NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SCSE20159- Speak your Mind: A Study of Decoding Silent Speech from Brain Signals Using Deep Learning

Presented by Saraf Ishita

Supervised by Prof Guan Cuntai & Dr Tushar Chouhan

Motivation

Neurological impairments such as brain strokes, paralysis, epilepsy, and amyotrophic lateral sclerosis (ALS) can result in loss of motor and communication abilities. Sudden loss of verbal communication abilities can make an individual's life devastating. Currently, patients rely on tools that make use of their residual head and eye movements or Brain Computer Interfaces (BCIs) that control cursors to spell out words letter-by-letter. Although greatly helpful, spelling-based approaches with an average of 10 words per minute cannot be compared to the ease of communicating verbally with 150 words per minute. Hence, we aim to cross the hurdle of spelling-based communication to develop a system that can decode speech directly from the patient's neural activity that remains intact even after such tragedies.

BCIs for Silent Speech Decoding

There are two broad methods for recording neural activity for the purpose of learning speech features. The first is **Electrocorticography (ECoG)**, an invasive recording technique, in which subjects go through the process of getting electrodes surgically implanted on the exposed surface of their brain to record brain signals from different parts of the cerebral cortex. The process yields brain signals of high resolution and effectively capture high-frequency gamma waves crucial for speech decoding. However, getting the electrodes surgically implanted makes ECoG a less feasible choice for patients. The second approach is using **Electroencephalography (EEG)**, a non-invasive method that uses electrodes placed on a subject's scalp to record electrogram of neural activities. Although a much more convenient choice for patients, the brain waves being recorded suffer from attenuation from the skull and scalp, resulting in a much poorer signal-to-noise ratio and temporal and spatial resolution. Hence, in order to effectively extract speech from EEG waves, we must process the raw waves using methods such as artifact removal, band-pass filtering, Laplacian filtering, and independent component analysis. After carefully weighing the pros and cons of both methods, we have decided to perform speech decoding on EEG data because of its great feasibility and usefulness for all patients.

Electrocorticography (ECoG)	Electroencephalography (EEG)
Higher temporal and spatial resolution	Lower temporal and spatial resolution due to attenuation by skull and scalp
Typically higher amplitude and contain high frequency bands such as gamma	Typically lower amplitude and do not contain high frequency bands clearly
Source of brainwave in the cortex can be clearly pinpointed	Source of brainwaves recorded is vague

Fig 1: ECoG vs EEG Brain Signals

Detailed experiments have been designed and conducted by several researchers to record brain signals for the purpose of speech classification. In most experiments, multiple electrodes are placed on different positions on the scalp, overlying parts of the brain involved in speech production. A word is displayed in front of the subject and the subject is required to imagine speaking the word, without any actual movement of the vocal and articulatory tract, while their brain waves are recorded through EEG devices. This EEG data, recorded across multiple subjects and multiple trials for each subject comprise the data that will be fed to the deep learning model.

