Modifying Dysarthric Vowels using Formant Transformation Technique and Vowels Synthesis

[1] Divya V, [2] Veena Karjigi
[1] Dept. of ECE, Siddaganga Institute of Technology, Tumakuru, Karnataka, India
[2] Associate Professor, Dept. of ECE, Siddaganga Institute of Technology, Tumakuru, Karnataka, India

Abstract: Speech is the most preferred way of communication. However, people with communication disorders find it difficult to communicate fluently. Dysarthria is one such disorder where person lacks control over the muscles and articulators. These people may not be able to utter some of the phonemes properly. This decreases speech intelligibility. This paper is based on formants transformation for dysarthric vowels, with the goal of improving intelligibility. Formants are extracted using PRATT software and these formants are modified to more closely approximate the desired targets using transformation function. In order to improve a level of understanding, the modification of formants of dysarthric speaker vowel portion with the known vowel portion formants of normal speaker. After all modifications, the synthesized vowels are far better than dysarthric speaker vowels. Objective evaluation is carried out, and results show that the transformed formants are approximately close to normal speech.

***Index Terms* - Dysarthria, Intelligibility, Improper articulation, Formant transformation, Synthesized speech**

I. INTRODUCTION

The word dysarthria is originated from dys and arthrosis, means difficult or imperfect articulation. Dysarthria is due to nervous system disorder such as strokes, brain injury, and throat or tongue muscle weakness. Impairments of physical production of speech is due to the damage of peripheral or central nervous system. These impairments reduce the normal control of vocal articulators but do not affect the regular comprehension. Damage in the laryngeal nerve reduces control of vocal fold vibration, which results in abnormal voice.

Dysarthria is described as impairment in one or more processes of speech production such as phonation, prosody, articulation. People with dysarthria have problem in controlling the pitch, loudness, and voice qualities of their speech. There are six main types of dysarthria: spastic, ataxic, flaccid, hyperkinetic, hypokinetic, and mixed dysarthria. In all types of dysarthria, phonatory dysfunction is impairment. Features that are related to phonatory dysfunction reduce the speaker intelligibility and even alter the naturalness of speech. The time duration taken by dysarthric speaker is more than that of non-dysarthric speakers. Hence, it is necessary to correct or improve dysarthric speech.

In this regard, the improvement of dysarthric speech can be achieved by following ways. The concatenation algorithm and a grafting technique [1] is used to correct wrongly pronounced phonemes. The grafting algorithm replaces the wrongly pronounced sound units following or preceding the vowel from a normal speaker. The word-level intelligibility of dysarthric speech [2] is improved by performing a short-term spectral level modification. A transformation system [3] is developed to correct pronunciation errors in dysarthric speech by correcting dropped and inserted phoneme errors which is done by tempo morphing, frequency morphing. The improvement of dysarthric speech [4] can be achieved by transforming the features of dysarthric speaker to closely match with that of a normal speaker. An HMM-based speech recognition system [5] is developed by training word models and testing with a sentence level network. The recognized text output serves as an input to a speaker-adaptive synthesis system. The resulting synthesized speech is more intelligible. The recognition rate of dysarthric speech [6] can be improved by modifying the formants, pitch values of dysarthric speaker with that of normal speaker.

The rest of the paper is organized as follows. Section II describes analysis of vowels for normal and dysarthric speech. Section III includes the transformation of formants. Transformation of formant trajectories is discussed in Section IV. Section V includes vowel synthesis. In Section VI, the objective evaluation of formant trajectories is discussed. Finally, the paper is concluded in Section VII.

I. ANALYSIS OF VOWELS FOR NORMAL AND DYSARTHRIC SPEECH

A. Overview

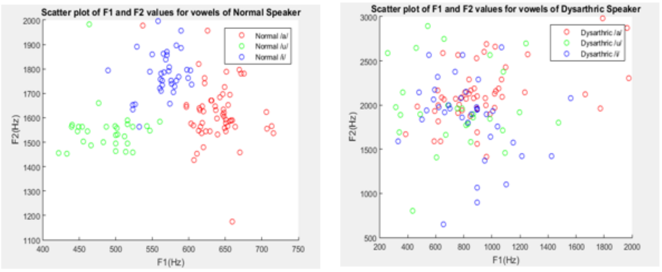
The objective of our work is to improve intelligibility of dysarthric speech by performing analysis, transformation, and synthesis. The transformation step consists of replacing the dysarthric speech features by means of normal speech features by means of a trained transformation function.

B. Database

Database is taken from “Dysarthric speech database for Universal Access Research” [7]. All subjects exhibit symptoms of spastic dysarthria. Along with dysarthric speaker the same utterances of normal speaker are also obtained. The 3 basic vowels /a/, /u/, /i/ are considered from different utterances for analysis.

C. Analysis

Formant trajectories were derived for vowels using the PRATT software from the utterances of UADSR database. Fig. 1 shows the scatter plot of formants F1 and F2 for both normal and dysarthric speaker of 3 vowels /a/, /u/, /i/ with 56, 30, 38 utterances respectively for a single frame.



(a) (b)

Fig. 1. Scatter plot of F_1 and F_2 values for (a) Normal speaker, (b) Dysarthric speaker

In case of normal speaker (Fig 1(a)) the formants of each vowel is well clustered so we can observe three clusters, but in case of dysarthric speaker the points are scattered (Fig 1(b)). Hence it is necessary to modify the dysarthric speech.

III. FORMANTS TRANSFORMATION

The transformation step consists of mapping the dysarthric speech formants to normal speech formants. So a Gaussian mixture model (GMM) is developed using the formants. This model is used in a transformation function to map the dysarthric formants to the formants of normal (target) speaker.

Gaussian mixture is one which is closer to the natural distribution. Hence, GMM is a statistical based model which has become the de facto standard approach in the correctness of the pronounced speech. Gaussian mixtures are a collection of finite mixture of multivariate Gaussian components. The Gaussian mixture uses the expectation maximization algorithm for fitting the mixture of Gaussian models. The multi-variate (dimension-2) Gaussian distribution function is given by,

$$N(x,\mu,\Sigma)=\frac{e^{-0.5(x-\mu)^{T}\Sigma^{-1}(x-\mu)}}{(2\pi)^{\frac{d}{2}}\sqrt{\det(\Sigma)}} \quad (1)$$

where μ is the mean (2 x 1) and Σ represents the covariance matrix (2 x 2) obtained by using formants(F_1 & F_2).

A. Transformation using GMM

Establishing a relationship between normal speech features (y) and dysarthric speech features (x). Formants that are obtained are used as features to build a GMM. The GMM posterior probability [8] for the input vector corresponding to the i^{th} Gaussian mixture is given by,

$$h_i(x)=\frac{\alpha_i N(x,\mu_i,\Sigma_i)}{\sum_{j=1}^{q}\alpha_j N(x,\mu_j,\Sigma_j)} \quad (2)$$

where α_i is the weight of the i^{th} Gaussian mixture and q is the number of mixtures. The mapping function for converting acoustic features of dysarthric speaker (x) to that of normal speaker (y) is given by,

$$F(x)=\sum_{i=1}^{q} h_i(x)[\mu_i^y+\Sigma_i^{yx}\Sigma_i^{xx^{-1}}(x-\mu_i^x)] \quad (3)$$

Mean and covariance is obtained for normal $(\mu_i^y,\Sigma_i^{yy},\alpha_i)$, dysarthric $(\mu_i^x,\Sigma_i^{xx},\alpha_i)$ and normal–dysarthric $(\mu_i^y,\mu_i^x,\Sigma_i^{yx},\alpha_i)$ speaker by training GMM.

IV. TRANSFORMATION OF FORMANT TRAJECTORIES

In this section we presented the results of formants transformation where only one set of formants from the mid of each vowel was considered for analysis. This section presents the transformation of formant trajectories from the three vowels /a/, /u/, /i/. Since the dysarthric speaker utterance is longer compared to that of normal speaker there is a mismatch in time alignment, so it is necessary to make the length of normal and dysarthric speaker utterances equal a dynamic time warping (DTW) algorithm is applied. Here we considered 2 basic formants F_1 and F_2. Since the first 2 formants are important in determining the vowel. The open vowel such as /a/ is having higher formant F_1 frequency and the close vowel such as /i/ or /u/ is having lower frequency. The front vowel such as /i/ is having higher formant F_2 frequency and a back vowel such as /u/ is having lower F_2 frequency.

