

Identifying Stuttering using Deep learning



Vedant Tibrewal, Mohammed Mobasserul Haque, Aman Pandey, Manimozhi M

Abstract: Stuttering is a prevalent neurodevelopmental speech disorder, wherein people suffer from disfluencies in speech production. Speech disorders such as stuttering affect a variety of other communication problems such as hearing and fluency. Common therapies of stuttering involve strategies to minimize stuttering but do not attempt to eliminate stuttering, Researchers have analyzed the root cause of stuttering tends to be neurological roots. Therefore, there needs to be a more generic therapy technique which is more adaptive. This paper proposes a deep learning and neural network-based algorithm for adaptive neurological stuttering by utilizing the potential of mirror neurons.

Keywords: ANN, Audio Feedback, MFCC, Signal Processing,, Stuttered speech diagnostics,

Post-processing and deep learning algorithms are used to remove the stuttering of the patients. Here when the person reads the paragraph the audio signal is processed. The algorithm identifies stuttering intervals from the audio signal and removes the stuttered part, it then combines the entire speech signal. When the patients hear speech in his own voice as output, his mirror neurons fires more as compared to the usual computerized method of text to speech. On undergoing this therapy regularly, the algorithm compares his improvement in stuttered speech and graph is plotted, which can be studied by the doctor. With regular sessions, the person stuttering will improve at a faster rate as compared to other techniques.

I. INTRODUCTION

Stammering or stuttering is one of the most common speech disorders. In which the patients suffer from the tense struggle to get words out, which is different from regular hesitation we all suffer while speaking. More than 70 million people worldwide are stutterers, that's one in every 100. In the US, more than 3 million people stutter. Commonly it involves getting stuck (silent blocking) or repeating or prolonging sounds of the words. Scientists have located the root cause for stammering is an error in neurological wiring. The anatomy of patients compared to a normal person is different. Genetics is also one of the main reasons. Another reason for stuttering is a psychological impact. The current therapies focus on helping them learn ways to minimize stuttering when they speak, such as by speaking more slowly, regulating their breathing, or gradually progressing from single-syllable responses to longer words and more complex sentences and these are diagnosed by a speech-language pathologist. We aim to develop an adaptive approach by using the potential of mirror neurons, a cognitive approach towards neurogenic stuttering and prolonged developmental stuttering. We developed a technique that is useful for the diagnostics of stuttered speech using therapy sessions. We initially accept patient parameters: input from the patient, on which letters he suffers from. After the patient verification, it requests the patient to read aloud a paragraph and simultaneously records the patient's speech, speech processing such as Mel Frequency Cepstral Coefficients [MFCC] are performed on the speech signal, to diagnose stuttering and segmentation of stuttered words.

II. EXPERIMENTAL APPROACH

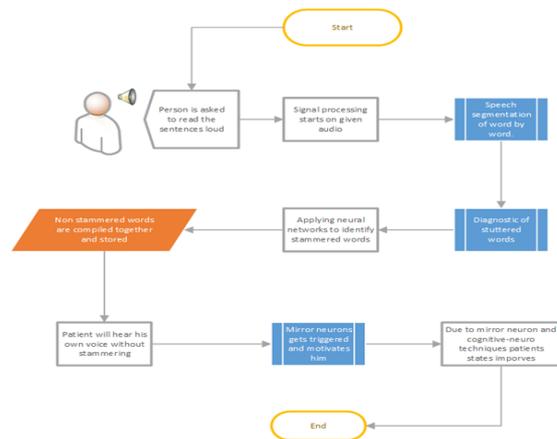


Fig: Flowchart of the process

We have developed an algorithm that is useful for the diagnostics of stuttered speech and therapy sessions. The patient reads aloud a paragraph and simultaneously the algorithm records the patient's speech, speech processing such as MFCC are performed on the speech signal, to diagnose stuttering and segmentation of stuttered words. Post-processing, neural networks are used to remove the stuttering of the patients. The algorithm identifies stuttering intervals from the audio signal and removes the stuttered part, it then combines the entire processed speech signal back and is delivered to the patient. When the patients hear speech in his own voice as output, his mirror neurons fires more as compared to the usual computerized method of text to speech. With regular sessions, the person stuttering will improve at a faster rate as compared to other techniques.

III. DEEP LEARNING MODEL

An Artificial Neural Network Model is developed for both training and test purposes where we have the human categorized dataset for training purposes.

Manuscript published on 30 September 2019.

*Correspondence Author(s)

Vedant Tibrewal Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering

Aman Pandey Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering

Mohammed Mobasserul Haque Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering

M. Manimozhi, Associate Professor in the department of Control and Automation, VIT University.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Identifying Stuttering using Deep learning

The preprocessed signal will act as input to the Keras model. ANN has a sequential model that has an input shape parameter initially, the further layers along the ANN automatically resolves and deal with shape interferences. The input shapes of 13 neurons are considered for the model. The learning curve of the model revolves around three important arguments: optimization, evaluation metrics, and the supervisory loss function.

