

Analysis of speech related EEG signals using emotiv epoc+ headset, fast fourier transform, principal component analysis, and K-nearest neighbor methods

Abstract

Electroencephalography (EEG) is a method used for measuring electrical impulses, generated in the cerebral cortex, by using electrodes located in different positions on the scalp. In this work, EEG signals related to imagined speech of letters a and b are acquired by the Emotiv EPOC+ headset - a relatively low cost electroencephalogram with 14 electrodes (channels) - from three subjects. A multi-step process is utilized to preprocess the raw EEG data. The data are first preprocessed using the digital signal processing (DSP) techniques of filtering to limit the effects of noise and artifacts, on the one hand and to determine the relevant channels to be used. The preprocessed signals are further analyzed with Fast Fourier Transform (FFT), principal component analysis (PCA) for feature extraction, and k-nearest neighbor (KNN) for classification. Additionally, a calibration test is developed to make the recognition process subject specific. The results for each step are presented and performance is evaluated.

Keywords: EEG, imagined speech, calibration, FFT, PCA, k-NN, emotiv epoc+

Volume 5 Issue 3 - 2019

Louiza Sellami, Theresa Neubig

Department of Electrical and Computer Engineering, United States Naval Academy, USA

Correspondence: Louiza Sellami, Department of Electrical and Computer Engineering, United States Naval Academy, Annapolis, MD, USA, Email sellami@usna.edu

Received: May 19, 2019 | **Published:** May 23, 2019

Introduction

Communication is essential for human beings to thrive on earth. Without communication, there is no practical possibility for growth, since communication, whether verbal or nonverbal, is the basis for the transfer of knowledge. Therefore, any technology which improves the ease of communication also improves societal growth. This research focused on developing a system of improved efficiency that is capable of reading ones thoughts and reproducing those thoughts either visually or audibly. The Emotiv Epoc+, a commercially available electroencephalography (EEG) headset, is utilized to gather a broader range of data and to determine which electrode locations are most effective for collecting brain waves specific to the brain processing thoughts. Previous research has shown that using EEG to read imagined speech is certainly possible. However, no calibration test has been developed to allow the same EEG device to accurately read the imagined speech of different people. This research developed and tested two different calibration methods: additional mini-training sets for each person, and a group of normalized training sets. There are numerous applications where communication via imagined speech (communication via thoughts) would be significantly useful. The medical community could use this imagined speech recognition to communicate with a stroke victim who is fully coherent but unable to produce speech. Similarly, the military is interested in developing a method of enabling soldiers to communicate with each other via thoughts while in combat.¹ While previous research has shown that "mind-reading" using EEG is absolutely possible, it has also shown that widespread use of it is impractical due to the differences in how people think. Namely, when one person thinks of the letter 'a' a specific set of neurons fire in her brain, while when another person thinks of the same letter, a different set of neurons fire, as described.² While research is currently being conducted to map out the specific sets of neurons firing during different brain activities, it is far from being completed. The two calibration tests that are explored during this research are on-the-go training sets and normalized training sets.

Previous work and background

Up until the 21st century, neuroimaging had been primarily used for medical purposes, such as diagnosing neurodegenerative diseases, including encephalopathies and Alzheimer's, diagnosing neurological diseases like epilepsy, and locating brain tumors.³ In 2008, the Defense Advanced Research Projects Agency (DARPA) invested in research to develop a synthetic telepathy device that would allow soldiers to silently talk to each other.¹ Recent technology has enabled companies such as Emotiv Systems, NeuroSky, and InteraXon, to design headsets that use EEG to control prosthetic limbs and video game characters.⁴ These headsets are commercially available to the public and, while they have the intriguing applications previously mentioned, are also advertised as re-search tools.

