The Thesis committee for Megha Parhi
Certifies that this is the approved version of the following thesis:

Classifying Imaginary Vowels from Frontal Lobe EEG via Deep Learning

APPROVED BY

SUPERVISING COMMITTEE:

Ahmed Tewfik, Supervisor

Joydeep Ghosh

Classifying Imaginary Vowels from Frontal Lobe EEG via Deep Learning

by

Megha Parhi

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE in ENGINEERING

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2019

Acknowledgments

I am indebted to my adviser Professor Ahmed Tewfik for giving me the opportunity to join his research group and advising me during this research project. I have learned from him how to approach research and how to understand high-level research perspectives. He amazes me with his abilities and I have been privileged to work under his supervision. I also would like to thank Professor Joydeep Ghosh for taking his time to be a reader for this thesis and for his Data Mining course.

I have had the privilege to meet many amazing people at UT. First, I would like to thank Anum, Preeti, Shalmali, Eirini, Ahmad, Rebal, Srilakshmi, Avro, Khurram, Travis, and Mai Lee for their words of wisdom and friendship during the tough times. Second, I would like to thank Manan, Alan, Ashish, Soumya, Vatsal, Shorya, Ronshee, Vutha, Nitin, and Rajat for the enlightening discussions and fun. Finally, I would like to thank Melanie Gulick and Karen Little, who made administrative work easy to deal with.

Classifying Imaginary Vowels from Frontal Lobe EEG via Deep Learning

Megha Parhi, M.S.E
The University of Texas at Austin, 2019

Supervisor: Ahmed Tewfik

Brain-Computer Interface (BCI) is a promising technology for individuals who suffer from motor or speech disabilities due to the process of decoding brain signals. This thesis uses a dataset for imagined speech to classify vowels based on the neurological areas of the brain. We demonstrate that by using the frontal region of the brain, we obtain higher than 85 percent accuracy using a CNN and LSTM. This accuracy is higher than previous studies that have classified the dataset using the entire brain region. This work shows great promise in using the physiological aspects of the brain associated with specific tasks.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Brain-Computer Interface (BCI) is a promising technology that shows great promise in improving the quality of life in clinical neurology and rehabilitation [24, 9]. BCI's objective is to aid people with disabilities to interact with their environment by decoding brain signals instead of relying on their muscle movement [63]. This technology could be very beneficial to people, for example, who suffer from Locked-in syndrome. Locked-in syndrome is a rare neurological disease where a person is completely paralyzed and are unable to move any of their muscles, but they are able to communicate with their eye blinks and eye movement [38]. BCI could help these individuals to communicate by using covert or silent speech. There are several different methods to measure brain signals for speech.

Electromyography (EMG) has shown promise in silent speech for healthy people by using EMG electrodes on the larynx and orofacial muscles [4, 16, 30]. EMG-based recordings are beneficial for speech deprived people, but this approach would not work for someone who is unable to move their muscles. A BCI would be very beneficial to aid such a person.

BCI signals can be measured by a plethora of instruments. These measurement methods include: magnetoencephalography (MEG), electrocorticography (ECoG), Local Field Potential (LFP), single-unit activity (SUA) and electroencephalogram (EEG) [9]. BCI measurements can be either invasive [26, 40] or non-invasive [64, 57]. EEG and MEG are non-invasive, while ECoG, LFPs, and SUAs are invasive methods. EEG and MEG record signals that are based on the average of the activity of millions of neurons near the electrode. MEG is a method where signals have higher spatiotemporal resolution than EEG data [43]. A disadvantage of the use of MEG is that it is very expensive to measure data on a MEG machine and there are very few machines at this point in time. ECoG requires a surgical procedure where the electrodes are placed on the cerebral cortex. ECoGs have also been used in BCI experiments for controller based experiments [48, 53, 40]. LFPs and SUAs also require a surgical procedure where electrodes are placed into the cerebral cortex. SUAs record data for a single neuron, while LFPs record signals based on the activity of 10-90 neurons. EEG measurements are more common as these are portable, non-invasive, and low-cost. EEG shows promise in applications for control and speech. For this thesis, we use EEGs as the method of measurement for experiments.

The first study for EEG was conducted by Hans Berger in 1929 [3]. EEG data can be collected in two ways. In one approach, electrodes are placed surgically in the brain. This is very costly and is an invasive method, which could lead to long-term complications. A non-invasive method involves dry EEG electrodes are placed on the head. Non-invasive EEG measurements are low-cost and have been shown to distinguish the differences in brain activity using electrical fields. The drawback of using EEG is that the data collected is challenging to analyze due to its high-dimensionality,low signal-to-noise ratio (SNR), and various artifacts from the participant as well as the instrument. Speech for BCI systems have been analyzed extensively in both physical and imaginary speech scenarios.

Speech is a vital sense for communication. There have been several studies towards understanding speech from both covert (or imaginary) speech and physical speech perspectives to solve this complex problem for different scenarios. Several studies have investigated how to classify individual speech into categories like English vowels, short words, and long words [12, 15, 33, 46]. The majority of this work has been carried out using classical machine learning (ML) models like Support Vector Machines. Neural Networks have shown great promise for speech processing and have achieved better accuracy than classical ML methods. Many of the experiments for speech use all the data from the electrodes and don't specifically pinpoint a certain location of the brain where activity occurs. The question lies in how do the neurological areas of the brain associate with the data and how a model can be learned that requires less computation leading to a low-cost speech-based BCI.

The following thesis addresses this question in the following manner. We show that by using the electrodes from the frontal lobe, i.e., the region responsible for speech in the brain, we can get the same, if not better, accuracy than using the measurements from all the electrodes. To the best of our knowledge, this is the first study of using a certain lobe to classify speech from the dataset of [46]. To show that the classification accuracy is as good, we analyze the data using subsets of the entire 64 electrodes. The subsets studied include all the electrodes and the electrodes in the frontal region of the brain where speech occurs. The classification process is modeled using two very well known and frequently used deep learning algorithms. These two algorithms are the Long Short-Term Memory (LSTM) and the Convolutional Neural Network (CNN). Our results show that by using the frontal electrodes, the accuracy of the data is above 90 percent for each participant. This shows that speech-based BCI signals can be classified using only the active parts of the brain, which would help in enabling less computation time and less hardware needed for BCI experiments.