A. Dynamic Time Warping (DTW)

The Dynamic Time Warping (DTW) is a dynamic programming algorithm that finds the optimal match between two temporal sequences. A similarity measure is performed between two utterance sequences that vary with timing and pronunciation. A warping path is produced by making a similarity between two sequences. These warped signals will be aligned in time. The signal with an original scale X(original), Y(original) is transformed to X(warped), Y(warped). Since the dysarthric speaker utterance is slower than normal speaker the time taken by the dysarthric speaker to utter is lengthy. Hence it is necessary to apply DTW.

1) Formant trajectory for vowel /A/ after applying DTW:

Trajectories of formants F1 and F2 of vowel /a/ from the word man for both normal and dysarthric speaker obtained before and after dynamic time warping are shown in Fig. 3.

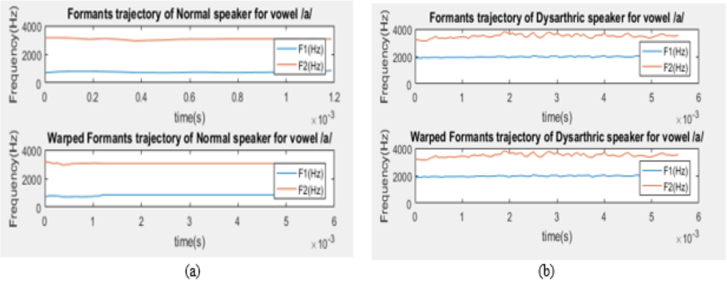


Fig. 3. (a) Formants trajectory of Normal speaker original and after warping, (b) Formants trajectory of Dysarthric speaker original and after warping

2) Formant trajectory for vowel /u/ after applying DTW:

Trajectories of formants F_1 and F_2 of vowel /u/ from the word look for both normal and dysarthric speaker obtained before and after dynamic time warping are shown in Fig. 4.

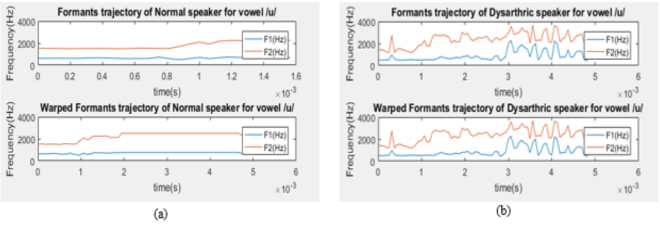


Fig. 4. (a) Formants trajectory of Normal speaker original and after warping, (b) Formants trajectory of Dysarthric speaker original and after warping

3) Formant trajectory for vowel /i/ after applying DTW:

Trajectories of formants F1 and F2 of vowel /i/ from the word kilo for both normal and dysarthric speaker obtained before and after dynamic time warping are shown in Fig. 5. By applying DTW, formants of normal and dysarthric speaker utterances are time aligned.

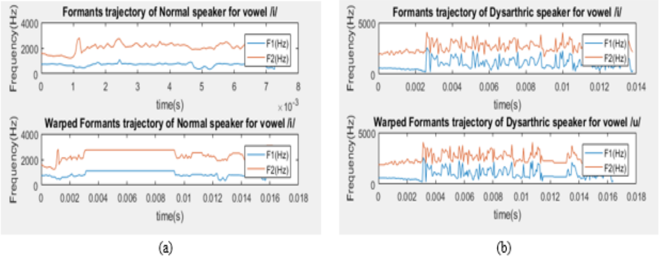


Fig. 5. (a) Formants trajectory of Normal speaker original and after warping, (b) Formants trajectory of Dysarthric speaker original and after warping

B. Formant trajectory transformation using DTW and GMM

DTW is a method that calculates an optimal match between two given sequences. The sequences are "warped" non-linearly in the time domain to determine a similarity measure using DTW and GMM is applied to obtain the model for normal, dysarthric, and normal-dysarthric feature.

1) Formant trajectory for vowel /a/ before and after transformation: The transformation function results of vowel /a/ for first 2 formant trajectory values of both normal, dysarthric and transformed results is shown in Fig. 6. As compared to that of dysarthric speaker formant trajectory the transformation formant trajectory looks similar to that of normal formant trajectory.