The concept of reading ones thoughts, often referred to as imagined speech, is not a novel concept. Extensive research has been done in this field, with the most success resulting from the use of an invasive technology called electrocorticography (ECoG). ECoG is a direct cortical reading of electrical activity in the brain via an invasive installation of subdural electrodes.⁵ One such ECoG experiment was led by Stephanie Martin and a team of six other researchers.⁶ While ECoG is often preferred over EEG due to its superior spatial resolution and temporal resolution, EEG is preferred due to its noninvasive installation. A noninvasive installation is of the utmost importance since the final product would ultimately have to be both portable and capable of being used upon multiple people. While ECoG typically is more successful than EEG at recording imagined speech, EEG research has had its own success. In 2010, Brigham and Kumar processed EEG data from 7 human subjects at the University of California, Irvine, as each subject imagined the syllables /ba/ and /ku/ without speaking or performing overt actions while wearing a 128 Channel Sensor Net.⁷ The effect of artifacts and noise were reduced through a multi-step process, with the primary step being the filtering of the EEG signals outside of the range 4-25Hz. This filtering process limited the effects

of artifacts, since most surface electromyography (sEMG) frequencies (artifacts resulting from facial muscle movements such as blinking) are typically above 25Hz, and the effects of the 60Hz noise line. In addition, the Hurst exponent can estimate of the predictability of a time signal, with a 0 indicating a completely unpredictable signal and a 1 designating a completely predictable signal⁸ was used to limit the effect of artifacts by discarding any EEG components that have Hurst values outside of the range 0.70-0.76. Finally, the EEG signals were divided into signal and noise subspaces through Principal Component Analysis (PCA) and sent through a Wiener filter to further reduce the noise effect. The processed EEG data was then characterized with a Univariate Autoregressive (AR) model shown in Equation (1),

$$x[n] = \sum_{k=1}^p a_k x[n-k] + e[n] \quad (1)$$

where $x[n]$ represented the EEG signal at time n , a_k represented the coefficients of the AR model of order p (the results showed that an order of 6 resulted in the highest accuracies), and $e[n]$ represented white noise. Once the AR coefficients were computed, the imagined syllables were classified using a 3-Nearest Neighbors classifier based on the Euclidean distance between the AR model coefficients in the training and testing set, and the process was repeated 100 times per trial (with 85 trials total) to obtain an average classification rate of 61percent overall.⁷

In 2012, the team of Esfahani & Sundarajan⁹ utilized EEG (they used an early version of the Emotiv 14channel neuroheadset) to classify five simple imagined shapes (cubes, spheres, cylinders, pyramids, and cones) to explore the possibility of using Brain-Computer Interfaces (BCI) as an interface with CAD systems.⁹ The team used independent component analysis (ICA) and the Hilbert-Huang Transform (HHT) to determine the marginal spectra of different frequency bands; a Mann-Whitney U-test was then used to rank the 14 EEG channels by relevance to form a vector to train a linear discriminant classifier. While this research produced an average accuracy of 44.6 percent which is higher than a blind guess, the high concentration needed to visualize a shape was difficult for the human subject to sustain repeatedly. Due to its high concentration requirement, many researchers reverted back to easily repeatable speech imagery rather than visual imagery.

A German student research group, led by Wester, used an EEG skull cap to recognize normal speech, whispered speech, silent speech, mumbled speech, and imagined speech. Short Time Fourier coefficients, delta coefficients, delta mean coefficients, and P300 evoked potentials were all used in the classification process. Originally, the results of this research were very promising with an average accuracy rate of four to five times the blind accuracy rate and was therefore thought to have high accuracy;¹⁰ however, the accuracy of the tests was compromised by the data collection procedure data was collected for the same word multiple times in a row, which caused temporal correlation on the EEG signals.¹¹ Another student research team from the University of Connecticut used the Emotiv Epoc+ to gather EEG data for imagined speech in order to establish a baseline to compare different modeling techniques for EEG signal processing.¹² Ultimately, other research projects obtained marginal success when only two words or syllables were tested (random chance has a 50 percent probability of accurately picking which word a person imagined, while one of the most successful studies so far has had an accuracy of 82.3 percent, using Compumedics Neuroscans SynAmps 2 EEG system).¹³ Most recently, in 2017, a team led by Nguyen, Karavas, and Artemiadis used a BrainProducts ActiCHamp

EEG System to infer imagined speech with a novel normalization method based on covariance matrix descriptors, more specifically, using Riemannian manifold features.¹⁴ The results of this research had extremely high accuracy, reaching 70 percent for a three-word test and 95 percent for a two word test. The high accuracy rates achieved demonstrated the potential for success associated with developing novel algorithms to process EEG data.

