Improvement of Continuous Dysarthric Speech Quality

Anusha Prakash[1]*, M. Ramasubba Reddy*[1]*, Hema A Murthy*[2]

[1]Dept of Applied Mechanics, Indian Institute of Technology Madras, India
[2]Dept of Computer Science & Engineering, Indian Institute of Technology Madras, India
am13s002@smail.iitm.ac.in, hema@cse.iitm.ac.in

Abstract

Dysarthria refers to a group of motor speech disorders as the result of any neurological injury to the speech production system. Dysarthric speech is characterised by poor speech articulation, resulting in degradation in speech quality. Hence, it is important to correct or improve dysarthric speech so as to enable people having dysarthria to communicate better.

The aim of this paper is to improve the quality of continuous speech of several people suffering from dysarthria. Experiments in the current work use two databases- Nemours database and speech data collected from a dysarthric speaker of Indian origin. Durational analysis of dysarthric speech versus normal speech is performed. Based on the analysis, manual modifications are made directly to the speech waveforms and an automatic technique is developed for the same. Evaluation tests indicate an average preference of 78.44% and 67.04% for the manually and automatically altered speech over the original dysarthric speech, thus emphasising the effect of durational modifications on the perception of speech quality. Intelligibility of speech generated by three techniques, namely, proposed automatic modification technique, a formant re-synthesis technique, and an HMM-based adaptive system, is compared.

Index Terms: continuous dysarthric speech, Indian dysarthric speaker, durational modifications, formant re-synthesis, HMM-based adaptive system

1. Introduction

The word dysarthria, originating from *dys* and *arthrosis*, means difficult or imperfect articulation. Speech of a person suffering from dysarthria is affected due to a neurological defect in the speech production system [1]. There is a lack of coordination amongst the various parts involved in speech production to produce understandable speech. Dysarthric speech is characterised by the poor articulation of phonemes, problems with speech rate, incorrect pitch trajectory, swallowing or drooling while speaking. As a result, people with dysarthria have problems with speaking most often. The main aim of the paper is to improve the speech quality of continuous dysarthric speech.

Several efforts on correcting dysarthric speech to make it more intelligible are available in the literature. In [2], dynamic time warping (DTW) is first performed across dysarthric and normal phoneme feature vectors for each utterance, and then a transformation function is determined to correct dysarthric speech. In [3] and [4], the intelligibility of vowels in isolated words spoken by a dysarthric person is improved by formant re-synthesis of transformed formants, smoothened energy and synthetic pitch contours. In [5] and [6], dysarthric speech is improved by correcting pronunciation errors based on given transcriptions and by morphing the waveform in time and frequency. The authors report that the morphing doesn't increase intelligibility of the dysarthric speech. Some corrections are made by using an HMM-based speech recogniser followed by a concatenation algorithm and grafting technique to correct wrongly uttered units [7], or by synthesising speech using HMM-based adaptation [8]. In [9], poorly uttered phonemes are replaced by phonemes from normal speech with discontinuities in short term energy, pitch and formant contours at concatenation points addressed.

The work carried out in this paper focuses on continuous speech and also unstructured text. A durational analysis is carried out across dysarthric and normal speech. Though dysarthria is mostly characterised by slow speech, there are studies reporting rapid rate of speech [1], [10]. Based on the analysis for every dysarthric speaker, manual modifications are made directly to the speech waveforms. An automatic technique is proposed to achieve the same. The effect of these durational modifications on the perceptual quality of speech is studied.

Nemours database [11], a standard database for dysarthric speech, is used in the experiments. Additionally, a dysarthric speech dataset collected from an Indian speaker is also used. Unlike the text in Nemours database, the text in the Indian speech data does not conform to any particular structure. Analysis and modifications are made to speech data of different speakers in the Nemours database and the Indian English dysarthric dataset. Results of subjective evaluation, comparing modified and original dysarthric speech are then presented.

Additionally, two other systems are developed to improve the intelligibility of dysarthric speech. The first is a formant re-synthesis method based on an earlier work [4]. The second is an HMM-based text-to-speech (TTS) synthesis system adapted to the dysarthric person's voice [8]. We assume that a recognition system having 100% recognition accuracy is already available to transcribe speech for synthesis. A word error rate test is conducted to assess the intelligibility of the speech produced by these two systems along with the proposed automatic technique.

The rest of the paper is organised as follows. Section 2 describes the databases used in the experiments. The formant re-synthesis method is described in Section 3 followed by the HMM-based speech synthesis system using adaptation in Section 4. Durational analysis performed on the data along with the proposed modifications are detailed in Section 5. Evaluation results are presented in Section 6. The work is concluded in Section 7.

2. Speech databases used

Standard databases available for dysarthric speech are Universal Access, TORGO and Nemours [11–13]. Universal Access database contains audiovisual isolated word recordings and is hence not suitable for our purpose. TORGO database consists of acoustic and articulatory data of non-words, short words, and complete sentences. However, complete sentences are fewer in number and they account for low phone coverage. Nemours database consists of 74 sentences for each dysarthric speaker. Experiments are therefore performed with Nemours database and Indian English dysarthric speech dataset [1]. The Indian English dysarthric speech data will be referred to as "IE" in this paper.

2.1. Nemours database

Nemours database [11] consists of dysarthric speech data of 11 male North American speakers. The degree of severity of dysarthria varies across speakers: mild (BB, FB, LL, MH), moderate (JF, RK, RL) and severe (BK, BV, SC). The speech data consists of 74 nonsense sentences for each speaker. The sentences follow the same format: "The X is Y'ing the Z", where X and Z are monosyllabic nouns and Y'ing is selected from a set of bisyllabic verbs. Along with the recording of each dysarthric speaker, the corresponding speech by a normal speaker is recorded. The normal speakers are appended with the prefix "JP". Transcriptions are available in terms of Arpabet labels [14].

Phone level segmentation is available for dysarthric speech while word level segmentation is available for normal speech. The procedure to obtain phone level segmentation for normal speech is described in the following section. Pauses within an utterance were already marked for speaker RK in the database but were not available for speakers BK, RL and SC. Hence for these three speakers, pauses were marked manually. Significant intra-utterance pauses are not present in the speech of other dysarthric speakers. For speaker KS, phonemic labeling is not provided. Hence, it is excluded from the experiments.

2.1.1. Segmentation of normal speech data at the phone level

Hidden Markov models (HMM) are used to segment normal speech data at the phone level. Word level boundaries and phone transcriptions for each word are available in the database. HMMs are used to model monophones in the data. Source and system parameters of speech are modeled by these HMMs. The source features are $logf0$ (pitch) values, along with their velocity and acceleration values. The system parameters are mel frequency cepstral coefficients (MFCC), along with their velocity and acceleration values. Instead of embedded training of HMM parameters at the sentence level, embedded re-estimation is restricted to the word boundary. This is inspired by [15], where phone level alignment is obtained from embedded training within syllable boundaries.

HMMs built using Carnegie Mellon University (CMU) corpus [16] were used as initial monophone HMMs instead of using the conventional flat start method to build HMMs, where the models were initialised with global mean and variance. This resulted in better phone boundaries. Data of American speaker referred to as "rms" in CMU corpus was used for this purpose.

[1]The Indian English dysarthric speech data can be found at the link: `www.iitm.ac.in/donlab/website_files/resources/IEDysarthria.zip`

2.2. Indian English dataset

2.2.1. Text selection

The text was chosen from CMU corpus [16]. 73 sentences were selected such that they ensured enough phone coverage. The phoneme transcriptions of the text were obtained from CMU pronunciation dictionary [17] and were later manually corrected when the word pronunciation varied. An additional label "pau" was added to account for pauses or silences.

2.2.2. Speech recording

The speech of an Indian male suffering from cerebral palsy, who is mildly dysarthric, was recorded. The speech was recorded in a low-noise environment and sampled at 16 kHz, with 16 significant bits. The recording was performed over several sessions, each session not exceeding half-an-hour. Frequent breaks were given during the sessions as per the convenience of the speaker so that fatigue didn't affect the quality of speech. About 11 minutes of speech data was collected. Frenchay dysarthria assessment (FDA) [18] was not performed due to unavailability of a speech pathologist.

2.2.3. Segmentation at the phone level

Before segmenting the dysarthric speech data, long silence regions (more than 100 ms) were removed from the speech waveforms by voice activity detection (VAD). 11 minutes of data then reduced to about 8.5 minutes. Segmentation was performed semi-automatically. HMMs were built from already available normal English speech data of an Indian (Malayalam) speaker "IEm" [19], as speaker IE is a native Malayalam speaker. These HMMs were used as initial HMMs to segment dysarthric speech data at the phoneme level. Segmentation was then manually inspected and corrected.

3. Formant re-synthesis technique

In reference [4], the intelligibility of dysarthric vowels in isolated words of CVC type (C- consonant, V- vowel) is improved. Borrowing from this work, a similar approach is adopted to improve intelligibility of continuous dysarthric speech in this paper. Formants F1-F4, pitch and short-term energy values are extracted from dysarthric and normal speech. Frame length of 25 ms and frame shift of 10 ms are considered. Formant transformation from dysarthric space to normal space is only carried out in the vowel regions. For this purpose, utterances segmented at the phone level are required. The transformation makes use of vowel boundaries and vowel identities. Then, formant values at the stable point of the vowels are determined [4]. The stable point (or region) is the vowel point (or region) that is least affected by context. A 4-dimensional feature vector represents each instance of a vowel- F1stable, F2stable, F3stable and vowel duration. In [4], formant transformation is achieved by training Gaussian mixture model (GMM) parameters using joint density estimation (JDE). This works well for data that

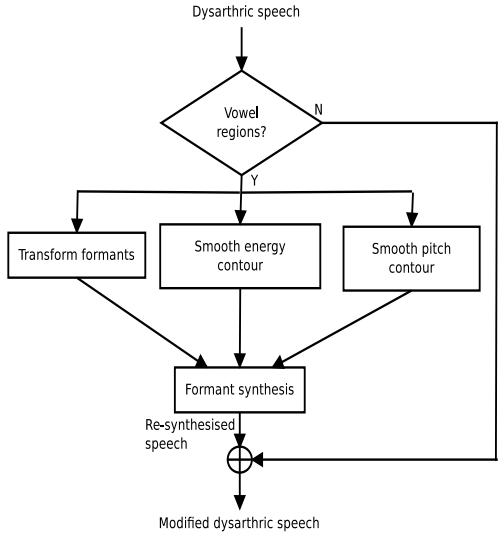

Figure 1: Formant re-synthesis of dysarthric speech

is phonetically balanced. The data used in the experiments in this paper suffers from data imbalance as the frequency of individual vowels in the database varies. To overcome this problem, a universal background model-GMM (UBM-GMM) [20] is trained and adapted to individual vowels of dysarthric and normal speech. Maximum a posteriori (MAP) is the adaptation algorithm used. The procedure to obtain adapted models is as follows:

- Each frame in a vowel region is represented by a 4-dimensional feature vector- formants F1-F3, and vowel duration. All the feature vectors, irrespective of stable points, are pooled together for all vowel instances across dysarthric and normal speech to train the UBM-GMM.
- The adaptation data for a vowel of dysarthric or normal speech is a 4-dimensional feature set (F1stable, F2stable, F3stable, vowel duration) across all instances of that vowel.
- A set of $(2 * number\_of\_vowels)$ models is obtained by adapting only the means of the UBM-GMM. This is a codebook of means for the same vowel across dysarthric and normal speech.

The dysarthric speech data is initially split into train (80%) and test data (20%). Normal speech corresponding to the dysarthric speech in Nemours database is used for obtaining the codebook. For the Indian dysarthric speech, speech of speaker "ksp" from CMU corpus [16] is used as normal speech. Adapted models are built using the train data. The codebook size of the UBM-GMM is 64. The procedure to re-synthesise dysarthric speech is shown in Figure 1. For test data, pitch ($F0$) and energy contours are smoothened. Smoothening is performed by using a median filter of order 3 and then low-passing using a Hanning window. This approach differs from the work carried out in reference [4], where a synthetic F0 contour is used for the dysarthric speech. Using the vowel boundaries, every vowel in the test utterance is represented by a 4-dimensional feature vector (stable F1-F3+vowel duration). Using the codebook of means for vowels across dysarthric and normal speech, the features of the dysarthric vowels are replaced by the means of their normal counterpart. The replaced or transformed stable point formants represent the entire vowel. Hence, the same stable point formant value is repeated across the duration of the vowel. Using the transformed formant contours, smoothened pitch and energy contours, speech is synthesised using a formant vocoder [21]. The modified dysarthric speech is then obtained by replacing non-vowel regions in the re-synthesised dysarthric speech by the original dysarthric speech.

4. HMM-based synthesiser using adaptation

An HMM-based TTS synthesiser (HTS) adapted to the dysarthric person's voice is developed [22], [8]. This is to evaluate the maximum intelligibility of synthesised speech that can be obtained given a recognition system for dysarthric speech that is 100% accurate. The purpose of using an HMM-based adaptive TTS synthesiser is two-fold: (1) not enough data to build a speaker-dependent system for every dysarthric speaker, and (2) to correct the pronunciation of the dysarthric speaker.

The HMM-based adaptive TTS can be divided into three phases- training, adaptation and synthesis. Audio files and corresponding transcriptions are available for training and adaptation data. In the training phase, mel-generalized cepstral (MGC) coefficients and $logf0$ values, along with their velocity and acceleration values are extracted from the audio files. Average voice models are then trained from speech features corresponding to the training data. In the adaptation phase, CSMAPLR+MAP adaptation (CSMAPLR- constrained structural maximum a posteriori linear regression) is performed to adapt the average voice models to the adaptation features. Speaker adaptive training (SAT) is performed to reduce the influence of speaker differences in the training data. In the synthesis phase, the test sentence is broken down into phones. Phone HMMs are chosen based on the context and concatenated to form the sentence HMM. MGC coefficients and $f0$ values are generated from the sentence HMM, and speech is synthesised using mel log spectrum approximation (MLSA) filter.

To build an adaptive TTS system for speakers in Nemours database, speech of two normal American male speakers, "bdl" and "rms" from the CMU corpus, is used as the training data. For the Indian English dysarthric data, the training data is speech of an Indian speaker "ksp" from the CMU corpus. 1 hour of speech data is available for every speaker in the CMU corpus. Dysarthric speech data is split into adaptation data (80%) and test data (20%). Synthesised speech of the sentences in the test data is used in the subjective evaluation. For developing the HMM-based adaptive TTS sythesiser, HTS version 2.3 software is used.

5. Proposed modifications to dysarthric speech

5.1. Durational analysis

A durational analysis across dysarthric and normal speech is performed. The following observations with respect to dysarthric speech are made:

- The average phone durations of dysarthric speech in the

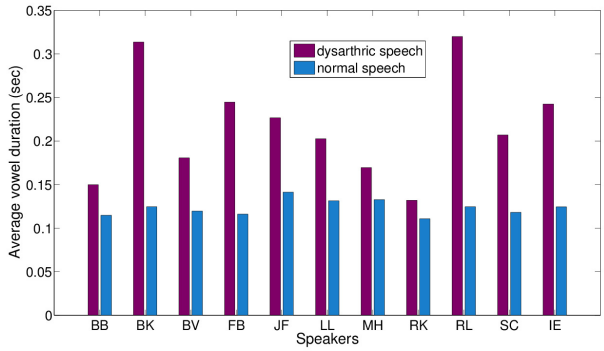

Figure 2: Average vowel durations across dysarthric and normal speakers

databases are longer than their normal speech counterparts [13, 23, 24]. As an example, average vowel durations are plotted in Figure 2.

- Standard deviations of vowel durations of dysarthric speakers are also longer (Figures 3 and 4), indicating that either the vowel is sustained for a longer duration or is hardly uttered.
- Speech data of the Indian dysarthric speaker IE is compared with the speech of different normal speakers. Four different nativities of Indian English (Hindi, Tamil, Telugu and Malayalam) in the Indic TTS corpus [19], and speech of an American speaker "rms" from CMU corpus [16] are the normal speech data considered. It is observed that the duration plot of speaker IE is clearly shifted with respect to that of normal speakers (Figure 5).
- For the same set of sentences spoken by dysarthric and normal speakers, the total utterance duration is longer for the dysarthric speaker. This indicates insertion of phones, intra-utterance pauses, etc. while speaking.

Based on the above analysis, if the duration is reduced closer to that of normal speech, the quality of dysarthric speech may improve. Reference [25] observes that as phone durations of dysarthric speech increase, the intelligibility of speech in terms of FDA score comes down. Taking this observation forward, in this paper, dysarthric speech is modified both manually and automatically to achieve this durational reduction.

5.2. Manual Modifications

The increase in phone duration is due to elongation of vowels, artifacts while producing sounds, or significant pauses within words. Hence, randomly increasing the speech rate of the utterance won't be useful, specific corrections are required. Each phoneme segment of the dysarthric speech is compared with its counterpart in normal speech. Elongations and artifacts are manually removed, keeping in mind not to degrade the intelligibility of speech. Steady regions of elongated vowels are spliced out. Segments are carefully deleted so as to not cause a sudden change in spectral content. For speaker IE, the recorded speech of Malayalam speaker "IEm" is considered as the reference. Original and corresponding manually modified waveforms are used in the subjective evaluation.

5.3. Proposed automatic method

A Dynamic Time Warping (DTW) algorithm is used to compare the similarity between MFCC features of dysarthric (test) and corresponding normal speech (reference). 39-dimensional MFCC features, including velocity and acceleration values are used. Wherever the slope of the DTW path is zero for a minimum number of frames, termed as *frameThres*, those frames are considered for deletion. When deleting frames, it is important to ensure that there is no sudden change in energy at the points of join, i.e., the energies between frames before and after deletion. It is observed that artifacts are introduced in places where the energy difference between frames at concatenation points is high. Therefore, the short-term energy (STE) difference is considered as an additional criterion for deletion. Whenever STE difference is less than a certain limit, *STEThres*, frames are deleted. In the experiments, *frameThres* and *STEThres* are set to 6 and 0.5 respectively. These thresholds are obtained empirically after testing with *frameThres* ranging from 4 to 10 and *STEThres* ranging from 0.3 to 2.5. This automatic procedure of deletion is illustrated in Figure 6.

The DTW paths of a sample utterance of dysarthric speaker RL before and after automatic modifications compared with respect to the same utterance of normal speaker JPRL is shown in Figure 7. It is observed that the DTW path is more diagonal in Figure 7b compared to Figure 7a, indicating that the modified dysarthric utterance is more similar to the normal utterance. It also results in a considerable reduction in number of frames or duration of the utterance. This method is referred to as the

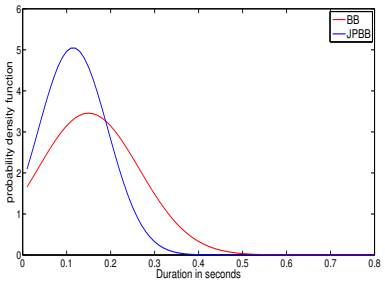

Figure 3: Duration plot for vowels of dysarthric speech BB and normal speech JPBB

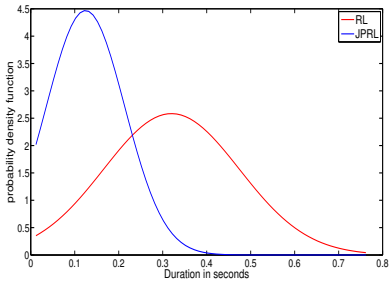

Figure 4: Duration plot for vowels of dysarthric speech RL and normal speech JPRL

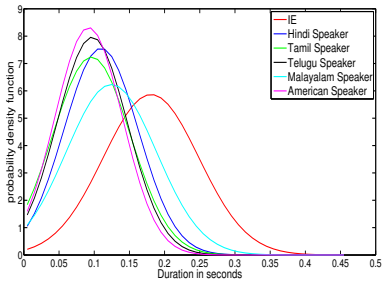

Figure 5: Duration plot for vowels of dysarthric speech IE and normal speech of other speakers

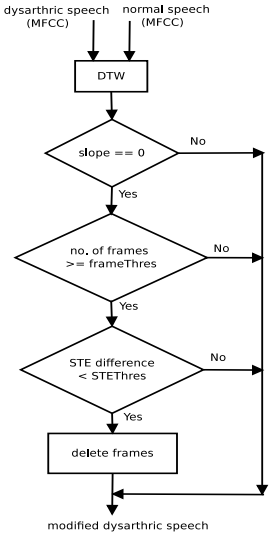

Figure 6: Flowchart of automatic (DTW+STE) method to modify dysarthric speech

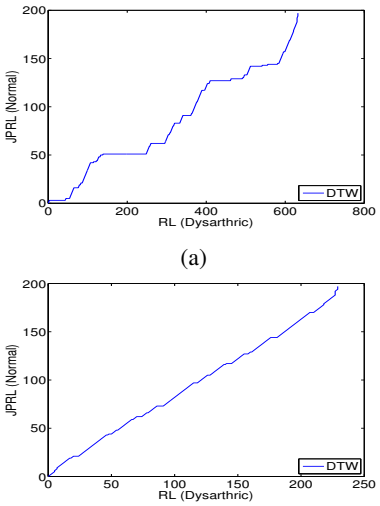

Figure 7: DTW paths of an utterance of speaker RL between: (a) original dysarthric speech and normal speech, and (b) modified dysarthric speech and normal speech

DTW+STE modification method.

6. Performance evaluation

Subjective evaluation is conducted to evaluate the techniques used. A pairwise comparison test is performed to assess the proposed modification techniques and a word error rate test to compare intelligibility across different methods. Naive listeners are used in the subjective tests rather than expert listeners in order to assess how a naive listener, who has little or no interaction with dysarthric speakers, evaluates the quality of dysarthric speech. Tests are conducted in a noise-free environment.

6.1. Pairwise comparison tests

A pairwise comparison test is conducted to compare the quality of speech modified by the proposed techniques and original dysarthric speech [26]. In the "A-B" test, A is played first and then B, and vice-versa in the "B-A" test to remove the bias in listening. "A" is the modified speech and "B" is the original speech in both the tests. Preference is always calculated in terms of the audio sample played first. The score "A-B+B-A" gives an overall preference for system A against system B and is calculated by the following formula:

$$\text{“}A-B+B-A'' = \frac{\text{“}A-B'' + (100 - \text{“}B-A'')}{2}$$

About 11 listeners evaluated a set of 8 sentences for each speaker. Results of the evaluation are shown in Figure 8. Results indicate a preference for the modified versions over original dysarthric speech in almost all cases. From Figure 8, it is evident that the manual method out-performs the DTW+STE (automatic) method. This is because manual modifications are hand-crafted carefully so as to produce better-sounding speech. For speakers BB and IE, who are mildly dysarthric, the performance of the DTW+STE method drops drastically due to artifacts introduced in the modified speech. This is true for speakers

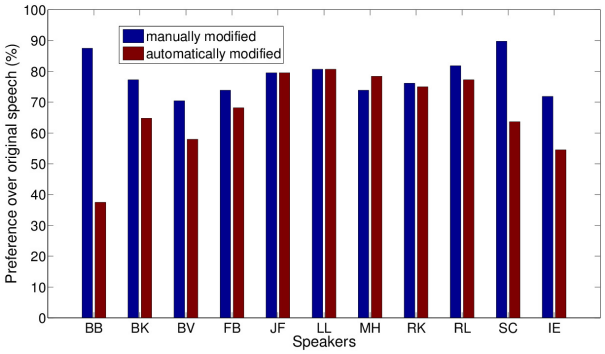

Figure 8: Preference for manually and automatically (DTW+STE) modified speech over original dysarthric speech of different speakers

BK and BV, where artifacts in the original speech are not eliminated by the DTW+STE technique. The drop in performance from the manual to the automatic technique is quite high for speaker SC because of the slurry nature of speech. Hence, in such cases identifying the specifics of dysarthria for individual speakers is vital to improving speech quality. Nonetheless, the performance of both methods is almost on par for speakers JF, LL, FB, MH, RL, RK who are mild to severely dysarthric.

Pairwise comparison tests were also conducted between original and formant re-synthesised speech, and between automatically modified and formant re-synthesised speech. About 10 listeners evaluated a set of 8 sentences for each speaker in each test. Preference was individually over 82% for the original dysarthric speech and DTW+STE method over the formant re-synthesis method.

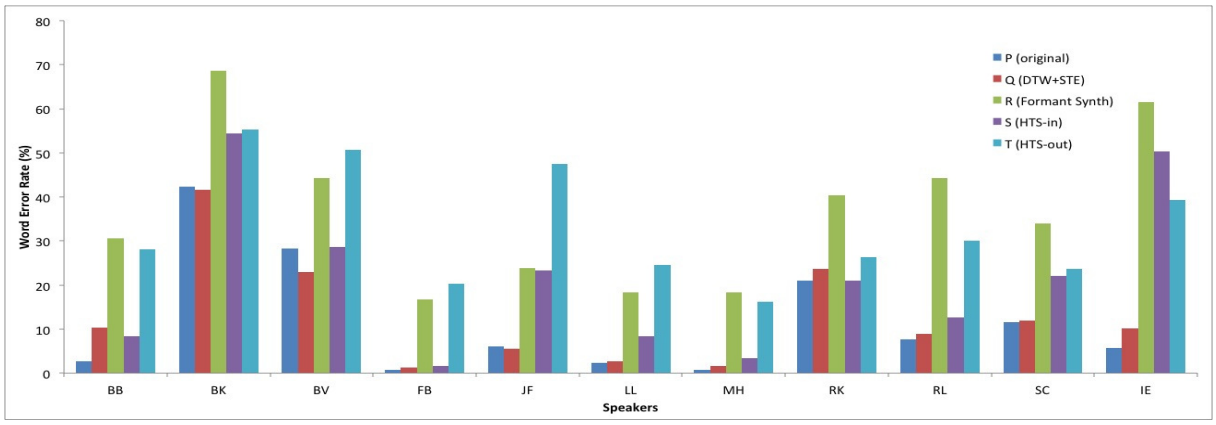

Figure 9: Word error rates for different types of speech across dysarthric speakers

6.2. Intelligibility tests

To evaluate intelligibility across different systems, a word error rate (WER) test was conducted. Based on the feedback on the pairwise comparison tests and the text in Nemours database containing nonsensical sentences, it is difficult to recognise words in dysarthric speech. Hence, given the text, listeners were asked to enter the number of words that was totally unintelligible. Though the knowledge of the pronounced word may have an influence on its recognition, this is a uniform bias that is present when evaluating all systems. About 10 listeners participated in the evaluation. The following types of speech were used in the listening tests:

P (original): original dysarthric speech
Q (DTW+STE): dysarthric speech modified using the DTW+STE method
R (Formant Synth): output speech of the formant re-synthesis technique
S (HTS-in): speech synthesised using the HMM-based adapted TTS for text in the database not used for training (held-out sentences)
T (HTS-out): speech synthesised using the HMM-based adapted TTS for text from the web

The results of the WER test are presented in Figure 9. It can be seen that the intelligibility of formant re-synthesis technique is poor for all speakers. For the DTW+STE method, WER is higher compared to original dysarthric speech for a majority of speakers. WER of HMM-based adaptive synthesiser on held out-sentences, i.e., sentences not used during training is high compared to original dysarthric speech in almost all cases. Intelligibility of sentences synthesised from the web is quite poor compared to that of held-out sentences for speakers in Nemours database. This is the opposite for Indian dysarthric speaker IE. This is due to the similar structure of held-out sentences and sentences used in training the HMM-based synthesiser in Nemours database, unlike the sentences in the Indian dysarthric dataset that are unstructured. Overall, the intelligibility of original dysarthric speech does not increase. However, for speakers BK, BV and JF, DTW+STE modified speech has the lowest WER. For speaker RK, the intelligibility of HMM-based adaptive synthesised speech is on par with that of original dysarthric speech. By informal listening, it is noted that some pronunciations of the dysarthric speaker do get corrected in the sentences synthesised using the HMM-based adaptive TTS system. This indicates that the technique used to increase intelligibility largely depends on the type and severity of dysarthria.

While the DTW+STE does not need segmented boundaries, it makes use of a reference for comparison. Only insertion of sounds are taken care of, deletion and substitution of phonemes are not addressed. Though this technique does not increase intelligibility for most speakers, the overall perceptual quality of the modified dysarthric speech is improved.

In the speech synthesis domain, the HMM-based adaptive synthesiser is a statistical parametric speech synthesiser (SPSS) and the DTW+STE technique is analogous to a unit selection speech (USS) synthesiser. The synthesised speech of the HMM-based synthesiser lacks the voice quality of the dysarthric speaker. Similar to the USS system, the speech output of the DTW+STE method has discontinuities but preserves the voice characteristics of the dysarthric speaker.

7. Conclusions

Continuous dysarthric speech quality is improved upon in the work. A durational analysis is performed by comparing dysarthric and normal speech for speakers in Nemours database and an Indian English speaker having dysarthria. Based on the analysis, dysarthric speech is directly modified manually, and an automatic method is developed to do the same. The intelligibility of dysarthric speech modified using different techniques is studied. Evaluations indicate an improvement in speech quality using the STE+DTW method. This emphasises the importance of duration in perceptual speech quality, indicating that this kind of modification may be used as a pre-processing step for improving dysarthric speech quality. Only durational attributes are analysed in this work, this can be extended to analyse other attributes that affect the speech of a dysarthric person.

8. Acknowledgements

The authors would like to thank Rajiv Rajan, Kalpana Rao and Namita Jacob of Vidyasagar, Chennai, with whose help, collecting dysarthric speech data was possible. The authors would also like to thank Jom, Shreya and other colleagues in the lab for their inputs and assistance in conducting the evaluations.

9. References

[1] American Speech Language Hearing Association, "Dysarthria," http://www.asha.org/public/speech/disorders/dysarthria/.

[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," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2003, pp. 924 – 927.

[3] 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*, June 2004, pp. 25–30.

[4] A. B. Kain, J.-P. Hosom, X. Niu, J. P. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelligibility of dysarthric speech," *Speech Communication*, vol. 49, no. 9, pp. 743 – 759, 2007.

[5] F. Rudzicz, "Acoustic transformations to improve the intelligibility of dysarthric speech," in *2nd Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*, July 2011, p. 11 21.

[6] ——, "Adjusting dysarthric speech signals to be more intelligible," *Computer Speech & Language*, vol. 27, no. 6, pp. 1163–1177, 2013.

[7] M. S. Yakcoub, S. A. Selouani, and D. O'Shaughnessy, "Speech assistive technology to improve the interaction of dysarthric speakers with machines," in *3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)*, March 2008, pp. 1150–1154.

[8] M. Dhanalakshmi and P. Vijayalakshmi, "Intelligibility modification of dysarthric speech using HMM-based adaptive synthesis system," in *2nd International Conference on Biomedical Engineering (ICoBE)*, March 2015, pp. 1–5.

[9] M. Saranya, P. Vijayalakshmi, and N. Thangavelu, "Improving the intelligibility of dysarthric speech by modifying system parameters, retaining speaker's identity," in *International Conference on Recent Trends In Information Technology (ICRTIT)*, April 2012, pp. 60–65.

[10] G. L. Dorze, L. Ouellet, and J. Ryalls, "Intonation and speech rate in dysarthric speech," *Journal of Communication Disorders*, vol. 27, no. 1, pp. 1 – 18, 1994.

[11] X. Menendez-Pidal, J. Polikoff, S. Peters, J. Leonzio, and H. Bunnell, "The Nemours database of dysarthric speech," in *Fourth International Conference on Spoken Language (ICSLP)*, vol. 3, October 1996, pp. 1962–1965.

[12] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. S. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research." in *Annual Conference of the International Speech Communication Association, INTERSPEECH*, September 2008, pp. 1741–1744.

[13] F. Rudzicz, A. K. Namasivayam, and T. Wolff, "The TORGO database of acoustic and articulatory speech from speakers with dysarthria." *Language Resources and Evaluation*, vol. 46, no. 4, pp. 523–541, 2012.

[14] Wikipedia, "Arpabet," https://en.wikipedia.org/wiki/Arpabet.

[15] S. A. Shanmugam and H. Murthy, "A hybrid approach to segmentation of speech using group delay processing and HMM based embedded reestimation," in *Annual Conference of the International Speech Communication Association, INTERSPEECH*, Singapore, September 2014, pp. 1648–1652.

[16] J. Kominek and A. W. Black, "The CMU arctic speech databases," in *5th ISCA Speech Synthesis Workshop*, June 2004, pp. 223–224.

[17] Carnegie Mellon University, "The CMU pronounciation dictionary," http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

[18] P. Enderby, "Frenchay Dysarthria Assessment," *International Journal of Language & Communication Disorders*, vol. 15, no. 3, pp. 165–173, December 2010.

[19] TTS Consortium, DeitY, Government of India, "Indic TTS," https://www.iitm.ac.in/donlab/tts/.

[20] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19–41, January 2000.

[21] L. Rabiner and R. Schafer, *Theory and Applications of Digital Speech Processing*, 1st ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2010.

[22] J. Yamagishi, T. Nose, H. Zen, Z. H. Ling, T. Toda, K. Tokuda, S. King, and S. Renals, "Robust speaker-adaptive HMM-based text-to-speech synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1208–1230, August 2009.

[23] V. Surabhi, P. Vijayalakshmi, T. S. Lily, and R. V. Jayanthan, "Assessment of laryngeal dysfunctions of dysarthric speakers," in *IEEE Engineering in Medicine and Biology Society*, Minnesota, September 2009, p. 2908 2911.

[24] H. Ackermann and I. Hertrich, "Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing," *Folia Phoniatrica et Logopaedica*, vol. 46, no. 2, pp. 70–78, 1994.

[25] P. Vijayalakshmi and M. Reddy, "Assessment of dysarthric speech and analysis on velopharyngeal incompetence," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, August 2006, pp. 3759–3762.

[26] P. Salza, E. Foti, L. Nebbia, and M. Oreglia, "MOS and pair comparison combined methods for quality evaluation of text to speech systems," in *Acta Acustica*, vol. 82, 1996, pp. 650–656.