This thesis is organized in the following manner. Chapter 2 discusses the past work in BCI. Chapter 3 explains the data used for the experiments. Chapter 4 explains the modeling of the data. Chapter 5 discusses and compares the results. Finally, Chapter 6 concludes the work and presents future work.

Chapter 2

Past Work

BCIs have been studied extensively for applications in speech and other tasks to understand what information is obtained from brain-signals using the many different measurement methods. The main objective of much of BCI is to aid in rehabilitation for people with motor disabilities.

2.1 Applications in BCI

BCI's target population is subdivided into three groups [47]. The first group involves $locked - in$ patients who have lost all motor-control. The second group involves patients who have some capability of movement. The third group is for healthy individuals, who have no real need for aid from BCI. The types of applications that BCI can be involved in include entertainment, motor restoration, enviromental control, locomotion, and communication. Entertainment has had quite a growth in applications for BCI. Gaming and controlling objects using BCI systems are some examples of the types of applications that are involved in BCI entertainment and environmental control. Gaming has been very popular recently with the idea of being able to control objects with your mind. Examples include the "mind-controlled" BCI quadcopter [37] and

BCI using virtual reality [11]. Pacman, Pong, and other such classic controller games have been played using motor imagery using BCI systems [34]. Pinball has also been played using non-invasive recordings to show the promise of using BCI systems with complex controller tasks [58]. However, entertainment using BCI is not going to aid in uncovering how to best create a system for someone who has poor motor-control. Communication is a vital application for BCI systems to aid in rehabilitation.

Communication in BCI has different applications. A common method of studying BCI is to communicate using a keyboard on a screen with a BCI system. For example, using a virtual keyboard, studies have developed a device that spells with such a system [5, 10, 17, 29]. Eye blinks were also shown to be able to use a virtual keyboard to communicate [10]. Another application in communication was based on internet browsers[31, 2]. Speech-based communication is another method of communication that has been studied extensively.

2.2 Classical Machine Learning in BCI

BCI data has been studied extensively before the promise of neural networks using classical ML algorithms like Support Vector Machines. Classification is one approach to understanding BCI data. Classification has been used to recognize the characteristics of the brain activity based on features. Traditionally, classification was accomplished by supervised learning. Nguyen *et. al* investigate an imaginary speech dataset using Riemannian manifold features classified by a Relevant Machine Vector [46]. Another example of supervised learning used frequency-following responses to project electrophysiological responses onto a low-dimensional spectral feature for two vowels [65]. Supervised learning, however, has many drawbacks due to a BCI system being non-stationary [55]. Supervised learning has many drawbacks especially in large datasets. For this reason, semi-supervised learning has been suggested for a speller system using BCI [41]. Semi-supervised learning, however, is not the most realistic method to look at BCIs using the brain signal when ground truth is not known.

Unsupervised learning or reinforcement learning is a good method to identify BCI data when all the data are unlabelled. One approach to unsupervised experiments is to have the user and BCI learning together [60, 42, 28]. Reinforcement learning, a type of supervised learning that is based on a $reward - learning$, has been used for classification by observing the neuron spikes when a person makes errors [56].

2.3 Speech-based clinical studies

BCI for communication has been studied extensively by both invasive and not-invasive methods. One method of study was to implement a "typewriter" approach where intracortical electrodes are implanted. One such experiment in by Kennedy was the first implant for a human subject with a chronic microelectrode to aid a paralyzed patient by an intracortical BCI [32]. This subject could make binary decisions until her death of 76 days after implantation. Another study involved an experiment by Donoghue [26]. This study involved human volunteers, who were implanted with the Utah microelectrode array-based system. At least two of these subjects learned to use a mouse cursor on a computer screen. Speech data has also been studied based on deep learning for EEG

2.4 Machine Learning for Speech

Speech is crucial for rehabilitation for locked-in patients or those with speech impairments. Several studies have been made for imaginary speech and physical speech. Speech imagery is much easier to be repeated than image and motor imagery. We use in this thesis a speech imagery data for vowels. Research in speech imagery has been investigated by phonemes and syllables without vocalizing. One such example, was by Wester *et al.*, who showed a system with high accuracy that was capable of recognizing imagined speech with high accuracy [62]. It was later found that this high accuracy was due to the way the data was collected creating temporal correlation in the EEG [49]. A study by DaSalla *et al.* showed 68-79 percent accuracy when classifying a, u, and rest using CSP [12]. This high accuracy was later found to be obtained due to CSP having discriminant channels Fz, C3,Cz, and C4. These four channels are related to motor imagery and not speech imagery. Deng *et al.* used Huang-Hilbert transform to get an accuracy of 72.6 percent by classifying ba and ku vowels [15]. These studies have primarily used signal processing algorithms and classical ML to determine the classification for imagined speech in vowel data. The accuracy for these studies has been reasonable, but not as well as in neural networks.

Several models have been used to gain an understanding of the information that lies in BCI EEG for speech processing with neural networks. Hidden Markov Models (HMMs) were the first models to be used with neural networks for speech processing [6, 61]. However, the combination of HMMs with neural networks has shown to not perform as well as deep neural networks. Neural networks gained even more recognition in the use of acoustic modeling [45]. Several papers have shown that speech recognition with recurrent neural networks (RNNs) avoid the misplaced alignments of HMMs and achieve better accuracy due to the larger state space [19, 20]. Long short-term memory models are a subgroup of recurrent neural networks that have shown to also have very good accuracy in categorizing speech [35, 52]. Some work has also been done using a hybrid neural network architecture with LSTM and Convolutional Neural Networks (CNN) speech system that has shown better results than using a single neural network model [51]. These methods all show that accuracy improves using a neural network compared to classical ML models. They all also classify speech signals based on using the entire brain's data. Our approach for this thesis is show that the physiological aspects of the brain play a big part in understanding the BCI data and that less computation is sufficient when using the parts of the brain that are active during speech.

