International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Prosodic Analysis and Enhancement of Dysarthric Speech

S. Ezhilkathir, Sheena Christabel Pravin

Abstract— ***Dysarthria is a disorder that is caused in the nervous system. It is caused by damage in some parts of the brain such as cerebellum. Because of the damage in brain it causes weakness in muscles used for speech therefore it happens as in mumbled, slurred or slow speech that human and the machine find it difficult to understand such slurred speech. The Automatic Speech Recognizers which were designed for speech intelligibility perform poorly on dysarthric speech recognition. This paper focuses on the transformation of voice for dysarthria to enhance its intelligibility Formant tracking, pitch and energy estimation with durational cues from dysarthric speech facilitate the modification of these trajectories to more closely approximate the desired intelligible target speech. The transformation of the speech is done by using formant re-synthesis, pitch change and duration morphing. The results of such transformation results indicate that the transformation of the pitch and duration step enhances the intelligibility of dysarthric speech and make it easy to understand for humans and machines.***

Index Terms—***Dysarthria, intelligibility-enhanced speech, Formant tracking, pitch and energy estimation, durational cues, formant synthesis.***

I. INTRODUCTION

Dysarthria is a speech disorder caused due to brain damage embodying itself a weakness of muscles controlled by the brain and nervous system which are used in the progress of production of speech, known as lips, jaw and throat .Patients with dysarthria produce harsh and breathy speech with abnormal prosodic patterns, such as very low speech rate or flat intonation, which makes their speech unnatural and difficult to comprehend. Damage to the nervous system is the main cause of dysarthria. It can happen as an effect of multiple possible neurological disorders such as cerebral palsy, brain stroke, dementia or brain cyst. Research into automatic recognition of dysarthric speech has reserved more concern due to the increasing fame and prospect of speech inputs, peculiarly due to voice-established communication is accessible for persons suffering with neuro-motor troubles as correlated to inputs of the keypad. Various approaches are engaged to advance the improvement of speech recognition achievement of the dysarthric speech. Such kind of techniques are the audile space enhancement, feature extraction, speaker adaptation, rhetorical model adjustment for an individual or as a whole solution.

Formant re-synthesis is done by modifications in the formant path and efficiency, for the dysarthric speech vowels showed more improvement in the evaluation of intelligibility of Consonant-Vowel-Consonant (CVC) utterances **(Rudzicz 2012)**. Audile space enhancement accomplished through sensual and frequency morphing enhanced automatic speech recognition of dysarthria along with biased assessment. The Mel-scale carrying form is large, which indicates that the Mel-filter is equivalent to an unclear system **(Min et.al 2015).** It can be detected that sensual transformation established on dysarthria harshness level enhanced the performance of automatic speech recognizer for dysarthric speech recognition at each level of hardness. Improvement can be achieved altering the vowels of a orator with dysarthria to closely test the vowel space of a healthy orator **(Kain et.al 2007).**

This paper characterizes examinations in which pitch morphing, tempo morphing and formant alterations are used to make the signals of dysarthric speech more identical to the speech from that of the natural people. The available standard speech databases for the dysarthric speech are of different types. There are UA, Nemours and TORGO. We have chosen the TORGO database that comprises of audile and articulatory input of words that are non-words, brief words, and entire terms. The paper is divided as the following. Section 2 describes the prosodic examination of dysarthric speech and the methodology used to extract and enhance speech characteristics for dysarthric speech realization and formant re-synthesis. Section 3 discusses the outputs and development. The Section 4 presents the conclusion and future work.

II.PROSODIC ANALYSIS

2.1 Pitch and Duration Analysis

Literature on comprehensibility enhancement of dysarthric speech show more focus on developing assistive devices to help people with communicative disabilities. Widely used techniques [4] include the use a speech transformation process between normative and dysarthric speech utterances for developing the comprehensibility of dysarthric speech. Various feature level transformation techniques including pitch, duration and formant morphing and signal processing algorithms have been found to increase the comprehensibility of dysarthric speech.

Manuscript published on November 30, 2019.
* Correspondence Author
S. Ezhilkathir*, ECE- Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.(Email: kathir.ezhil14@gmaill.com)
Sheena Christabel Pravin, Assistant Professor, ECE- Rajalakshmi Engineering College, Chennai, Tami Nadu, India.(Email: sheena.s@rajalakshmi.edu.in)

2.1.1 Tempo morphing

In tempo morphing the modification was performed on time domain. Sonorants in dysarthric speech are quite long and those phonemes are contracted in time to intensify the comprehensibility of the dysarthric speech [3]. Thus, the dysarthric utterances are decreased in time domain to make the dysarthric speech intelligible for the listeners. In order to preserve the pitch and frequency characteristics in this transformation, phase vocoder was used.

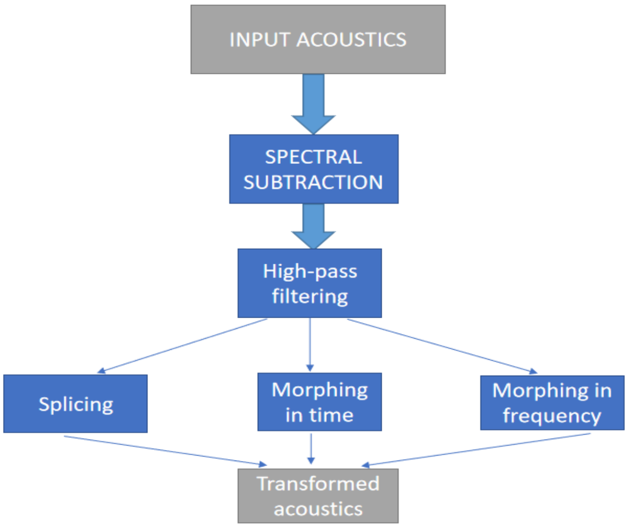


Figure 1. Schematic architecture to transform dysarthric speech to intelligible speech

2.1.2 Frequency morphing

In frequency morphing, the formant trajectories of the dysarthric speech were replaced with values identified from the known vowel segment [3], obtained from the database containing healthy speakers' speech data. By using this, the frequency of the speech is changed to another frequency. Here two speakers are told to say the same sentence and the frequency of the speech is changed such that it is changed from fundamental frequency of one speaker 1 to the fundamental frequency of another speaker. This same way is used for enhancing the dysarthric speech by changing the frequency of the dysarthric speech to a frequency of the healthy speech by changing the formants.

2.2 Formant Analysis

The formant arrangement of a vowel is an essential acoustical associate of vowel integrity. The first formant (F1) is conversely analogous to the impression of elevation of the vowel, and the difference between the first formant (F1) and the second (F2) formant is relevant to the impression of vowel. Formant study may benefit us to figure out the acoustical and articulatory sense the damage of the comprehensibility of vowel resonation in dysarthric speech. By matching the formant design of vowels in dysarthric speech with that in natural speech, we also await to encounter a method to better the comprehensibility of dysarthric speech. Here the comparison is done between F1 and F2 formants. This is partly because F1 and F2 are intuitively far more meaningful than other formants partially because the formant detector we used could contribute more predictable F1 and F2 paths than other formant paths.

To pact with the hard level of the assignment, we narrowed ourselves to researching on the context of the consonant-vowel-consonant. The key concept of our way is to enhance the comprehensibility by testing, altering, and synthesis of a limited set of intuitive-related speech characteristics. The alteration step dwells of arranging the dysarthric features against known healthy speech features by means of a competent feature alteration. The number of training parameters used in the formant transformation must be kept limited in order to allow practicing of the alteration purpose with a relatively limited extent of training inputs. Because of these logics, the expected speech features and synthesis approach have found to produce comprehensible and manageable speech from a compressed form.

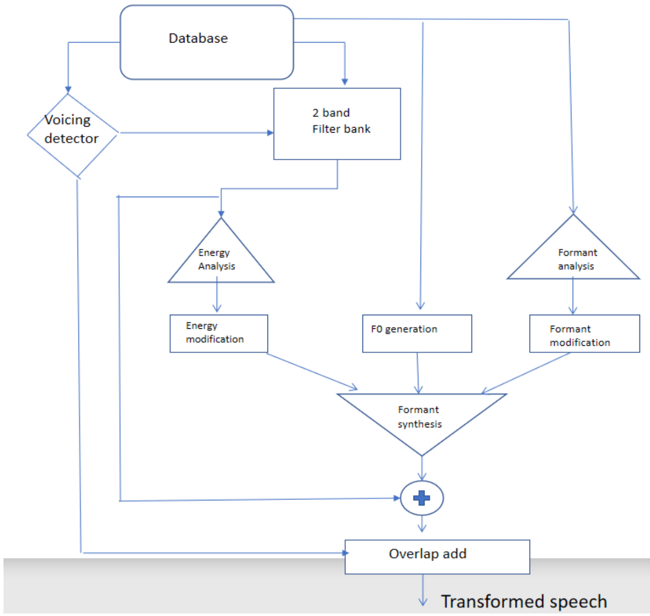


Figure 2. Flowchart of the modification system

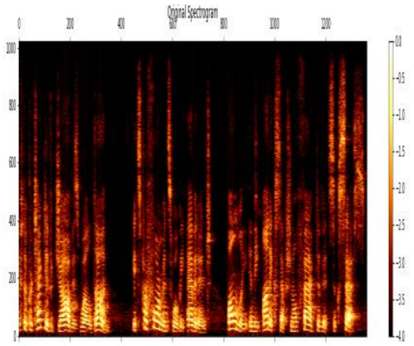
The intake of the dysarthric speech waveform is forced to various tests that condenses the energy, F0 and formant repot. For voiced zones, these features are finally altered or, as is in the crisis for F0, fully reconstructed. The transformed features are then used as data to a formant synthesizer. Finally, the initial unvoiced zones and the transformed voiced zones are put together into an output speech waveform. Observe that the scheme does not cause any modifies to phoneme continuation. Further, no experiments are made to use supplementary, more traditional transformations such as compression and enlargement of the speech passage or amplifications of certain frequency regions.

III.RESULTS AND DISCUSSION

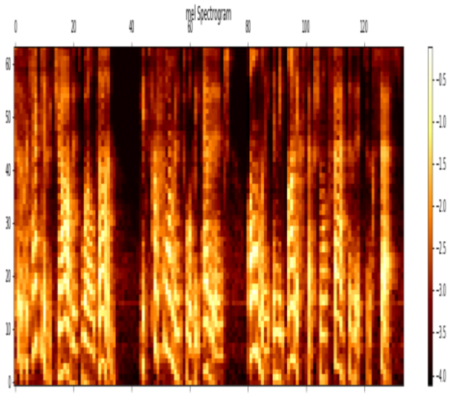
The spectrogram is created using Mel-scaled filter Figure 3 shows morphing of pitch from one waveform to another on a word-by-word level in increments of 33% (33%, 66%, 100%). Note that the morph adjusts the temporal dimension of the target signal to fit the duration of the source signal (the source and generated contours are equally shorter than the target contour). This is reflected on the complete speech utterance unless and until the user specifies an equal number of segments to align

Retrieval Number: D5264118419/2019©BEIESP
DOI:10.35940/ijrte.D5264.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



(a)



(b)

Fig. 3. Outputs for Spectrograms of (a) the original waveform and (b) the Mel frequency-modified spectrogram

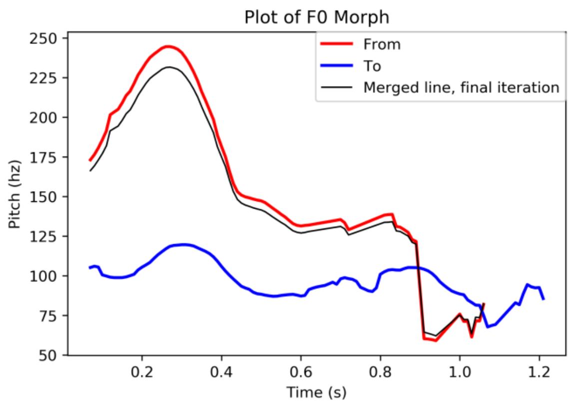


Figure 4. Pitch morphing

The deviation of the morphed pitch from the original pitch is presented in Figure 4. Thus, the pitch inconsistencies in dysarthric speech is rectified by pitch morphing. Figure 5 shows morphing of duration from one speaker(female) to another speaker(male) on a word-by-word basis in increments of 33% (33%, 66%, 100%). This process can operate over an entire file or, similar to pitch morphing, with annotated segments.

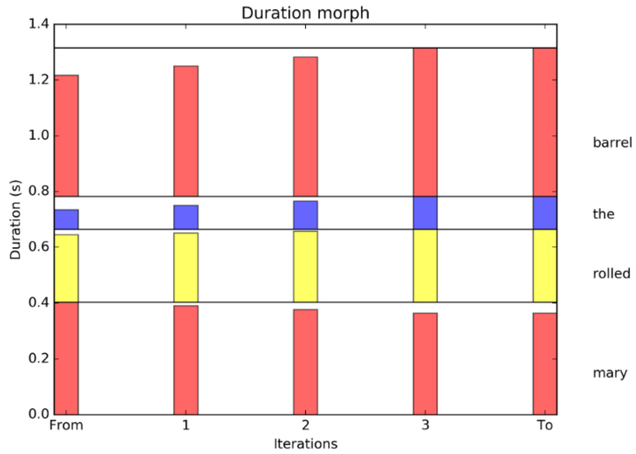


Figure 5. Duration morph

By using the modified formant contours, flattened pitch and energy outline, speech is synthesized using a formant vocoder. The change in frequency and the change in bandwidth is employed by portioning the voiced and unvoiced speech. Synthesized speech is constructed by linking the pieces of documented speech that are stocked in a database. A re-synthesized speech utterance is presented in Figure 6.

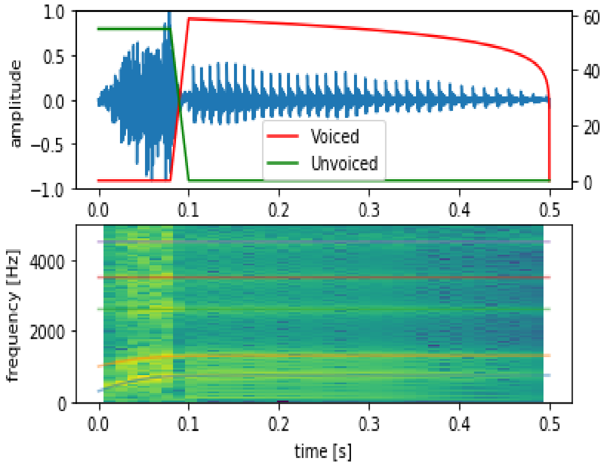


Figure 6. Formant re-synthesis

IV. CONCLUSION AND FUTURE WORK

In this work, various feature transformation techniques used to enhance the intelligibility of dysarthric speech were explored and discussed. Majority of techniques in literature enhances the quality of the dysarthric speech through modification in the phonemes/vowels, temporal, prosodic and frequency characteristics of speech. There is a significant improvement in intelligibility scores, when more than one transformation is applied on the dysarthric speech.

Our future work will focus towards implementing deep learning generative models like the Variational AutoEncoders (VAEs) to enhance the intelligibility of dysarthric speech.

REFERENCES

1. M. J. Kim, J. Yoo, and H. Kim, "Dysarthric speech recognition us-ing dysarthria-severity-dependent and speaker-adaptive models." in Proc. Interspeech, 2013, pp. 3622–3626.
2. H. Christensen, I. Casanueva, S. Cunningham, P. Green, and T.Hain, "Automatic selection of speakers for improved acoustic modelling: recognition of disordered speech with sparse data," in IEEE Spoken Language Technology Workshop (SLT), Dec 2014, pp. 254–259.
3. S. A. Borrie, M. Baese-Berk, K. Van Engen, and T. Bent, "A relationship between processing speech in noise and dysarthric speech," The Journal of Acoustical Society of America, vol. 141, no. 6, pp. 4660–4667, 2017.
4. M. B. Mustafa, S. S. Salim, N. Mohamed, B. Al-Qatab, and C.Siong, "Severity-based adaptation with limited data for ASR to aid dysarthric speakers," PLoS One, 2014.
5. C. Bhat, B. Vachhani, and S. Kopparapu, "Improving recognition of dysarthric speech using severity based tempo adaptation," in SPECOM 2016, Budapest, Hungary, August 23-27, 2016, Pro-ceedings, 2016, pp. 370–377.
6. H. V. Sharma and M. Hasegawa-Johnson, "Acoustic model adaptation using in-domain background models for dysarthric speech recognition," Computer Speech & Language, vol. 27, no. 6, pp. 1147 – 1162, 2013, special Issue on SLPAT.
7. J. L. Ming Tu, Visar Berisha, "Interpretable objective assessment of dysarthric speech based on deep neural networks," in Proc. In-terspeech, 2017, pp. 1849–1853.
8. E. Yılmaz, M. Ganzeboom, C. Cucchiarini, and H. Strik, "Multi-stage dnn training for automatic recognition of dysarthric speech," in Proc. Interspeech, 2017, pp. 2685–2689.
9. S. Selva Nidhyananthan, R. Shantha Selva kumari, and V. Shen-bagalakshmi, "Assessment of dysarthric speech using elman back propagation network (recurrent network) for speech recognition," International Journal of Speech Technology, vol. 19, no. 3, pp. 577–583, Sep 2016.
10. C. Bhat, B. Vachhani, and S. K. Kopparapu, "Recognition of dysarthric speech using voice parameters for speaker adaptation and multi-taper spectral estimation," in Proc. Interspeech, 2016, pp. 228–232.
11. H. Christensen, S. Cunningham, C. Fox, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in Proc. Interspeech, 2012, pp. 1776–1779.
12. T. Nakashika, T. Yoshioka, T. Takiguchi, Y. Ariki, S. Duffner, and C. Garcia, "Dysarthric speech recognition using a convolu-tive bottleneck network," in 2014 12th International Conference on Signal Processing (ICSP), Oct 2014, pp. 505–509.
13. F. Rudzicz, "Adjusting dysarthric speech signals to be more intelligible," Computer Speech & Language, vol. 27, no. 6, pp. 1163 – 1177, 2013, special Issue on SLPAT.
14. F. Rudzicz, "Learning mixed acoustic/articulatory models for disabled speech," in Proc. NIPS, 2010, pp. 70–78.
15. A. B. Kain, J.-P. Hosom, X. Niu, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelli-gibility of dysarthric speech," Speech Commun., vol. 49, no. 9, pp. 743–759, Sep. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.specom.2007.05.001

AUTHORS' PROFILE

Ms.S.Ezhilkathir is currently pursuing Masters in Engineering at Rajalakshmi Engineering College under Anna University, Chennai. She completed her Bachelor in Engineering in the stream of Electronics and Communication. Her research interests are in the field of disordered speech analysis, enhancement and recognition.

Ms. Sheena Christabel Pravin is currently working as Assistant professor in the department of ECE, Rajalakshmi Engineering College Chennai. She completed her Masters' degree in Communication System and Bachelor degree in Electronics and Communication Engineering under Anna University Chennai India. Currently she is pursuing her research in the field of Linguistic and Disfluent Speech processing in Anna University Chennai India. Her areas of interest include pathological speech analysis, Speech Recognition, Enhancement and speech re-synthesis.