Data acquisition and processing

The process for using EEG data to predict which letter a human test subject has imagined is as follows:

- I. Data Collection
- II. Preprocessing
- III. Feature Extraction
- IV. Speech Classification

First, the raw EEG data was gathered for an unknown imagined letter using the Emotiv Epoc+. Then the data was preprocessed with a band-pass filter and normalization calculations. This was followed by feature extraction through principal component analysis and Fourier transform. Finally, speech classification was performed using a k nearest neighbors test (k -NN). A visual representation of this process is shown in Figure 1.

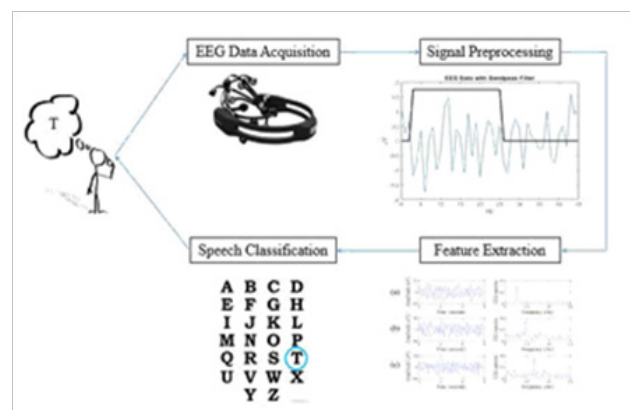


Figure 1 Experimental block diagram.

Data collection

In order to gather the baseline training set, data were collected from three human subjects, for imagining letters "a" and "b", five times for a total of 15EEG signals per letter. The test subject sat down in a chair in a room with minimal distractions. The Emotiv Epoc+, shown in Figure 2, was then positioned on the test subject's head. The test subject was asked to empty his or her mind. The recording started, and at five seconds the test subject was asked to focus on a letter, known to the testers, for 15more seconds. In order to test the classification algorithm and the calibration test, the test subject was asked to focus on one of two letters and the data was then processed to determine which letter the subject imagined

Preprocessing

The effect of artifacts and noise were reduced through a multi-step process, with the primary step being the filtering of the EEG signals outside of the range 5-25Hz using a band-pass filter. This filtering process limited the effects of artifacts, since most surface

electromyography (sEMG) frequencies (artifacts resulting from facial muscle movements such as blinking) are typically above 25Hz, and the effects of the 60 Hz noise line, as shown in Figure 3. In addition, the data were normalized from -1 to 1 in order to make the peaks more prominent.



Figure 2 Emotiv Epoc+.³

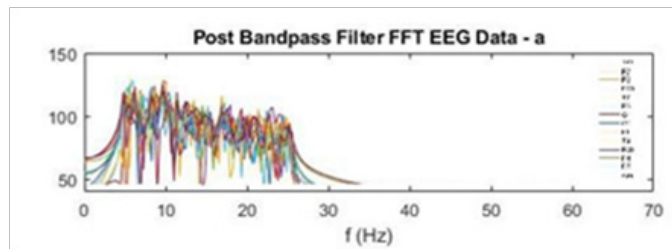


Figure 3 Example EEG Data for the Letter 'a' Post Bandpass Filter.

