A Synopsis

On

Real Time Feedback System for Speech Therapy

By

Prathwi Raj Jain 4SO15EC055
Reshveena Melisha Lobo 4SO15EC060
Ashish Kini 4SO15EC016
Shrijith P 4SO15EC080

Under the Guidance of
Ms Jennifer Charlotte Saldanha, B.E, M.Tech
Assistant Professor

ST JOSEPH ENGINEERING COLLEGE
Vamanjoor, Mangaluru - 575028, India
INTRODUCTION

Speech communication is the transfer of information from one person to another via speech. Stuttering and cluttering (also known as stammering) is a speech disorder which makes the person speak in a disordered way where the flow of speech is interrupted by involuntary repetitions and prolongations of sounds, syllables, words, phrases including involuntary silent pauses or blocks where the person with this is unable to produce sounds. The main aim of this project is mainly divided into two parts. First part involves designing a system which implements three therapy techniques used by speech language pathologists to treat patients. The second part involves measuring the improvement of stuttered speech by comparing it with the normal speech.

Keywords: Delayed Auditory Feedback, Frequency Altered Feedback, Metronome, Mel Frequency Cepstral Coefficients, Prolongation, Repetition, Silence Finding.

OBJECTIVES

The objective of the project is to develop a real time device which facilitates the people suffering from stuttering and cluttering to reduce their disorder and improve their speech. The developed device is able to help the people with speech disorder in following ways:

- **Provide delayed auditory feedback**: Take in the speech sample, provide time delay and give feedback to the user.
- **Provide frequency altered feedback**: Take in the speech sample, up/down shift the frequency of the speech and give feedback to the user.
- **Provide a metronome**: A continuous impulse signal is generated at regular intervals according to which the person is made to speak.
- **Measuring the improvement**: The speech of stutterers and that of normal speakers are used to train a system which should be able to measure the improvement during the therapy.

METHODOLOGY

1) THERAPY TECHNIQUES:

The general block diagram of the system is as shown in Fig.1.

1. **Delayed Auditory Feedback (DAF)**: The speech input is given to the processor via a microphone with good quality reception. With the help of a library called SoX, the signal was shifted in the time domain causing a delay. The delay is kept variable which will be different for different people.
2. **Frequency Altered Feedback (FAF):** The noise free signal is Fourier transformed to obtain frequency domain signal. The frequency can be up shifted or down shifted. The output is then inverse Fourier transformed to get back the time domain signal. The time domain audio signal is then sent back to the user as the feedback.

3. **Metronome Generation:** A sine wave of particular frequency is generated for 0.1s and the same sine wave is repeated after specified interval of time. The user is made to utter a word according to the metronome generated.

The three techniques Delayed Auditory Feedback, Frequency Altered Feedback and Metronome were implemented using Raspberry Pi and Audio Processing Library (SoX).

2) **MEASUREMENT OF IMPROVEMENT:**

This project tries to classify normal speech and stuttered speech using the following features:

1) Repetition
2) Prolongation
3) Silence Finding
4) Mel Frequency Cepstral Coefficients

The repetition is found out by finding the number of syllables of the given input speech. If the input speech has more number of syllables than that of normal speech, it is classified as stuttered speech. Prolongation is found out by finding the power of each frame and comparing the power
with the next consecutive frame. If the power is similar (power difference between consecutive frames is less than a threshold value) for a certain number of frames, prolongation can be easily found out. In silence finding, power spectrum is calculated for each frame. Mean and standard deviations of the whole speech is taken using mahalanobis distance algorithm and the number of silence frames present in the speech is calculated. If the number of silence frames are more than a certain threshold, the it is classified as stuttered speech.

**RESULTS**

In the first part of the project, three techniques namely Delayed Auditory Feedback, Frequency Altered Feedback and Metronome using audio processing library called SoX were implemented. The working interface was setup using shell script and Tkinter - Python based GUI.

![Image](image1.png)

**Fig.4. Tkinter - Python based GUI**

![Image](image2.png)

**Fig.4. Prolongation Graph**

The above figure (Fig. 4) represents the speech signal along with the sawtooth waveform of the prolongation. The sawtooth waveform represents the number of frames the prolongation is
occurring. Prolongation is considered if a minimum of 25 consecutive frames of the speech signal has similar power difference.

In total 7 silence is found by the Mahalanobis distance algorithm as shown in Fig. 5. Each frame is of the length 25 ms. A graph for silence found in the frame is plotted.

![Graph showing silence detection](image)

**Fig.4. Detection of Repetition using Syllable Detection**

**CONCLUSION**

The three techniques namely Delayed auditory feedback, Frequency altered feedback and Metronome were implemented using Raspberry Pi as the processor and audio processing library called SoX. To measure the improvement in the speech, the feature vectors of the speech samples are extracted using MFCC. To further improve the classification, additional features like Prolongation, Repetition, and Silence Finding was implemented.

**FUTURE SCOPE**

This project can further be improved by developing an android application which implements the three therapy techniques i.e. Delayed Auditory Feedback, Frequency Altered Feedback and Metronome Generation. The accuracy can probably further be increased by using additional features alongside MFCC like Delta, Delta - Delta, etc. The features extracted can be combined to obtain a robust system. Other classifiers such as SVM, HMM, ANN can be used to improve the accuracy of the system. All types of stuttered speech samples should be taken to get better results.