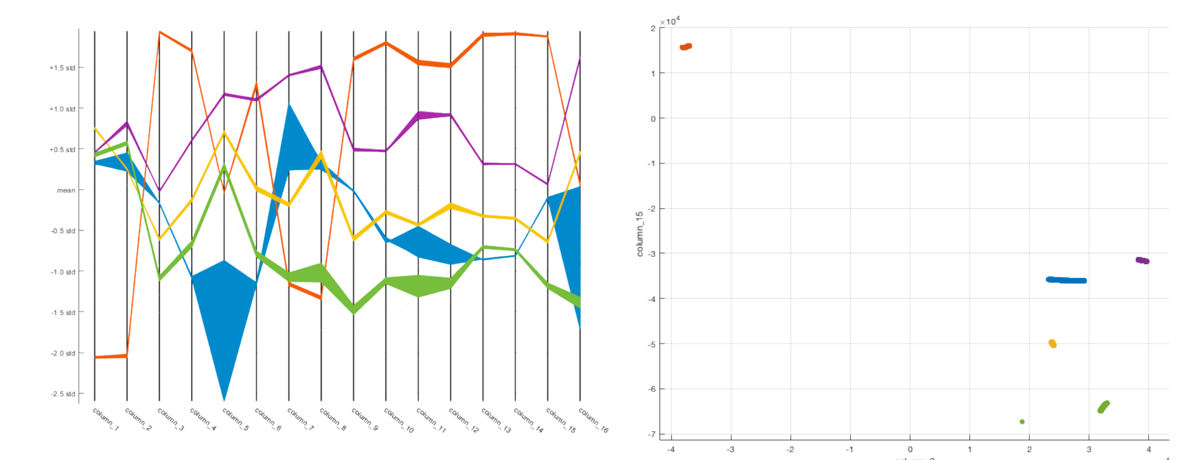


4.1 Music (Auditory)

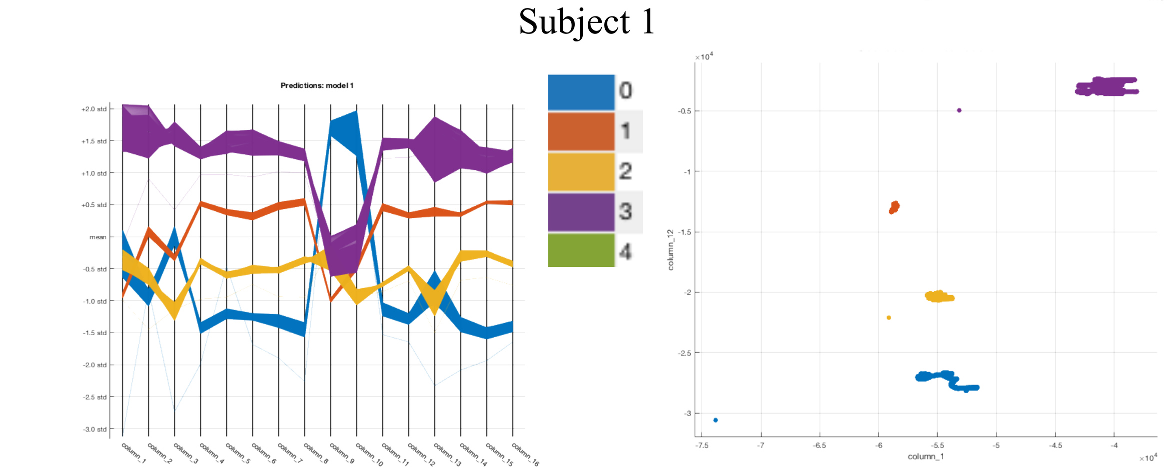
Five songs were analyzed

- Overture by Pyotr Ilyich Tchaikovsky (0)
- Asturias by Isaac Albeniz (1)
- Op Op by Unknown (2)
- The Sorceress Apprentice by Paul Dukas (3)
- Suite española by Isaac Albeniz (4)

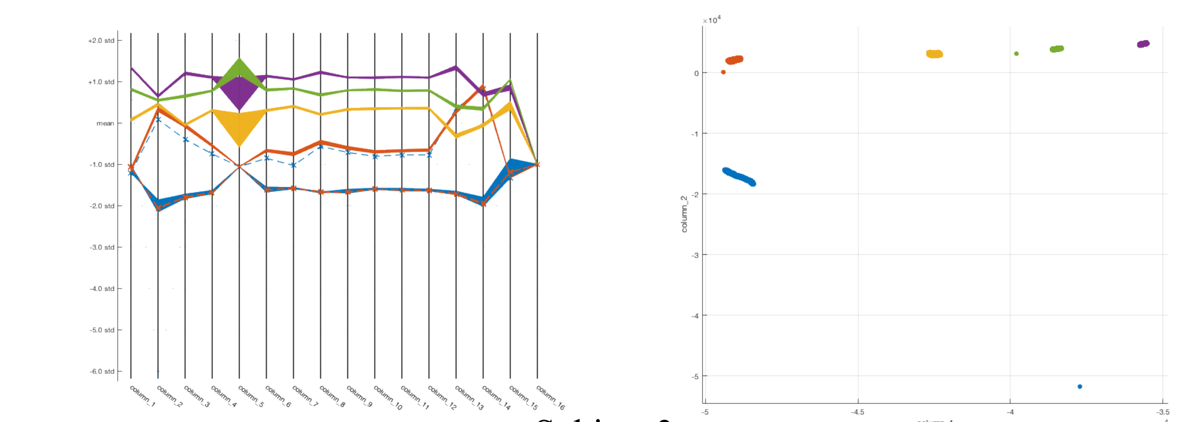
4.1 Music *(continued)*



Subject 1



Subject 2



Subject 3

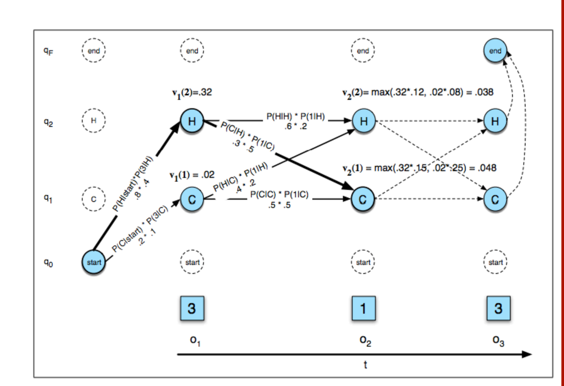
Utilized PCA to explain up to **95% of the variance**

	Cubic KNN Model Accuracy	Prediction accuracy with trial 1	Prediction accuracy with trial 2
Subject 1	100%	96.4%	95.3%
Subject 2	100%	94.3%	94.6%
Subject 3	99.7%	94.1%	93.7%

Hidden Markov Model

Vitebri Algorithm Implementation

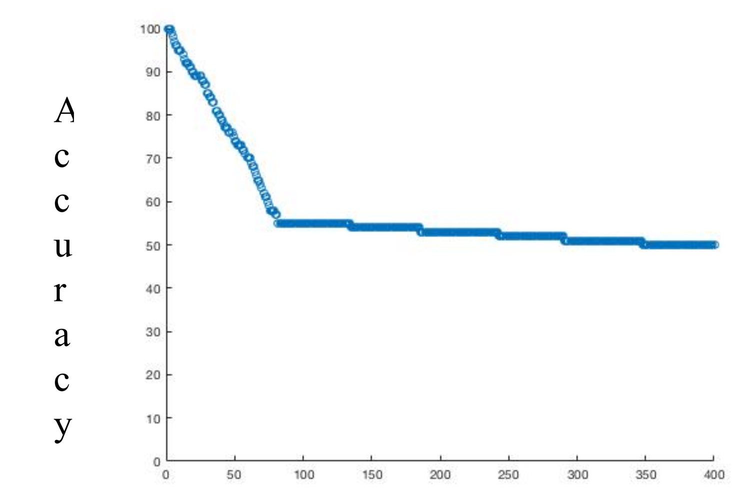
```
function VITERBI(O, S, Π, Y, A, B) : X
    for each state i ∈ {1, 2, ..., K} do
        T1[i,1] ← π_i · B_{iy1}
        T2[i,1] ← 0
    end for
    for each observation i = 2, 3, ..., T do
        for each state j ∈ {1, 2, ..., K} do
            T1[j, i] ← max_k (T1[k, i − 1] · A_kj · B_{jy_i})
            T2[j, i] ← arg max_k (T1[k, i − 1] · A_kj · B_{jy_i})
        end for
    end for
    z_T ← arg max_k (T1[k, T])
    x_T ← s_{z_T}
    for i ← T,T-1,...,2 do
        z_{i-1} ← T2[z_i,i]
        x_{i-1} ← s_{z_{i-1}}
    end for
    return X
end function
```



$$v_t(j) = \max_{i=1}^{N} v_{t-1}(i)\, a_{ij}\, b_j(o_t)$$

4.2 Music *(continued)*

Accuracy of the Model with Markov Models



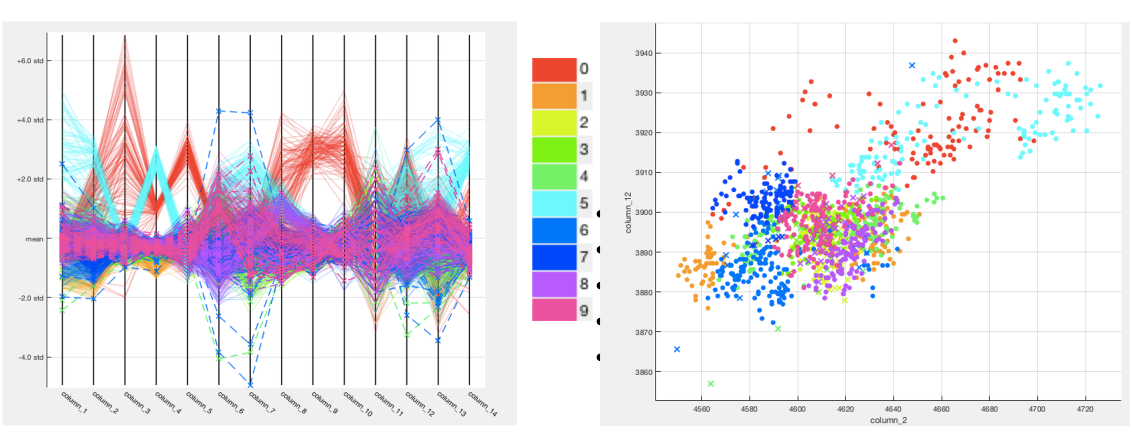
Vitebi Sets

Shows that music is processed relatively the same way among humans.

4.2 Pizza Toppings (Visual)

Collected data while subject viewed these toppings on the following pizza:

- Anchovies (0)
- Bacon (1)
- Canadian Bacon (2)
- Grilled Chicken (3)
- Ground Beef (4)
- Ham (5)
- Italian Sausage (6)
- Pepperoni (7)
- Prosciutto (8)
- Spicy Italian Sausage (9)



Both PCA and T-SNE would be valid. Due to low cost, PCA was chosen with an explained variance of 95%

Quadratic SVM returned an accuracy rate of **94.1%**

$$G(x_1, x_2) = (1 + x_1' x_2)^p.$$

Model tested with second take showed an accuracy rate of **89.2%**

5. Future Works and Conclusion

- Depending on data criteria, these models can be tuned for an individual and used for thought classification in that context.
- Ugly Duckling Theorem can be easily put in check with the methods presented.
- Possible usage: communicative device for patients suffering from Broca's Aphasia, neuroscience research, and expanded development of endogenous and exogenous BCI.

6. References

- Heyden, M. (2016). Classification of EEG data using machine learning techniques. LUP Student Papers. Retrieved December 4, 2017.
- Hoefle, S., Engel, A., Basilio, R., Alluri, V., Toiviainen, P., Cagy, M., & Moll, J. (2018). Identifying musical pieces from fMRI data using encoding and decoding models. Scientific Reports, 8(1). doi:10.1038/s41598-018-20732-3