Chapter 3

Data

Data plays an important role in all machine learning problems. The data used for this thesis is from a publicly available imaginary speech dataset [46]. This data was collected at the Human-Oriented Robotics and Control Lab of the Arizona State University. There were 15 subjects in this study, however data for seven subjects were made publicly available. The subjects for the dataset are all right-handed except Subject 13. During the experiments, the subjects were instructed to pronounce words in their mind with cues from a computer monitor. For each trial, a beep would appear when the trial started and a visual cue would be prompted. The beeping sound was repeated with a period of T. T is 1 second for the vowel data. The trials would then end with a rest period of approximately two seconds. The reasoning for using this approach to run the experiments was to make sure that the areas of the brain where sound occurs are not activated. This dataset has data for short words, long words, and vowels for EEG data that follows the international 10/20 system placement, which is explained more in Section 3.1 [54]. For this thesis we focus on the vowel data. The vowel data consists of data for a, i, and u. An example of the data for an EEG can be seen in Figure 3.1 for Subject 11.

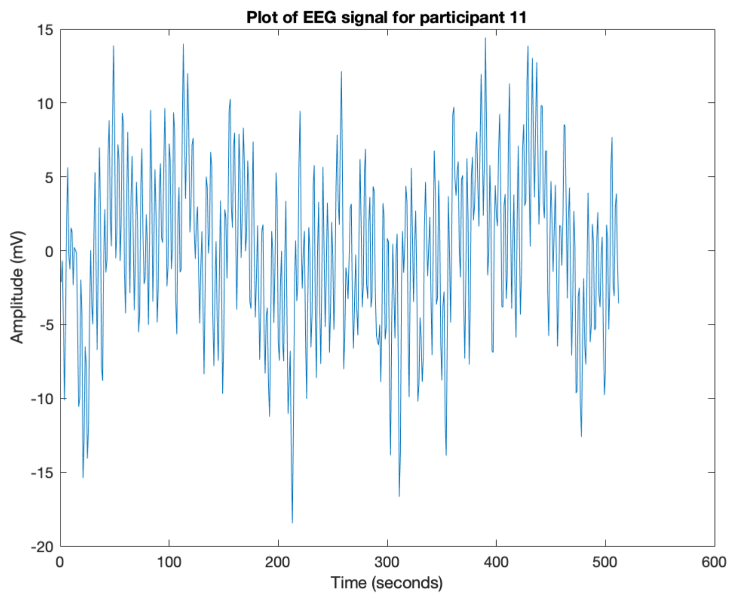

Figure 3.1: Example of EEG data from dataset for Participant 11.

3.1 10/20 System

The 10/20 system is an internationally recognized system based on the location between an electrode and the cerebral cortex. The numbers ten and twenty refer to the distances between adjacent electrodes. An example of the 10/20 system is shown in Figure 3.2. Each electrode has a letter identifying the lobe in the brain where the electrode resides. Table 3.1 summarizes what region of the brain each electrode resides in. Even numbered electrodes represent electrodes on the right hemisphere of the brain. Similarly, odd numbers represent the left hemisphere. Each region of the brain pertains to a specific function. Table 3.2 summarizes the function of each electrode. This can be observed better in 3.3 from [1].

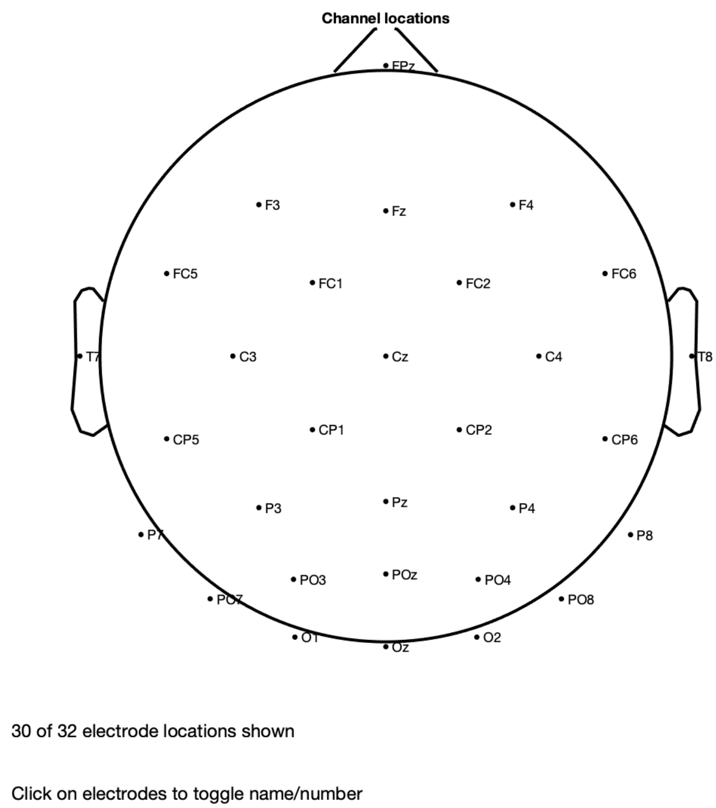

Figure 3.2: Example of electrode position in the 10-20 system

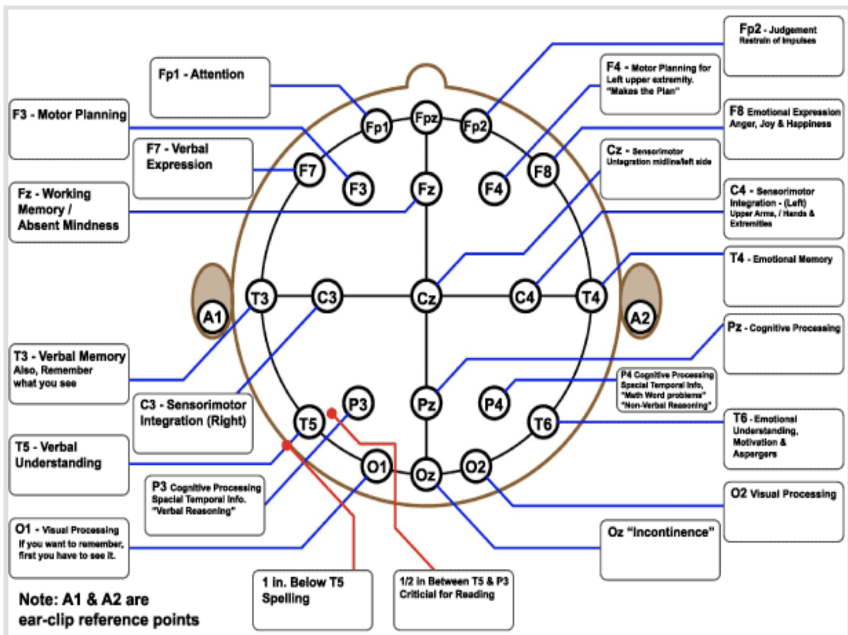

Figure 3.3: Example of the neurological function of each electrode and placement using the 10-20 system with brain function [1].

Table 3.1: Lobes of the brain pertaining to electrode abbreviation

Electrode abbreviation	Lobe in Brain
F	Frontal
T	Temporal
C	Central
P	Parietal
O	Occipital

Table 3.2: Summary of the neurological function of each electrode

Electrode	Function
Fp1	Attention
Fp2	Judgement
F3,F4	Motor Planning
F7	Verbal Expression
F8	Emotional Expression
Fz	Working Memory/Absent mindedness
T3	Verbal Memory
T4	Emotional Memory
T5	Verbal Understanding
T6	Emotional Understanding
C3,C4,Cz	Sensorimotor Integration
P3,P4,Pz	Cognitive Processing
O1,O2	Visual Processing

3.2 Preprocessing

The following data has been preprocessed as listed below:

- Bandpass filter at 8-70 Hz using a 5th order Butterworth filter.
- Notch filter at 60 Hz for power line signal.
- EOG artifact removal by the ADJUST algorithm [44] to remove muscle

and eye blinks.

- Downsample from 1000 Hz to 256 Hz.

After pre-processing the dataset, the cross-correlation matrices are calculated. This process is explained in Section 3.3.

3.3 Cross-Correlation matrix

The features for the deep learning models chosen are the cross-correlation matrices between the electrodes for each trial for each of the seven subjects. We calculate the cross-correlation matrices for each subject. Each cross-correlation matrix is calculated in the following manner with e_i representing electrode i and e_j representing electrode j. X represents the EEG data and t represents the time. The covariance between each electrode is measured:

$$Cov(X_{e_i}, X_{e_j})(t) = E[(X_{e_i}(t) - E(X_{e_j}(t))(X_{e_i}(t+\tau) - E(X_{e_j}(t+\tau)]$$

The cross-correlation matrix between each electrode is then calculated as follows using the covariance:

$$R_{e_ie_j} = \frac{Cov_{e_ie_j}}{\sqrt{Cov_{e_ie_i}Cov_{e_je_j}}}$$

The matrices are calculated for each trial for each subject. The function 'corrcoef' from python was used in this calculation. Figure 3.4 represents an example of a correlation matrix from the data between a and i for subject 8e. The white diagonal represents the correlation between two

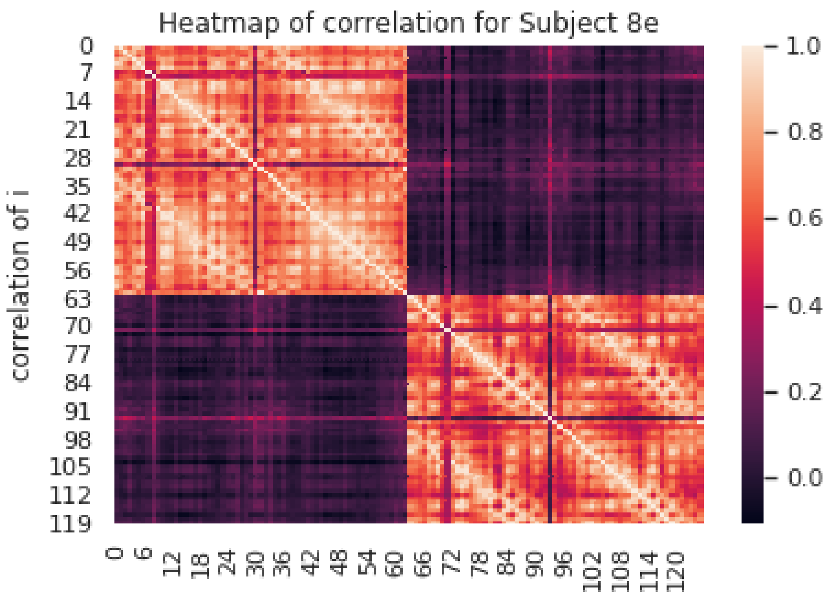

Figure 3.4: Example of heatmap of cross-correlation matrix for Subject 8e between a and i. The first half of the electrodes correspond to a and the second half to i.

of the same electrodes. The correlation matrices are symmetric. We can observe that there are some highly correlated electrodes and a few electrodes that have low correlation. This correlation matrix was calculated for data with 64 electrodes, which creates a 128×128 matrix. The top right and bottom left 64×64 matrix or the heavily purple section of the heat map represents the cross-correlation of a and i. The top left and bottom right 64×64 matrix represents the correlation of the vowel and itself. A total of 191 such matrices are calculated for each subject and used as features for all electrodes. The less correlated electrodes show that those electrodes may provide more discrimination than the electrodes that are highly correlated with each other. A subset of all the 64 electrodes is the 20 frontal electrodes where speech occurs. The cross-correlation matrices for the frontal electrodes are calculated in the same manner as the electrodes. The difference between the frontal electrodes correlation matrices and all the electrodes is the number of channels and the dimension of the frontal cross-correlation matrices. The frontal electrodes consist of 20 channels. The dimension of the cross-correlation matrices is 40×40. There are 191 matrices calculated for the frontal electrodes also. Labeling the data for the frontal and all the electrodes is the same.

3.4 Labeling

The correlation matrices vowel data consists of 191 matrices. Each subject has ten trials of data for each vowel. The trial data consists of preprocessed EEG data. The cross-correlation matrices are calculated as explained in Section 3.3. This data is then used to calculate the 191 matrices with labels of a,i, and u. The matrices are labeled by taking the first trial vowel with whatever vowel it is correlated with. For example, corr0 is labeled as a due to the fact that we take the first trial of a and then correlate it with i. This method aids in understanding what part of the data we are correlating. All the other matrices are calculated in a similar manner for each subject. For the frontal electrodes, we label the data in the exact same manner. Table 3.3 shows which correlation matrices (corr) pertain to which vowel with respect to both all the electrodes and the frontal electrodes. There are 75 matrices

Table 3.3: Labeling of Correlation Matrices for both frontal and all electrodes

Vowel label	Correlation Matrix
a	corr0–18,corr55–70,corr100–112,corr136–145,corr163–169,corr181–190
i	corr19–37,corr71–85,corr113–124,corr146–154,corr170–175
u	corr38-54,corr86–99,corr125–135,corr155–162,corr176–180

labeled as a, 61 labeled as i, and 55 labeled as u. This shows this dataset is not an unbalanced dataset.

