**Speech Recognition for Neurological disordered patients who can’t pronounce letters clearly**

In market today there are various speech recognition tools that can translate your speech to words & sentences but there is no framework exist that can decipher personalized paralyzed speech to text.

The following tool is made to help such patients recognize their lost voice to speech.

**Approach:**

In order to understand how sound is produced and how speech spectrum can be decompositioned, please refer to this great video:

https://www.youtube.com/watch?v=4\_SH2nfbQZ8

* The frequency produced in the glottal pulse results from the vibration of the vocal cords resonating against the [larynx](https://www.wisegeek.com/what-is-the-larynx.htm).
* Depending how you shape your vocal tract, you’ll get different speech signal from similar Glottal pulses.
* Vocal tract finally produces Spectral envelope of sound that’s nothing but personalized phenoms (vowels/consonants)

Diagram

Description automatically generated

Diagram

Description automatically generated

We can say ->

Speech = Convolution of Vocal tract frequency response with Glottal Pulse

But to identify the identity of speech (like phonemes, timbre) we need the Spectral envelope part, not the Glottal Pulse.

So how can extract Spectral envelope?

MFCCs (Mel Frequency Cepstral Coefficients) comes into rescue. MFCC considers human perception for sensitivity at appropriate frequencies by converting the conventional frequency to **Mel Scale**, and are thus suitable for speech recognition tasks quite well (as they are suitable for understanding humans and the frequency at which humans speak/utter).

Mel Frequency scale provides the clear distinction between vocal tract phoneme responses. The diagram below shows ‘G’ and ‘E’ of a neuro patient with 13 MFCC frequency band energy.

Chart, PowerPoint

Description automatically generated

Data Gathering:

* Record A-Z each letter minimum 100 times. The recording was done in .mp3 format.

Data Preparation:

* Change mp3 to .wav format
* Reduce noise of the sound. Remove sound signal less than 15dB
* Create equal Size chunks audio files for each letter.
* Create multiple files with audio clips of same length.

Analyze:

* Extracted 39 MFCC features from each of the samples. We’ll run CNN modeling on the samples of MFCC.

A picture containing text

Description automatically generated

Graphical user interface, application

Description automatically generated

Training Model:

Feed the training set to CNN network with different hyperparameters.

Deployment:

Implemented a simple single page Flask app with Ajax calls with recording to speech transformation. This app used common Bag Of Words while deciphering the uttered letters of the patient. Probabilistic calculations can be applied towards finding most frequent words.

A picture containing application

Description automatically generated