Using Deep Learning and EMG to predict non-audible speech

Rommel T. Fernandes
Seaver College of Science and Engineering
Loyola Marymount University
Los Angeles, CA 90045
Email: rferna16@lion.lmu.edu

***Abstract*—Many post-stroke victims deal with physiological problems such as speech impediments due to aphasia. With the advancement of Human-Computer Interface (HCI) research, this paper aims to create a project plan based on Silent Speech Interfaces that use Deep Learning and Machine Learning to predict non-audible speech. We will first briefly introduce HCI systems such as Silent Speech Interfaces, and go over how Deep Learning and Machine Learning can be used to predict speech. Next, we will go over our project plan that will investigate Deep Learning and Silent Speech Interfaces and go over the research problem, which will identify the main goal and objectives. Finally, we explain our research plan which will dive into the sub goals required to complete the main objective of building a silent-speech recognition system using Electromyography (EMG) and Deep Learning.**

I. Introduction

Over the past few years, HCI has been an increasing field of study. Human Computer Interaction can be described a as feedback loop between a human and computing interface. With the increased usage of sensors worn on humans, such as watches, heart rate monitors, and other smart sensors, researchers are trying to extract bio-signal information and classify typical human activities. One of the ways HCI is used is for Silent Speech Interfaces (SSI). SSI aims to use signal extracting systems like electromyography (EMG) and electroencephalography (EEG) to convert signals of silent or non-audible speech and use a machine to classify the results. This feedback loop involves feature extraction, model training, and activity inference [1].

This paper is motivated by recent work done that uses machine learning and biosignals such as electrocardiogram (ECG) to detect irregular heartbeats [2], electroencephalography (EEG) [3] and EMG [4], to predict movements associted with the human body. As the field of Deep Learning continues to expand, bio-signals from the human body will continue to be used to help improve the lives of people who have trouble with daily tasks, such as walking and communicating.

For the purpose of this research, non-audible speech can be classified as the inability to verbalize words or sentences through the use of sound in an effective way. SSI systems are not new; they have been proposed in the past. What's new is the computing resources and type of algorithms used to classify speech in SSI systems. In the past, machine learning algorithms such as decision tree, support vector machine, Naïve Bayes, and hidden Markov models were used as classifiers for speech. Though powerful, they require extensive feature extraction from EMG signals. Typically, with traditional feature extraction, only shallow features can be learned from those approaches, leading to undermined performance. Recently, we have witnessed the incremental development of Deep Learning, which alleviates the issue of feature engineering, since the models extract valuable information through several iterations. Therefore, using EMG signals to classify non-verbal speech does not have to a be a laborious task.

The rest of this paper will discuss the research problem we are trying to achieve. The paper is organized as follows. Section II discusses the related work to the problem. Section III discusses the experimental setup used to capture, analyze and model the data. Section IV will discuss the results from our analysis and compare that to other work. Section V will conclude with discussions about our work and review future possibilities.

II. Related Work

The research conducted using EMG to predict speech for SSI systems has been going on for over two decades. Before machine learning became very popular, using EMG to predict speech involved heavy feature extraction of the data, along with discrete mathematical modeling. Along the way, there have been well documented results that have used a combination of mathematical modeling and deep learning to to predict speech using EMG. Some of the research addresses syllable and single word based prediction [5], [6]. Other research has addressed using EMG to predict entire phrases [7], [8].

One of the earliest attempts to use EMG to predict speech was done in [6]. The goal was to predict isolated word recognition, which was performed on a vocabulary consisting of the ten English digits, "zero" to "nine". Seven electrodes were positioned on the face to extract bio-signals from the subjects. Hidden Markov Models (HMM) with Gaussian Mixture Models were used as classifiers. The study was able to get an average word accuracy of 97.3%.

In [9], newer machine learning models were introduced, such as Restricted Boltzmann Machine algorithms. Their

corpus consisted of 25 sessions from 20 speakers comprising of 200 read English-language utterances such, as phones, consonants, and vowels. In the results, the vowel features achieved relatively low ac curacies (around 40%).

The work performed by [7] carried over some of the primary research done by [9]. This research focused on using modern Deep Learning techniques such as Long Short Term Memory (LSTM) and comparing it with GMM. Their results showed that LSTM models performed better than that of GMM, with a mean MCD score of 5.46 versus 5.69, where MCD is a measure of distance, and lower numbers represent better results.

In [8] built a proof of concept SSI system that uses a one-dimensional Convolutional Neural Network (CNN) as a classifier. Seven electrodes where placed around the throat face. In their quantitative results they got an average accuracy over all runs for all the users of 92.01%. Their corpus included individual words, and short phrases.

Finally, [2] uses latest methods to classify ECG signals by converting the bio-signals to scalograms and passing them through CNN models. This method is used to predict irregular heartbeats. Similar techniques can be adopted using this method for EMG bio-signals.

The number of electrodes and bipolar channels are far greater in [8], [9], [7], [6] than compared to our research, where we are only using three channels. This number was based on [6] which stated, EMG based speech processing requires the very least signals from the cheek area and the throat. Our research will attempt to reproduce existingwork for LSTM models and One-Dimensional Constitutional Neural Networks. We will then attempt to experiment with similar techniques presented in [2] for ECG signals, but using it for EMG signals to predict speech.

REFERENCES

[1] L. Masuch, "Deep Learning - The Past, Present and Future of Artificial Intelligen…." [Online]. Available: https://www.slideshare.net/LuMa921/deep-learning-the-past-present-and-future-of-artificial-intelligence?ref=https://www.analyticsindiamag.com/popular-presentations-on-artificial-intelligence-and-machine-learning/

[2] "Classify Time Series Using Wavelet Analysis and Deep Learning - MATLAB & Simulink Example." [Online]. Available: https://www.mathworks.com/help/wavelet/examples/signal-classification-with-wavelet-analysis-and-convolutional-neural-networks.html

[3] A. Eltvik, "Deep Learning for the Classification of EEG Time-Frequency Representations," p. 122.

[4] A. A. Altamirano, "EMG Pattern Prediction for Upper Limb Movements Based on Wavelet and Hilbert-Huang Transform," p. 134.

[5] E. Lopez-Larraz, O. M. Mozos, J. M. Antelis, and J. Minguez, "Syllable-based speech recognition using EMG," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug. 2010, pp. 4699–4702.

[6] L. Maier-Hein, F. Metze, T. Schultz, and A. Waibel, "Session independent non-audible speech recognition using surface electromyography," in *IEEE Workshop on Automatic Speech Recognition and Understanding, 2005.* San Juan, Puerto Rico: IEEE, 2005, pp. 331–336. [Online]. Available: http://ieeexplore.ieee.org/document/1566521/

[7] M. Janke and L. Diener, "EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2375–2385, Dec. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8114359/

[8] A. Kapur, S. Kapur, and P. Maes, "AlterEgo: A Personalized Wearable Silent Speech Interface," in *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*. Tokyo, Japan: ACM Press, 2018, pp. 43–53. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3172944.3172977

[9] M. Wand and T. Schultz, "Pattern learning with deep neural networks in EMG-based speech recognition," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago, IL: IEEE, Aug. 2014, pp. 4200–4203. [Online]. Available: http://ieeexplore.ieee.org/document/6944550/