Chapter 4

Models

Figure 4.1 summarizes the basic modeling of the experiments with the classifier depicting the deep learning model chosen for classification of the vowel data. In the case of the experiments, the two models chosen are LSTM and CNN. They were chosen primarily for their promising results in classification tasks [14, 22, 36]. The data input has 191 correlation matrices representing

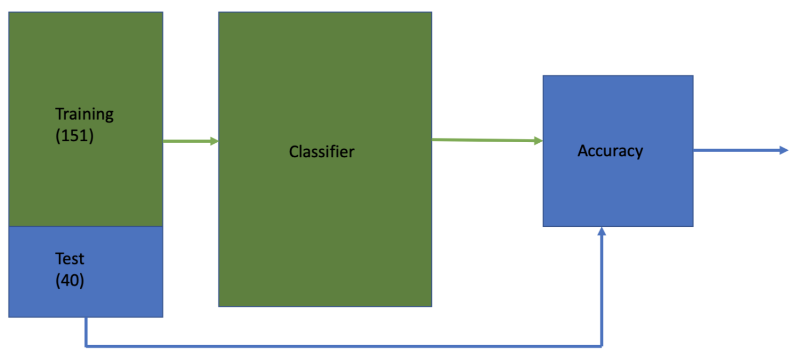

Figure 4.1: Basic summary of the training and testing block diagram for classification.

the features of the input to the classifiers for each subset of the electrode. Each correlation matrix is a 128×128 matrix for all the electrodes. For the subset of all the electrodes, the frontal electrodes consist of 40×40 with the same number of 191 correlation matrices for each subject. Out of the 191, 151 matrices are used as the training data and 40 as the testing data for each subject. For the experiments considered, we take all the electrodes in the brain, and the twenty electrodes from the frontal region for both the left and right hemisphere. We calculate the accuracy for each of the seven subjects.The vowels are *one–hot encoded*. The test accuracy is calculated based on how a vowel is calculated based on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This is calculated as:

$$Accuracy \ = \ \frac{TP + TN}{TP + TN + FP + FN}$$

The loss is calculated by categorical cross entropy. Categorical cross entropy is where the true class is represented as a one-hot encoded vector. The outputs are compared to the one-hot encoded vector, which will then determine how low the loss is. This can be measured in the following manner where $\hat{y}$ is the predicted output, which is the output from the softmax:

$$L(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}(y_{ij}log(\hat{y}_{ij}))$$

4.1 LSTM

Long short-term memory (LSTM) is a recurrent neural network (RNN) with the capability of long-term memory units [27]. It has shown promising results in speech systems. LSTMs have four gates that interact in a specific way. RNNs typically have one neural network layer. Traditional LSTMs incorporate sequential data as their input data including time-series data. Adding layers to an LSTM can add a certain-level of abstractness to the results. There are several different LSTM models. A stacked LSTMs is one type of LSTM that was proposed by Hinton et, al. for speech [20]. In this thesis, we use the same stacked LSTM as [20]. One layer of the LSTM used is shown Figure **??**. We

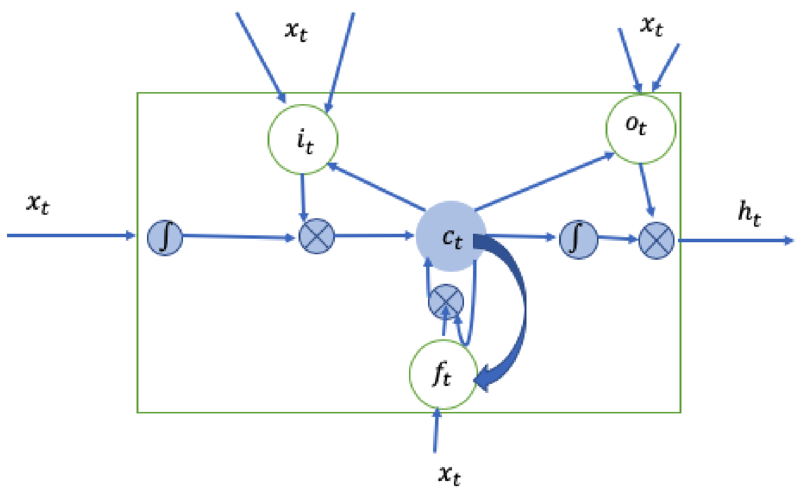

Figure 4.2: One layer of LSTM architecture

run the data with correlation matrices using a stacked LSTM with two layers. In a typical RNN with an input sequence $\mathbf{x} = (\mathbf{x_1}, ...\mathbf{x_T})$, $\mathbf{h} = (\mathbf{h_1}, ..., \mathbf{h_T})$ is the hidden vector, $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, ...\mathbf{y_T})$ is the output vector, and the time series is from $t = 1, 2, ..., T$. The RNN is formulated as follows:

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h)$$

$$y_t = W_{yh}h_t + b_y$$

where W represents the weight matrices, b represents the bias vector, and $\mathcal{H}$ denoted the hidden layer function. The LSTM comprises of four gates: input gate, forget gate, output gate, and cell activation vector. These are all the same size as h from RNN. The LSTM can be written as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_t - 1 + W_{cf}c_{t-1} + b_f)$$

$$c_t = f_tc_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_c)$$

$$h_t = o_t tanh(c_t)$$

where σ represents the logistic sigmoid function, i, f, o, and c represents the input gate, forget gate, output gate, and cell activation. The weight matrices from the cell to gates are diagonal W_{si}.

4.2 CNN

Convolutional Neural Networks (CNN) have shown great promise in classifying images [36, 39]. A CNN is a multi-layer neural network with several convolution-pooling layer pairs and fully-connected layers at the output. It can take an input image and be able to differentiate from other images. For this thesis, our correlation matrices are the images. Pre-processing in CNNs are much lower than other classification algorithms, which makes it easier to work with. First, a CNN has an input tensor in the case of our data for all the electrodes, we have an input tensor with size 191 by 128 by 128 consisting of each image or cross-correlation matrix. An individual image is of size 1 by 128 by 128. This image is an RGB or Red, Blue, and Green image. Next, the convolution layer or kernel is run with certain image dimensions. In our case, we take a 3*3*1 image. The kernel then shifts through the entire image. This is known as the convolution step, which is used to extract high-level features. We use two convolutional layers in our experiment. After convolving, a pooling layer is run, which is used to decrease the computation needed. We use a max pooling layer in our case. After the CNN has output, an activation and softmax layer is calculated in parallel. This is then used to find the cross-entropy loss. This step is also done for the LSTM.