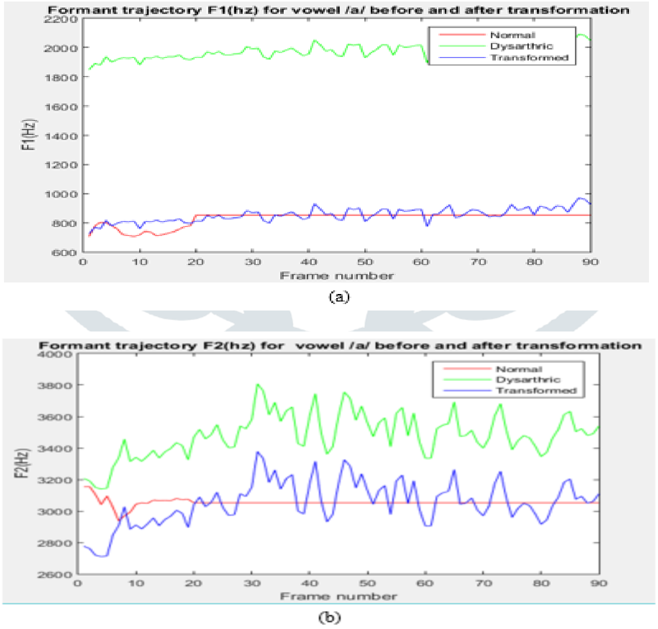
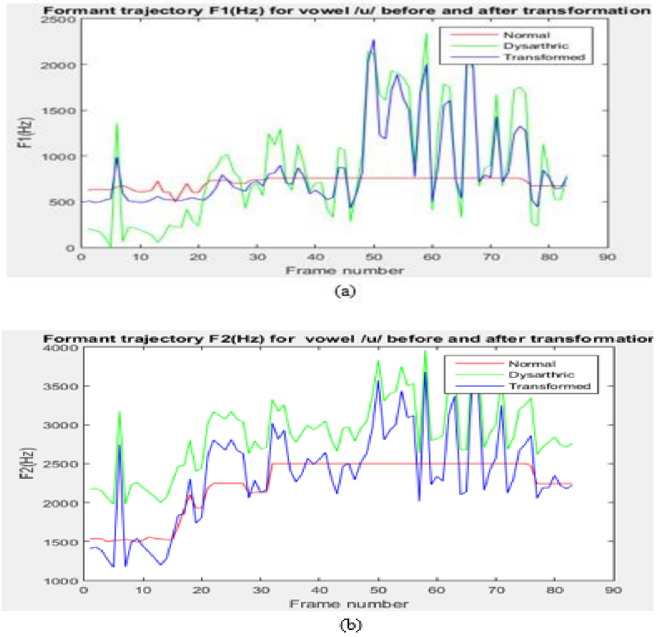


Fig. 6. (a) Comparing Normal and Dysarthric speaker utterance formant trajectory (F1) before and after Transformation for vowel /a/, (b) Comparing Normal and Dysarthric speaker utterances formant trajectory (F2) before and after Transformation for vowel /a/

2) Formant trajectory for vowel /u/ before and after transformation: The transformation function results of vowel /u/ for first 2 formant trajectory values of both normal, dysarthric and transformed results is shown in Fig.7. As compared to that of dysarthric speaker formant trajectory the transformation formant trajectory looks similar to that of normal formant trajectory.

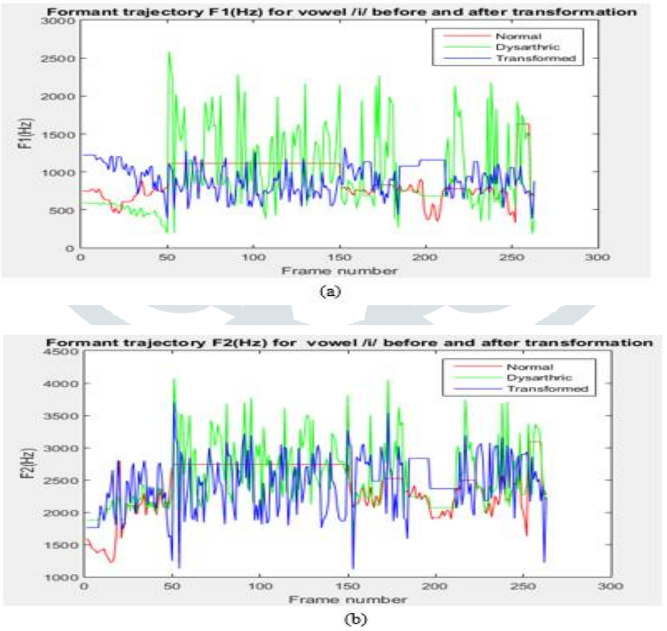


(a)