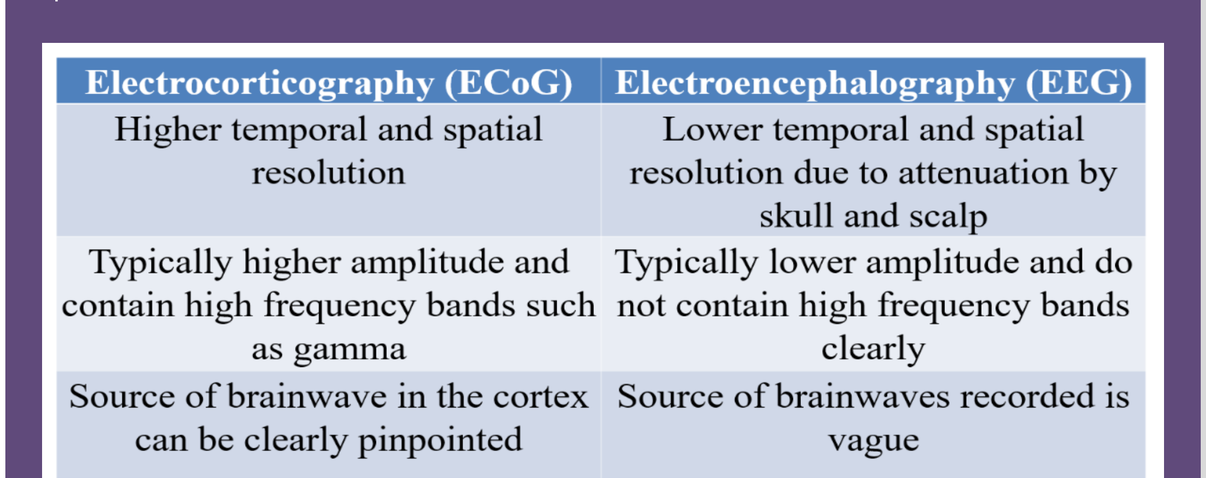

Fig 2: Methodology followed

Literature Review

Deep learning models are based on neural network architectures, designed specific to the learning goal. There are several state-of-the-art models designed by researchers for the task of speech decoding from brain waves. Anumanchipalli et al., 2019 suggested a two-stage decoder, which introduce an intermediate stage of decoding brain signals to articulatory-kinematic features of the vocal tract and then use these features to decode speech audio, as depicted in Fig 3. Others have considered speech decoding as a task of machine translation, exploiting encoder-decoder networks (Makin et al., 2021). After extensive research, we have found a few interesting approaches interesting. Speech production is a complex task involving frequent information exchange between different parts of the brain. Saha et al., 2019 proposed to capture this information exchange by computing the cross-covariance matrix among EEG data from all channels. They use this channel-cross-covariance to train their four-stage hierarchical model. As shown in Fig 4, the first level of the hierarchy consists of a Convolutional Neural Network (CNN) along with a parallel Long Short-Term Memory (LSTM) network. The second layer is an unsupervised deep autoencoder (DAE) followed by an Extreme Gradient Boost classification layer (XG Boost), which performs well on structured data such as EEG, as the last level of the hierarchy.

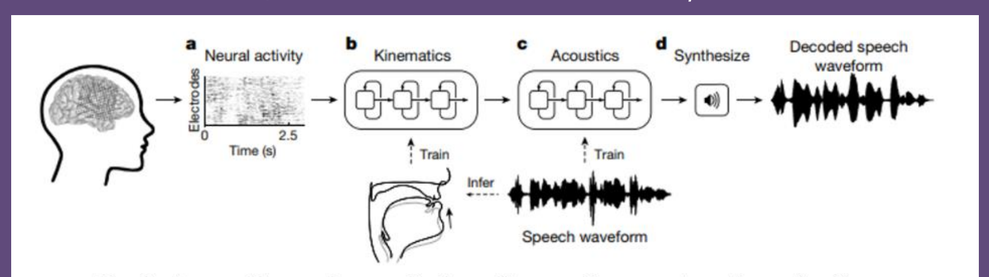

Fig 3: An end to end speech decoding and reconstruction pipeline (Anumanchipalli et al., 2019)

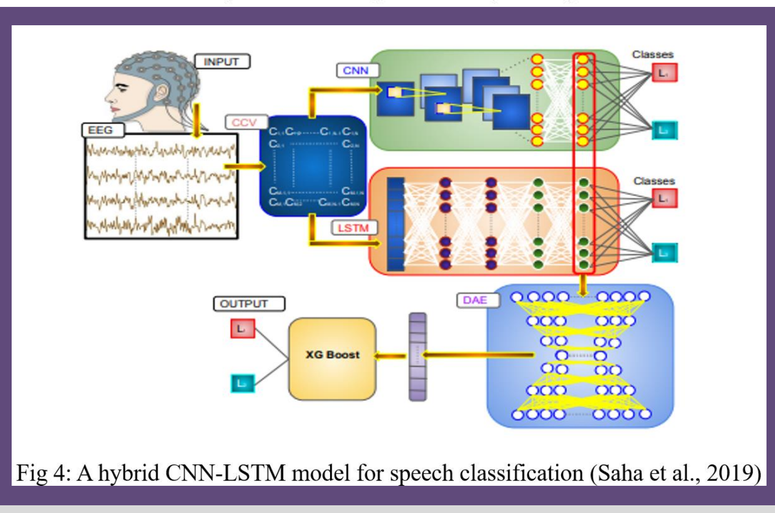

Fig 4: A hybrid CNN-LSTM model for speech classification (Saha et al., 2019)

Remarks and Future Work

The encoder-decoder model proposed by Makin et al., 2021 is not merely a sentence classifier but it is an end-to-end speech decoding model that that learns speech features from brain signals data and maps it to audio (Mel-Frequency Cepstral Coefficients) as well as text (word sequences). Their model predictions have achieved word error rates between 0% and 5% for 3 out of 4 subjects. The authors in Saha et al., 2019 report that their four-stage hierarchical model is found to provide better accuracy (~21%) over other state of the art classifiers. However, this remains to be validated and is part of our on-going efforts. Hence, through our research we aim to create a deep learning model which will obtain a better accuracy for classifying EEG data into words. We have thus reported a few potential approaches to speech decoding for EEG-BCIs. As part of our on-going and future works, we aim to verify and reproduce the results reported in these works which will be a step further in building high performance speech decoders using EEG. We believe that precise categorization of subject-independent EEG data into finite words or phonemes is the fundamental step required to build full-fledged speech decoders that will have great usefulness for patients suffering from loss of verbal communication abilities.

References

1. Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature, 568*(7753), 493–498. https://doi.org/10.1038/s41586-019-1119-1
2. Saha, P., Fels, S., & Abdul-Mageed, M. (2019). Deep learning the EEG manifold for phonological categorization from active thoughts. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2019.8682330
3. Makin, J. G., Moses, D. A., & Chang, E. F. (2021). Speech decoding as machine translation. *SpringerBriefs in Electrical and Computer Engineering*, 23–33. https://doi.org/10.1007/978-3-030-79287-9_3