Chapter 5

Results

To test whether the hypothesis that the selected physiological regions of the brain creates higher accuracy than using the entire brain, we ran experiments using all 64 electrodes and a subset of the 64 electrodes containing 20 electrodes that correspond to the frontal electrodes. These are known as physiological areas of the brain where speech occurs. We ran all our experiments using Google CoLab. This chapter discusses the parameters used and results of the experiments.

5.1 Model Parameters

Parameter tuning is an art when it comes to training models. For the experiments that are run on the correlation matrices, Table 5.1 summarizes the parameters for each model for each subset of the electrodes.

5.2 Experimental Results

5.2.1 LSTM

Table 5.2 shows the accuracy of the LSTM based on all 64 and subset containing 20 electrodes. We can observe that utilizing the brain signals in

Table 5.1: Parameters of Models

Model Parameter	CNN	LSTM	frontal CNN	frontal LSTM
Epochs	100	50	100	50
Batch Size	150	100	150	100
Total Layers	2	2	2	2
Number of Hidden Layers	1	1	1	1
Activation	ReLu	ReLu	ReLu	ReLU
Optimizer	Adam	Adam	Adam	Adam

the frontal electrodes shows higher accuracy than learning from the entire 64 electrodes for these eight subjects. Using all the electrodes, the accuracy using LSTM never reaches above 80 percent, while when the frontal electrodes are used, the accuracy significantly improves to above 90 percent for each subject. This shows the promise of using EEG data in certain regions of the brain based on activities.

Table 5.2: Test Accuracy of LSTM

Participant	Accuracy (all 64 electrodes)	Accuracy (frontal 20 electrodes)
Subject 8	74.4	96.2
Subject 8e	76.9	100
Subject 9	33.3	92.1
Subject 11	12.8	90
Subject 12	51.3	99.2
Subject 13	33.3	100
Subject 15	69.2	100

5.2.2 CNN

The test accuracy for CNN is shown below in Table 5.3 using the parameters from 5.1.

Table 5.3: Test Accuracy of CNN

Participant	Accuracy (64 electrodes)	Accuracy (20 frontal electrodes)
Subject 8	68.5	92.1
Subject 8e	73.4	95.1
Subject 9	28.2	89.5
Subject 11	15.2	85
Subject 12	53.6	94.5
Subject 13	27.4	98.4
Subject 15	73.4	100

Similar to LSTM, we observe that there is significant improvement by just using the frontal electrodes. The CNN accuracy is slightly worse than LSTM, but it shows significant improvement. Now let's observe how well the accuracy achieved is compared to past work using the same dataset.

5.3 Comparison of Results with Past Work

We compare the results from Table 5.2 and Table 5.3 with results from [50] ans [46]. We compare with the two literatures due to them using the same dataset. Saha *et al.* proposed a hybrid LSTM and CNN hybrid-based model. They calculated the channel cross-covariance of the electrodes to determine the accuracy using the Nguyen *et al.* dataset, which we used in our experiments as well. Unlike Saha *et al.*, we calculate the cross-correlation matrices between the two electrodes for our features. Another difference between the two studies is that we pinpoint the frontal electrodes as the region where the most activity in the brain occurs during speech. Both Saha and Nguyen use all 64 electrodes of the brain region. Table 5.4 summarizes the results from [50], which is compared with [46]. Nguyen *et al.* compute the covariance matrix as the feature vector and use a Relevance Vector Machine to classify the vowel data using Riemannian Manifold features [46]. Comparing Table 5.2, Table 5.3,and

Table 5.4: Results of test accuracy for vowel data from Saha *et al*. [50] and [46]

Subjects	Nguyen *et al.* [46]	Saha *et al.* [50]
Subject 8	51	73
Subject 11	53	75
Subject 12	51	79
Subject 13	46.7	69
Subject 15	48	84

Table 5.4, we observe that the results from [46] are very low in accuracy compared with Saha *et al.* [50]. The accuracy for [46] is approximately at 50 percent for all the subjects. This shows that for our Subjects 9 and 13 get worse accuracy using all the electrodes using neural network models. This also demonstrates the difference between traditional ML methods and the accuracy that can be obtained with neural networks for classification.

The approach in Saha *et al.* [50] achieves accuracy less than 85 for all the subjects. This is slightly better than our results for all electrodes since we achieve less than 80 percent. From our results, for all 64 electrodes our results have similar accuracy to [50] for the Subject 8. When we use just the frontal electrode, we have accuracy above 85 percent for all subjects using both LSTM and CNN, which are better than the hierarchical model using simple LSTM and CNN architectures. This shows that by understanding the physiological aspects of the brain, we can better understand and classify brain signals to aid in computation time and accuracy.

Overall, we were able to prove that by pinpointing the neurological area of the brain that is active, one is able to obtain higher accuracy than using the data from the entire brain. This creates better understanding of brain signals and shows the need to understand the physiology of the brain associated with specific tasks.

Chapter 6

Conclusion and Future Work

We demonstrate in this thesis that by using the correlation data from the frontal region of the brain that we are able to obtain an accuracy that is above 90 percent, while using the entire brain region the accuracy tends to be below 80 percent using LSTM and CNN. This demonstrates that the neurological parts of the brain where the brain is active could be the only regions needed to gain information to aid non-vocal patients. This would significantly reduce hardware as well as computational time in BCI experiments

There are several paths that could be promising from this initial work. One path could be to run more sophisticated models, such as a hybrid model, on the data to determine how the accuracy is for each region of the brain. This would allow in getting a full understanding of each part of the brain. These results could then be compared with these initial results. There are several other methods that could be studied to get a general sense of how one could use the physiology of the brain to get a better understanding of how BCI could hopefully one day have a state of the art BCI system to aid those that are unable to speak.