Feature extraction

The preprocessed EEG data were then put through a series of feature extraction involving principal component analysis (PCA) and Fast Fourier Transform. PCA uses a linear trans-formation to convert a set of correlated observations into a set of uncorrelated principal components, thereby reducing the dimension of the data set.⁷ This is done by determining the directions of greatest variance within the data, then rotating the axes to point in these directions. This makes decision boundaries more apparent and therefore makes it easier to classify the data. In this research, PCA was used to convert the preprocessed EEG data into fourteen components, one for each channel. A visual representation of the rotation of axes for PCA is shown in Figure 4. The Fourier Transform of the signal was also used to visually detect which channels would provide the most information. The Fourier Transform described in Equation 2 converts a time-domain signal into the frequency domain where $g(t)$ is the time domain signal, f is the frequency, and t is time. This allowed the data to be analyzed based on which frequencies were present in a signal. An example of using frequency to determine which letter would be best to use is shown in Figure 5. The channel which appeared to have the greatest consistency in frequency peaks were used initially, but eventually all 14 channels were tested.

$$FT(g(t)) = \int g(t) e^{-j2\pi ft} dt \quad (2)$$

Speech classification

Once the PCA of the preprocessed EEG data was completed, the k-NN algorithm was used for imagined speech classification. The k-NN test is a classification method that uses Euclidean distances

between the unknown datum and its known neighbors. The closest neighbor to the unknown datum was used to classify that datum. For example, if the nearest neighbor was classified as the letter a, then the datum was also classified as the letter a. An example of k-NN classification is depicted in Figure 6.

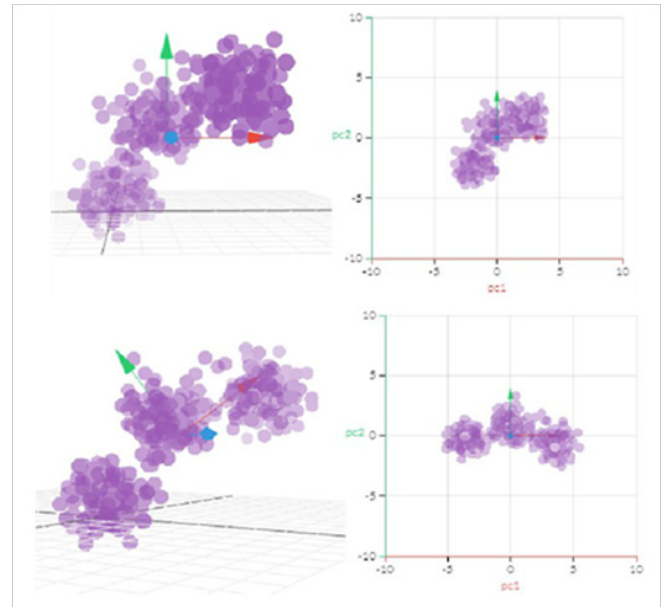


Figure 4 PCA Rotation of Dimensions for Feature Extraction.

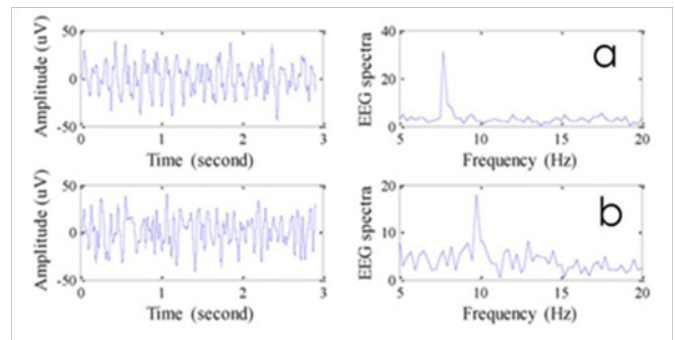


Figure 5 Example of Data Shown in the Frequency Domain via FFT for Channel Selection.

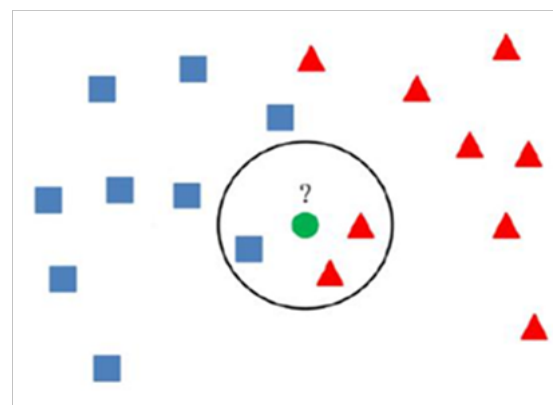


Figure 6 k-Nearest Neighbors Classification: the unknown green circle datum would be classified as a red triangle if a 3-NN classification test were to be used.

Simulation results and analysis

Table 1 & Table 2 show the accuracy rate of the speech classification process for two test subjects. The two tables are divided by letter. Three classification methods were used: k-NN based on the average power of the signal, k-NN based on the PCA of all 14 channels, and k-NN based on a single channel. Each method was tested using the same subject as the training and testing data sets, and crossing the subjects for training and testing sets. The channels which resulted in the highest accuracy rates were channels F3, F8, and AF4 - all of which achieved an over-all accuracy rate of 87.5%. These results could demonstrate that for future testing only one channel would be necessary to collect the raw EEG data, which in turn could result in cheaper and more accessible headsets. In addition, it is not that surprising that these channels are the ones that produced the best results. Channels F8 and AF4 are both located on the right frontal lobe - the part of the brain who's primary function is focus. It makes sense then that these channels were among the most stimulated considering how much focus was required for imagining a specific letter for 15seconds. However, what is even more interesting is the F3 channel which is located in the left frontal lobe right over the Broca's area, which is depicted in Figure 7. The Broca's area is the part of the brain solely responsible for speech production. This implies that a single electrode located on the Broca's area, or even a group of micro electrodes located here could be sufficient for future imagined speech recognition. The Hurst exponent is an estimate of the predictability of a time signal, with a 0 indicating a completely unpredictable signal and a 1 designating a completely predictable signal. The use of the Hurst exponent is proposed for future work to limit the effect of artifacts by discarding any EEG components that have Hurst values outside of the range 0.70-0.76, following.⁸ In addition, future work should include investigating the effect of the use of a normalization process such as the one mentioned earlier that was proposed.¹⁴

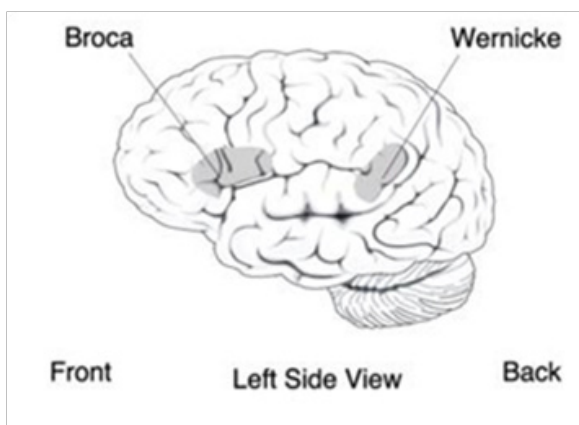


Figure 7 Broca's Area Location on the Right Frontal Lobe.

A training set is a set of data against which a datum is compared for imagined speech classification. In general, the larger the training set the more accurate the classification. These training sets are specific to an individual person because different people produce different brain activity patterns while thinking of the same thing. Due to this reason, it is rather difficult to classify one person's datum by comparing it to another person's training set. However, part of the completed research involved testing out a calibration test that classified one subject's imagined letter using another subject's trained model. This achieved an average accuracy rate of 69%. In an attempt to improve

this accuracy rate, two calibration tests were designed to be tested in the future. These two calibration tests have been titled "On-The-Go Training Sets" and "Normalized Training Sets".