The Neural Network consists of 13 neurons in its first layer which is densely connected to further layers which have 6,3,2 layer respectively, where activation function is chosen ReLu for each layer. The output layer has single neuron and sigmoid as its activation function, the sigmoid function gives a value between 0 and 1. A threshold value is set which will decide whether the word is stuttered or unshuttered. All the stuttered words will playback.

IV. RESULTS AND ANALYSIS

- 1) We have used signals processing; it is used to reduce the noise. here we filter the data sequence of one dimension.

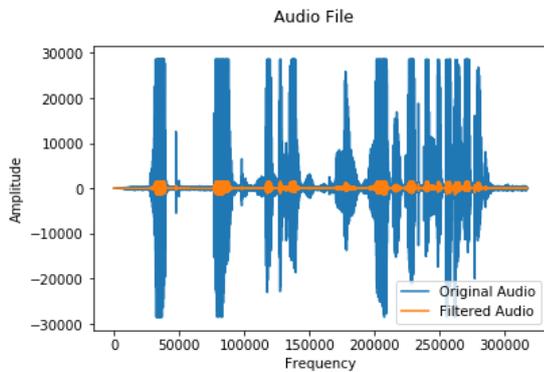


Fig: Amplitude vs Frequency graph for the voice signal

- 2) Next step involves Word Segmentation where words are extracted from the sentences so that they can be processed later. The graph of Amplitude is plotted against frequency for Word file profile.

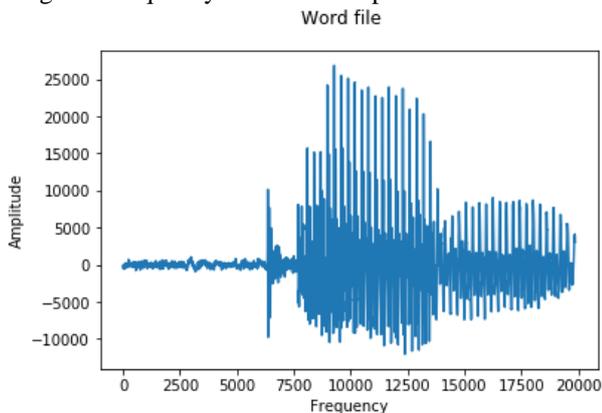


Fig: Amplitude vs Frequency graph for word file

- 3) Each point of the signal is overlapped with other parts of the signal so that the data can be retrieved from any single point. Hence the frames word file profile is plotted.

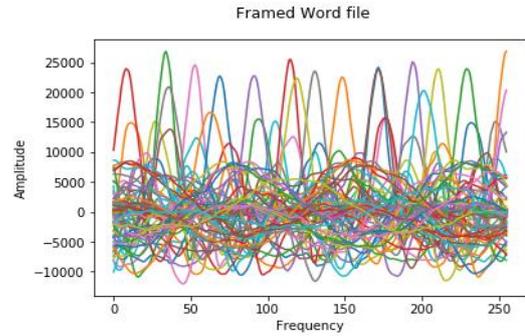


Fig: Amplitude vs Frequency graph for the framed word file

- 4) As after using buffer function noise of the signal is also increased, so to remove the side noises from the buffered signal we used hamming window.

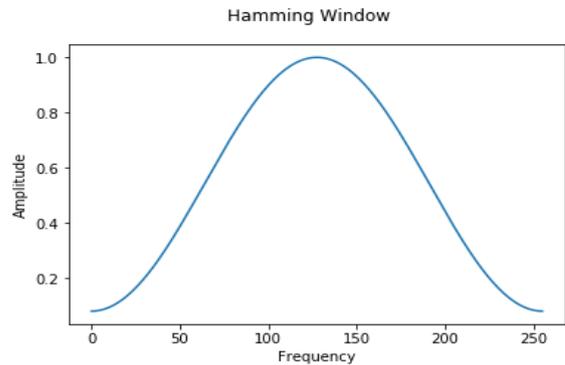


Fig: Amplitude vs Frequency graph for Hamming Window

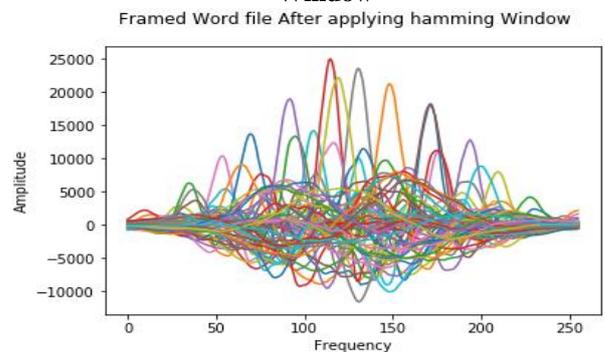


Fig: Amplitude vs Frequency graph for the framed word file after the Hamming window

- 5) Fast Fourier Transform is used for changing from the time domain to the frequency domain.