(b)

Fig. 7. (a) Comparing Normal and Dysarthric speaker utterance formant trajectory (F1) before and after Transformation for vowel /u/, (b) Comparing Normal and Dysarthric speaker utterances formant trajectory (F2) before and after Transformation for vowel /u/

3) Formant trajectory for vowel /i/ before and after transformation: The transformation function results of vowel /i/ for first 2 formant trajectory values of both normal, dysarthric and transformed results are shown in Fig. 8. As compared to that of dysarthric speaker formant trajectory the transformation formant trajectory looks similar to that of normal formant trajectory.



(a)

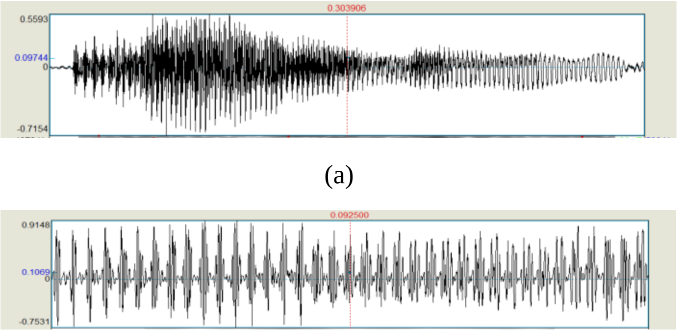
(b)

Fig.8. (a) Comparing Normal and Dysarthric speaker utterance formant trajectory (F1) before and after Transformation for vowel /i/, (b) Comparing Normal and Dysarthric speaker utterances formant trajectory (F2) before and after Transformation for vowel /i/

V. VOWEL SYNTHESIS

A. Feature modification and Generation

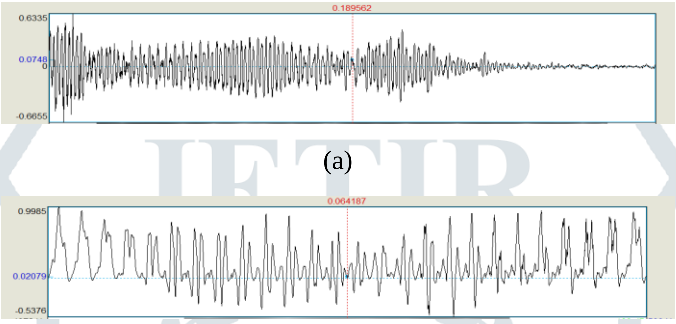
The intelligibility of dysarthric speech is achieved by modifying the vowel region of dysarthric speaker by the vowel region of normal speaker formant values [9]. Since the vowel plays a very important role in understanding. Formants (F1-F3) are extracted from the utterances of normal speaker of UADSR database (speaker CF02) which exhibits the spastic dysarthria for the vowel duration using PRATT software. The pitch (F0) of the dysarthric speaker is taken for that particular vowel duration. The frame length of 25ms and a 50% overlap is considered. By replacing the formant values of dysarthric speaker vowel portion with the formants of normal speaker vowel portion. A new synthesized vowel portion is obtained. This synthesized speech is far better than the dysarthric speaker utterances. The synthesized basic vowels /a/, /u/, /i/ are shown in fig 9, 10, 11 respectively.



(a)

(b)

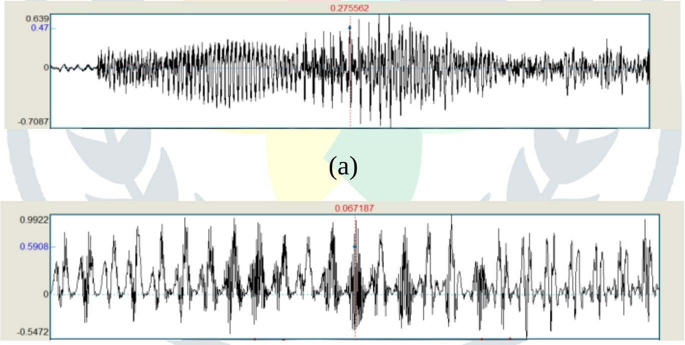
Fig. 9. (a)Waveform of Dysarthric vowel /a/, (b) Waveform of Synthesized vowel /a/



(a)

(b)

Fig. 10. (a)Waveform of Dysarthric vowel /u/, (b) Waveform of Synthesized vowel /u/



(a)

(b)

Fig. 11. (a)Waveform of Dysarthric vowel /i/, (b) Waveform of Synthesized vowel /i/

VI. OBJECTIVE EVALUATION OF FORMANT TRANSFORMATION

The Root-mean-square error (RMSE) is used to find the differences between values that are predicted by a model. The RMS (root mean square) values of formant trajectories between normal and dysarthric and between transformed speech for formants F1 and F2 are evaluated and are shown in Table II and Table III.

Table II: RMSE between formants of normal and original dysarthric speech and normal and dysarthric speech after transformation for formants trajectory F1 (Hz)

vowels	Normal-Dysarthric formants F_1(Hz)	Normal-Transformed formants F_1(Hz)
/a/	1145.13	474.49
/u/	603.92	461.27
/i/	554.02	318.79

Table III: RMSE between formants of normal and original dysarthric speech and normal and dysarthric speech after transformation for formants trajectory F2 (Hz)

vowels	Normal-Dysarthric formants F_2(Hz)	Normal-Transformed formants F_2(Hz)
/a/	452.38	145.86
/u/	719.63	423.55
/i/	596.52	586.02

It can be observed from tables II and III that RMSE value has decreased after transformation.

VI. CONCLUSION

Dysarthria is a disorder that impairs the physical production of speech. For the proper communication with the society, it is necessary to improve the intelligibility of dysarthric speech. The proposed work transforms the formants of dysarthric speech to match the formants of normal speakers which will be helpful in improving the intelligibility of dysarthric speech. Formants of normal and dysarthric speakers are extracted using PRATT software, a Gaussian mixture model is built for transformation. A vowel is synthesized by replacing the formants of dysarthric speaker vowel portion with that of formants of normal speaker vowel portion. This synthesized vowel is far better than the dysarthric speaker vowel. Thus there is an improvement in dysarthric vowels.

REFERENCES

[1] M. S. Yakcoub, S. A. Selouani, and D. O'Shaughnessy, "Speech assistive technology to improve the interaction of dysarthric speakers with machines," communications, control and signal processing, pp. 1150-1154, 2008.

[2] J. P. Hosom, A. B. Kain, T. Mishra, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Intelligibility of modifications to dysarthric speech," Acoustics, Speech and Signal Processing, vol. 1, pp. 924-927, 2003.

[3] F. Rudzicz "Adjusting dysarthric speech signals to be more intelligible," Computer speech and language, vol. 27, pp. 1163-1177, 2013.

[4] A. B. Kain, J. P. Hosom, X. Niu, Jan P. H. van Santen, M. F. Oken and J. Staehely, "Improving the intelligibility of dysarthric speech," vol. 49, pp. 743-759, 2007.

[5] M. Dhanalakshmi and P. Vijayalakshmi, "Intelligibility modification of Dysarthric speech using HMM-based adaptive synthesis system," pp. 1- 5, 2015.

[6] H. Tolba and A.S. Torgoman, "Towards the Improvement of Automatic Recognition of Dysarthric Speech," Computer Science and Information Technology, Beijing, pp. 277-281, 2009.

[7] H. Kim, H. M. Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin and S. frame, "Dysarthric speech database for universal access research," 2008.

[8] A. Kain and M. W. Macon "Spectral voice conversion for text to speech synthesis," Acoustics, Speech and Signal Processing, pp. 841-844, 1998.

[9] A. Kain, X. Niu, J. Hosom, Q. Miao, and J. P. H. van Santen, "Formant re-synthesis of dysarthric speech," in Fifth ISCA ITRW on Speech Synthesis, pp. 25–30, June 2004.