On-the-go training sets

For the on-the-go training sets, an initial complete training set was collected for a single person. Then each additional test subject completed a miniature training set that was combined with the full training set. The new test subject's datum was then classified via comparison with this combined training set. The mini-training sets required less than 5 minutes of data collection, and therefore can be completed on-the-go (Figure 8).



Figure 8 On-The-Go Calibration Test Diagram.

Normalized training sets

For the normalized training sets, initial complete training sets were collected for five people (five complete training sets total). These five training sets were then normalized and combined via covariance matrix descriptors such as Riemannian manifold features, similar to the novel normalization process described.¹⁴

Conclusion

In this paper we presented a method for acquiring, preprocessing, and classifying raw EEG signals related to imagined speech for letters "a" and "b". Also presented are two methods for calibrating the test to a particular subject (Figure 9). To date, the results have shown an average accuracy rate of 87.5%, which is greater than an untrained predictor accuracy rate of 50%, therefore making this research project a success. Future work that includes utilizing the Hurst exponent and a more sophisticated normalization process for signal preprocessing could further increase the accuracy rate. In addition, based on the current results, the F3 channel is the best channel to look at for imagined speech recognition. Future research should focus on isolating this channel. Finally, the two proposed calibration tests (on-the-go and normalized) should be explored in order to universalize the imagined speech recognition classifier. This research has shown that silent communication via reading a person's thoughts is certainly possible, and could even be widespread within the next decade.

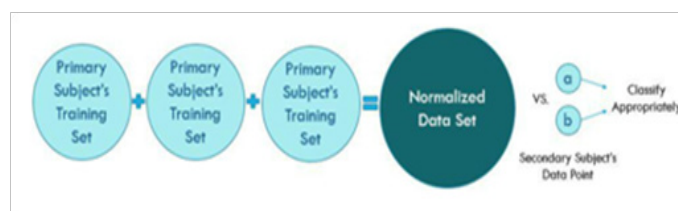


Figure 9 Normalized Calibration Test Diagram.

Acknowledgments

None.

Conflicts of interest

Author declares that there is no conflict of interest.

References

1. Tucker P. Actual Telepathy is One Step Closer to the Battlefield. 2014.
2. Goldblum N. The Brain Shaped Mind. Cambridge. Cambridge University Press; UK: 2001.
3. Small GW, Bookheimer SY, Thompson PM, et al. Current and future uses of neuroimaging for cognitively impaired patients. *The Lancet of Neurology*. 2008;7:161–172.
4. Emotiv EPOC+. 2018.
5. Thesan T. Electroencephalography. 2018.
6. Martin S, Brunner P, Pasley N. Word Pair Classification During Imagined Speech Using Direct Brain Recordings. 2016.
7. Brigham K, Kumar AVKV. Imagined speech classification with eeg signals for silent communication: A preliminary investigation into synthetic telepathy. *Proceeding of the International Conference on Bioinformatics and Biomedical Engineering*. 2010:1–4.
8. Mansukhani S. The Hurst Component: Predictability of Time Series. 2010.
9. Esfahani ET, Sundararajan V. Classification of primitive shapes using brain-computer interfaces. *Computer-Aided Design*. 2012;44:1011–1019.
10. Wester M. Unspoken Speech - Speech Recognition Based on Electroencephalography. Mellon Pittsburgh; University: 2006.
11. Porbadnigk A, Wester M, Schultz T. Eeg-based speech recognition impact of temporal effects. *Proceeding of the International Conference on Bio-inspired Systems and Signal Processing*. 2009;1:376–381.
12. Hu Y, Li S, Palmer A. Speak Your Mind: Extracting Imagined Words and Letters from EEG Signals. 2018.
13. Wang L, Zhang X, Zhang Y. Analysis and classification of speech imagery eeg for bci. *Biomedical Signal Processing and Control*. 2013;8:901–908.
14. Nguyen CH, Karavas GK, Artemiadis P. Inferring imagined speech using eeg signals: A new approach using reimannian manifold features. *Journal of Neuroengineering*. 2017;15:1–17.