One step to allow for a benchmark would be to create a clinically public datatset. A possibility for this is to use the resources at UTs medical school or cognitive psychology labs to determine how using the physiology of the brain can help locked-in people or people who are unable to speak. This would really help in comparing these individuals with healthy participants to determine how well this method would work. It would be great if this could be a large study to gain a better understanding and determine how well such a process to get closer to rehabilitation and have a benchmark with other researches on the best methods for this main goal.

Another possibility is to use the collected data and figure out how the system being non-stationary could affect the data and system as a whole. This could potentially help in understanding what algorithms need to be created that have potential to create one day an online system for people who suffer from speech disorders.

Bibliography

[1] Edmonton neurotherapy: Qeeg brain mapping. https://www.edmontonneurotherapy.com/edmonton_neurotherapy_qeeg_brain_mapping.html. Accessed: 2019-10-25.

[2] Michael Bensch, Ahmed A Karim, Jürgen Mellinger, Thilo Hinterberger, Michael Tangermann, Martin Bogdan, Wolfgang Rosenstiel, and Niels Birbaumer. Nessi: an eeg-controlled web browser for severely paralyzed patients. *Computational intelligence and neuroscience*, 2007, 2007.

[3] Hans Berger. Über das elektrenkephalogramm des menschen. *European archives of psychiatry and clinical neuroscience*, 87(1):527–570, 1929.

[4] Bradley J Betts, Kim Binsted, and Charles Jorgensen. Small-vocabulary speech recognition using surface electromyography. *Interacting with Computers*, 18(6):1242–1259, 2006.

[5] Niels Birbaumer, Nimr Ghanayim, Thilo Hinterberger, Iver Iversen, Boris Kotchoubey, Andrea Kübler, Juri Perelmouter, Edward Taub, and Herta Flor. A spelling device for the paralysed. *Nature*, 398(6725):297, 1999.

[6] Herve A Bourlard and Nelson Morgan. *Connectionist speech recognition: a hybrid approach*, volume 247. Springer Science & Business Media, 2012.

[7] David M Brandman, Sydney S Cash, and Leigh R Hochberg. human intracortical recording and neural decoding for brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1687–1696, 2017.

[8] Jonathan S Brumberg, Philip R Kennedy, and Frank H Guenther. Artificial speech synthesizer control by brain-computer interface. In *Tenth Annual Conference of the International Speech Communication Association*, 2009.

[9] Jonathan S Brumberg, Alfonso Nieto-Castanon, Philip R Kennedy, and Frank H Guenther. Brain–computer interfaces for speech communication. *Speech communication*, 52(4):367–379, 2010.

[10] Brijil Chambayil, Rajesh Singla, and Rameshwar Jha. Virtual keyboard bci using eye blinks in eeg. In *2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 466–470. IEEE, 2010.

[11] Christopher G Coogan and Bin He. Brain-computer interface control in a virtual reality environment and applications for the internet of things. *IEEE Access*, 6:10840–10849, 2018.

[12] Charles S DaSalla, Hiroyuki Kambara, Makoto Sato, and Yasuharu Koike. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural networks*, 22(9):1334–1339, 2009.

[13] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[15] Siyi Deng, Ramesh Srinivasan, Tom Lappas, and Michael D'Zmura. Eeg classification of imagined syllable rhythm using hilbert spectrum methods. *Journal of neural engineering*, 7(4):046006, 2010.

[16] Michael J Fagan, Stephen R Ell, James M Gilbert, E Sarrazin, and Peter M Chapman. Development of a (silent) speech recognition system for patients following laryngectomy. *Medical engineering & physics*, 30(4):419–425, 2008.

[17] Lawrence Ashley Farwell and Emanuel Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523, 1988.

[18] Jaime Gomez-Gil, Israel San-Jose-Gonzalez, Luis Fernando Nicolas-Alonso, and Sergio Alonso-Garcia. Steering a tractor by means of an emg-based human-machine interface. *Sensors*, 11(7):7110–7126, 2011.

[19] Alex Graves. Supervised sequence labelling. pages 5–13, 2012.

[20] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks, 2013.

[21] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6645–6649, May 2013.

[22] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

[23] Frank H Guenther, Jonathan S Brumberg, E Joseph Wright, Alfonso Nieto-Castanon, Jason A Tourville, Mikhail Panko, Robert Law, Steven A Siebert, Jess L Bartels, Dinal S Andreasen, et al. A wireless brain-machine interface for real-time speech synthesis. *PloS one*, 4(12):e8218, 2009.

[24] Christian Herff and Tanja Schultz. Automatic speech recognition from neural signals: a focused review. *Frontiers in neuroscience*, 10:429, 2016.

[25] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick Van Der Smagt, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372, 2012.

[26] Leigh R Hochberg, Mijail D Serruya, Gerhard M Friehs, Jon A Mukand, Maryam Saleh, Abraham H Caplan, Almut Branner, David Chen, Richard D Penn, and John P Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164, 2006.

[27] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[28] David Hübner, Thibault Verhoeven, Konstantin Schmid, Klaus-Robert Müller, Michael Tangermann, and Pieter-Jan Kindermans. Learning from label proportions in brain-computer interfaces: Online unsupervised learning with guarantees. *PloS one*, 12(4):e0175856, 2017.

[29] Beata Jarosiewicz, Anish A Sarma, Daniel Bacher, Nicolas Y Masse, John D Simeral, Brittany Sorice, Erin M Oakley, Christine Blabe, Chethan Pandarinath, Vikash Gilja, et al. Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. *Science translational medicine*, 7(313):313ra179–313ra179, 2015.

[30] Chuck Jorgensen, Diana D Lee, and Shane Agabont. Sub auditory speech recognition based on emg signals. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 4, pages 3128–3133. IEEE, 2003.

[31] Ahmed A Karim, Thilo Hinterberger, Jürgen Richter, Jürgen Mellinger, Nicola Neumann, Herta Flor, Andrea Kübler, and Niels Birbaumer. Neu-

ral internet: Web surfing with brain potentials for the completely paralyzed. *Neurorehabilitation and Neural Repair*, 20(4):508–515, 2006.

[32] P R Kennedy and R AE Bakay. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport*, 9(8):1707–1711, 1998.

[33] Jongin Kim, Suh-Kyung Lee, and Boreom Lee. Eeg classification in a single-trial basis for vowel speech perception using multivariate empirical mode decomposition. *Journal of neural engineering*, 11(3):036010, 2014.

[34] Roman Krepki, Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. The berlin brain-computer interface (bbci)–towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, 33(1):73–90, 2007.

[35] Gautam Krishna, Co Tran, Jianguo Yu, and Ahmed H. Tewfik. Speech recognition with no speech or with noisy speech. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1090–1094, May 2019.

[36] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[37] Karl LaFleur, Kaitlin Cassady, Alexander Doud, Kaleb Shades, Eitan Rogin, and Bin He. Quadcopter control in three-dimensional space using

a noninvasive motor imagery-based brain–computer interface. *Journal of neural engineering*, 10(4):046003, 2013.

[38] Steven Laureys, Frédéric Pellas, Philippe Van Eeckhout, Sofiane Ghorbel, Caroline Schnakers, Fabien Perrin, Jacques Berre, Marie-Elisabeth Faymonville, Karl-Heinz Pantke, Francois Damas, et al. The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress in brain research*, 150:495–611, 2005.

[39] Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*, pages 319–345. Springer, 1999.

[40] Eric C Leuthardt, Gerwin Schalk, Jonathan R Wolpaw, Jeffrey G Ojemann, and Daniel W Moran. A brain–computer interface using electrocorticographic signals in humans. *Journal of neural engineering*, 1(2):63, 2004.

[41] Yuanqing Li, Cuntai Guan, Huiqi Li, and Zhengyang Chin. A self-training semi-supervised svm algorithm and its application in an eeg-based brain computer interface speller system. *Pattern Recognition Letters*, 29(9):1285–1294, 2008.

[42] Shijian Lu, Cuntai Guan, and Haihong Zhang. Unsupervised brain computer interface based on intersubject information and online adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(2):135–145, 2009.

[43] Jürgen Mellinger, Gerwin Schalk, Christoph Braun, Hubert Preissl, Wolfgang Rosenstiel, Niels Birbaumer, and Andrea Kübler. An meg-based brain–computer interface (bci). *Neuroimage*, 36(3):581–593, 2007.

[44] Andrea Mognon, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti. Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2):229–240, 2011.

[45] A. Mohamed, G. E. Dahl, and G. Hinton. Acoustic modeling using deep belief networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):14–22, 2012.

[46] Chuong H Nguyen and Panagiotis Artemiadis. Eeg feature descriptors and discriminant analysis under riemannian manifold perspective. *Neurocomputing*, 275:1871–1883, 2018.

[47] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *sensors*, 12(2):1211–1279, 2012.

[48] Tejaswy Pailla, Werner Jiang, Benjamin Dichter, Edward F Chang, and Vikash Gilja. Ecog data analyses to inform closed-loop bci experiments for speech-based prosthetic applications. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5713–5716. IEEE, 2016.

[49] Anne Porbadnigk, Marek Wester, and Tanja Schultz Jan-p Calliess. Eeg-based speech recognition impact of temporal effects. 2009.

[50] Pramit Saha and Sidney Fels. Hierarchical deep feature learning for decoding imagined speech from eeg. *arXiv preprint arXiv:1904.04352*, 2019.

[51] Pramit Saha, Sidney Fels, and Muhammad Abdul-Mageed. Deep learning the eeg manifold for phonological categorization from active thoughts. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2762–2766. IEEE, 2019.

[52] Madhumitha Sakthi, Ahmed Tewfik, and Bharath Chandrasekaran. Native language and stimuli signal prediction from eeg. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3902–3906. IEEE, 2019.

[53] Gerwin Schalk and Eric C Leuthardt. Brain-computer interfaces using electrocorticographic signals. *IEEE reviews in biomedical engineering*, 4:140–154, 2011.

[54] F. Sharbrough, G.E. Chatrian, Ronald Lesser, H. Luders, M. Nuwer, and Terence Picton. American electroencephalographic society guidelines for standard electrode position nomenclature. *Clinical Neurophysiology*, 8:200–202, 01 1991.

[55] Pradeep Shenoy, Matthias Krauledat, Benjamin Blankertz, Rajesh PN Rao, and Klaus-Robert Müller. Towards adaptive classification for bci. *Journal of neural engineering*, 3(1):R13, 2006.

[56] Martin Spüler, Wolfgang Rosenstiel, and Martin Bogdan. Online adaptation of a c-vep brain-computer interface (bci) based on error-related potentials and unsupervised learning. *PloS one*, 7(12):e51077, 2012.

[57] Patrick Suppes, Zhong-Lin Lu, and Bing Han. Brain wave recognition of words. *Proceedings of the National Academy of Sciences*, 94(26):14965–14969, 1997.

[58] Michael W Tangermann, Matthias Krauledat, Konrad Grzeska, Max Sagebaum, Carmen Vidaurre, Benjamin Blankertz, and Klaus-Robert Müller. Playing pinball with non-invasive bci. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, pages 1641–1648. Citeseer, 2008.

[59] Meel Velliste, Sagi Perel, M Chance Spalding, Andrew S Whitford, and Andrew B Schwartz. Cortical control of a prosthetic arm for self-feeding. *Nature*, 453(7198):1098, 2008.

[60] Carmen Vidaurre, Claudia Sannelli, Klaus-Robert Müller, and Benjamin Blankertz. Machine-learning-based coadaptive calibration for brain-computer interfaces. *Neural computation*, 23(3):791–816, 2011.

[61] Oriol Vinyals, Suman V Ravuri, and Daniel Povey. Revisiting recurrent neural networks for robust asr. In *2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 4085–4088. IEEE, 2012.

[62] Marek Wester. Unspoken speech. 2006.

[63] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

[64] Jonathan R Wolpaw, Dennis J McFarland, and Theresa M Vaughan. Brain-computer interface research at the wadsworth center. *IEEE Transactions on Rehabilitation Engineering*, 8(2):222–226, 2000.

[65] Han G Yi, Zilong Xie, Rachel Reetzke, Alexandros G Dimakis, and Bharath Chandrasekaran. Vowel decoding from single-trial speech-evoked electrophysiological responses: A feature-based machine learning approach. *Brain and behavior*, 7(6):e00665, 2017.

Vita

Megha Parhi received her B.E.E at the University of Minnesota–Twin Cities in 2015.

Permanent address: mparhi@utexas.edu

This thesis was typeset with LaTeX† by the author.

†LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.