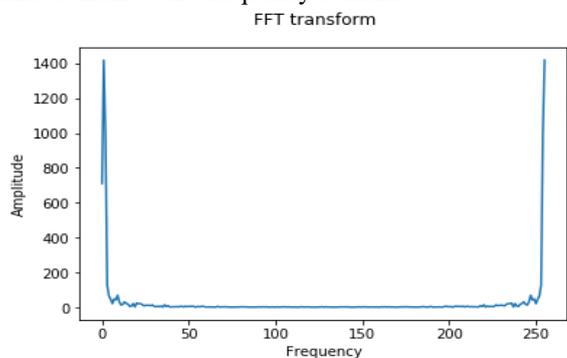


Fig: Amplitude vs Frequency graph for FFT

6) The shape of the vocal tract manifests itself in the envelope of the short-time power spectrum, and the job of MFCCs is to accurately represent this envelope. It is used to extract the unique feature of the audio signal which can be used for processing.

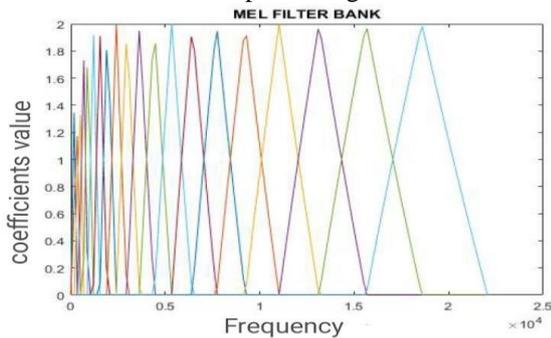


Fig: Coefficient Value vs Frequency

7) MFCC returns 13-26 coefficients value which will be converted into Data frame and will be used as input for the Deep Learning model.

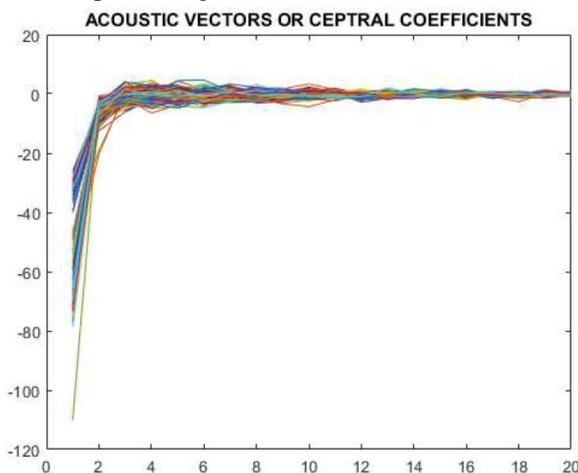


Fig: Cepstral Coefficient graph

14. https://www.researchgate.net/publication/10874232_Speaking_with_a_mirror_Engagement_of_mirror_neurons_via_choral_speech_and_its_derivatives_induces_stuttering_inhibition

AUTHORS PROFILE



Vedant Tibrewal Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering



Aman Pandey Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering



Mohammed Mobasserul Haque Student of Vellore Institute of Technology, 3rd-year Electronics and Communication Engineering



M. Manimozhi is currently an Associate Professor in the department of Control and Automation, VIT University. Her research work focused in the areas of Process Control and Instrumentation, State Estimation, Sensors and Signal Conditioning.

REFERENCES

- <https://www.stammering.org/help-information/topics/what-stammering/basic-information-stammering>
- https://www.researchgate.net/publication/323280189_Lisp_Detection_and_Correction_based_on_Feature_Extraction_and_Random_Forest_Classifier
- Speech Recognition and Correction of a Stuttered Speech :
- <https://ieeexplore.ieee.org/document/8554455>
- Speech Recognition and Correction of a Stuttered Speech: <https://ieeexplore.ieee.org/document/8554455>
- Speech stuttering assessment using sample entropy and Least Square Support Vector Machine: <https://ieeexplore.ieee.org/document/6194726>
- Smartphone support for persons who stutter: <https://ieeexplore.ieee.org/document/6846769>
- MFCC based recognition of repetitions and prolongations in stuttered speech using k-NN and LDA: <https://ieeexplore.ieee.org/document/5443210A> Comparative Study of the Techniques for Feature Extraction and Classification in Stuttering: <https://ieeexplore.ieee.org/document/8473099>
- Delayed Auditory Feedback for speech disorders: <https://ieeexplore.ieee.org/document/6178984>
- Pattern search in disfluent speech: <https://ieeexplore.ieee.org/document/6349744/>
- Mirror neurons as a model for the science and treatment of stuttering: <https://www.ncbi.nlm.nih.gov/pubmed/26565808>
- <http://e-cacd.org/journal/view.php?number=21>
- .Speaking with a mirror: engagement of mirror neurons via choral speech and its derivatives induces stuttering inhibition:

