International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 31-35

Speech Enhancement of Dysarthric Speech Signal

Anushya Sivanupandian*, Jesu Jennifer D$
School of Electronics Engineering, Vellore Institute of Technology-Vellore, TamilNadu, India
$*School of Electronics Engineering, Vellore Institute of Technology-Vellore, TamilNadu, India*

Abstract
The aim and objective of our project is speech enhancement of Dysarthric speech. Dysarthria is a condition in which the person is unable to speak clearly and audibly because of weakness in the muscles controlling the speech of a person. Dysarthria often is characterized by slurred or slow speech that can be difficult to understand. For such people affected by this, communication is very difficult. Therefore we aimed at a creating a method by which the certain parts of the speech becomes clearer that would help them in their day to day communications

Keywords: *Dysarthric, Slow speech, Speech Improvement, MATLAB, Digital signal Processing, Linear Predictive Coding, The Speech Spectrum, Formant Analysis Technique, Audio signal input, Formant Frequency, Praat-Vocal toolkit*

Date of Submission: 28-04-2022 Date of acceptance: 09-05-2022

I. INTRODUCTION

Dysarthria is a neuro-motor disorder in which the muscles used for speech production and articulation are severely affected. Dysarthric patients are characterized by slow or slurred speech that is difficult to understand. This work aims at enhancing the intelligibility of dysarthric speech by developing an effective speech therapy tool. Feature level transformation techniques based on linear predictive coding (LPC) coefficient mapping and frequency warping of LPC poles are done here. Speech utterances from Nemours dataset with mild and moderate dysarthria are used to study the effectiveness of the proposed algorithms. The quality of the transformed speech is evaluated using subjective and objective measures. A significant improvement in the intelligibility of speech was observed [7].

Our method henceforth could be used to enhance the effectiveness of speech therapy, by encouraging the Dysarthric patients to talk more, thus helping in their fast rehabilitation. Since much research is going on on this topic, we decided to work on this project. One of the methods to increase the intelligibility of the speech signal is to increase the energy of the signal. This makes the speech clearer and more audible. The best way to increase the energy of the signal is to increase the energy of the vowel part of the signal. This is implemented by Linear predictive coding (LPC).

1.1.1 Method

We followed the following steps for increasing the intelligibility of the Dysarthric speech.

- The Dysarthric audio signal was collected.
- The audio signal was divided into 2 parts- the voiced parts and unvoiced parts to remove any gaps or spaces or pauses from the signal.
- The spectral analysis of the voiced signal was performed.
- The signal is passed through a filter so that frequencies close to the normal frequency ranges in a voice signal are retained.

1.1.2 Formant Analysis Technique and Linear Predictive Coding

We then performed the energy analysis on this signal using the formant analysis technique. This technique involves the division of the audio signal into frames and then measuring the pitch frequency f0 and the side frequencies f1 and f2. The pitch frequencies and the formant frequencies are then further used in LPC(Linear Predictive Coding). LPC is an in-built MATLAB command that helps to identify the frequencies of a particular vowel and vowel-like sound. The frequencies so gained help us increase the energy of the vowels part of the speech. After this again the signal is passed on through a high pass filter and then the two signals – the vowel signal is imposed or overlapped on the previous signal. Thus by this method, the intelligibility of the speech signal increases by around 35-40% [2].

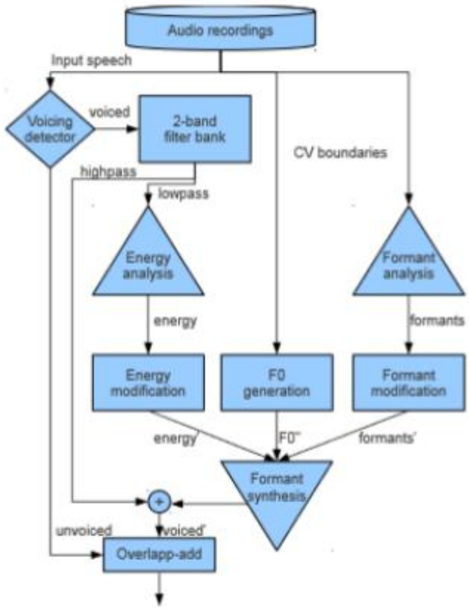


Figure1: Modification system model Layout [1]

1.2 The Speech Spectrum

Formant Analysis Speech and singing contain a mixture of voiced and unvoiced sounds. Voiced sounds are associated with the vowel portions of words, while unvoiced sounds are produced when uttering consonants like "s." The spectrum of a voiced sound contains characteristic resonant peaks called formants, and are the result of frequency shaping produced by the vocal tract (mouth as well as nasal passage), a complex time-varying resonant cavity. Speeches by different persons are distinguished by the short-term spectral characteristics and temporal variations of the spectral characteristics. This frequency is known as the fundamental frequency or pitch. In speech science and phonetics, a formant is also used to mean an acoustic resonance of the human vocal tract. Linear prediction-based spectrum calculation is used for formant extraction.

Formants are local maxims in the speech spectrum. For a normal speech, about five formant frequencies can be extracted. Mainly F1 and F2 frequencies represent the variations in the pronunciation of phones. These frequencies are used to analyze the changes in the articulation problems of the speech. These two frequencies are needed for disambiguating vowels. These two formants determine the quality of vowels in terms of the open/close and front/back dimensions. During the speech, the phonemes the formant frequency value changes. The time course of these changes in vowel formant frequencies is referred to as formant transitions. The frequency spectrum of normal speech has a more formal transition spectral slope due to frequent formant transitions. The Dysarthric speech spectrum has negligible formant transition frequency resulting in a small spectral slope [6].

1.2.2 Formant Trajectory Refinement system

Dysarthric speech enhancement using a formant trajectory refinement system initially extracts the voiced speech and unvoiced speech separately from the Dysarthric speech data using ’ the Praat-Vocal toolkit’. Voiced speech in this context is the data with information i.e. vowels along with consonants. Unvoiced speech refers to fillers, noise, gaps, etc. in a speech.

The voiced speech is used for the extraction of formant frequency. The LPC filter of the Praat tool is used for extracting the formant values from the extracted voiced data as per the burg algorithm. The extracted format values are stored as object text files. The F1 and F2 formant frequency values alone of each frame of the object text file are extracted separately using MATLAB. These extracted frequencies from each frame are then filtered using a 4-order high pass filter, thereby inducing slope. The changed frequency values after inducing slope are reflected on the object text files. These formant frequencies are then converted to LPC. After that inverse filtering is applied to the Formant Frequencies for normal frequencies [3].

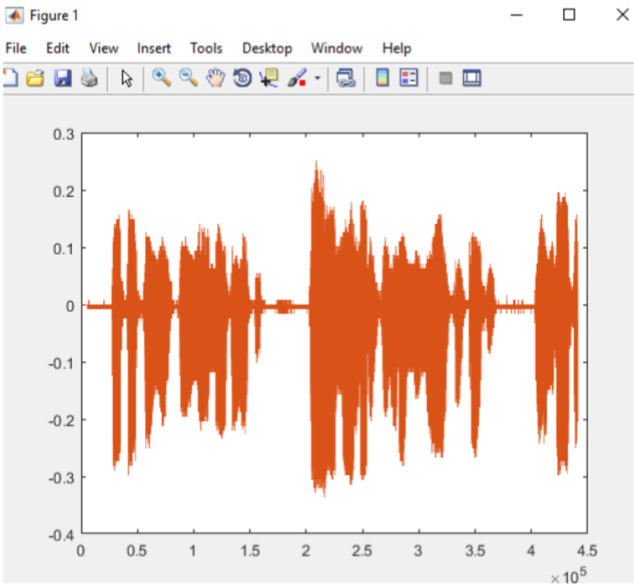
Phonetic Symbol	Example Word	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
/ow/	bought	570	840	2410
/oo/	boot	300	870	2240
/u/	foot	440	1020	2240
/a/	hot	730	1090	2440
/uh/	but	520	1190	2390
/er/	bird	490	1350	1690
/ae/	bat	660	1720	2410
/e/	bet	530	1840	2480
/i/	bit	390	1990	2550
/iy/	beet	270	2290	3010

Figure 2: Formant Frequencies for normal frequencies [4]

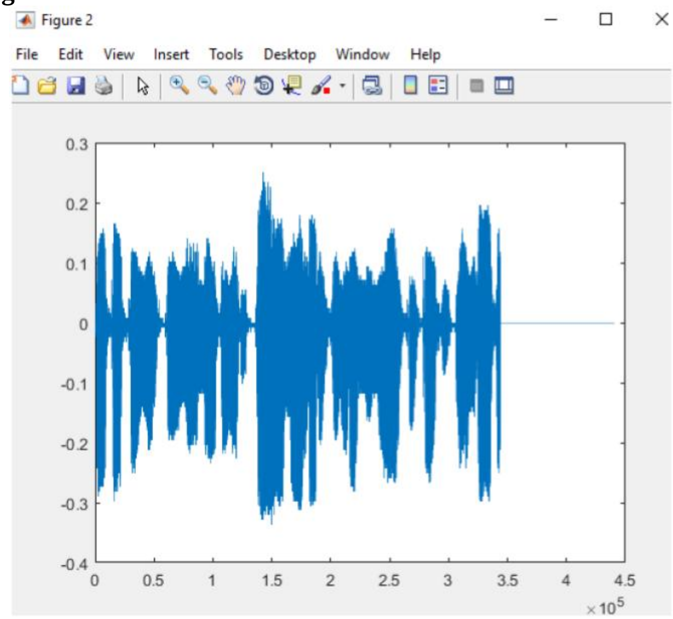
II. THE SIMULATION RESULT AND DISCUSSION

The results obtained are as discussed below

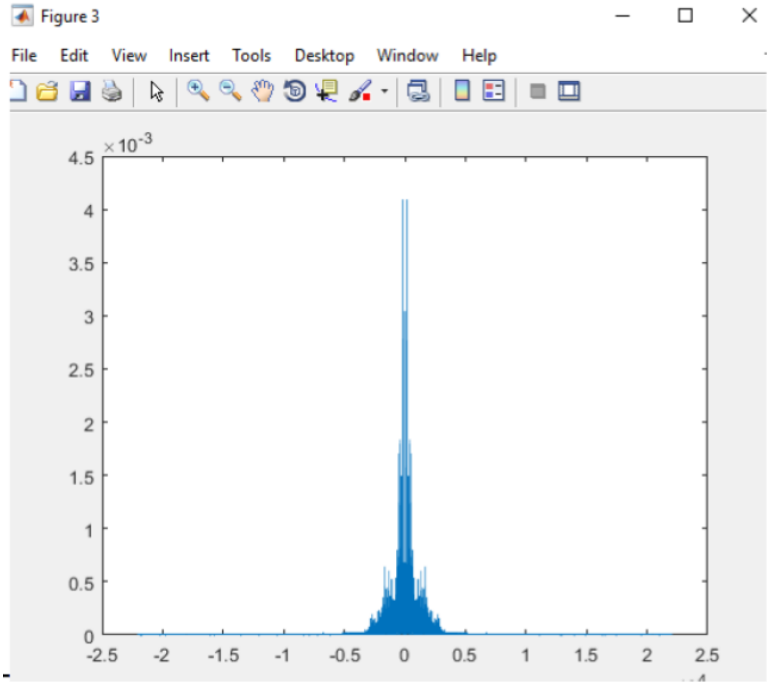
1.3.1 Audio signal input



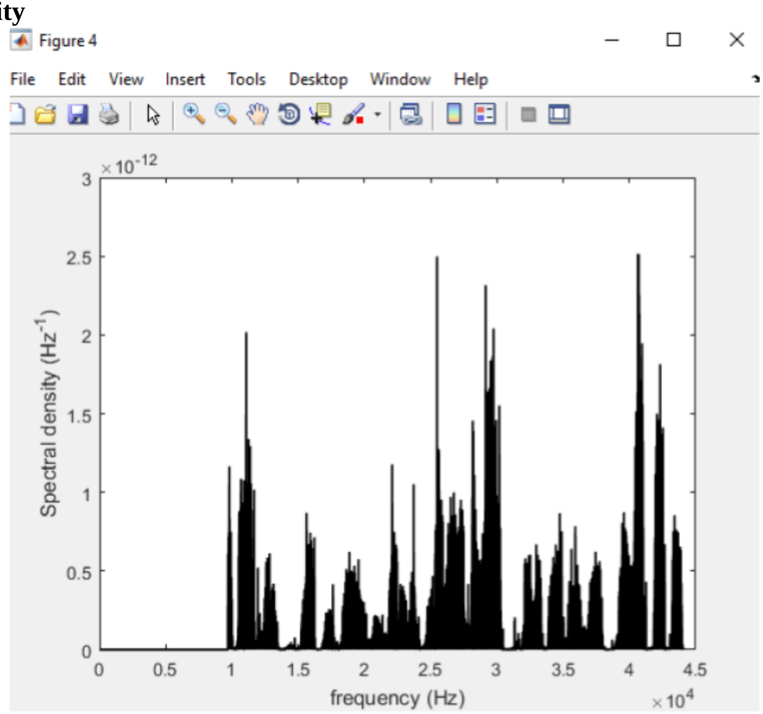
1.3.2 Voiced signal



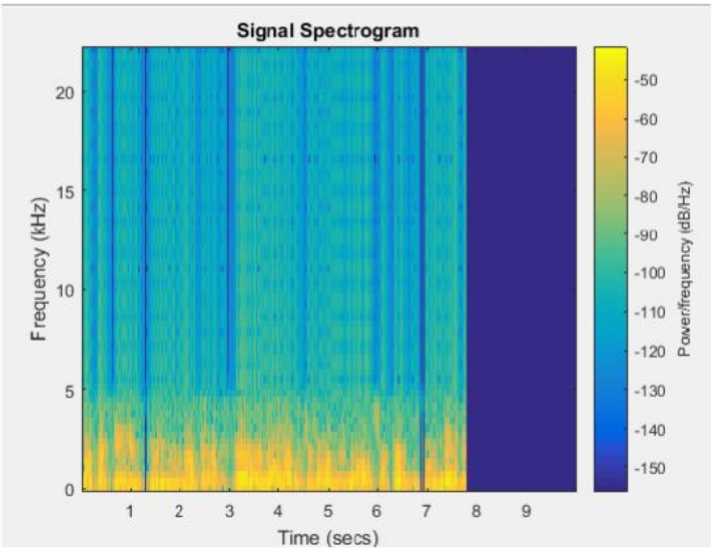
1.3.3 FFT



1.3.4 Spectral density



1.3.5 Spectrogram



III. CONCLUSION

In this project, LPC mapping and frequency warping of LPC poles are used to enhance the intelligibility of Dysarthric speech toward building a better and more effective therapy tool. This methodology is mainly focused on providing enhanced auditory feedback with a delay to assist Dysarthric patients to improve progressively leading to fast rehabilitation. The quality of the enhanced speech is evaluated using subjective and objective measures. A significant intelligibility improvement of Dysarthric speech with a mild and moderate disability was achieved using the proposed algorithm

REFERENCES

[1]. Xavier Menkndez-Pidal, James B. Poliko, Shirley M. Peters, Jennie E. Leonzio, H. T.Bunnel1 (1962–1965) "The Nemours Database of Dysarthric Speech" in Proceedings of Fourth International Conference on Spoken Language Processing. ICSLP'96. IEEE, 1996, vol. 3.

[2]. John-Paul Hosom, Alexander B. Kain, Taniya Mishra, Jan I H. van Santen, Melanie Fried-Oken, and Janice Staehely (April 2003) "Intelligibility of modifications to dysarthric speech" in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)", volume 1, pp 924-927.

[3]. Hesham Tolba and Ahmed S. EL Torgoman "Towards the Improvement of Automatic Recognition of Dysarthric Speech" in Iwanami Lecture Series: Language Sciences, Iwanami Shoten Publishers,1997.

[4]. OpenstaxCnx.,"Formant(Vowel)Synthesis," https://cnx.org/contents/nTOPtfsM@2.2:e8NdvtVS@2/Formant-Vowel-Synthesis.

[5]. Frank Rudzicz. (2011) "Acoustic transformations to improve the intelligibility of dysarthric speech. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies"*, pp 11–21, Edinburgh, Scotland, UK. Association for Computational Linguistics.

[6]. Sid-Ahmed Selouani, Mohammed Sidi Yakoub, and Douglas O'Shaughnessy, 'Alternative speech communication system for persons with severe speech disorders', EURASIP Journal on Advances in Signal Processing, 2009, p.6.

[7]. P.G Blanchet and P.R Hoffman, 'Factors Influencing the Effects of Delayed Auditory Feedback on Dysarthric Speech Associated with Parkinsons Disease', Journal of Communication Disorders, Deaf Studies Hearing Aids, Vol. 2, 2014.