International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 4, August 2023, pp. 4041~4051
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i4.pp4041-4051

Improving the intelligibility of dysarthric speech using a time domain pitch synchronous-based approach

Mahesh Giri, Neela Rayavarapu
Department of Electronics and Telecommunication Engineering, Symbiosis Institute of Technology,
Symbiosis International (Deemed University), Pune, India

Article Info

Article history:

Received Jul 14, 2022
Revised Oct 14, 2022
Accepted Oct 20, 2022

Keywords:

Acoustic parameters of speech
Digital signal processing
Dysarthria
Signal processing
Speech intelligibility enhancement

ABSTRACT

Dysarthria is a motor speech impairment that reduces the intelligibility of speech. Observations indicate that for different types of dysarthria, the fundamental frequency, intensity, and speech rate of speech are distinct from those of unimpaired speakers. Therefore, the proposed enhancement technique modifies these parameters so that they fall in the range for unimpaired speakers. The fundamental frequency and speech rate of dysarthric speech are modified using the time domain pitch synchronous overlap and add (TD-PSOLA) algorithm. Then its intensity is modified using the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT)-based approach. This technique is applied to impaired speech samples of ten dysarthric speakers. After enhancement, the intelligibility of impaired and enhanced dysarthric speech is evaluated. The change in the intelligibility of impaired and enhanced dysarthric speech is evaluated using the rating scale and word count methods. The improvement in intelligibility is significant for speakers whose original intelligibility was poor. In contrast, the improvement in intelligibility was minimal for speakers whose intelligibility was already high. According to the rating scale method, for diverse speakers, the change in intelligibility ranges from 9% to 53%. Whereas, according to the word count method, this change in intelligibility ranges from 0% to 53%.

Corresponding Author:

Mahesh Giri
Department of Electronics and Telecommunication Engineering, Symbiosis Institute of Technology,
Symbiosis International (Deemed University)
Lavale, Pune, 412115, India
Email: Officeofanandi@gmail.com

1. INTRODUCTION

Dysarthria is a motor-speech disorder that arises because of neurological damage to the motor-speech system and is indicated by a debilitated articulation of phonemes [1], [2]. Dysarthria could occur for several reasons, such as brain damage, multiple sclerosis, or Parkinson's disease [3]. Based on where the damaged speech muscles are located, dysarthria can be categorized as flaccid, spastic, hypokinetic, hyperkinetic, and ataxic [4], [5]. Dysarthric speech enhancers for individuals with dysarthria are a relatively untapped field of study. Dysarthric patients have trouble communicating freely with others in daily life and consequently suffer from low confidence levels. By emphasizing the articulation process, the speech therapist can improve the intelligibility of dysarthric speech. However, therapy treatment can only enhance dysarthric speech intelligibility to a limited extent and is time-consuming. Also, both the patient and the therapist need to devote time to the therapy treatment. Moreover, elderly patients have poor therapeutic responses [6]. Therefore, a method to improve the intelligibility of dysarthric speech must be investigated.

Speech parameters are divided into three categories: acoustic parameters of speech, articulatory parameters of speech, and auditory parameters of speech [7], [8]. Acoustic parameters comprise the fundamental frequency, formant frequency, and intensity of speech signals [7], [9]. It has been observed that the acoustic properties of dysarthric speech differ from those of unimpaired speakers. The normal range of the fundamental frequency for male speakers is 65 to 155 Hz; for female speakers, it is 165 to 255 Hz; and for children between the ages 6 and 18 years, it is 250 to 300 Hz or higher. The fundamental frequency is negatively related to the height and age of the child. That is, as children get older and taller, the fundamental frequency decreases. Furthermore, for children, it is higher for girls than for boys [7], [10]. However, the fundamental frequency of many dysarthric speakers is observed to be outside the normal range [11]. Additionally, in several cases, the intensity level of speech of dysarthric speakers was found to be higher than that of an unimpaired speaker [9], [12]. Similarly, the speech rate of many dysarthric speakers was affected compared to that of an unimpaired speaker [13], [14].

Some researchers have attempted to enhance the intelligibility of dysarthric speech. Prakash *et al.* [15] attempted to modify the durational attributes of dysarthric speech and tried to make it more similar to that of normal speech. However, the fundamental frequency and intensity were not considered in their research. Prema *et al.* [16] claimed to enhance the intelligibility of dysarthric speech. In this technique, speech samples from one male speaker were recorded and analyzed for enhancement purposes. The analysis was done with the Pratt tool. A vowel space manipulation approach was used for designing the algorithm. The formants of the impaired speech were also modified for enhancement. The shift in formant frequencies was measured and used to assess the change in intelligibility. However, only the formant shift was evaluated to determine the change in intelligibility. Other proven methodologies, such as subjective or objective intelligibility assessment methods, were not used. Also, this research was performed only on a single dysarthric speaker. Kain *et al.* [17] proposed an approach for enhancing the intelligibility of vowels of dysarthric speech. They attempted to modify the vowels of dysarthric speech and bring them closer to those of a normal speaker. Enhancement in the intelligibility of vowels from 48% to 54% was observed by them. However, the enhancement of the intelligibility of words or sentences was not the scope of this research. Further, Tolba and El_Torgoman [18] modified the first and second formant frequencies of dysarthric speech and tried to enhance the intelligibility of dysarthric speech. However, this study was focused on Arabic language speech. Sivanupandian and Jennifer [19] recently conducted a study to improve the intelligibility of dysarthric speech. The linear predictive coding (LPC) coefficient mapping and frequency warping of the LPC poles approach was applied to enhance the intelligibility of dysarthric speech. These researchers claimed nearly 35% to 45% enhancement in intelligibility. However, the details of the experiment, such as the total number of speakers on whom it was conducted and the method used to evaluate intelligibility, are left absent.

Therefore, in this work, speech parameters such as the fundamental frequency, intensity, and speech rate have been modified to enhance the intelligibility of dysarthric speech. Using the Pratt tool, these parameters were evaluated for the dysarthric speech of every dysarthric speaker, and optimized scaling factors were selected for every speaker [20]. The time domain pitch synchronous overlap and add (TD-PSOLA) method, which is a time domain approach that suits more single-pitch signals, is then applied to scale the fundamental frequency and speech rate of the dysarthric speech signal [21], [22]. Later, a method based on the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) was utilized to change the intensity of dysarthric speech. Following enhancement, subjective approaches, such as the rating scale method and the word count method, were used to evaluate the intelligibility of the speech. Based on the results of the rating scale and word count method, it was concluded that the intelligibility of dysarthric speech has increased.

The outline of the remaining paper is as: the selection of parameters to be modified is detailed in section 2. The methodology employed for the modification of speech parameters is discussed in section 3. The intelligibility tests for impaired and enhanced dysarthric speech and their results are covered in section 4. Finally, the conclusion of the study is presented in section 5, and the future scope of this study is discussed in section 6.

2. SELECTION OF PARAMETERS

The preceding section describes the different types of speech parameters. From that, the articulatory parameters are associated with the speech production mechanism, which involves tongue movement, lip position, vocal cord movement, and vocal tract. Typically, therapists focus on these articulatory aspects to enhance the intelligibility of dysarthric speech [8]. Once the speech is produced and transmitted over the air, the associated parameters are referred to as acoustic parameters of speech [7], [9]. Acoustic parameters include fundamental frequency, the intensity of speech, speech rate, formant frequencies, and prolonged vowel duration. This study primarily focuses on the fundamental frequency, intensity, and speech rate of a dysarthric speaker.

In this study, the above-said parameters are evaluated for the database of ten dysarthric speakers. Six dysarthric speaker's speech samples from a total of 10 speakers have been collected from the Ali Yavar Jung National Institute for The Hearing Handicapped (AYJNIHH), Mumbai, India, and the All India Institute of Physical Medicine and Rehabilitation (AIIPMR), Mumbai, India [11]. The dysarthric speech samples of the remaining four speakers have been collected from the database developed by the University of Illinois, Chicago [23]. Like the other databases, this database contains speech samples of the flaccid, hypokinetic, spastic, and ataxic kinds of dysarthria that have commonly used words and numerals [24]–[26]. A summary of the recorded database, including the speaker's age, gender, type of dysarthria, and severity, is provided in Table 1. A software tool, 'Audio recorder editor', was used to record dysarthric speech samples. Throughout the recording sessions, the audio recorder editor was configured with a mono channel, 16 KHz sampling rate, and 16-bit resolution [11].

Table 1. Dysarthric speakers' details

Speaker Number	Type of Dysarthria	Age	Gender	Institute	Original Intelligibility
Speaker 1	Hypokinetic	65	Male	AYJNIHH, Mumbai	Low
Speaker 2	Spastic	18	Male	AYJNIHH, Mumbai Hospital, Mumbai	Medium
Speaker 3	Flaccid	50	Male	AIIPMR, Mumbai	Medium
Speaker 4	Hypokinetic	78	Male	AYJNIHH, Mumbai Hospital, Mumbai	Medium
Speaker 5	Ataxic	35	Male	AIIPMR, Mumbai	High
Speaker 6	Hypokinetic	64	Male	AYJNIHH, Mumbai	High
Speaker 7	Spastic	30	Female	University of Illinois, Chicago	Low
Speaker 8	Spastic	21	Male	University of Illinois, Chicago	High
Speaker 9	Ataxic	48	Male	University of Illinois, Chicago	Low
Speaker 10	Spastic	35	Male	University of Illinois, Chicago	Medium

The above parameters that need to be modified are analyzed for impaired and unimpaired speech. A well-known speech processing tool, the Pratt tool, is used to determine the fundamental frequency and intensity of impaired and unimpaired speech. The comparison of intended parameters for impaired and unimpaired speakers is presented in Table 2. In the table, although parameter values for two utterances are displayed, the final inference is drawn by comparing parameter values for a minimum of five phrases from each speaker.

Table 2. Comparison of fundamental frequency and speech intensity for dysarthric and normal speakers

Speaker Number	Utterance	Fundamental Frequency in Hz		Intensity in dB		Comment on Speech Rate
		Normal	Impaired	Normal	Impaired	
Speaker 1	One	119	180	62.10	70.07	High
	Ten	111	169	63.34	67.10	High
Speaker 2	One	119	196	62.10	47.38	Low
	Ten	111	178	63.34	52.71	Low
Speaker 3	One	119	168	62.10	71.76	High
	Ten	111	169	63.34	77.00	High
Speaker 4	One	119	183	62.10	81.97	High
	Ten	111	216	63.34	78.55	High
Speaker 5	One	119	124	62.10	69.49	Normal
	Ten	111	130	63.34	73.89	Normal
Speaker 6	One	119	124	62.10	79.07	Normal
	Ten	111	115	63.34	71.90	Normal
Speaker 7	Copy	193	252	52.14	73.77	Low
	Backspace	222	285	52.12	69.72	Low
Speaker 8	Copy	121	197	44.97	75.93	Low
	Backspace	117	156	61.88	73.10	Low
Speaker 9	Copy	121	210	44.97	77.37	Low
	Backspace	117	159	48.84	76.52	Low
Speaker 10	Copy	121	163	44.97	74.94	Low
	Backspace	117	136	48.84	78.27	Low

From Table 2, it can be inferred that speakers 5, 6, and 10 have a fundamental frequency within the range of the fundamental frequency of unimpaired speakers. Table 2 also shows that the fundamental frequencies of the remaining seven speakers are higher than the fundamental frequency of the unimpaired speaker. All dysarthric speakers, except speaker 2, have a higher intensity than the unimpaired speakers. Furthermore, after listening to several utterances made by each speaker, it was noted that, except for speakers 5 and 6, the rest of the speaker's speech sounded slightly faster or slower. At the time of recording sessions, it was noted that many of these speakers talked loudly and became weary more quickly. As a result, they lose the ability to speak for a long time, which eventually deteriorates their speech intelligibility.

3. METHODOLOGY OF ENHANCEMENT

According to Table 2, the fundamental frequency, speech rate, and intensity of dysarthric speech are different from those of unimpaired speech [27]. Therefore, by modifying these parameters, the intelligibility of dysarthric speech can be improved. Time and frequency domain techniques can be employed to modify fundamental frequency and speech rate. The TD-PSOLA algorithm is an example of a time-domain method. In contrast, the phase vocoder is an example of a frequency-domain method. The TD-PSOLA is effective for single-pitch signals, and its computational complexity is less than that of the phase vocoder. On the other side, speech is a single-pitch signal. In the future, this dysarthric speech enhancement technology is supposed to be used in real-time. Therefore, in this research, to reduce computational complexity, the TD-PSOLA technique is used. The flow of steps to be followed for selecting the optimal fundamental frequency, intensity, and speech rate scaling factor is provided in Figure 1.

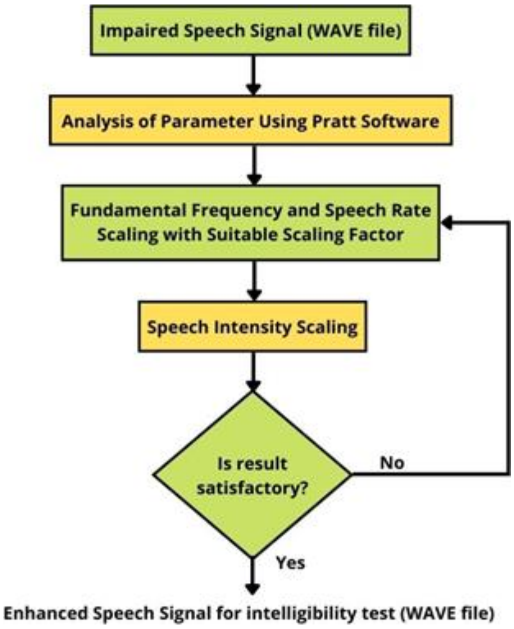

Figure 1. Flow diagram for selecting the suitable scaling factors

3.1. Methodology for modification of fundamental frequency and speech rate

The flowchart of the TD-PSOLA technique to modify the fundamental frequency and speech rate of dysarthric speech is depicted in Figure 2. First, the speech signal that has to be improved is pre-processed using low pass filtering and center clipping. The fundamental frequencies for impaired speech are then computed. Later, based on scaling factors, the fundamental frequencies and speech rate are synthesized, and then a modified speech signal is produced as the system's output. A detailed description of each block in Figure 2 is provided.

3.1.1. Impaired speech

The impaired dysarthric speech is provided as input to the system. In this system, the pre-recorded speech samples from the database are used as the input. These samples are available in WAVE format. Random inputs are provided to the system from each speaker, and then the system's performance for those samples is analyzed.

3.1.2. Low pass filter

Accurate fundamental frequency identification is critical for subsequent processing in the TD-PSOLA. On the other hand, the speech signal has a large number of harmonic components. Generally, the fundamental frequency range is between 65 and 300 Hz. However, fundamental frequencies of 500 Hz are also reported in some instances. The speech signal has many harmonic components, and the fundamental frequency component is often not the strongest. Therefore, the low pass filter is essential to avoid the deterioration of fundamental frequency from harmonics. The accuracy of fundamental frequency calculation increases because of low pass filtering. In this case, the cut-off frequency of the low pass filter is set at 900 Hz.

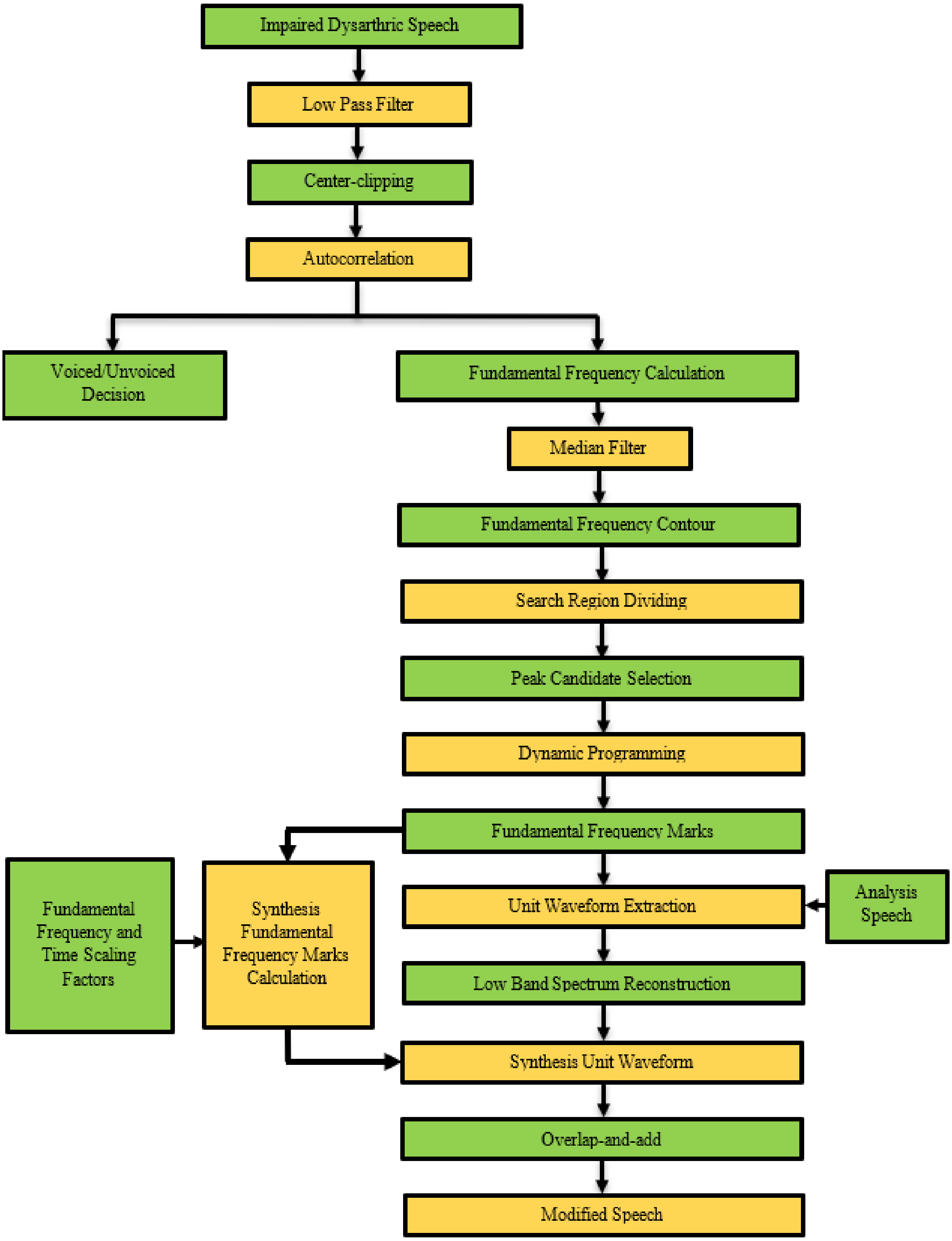

Figure 2. Flowchart of the TD-PSOLA technique

3.1.3. Center clipping

This research uses the autocorrelation method to estimate the fundamental frequency. Due to the formant frequencies, this method may encounter noise issues. The center clipping method is applied to get over this problem. The center clipping is applied using (1) [7], [28], [29]. In (1), C_L stands for clipping threshold, where its value is set to 30% of the maximum magnitude of the signal.

$$\left.\begin{array}{l} C[x(n)] = x(n) - C_L \; if \; x(n) > C_L \; or \\ C[x(n)] = 0 \; if \; |x(n)| \le C_L \; or \\ C[x(n)] = x(n) + C_L \; if \; x(n) < -C_L \end{array}\right\} \tag{1}$$

3.1.4. Autocorrelation

The fundamental frequency is estimated using this method. This method of autocorrelation is also utilized to make voice and non-voice decisions. The basic equation of the autocorrelation function is depicted in (2) [7], [29]. In this study, the autocorrelation of a signal is determined by first performing an FFT on the signal and then performing an IFFT on the result of the FFT. The equation to compute this is provided in (3) [7], [29], [30].

$$r(\tau) = \sum_{i=1}^{W-\tau} X_i X_{i+\tau} \tag{2}$$

$$r(n) = iFFT\{|FFT[x(k)]|^2\} \tag{3}$$

3.1.5. Fundamental frequency calculation

The low pass filtered signal from the previous stage is divided into frames of size 30 milliseconds. First, the fundamental frequency of each frame is calculated using the autocorrelation method. Then, the median filter is applied to the fundamental frequency of each frame to get the fundamental frequency contour [7], [29], [30].

3.1.6. Search region dividing

The accuracy of fundamental frequency marking plays a crucial role in its scaling. The fundamental frequency contour of the previous step is divided into several small search regions. Then, for each search region, three fundamental frequency marks are calculated. The (4) is used to determine the search region in which t_m denotes the global maxima location, T_0 denotes the fundamental frequencies period, and f is a parameter to calculate the width of the search region. In this research, the value of f is set to 0.7 [29].

$$SR = [t_m + f.T_0; t_m + (2 - f).T_0] \tag{4}$$

3.1.7. Peak candidate selection and dynamic programming

From each search region, eight fundamental frequency mark candidates are selected for the next stage. Finally, one of the eight potential fundamental frequency mark candidates from each search region is selected as the final fundamental frequency mark. In the above-said process, a dynamic programming method is applied. In dynamic programming, the two below-listed criteria are used to determine the final fundamental frequency mark candidate.

a. Relative candidate height

Out of eight different fundamental frequency marks, the fundamental frequency mark having maximum amplitude is considered under this criterion. The (5) is used to calculate this probability. In (5), $h_{(j)}$ is the amplitude of j^{th} frame, whereas the h_{min} and the h_{max} are the j^{th} frame's minimum and maximum amplitudes [29].

$$s(j) = \left(\frac{h(j) - h_{min}}{h_{max} - h_{min}}\right) \times \alpha \tag{5}$$

b. Distance between two fundamental frequency mark candidates in two relative search regions

In addition, it must ensure that the distance between two fundamental frequency marks in consecutive frames is not too low. The (6) is used to calculate this probability. In order to compute this probability, the transition likelihood is calculated using (5), in which T_0 is the pitch period of the frame containing i and j; d_{ij} is the distance between i and j. The α from (5), β and γ from (6) are three adjustment parameters adapted for each speaker [29].

$$t(i,j) = \left(\frac{1}{1+\beta|T_0 - d_{ij}|}\right) \times \gamma \tag{6}$$

Dynamic programming is applied to get the optimal pitch mark that maximizes the likelihood $P(k,j)$. Here, k infers to the current search region, j implies pitch mark candidate, $logt_k(i,j)$ is transition likelihood between pitch mark candidate i and j in region $k-1$. Finally, the maximum likelihood is obtained using (7) [29]. This fundamental frequency mark signal is then fed to the next stage.

$$P(k,j) = \max[P(k-1,i) + log t_k(i,j)] + \log s_k\ (j) \quad (7)$$

3.1.8. Synthesis fundamental frequency mark calculations and Waveform mapping

In this research, the synthesis fundamental frequency epochs are calculated using the analysis epochs. The accuracy of the epoch sequence is critical for the precision in fundamental frequency and speech rate modification [29], [31]. The waveform which is obtained from synthesized fundamental frequency marks is not smooth. As a result, it is necessary to smooth it out. Waveform mapping improves the smoothness of the synthesis waveform by employing a linear interpolation technique.

3.1.9. Low band spectrum reconstruction

The synthesis waveform is modified in the time domain. However, some of its frequency domain parameters are impacted, resulting in perceptual speech distortion. Low-band frequency reconstruction attempts to overcome this distortion [32]. After that, the overlap and add step is performed to obtain an enhanced speech signal [28]. The pre-recorded speech utterances of different dysarthric speakers are fed as input to the TD-PSOLA algorithm. The fundamental frequency for these utterances before and after applying the algorithm is provided in Table 3.

Table 3. Fundamental frequency comparison before and after enhancement

Speaker Number	Utterance	Fundamental Frequency Before Modification	Fundamental Frequency Scaling Factor	Fundamental Frequency After Modification		Accuracy in Percentage
				Theoretical	Practical	
Speaker 1	One	180	0.75	126	133	94.44
	Ten	169	0.75	135	126	93.33
Speaker 2	One	196	0.95	186	180	96.77
	Ten	178	0.95	169	164	98.22
Speaker 3	One	168	0.90	152	145	95.39
	Ten	169	0.90	152	147	96.71
Speaker 4	One	183	0.90	164	162	98.78
	Ten	216	0.90	194	190	97.93
Speaker 5	One	124	0.95	118	121	97.45
	Ten	130	0.95	124	132	93.54
Speaker 6	One	124	0.80	99	104	99.04
	Ten	115	0.80	92	90	95.19
Speaker 7	One	233	0.85	198	187	94.45
	Ten	285	0.85	242	243	99.58
Speaker 8	Copy	171	0.90	154	156	98.70
	Backspace	150	0.90	135	140	96.29
Speaker 9	Copy	210	0.75	155	161	99.61
	Backspace	219	0.75	164	168	97.56
Speaker 10	Copy	163	0.85	138	132	95.65
	Backspace	155	0.85	132	138	95.45
Average accuracy in percentage						**96.70**

3.2. Methodology for modification of speech intensity

In Table 2, intelligibility was recognized as high for a few dysarthric speakers. However, during recording sessions, it was observed that they spoke loudly and hence felt fatigued more quickly. Furthermore, it can be inferred from Table 2 that, except for speakers 1 and 2, the intensity level of the dysarthric speaker is found to be higher than that of an unimpaired speaker. As a result of this intensive speaking, many speakers have trouble pronouncing the words after continuous loud speaking.

Therefore, in this study, the speech intensity is adjusted after scaling the fundamental frequency and speech rate, and an FFT and IFFT-based technique is employed for this purpose [11]. Due to speech intensity scaling, the louder speakers are expected to speak with a softened voice, so their capacity to speak for a long time will rise. Then this softened speech can be amplified after fundamental frequency and speech rate modifications. The flow diagram for modifying speech intensity is shown in Figure 3.

Different values of fundamental frequency and time scaling factors were used to test the accuracy of the TD-PSOLA method. The observed values of fundamental frequency before and after applying the TD-PSOLA algorithm are shown in Table 3. First, Table 3 provides the theoretical value of fundamental frequency and its scaling factor. Then, the fundamental frequency of the modified speech sample computed using the Pratt tool is provided. Finally, the accuracy of the TD-PSOLA algorithm for fundamental frequency modification is assessed by comparing the modified signal's measured fundamental frequency to the modified signal's theoretically calculated fundamental frequency. Then, by averaging this accuracy, it can be inferred that the TD-PSOLA algorithm's accuracy for fundamental frequency modification is very high, almost close to 97%. Also, after listening to the modified speech, significant positive changes in speech rate and intensity of the modified signal are observed.

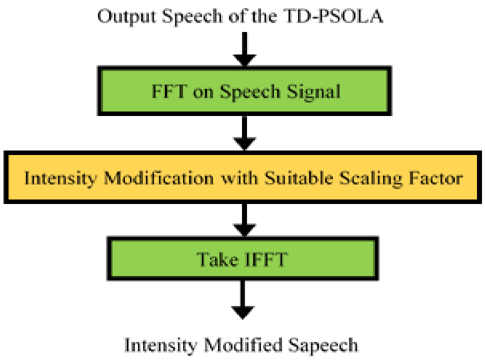

Figure 3. Intensity modification of speech signal

4. INTELLIGIBILITY TEST

Traditionally, speech intelligibility is defined as "the degree to which the speaker's intended message is recovered by the listener" [33]. In other words, speech intelligibility measures the comprehensibility of speech in given circumstances [34], [35]. Speech intelligibility can be assessed with subjective as well as objective methods. The subjective approach is a conventional and extensively used technique for evaluating speech intelligibility. In the subjective technique, various listeners are provided with impaired dysarthric speech for the listening test. Based on the feedback or scores from the listeners, the intelligibility of the dysarthric speech is assessed. This method is a little time-consuming, but it is inexpensive. The rating scale and word count methods, both of which are subjective methodologies for measuring intelligibility, are utilized in this study to determine the intelligibility of enhanced dysarthric speech [36]–[38].

4.1. Rating scale method

The rating scale method is a scaling-based method in which listeners listen to the impaired and enhanced speech samples of speakers and, based on that, give a score for the intelligibility of speech [37]. Every listener repeats this intelligibility test for all ten dysarthric speakers. As per the literature, the 6-point scale is more sensitive than the 7-point scale [36]. Therefore, the 6-point scale is used in this research, and this 6-point intelligibility scale is provided in Table 4.

Table 4. Speech intelligibility scale

Intelligibility Rating	Description
5	Normal Speech
4	Speech is understood without difficulty but sounds abnormal
3	Speech is understood with a little effort. Repetitions needed occasionally
2	Can be understood with attentiveness and effort by a listener, requires more than one repetitions
1	Can be understood with effort if the context is known
0	Cannot be understood even if the context is known

After performing the above task, the average of every listener's rating for each speaker's impaired and enhanced speech is calculated individually. On this basis, the enhancement in intelligibility for each speaker is computed. This test was conducted with a total of 25 listeners.

Figure 4 is a graphical representation of the results of this rating scale test. In Figure 4, the average rate scale value of impaired speech is represented by a bar filled with dot patterns and enhanced speech by a bar filled with slanted lines. Finally, for every speaker, the percentage change in intelligibility is shown with a bar filled with check patterns. The rating scale test method is simple to use and quick to complete.

4.2. Word count method

A word count method is an alternative way to measure speech intelligibility [37]. In this research, a word count method also validates the results of the rating scale method. The rating scale method is universally accepted. However, the word count method provided a more objective measure of intelligibility than the rating scale method [32]. This word count method is based on the total number of correctly recognized words from the set of impaired and enhanced dysarthric speech samples. In this test, every listener listens to the impaired speech sample and notes down the utterances they hear. Then, the listener listens to an enhanced speech sample of the same speaker and notes down the utterances they hear. For every listener, this process is repeated for all ten speakers. First, the number of correctly identified words from impaired and enhanced utterances is counted for each speaker and logged in the observation table. Then, the change in intelligibility for each speaker is calculated using the number of correctly recognized words from the impaired and enhanced speech sets. Following that, each speaker's overall change in intelligibility is calculated by averaging the changes in intelligibility obtained from all 25 listeners. The result of the word count test is shown in Figure 5, in which the average change in intelligibility for each speaker is shown by a bar.

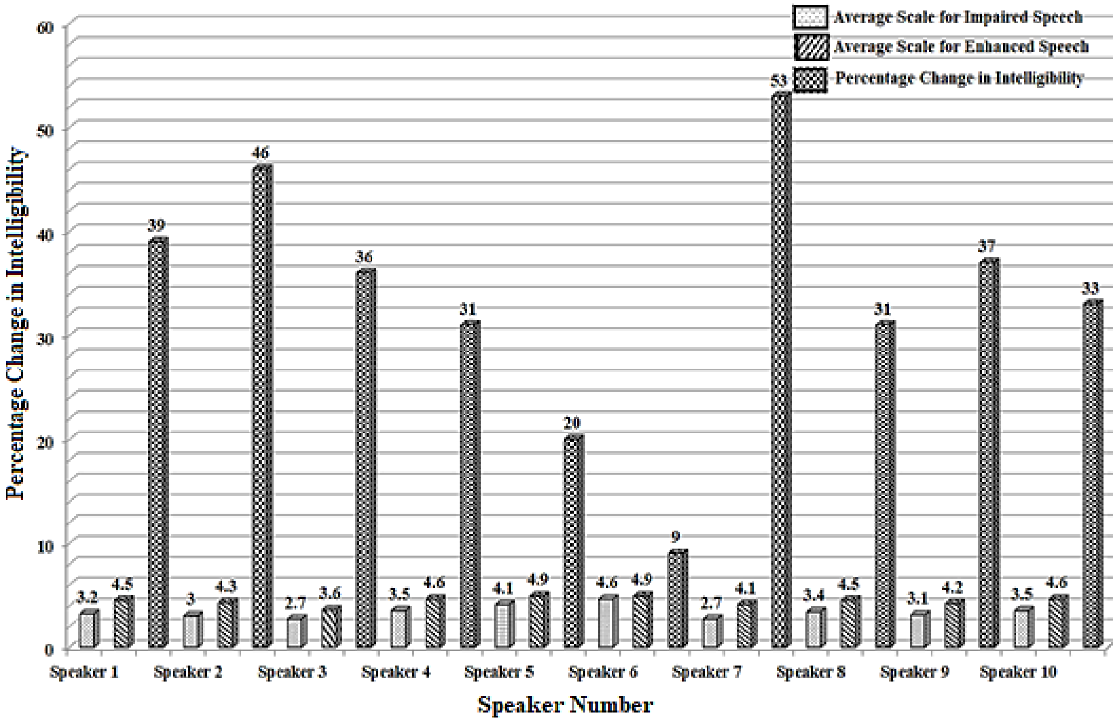

Figure 4. Result of rating scale method

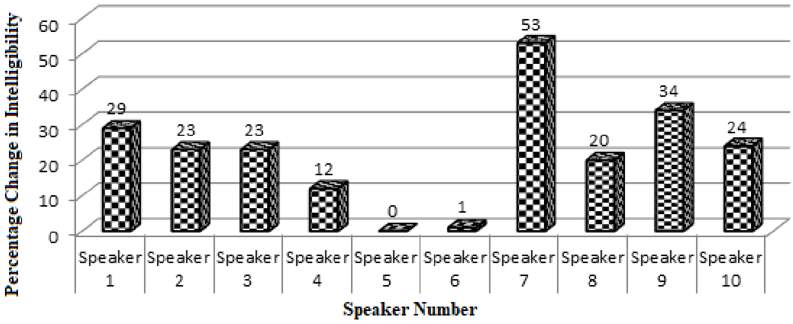

Figure 5. Result of word count method

5. CONCLUSION

Following the study of the acoustic parameters of unimpaired and impaired dysarthric speech, it can be inferred that dysarthric speakers' fundamental frequency, intensity, and speech rate differ from those of unimpaired speakers. In this research, an attempt to enhance the intelligibility of dysarthric speech is made by modifying these parameters. First, the fundamental frequency and intensity of dysarthric speech are modified using the TD-PSOLA algorithm. Then, the intensity of dysarthric speech is modified with FFT and IFFT-based approaches. The TD-PSOLA algorithm's accuracy for modification of fundamental frequency is evaluated and observed to be 97%. The intelligibility modifications are then evaluated using subjective listening tests. The rating scale test and word count test were carried out for a total of 25 listeners. For ten dysarthric speakers, the change in intelligibility computed using the rating scale and word count tests are satisfactory. The change in intelligibility using the rating scale method ranges from 9% to 53%. Whereas for the word count test, this change in intelligibility ranges from 0% to 53%. It has been observed that the change in intelligibility is less in speakers whose original intelligibility is good. However, for speakers with poor intelligibility, in their case, the change in intelligibility is satisfactory. As a result, it can be inferred that by modifying the fundamental frequency, intensity, and speech rate, the intelligibility of dysarthric speech can be enhanced.

6. FUTURE WORK

In the future, real-time implementation of this method needs to be the focus. The processing time is the primary constraint for real-time systems. Therefore, processing time must be kept to a minimum during real-time implementation. In addition, after real-time implementation, the intelligibility assessment for impaired and enhanced speech also needs to be carried out.

ACKNOWLEDGEMENT

The authors are thankful to the Department of Audiology at Ali Yavar Jung National Institute for the Hearing Handicapped, Mumbai, and the Audiology and Speech Therapy Department at All India Institute of Physical Medicine and Rehabilitation, Mumbai, for their assistance in creating a dysarthric speech database. Additionally, the authors would like to thank Prof. Mark Hasegawa-Johnson of the ECE Department at the University of Illinois in Chicago for permitting us to use their database for research purposes.

REFERENCES

[1] P. Giulio, "Dysarthria: Definition, clinical contexts, neurobiological profiles and clinical treatments," *Archives of Community Medicine and Public Health*, pp. 138–141, Jul. 2020, doi: 10.17352/2455-5479.000094.

[2] A. A. Joshy and R. Rajan, "Automated dysarthria severity classification using deep learning frameworks," in *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, pp. 116–120, doi: 10.23919/Eusipco47968.2020.9287741.

[3] R. Patel, "Identifying information-bearing prosodic parameters in severely dysarthric vocalizations," Doctoral Thesis, University of Toronto, 2000.

[4] T. Ashizawa and G. Xia, "Ataxia," *CONTINUUM: Lifelong Learning in Neurology*, vol. 22, no. 4, pp. 1208–1226, Aug. 2016, doi: 10.1212/CON.0000000000000362.

[5] L. Rampello, L. Rampello, F. Patti, and M. Zappia, "When the word doesn't come out: A synthetic overview of dysarthria," *Journal of the Neurological Sciences*, vol. 369, pp. 354–360, Oct. 2016, doi: 10.1016/j.jns.2016.08.048.

[6] R. Palmer and P. Enderby, "Methods of speech therapy treatment for stable dysarthria: A review," *Advances in Speech Language Pathology*, vol. 9, no. 2, pp. 140–153, Jan. 2007, doi: 10.1080/14417040600970606.

[7] I. V. McLoughlin, *Speech and audio processing*. Cambridge University Press, 2016.

[8] F. Rudzicz, "Articulatory knowledge in the recognition of dysarthric speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 947–960, May 2011, doi: 10.1109/TASL.2010.2072499.

[9] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, and J. R. Duffy, "Acoustic studies of dysarthric speech," *Journal of Communication Disorders*, vol. 32, no. 3, pp. 141–186, May 1999, doi: 10.1016/S0021-9924(99)00004-0.

[10] B. Linders, G. G. Massa, B. Boersma, and P. H. Dejonckere, "Fundamental voice frequency and jitter in girls and boys measured with electroglottography: influence of age and height," *International Journal of Pediatric Otorhinolaryngology*, vol. 33, no. 1, pp. 61–65, Aug. 1995, doi: 10.1016/0165-5876(95)01197-J.

[11] M. P. Giri and N. Rayavarapu, "Assessment on impact of various types of dysarthria on acoustic parameters of speech," *International Journal of Speech Technology*, vol. 21, no. 3, pp. 705–714, Sep. 2018, doi: 10.1007/s10772-018-9539-0.

[12] H. Dahmani, S.-A. Selouani, D. O'shaughnessy, M. Chetouani, and N. Doghmane, "Assessment of dysarthric speech through rhythm metrics," *Journal of King Saud University-Computer and Information Sciences*, vol. 25, no. 1, pp. 43–49, Jan. 2013, doi: 10.1016/j.jksuci.2012.05.005.

[13] H. Martens, T. Dekens, L. Latacz, G. Van Nuffelen, W. Verhelst, and M. De Bodt, "Automated assessment and treatment of speech rate and intonation in dysarthria," in *Proceedings of the ICTs for improving Patients Rehabilitation Research Techniques*, 2013, pp. 382–384, doi: 10.4108/icst.pervasivehealth.2013.252366.

[14] S. Yang *et al.*, "The physical significance of acoustic parameters and its clinical significance of dysarthria in Parkinson's disease," *Scientific Reports*, vol. 10, no. 1, Jul. 2020, doi: 10.1038/s41598-020-68754-0.

[15] A. Prakash, M. R. Reddy, and H. A. Murthy, "Enhancement of continuous dysarthric speech," in *Voice Technologies for Speech Reconstruction and Enhancement*, De Gruyter, 2020, pp. 35–66.

[16] P. Prema, S. Vidhya, and T. L. Mathew, "Prototype for dysarthric speech enhancement," *Journal of Signal and Image Processing*, vol. 2, no. 2, pp. 50–53, 2011.

[17] A. B. Kain, J.-P. Hosom, X. Niu, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Improving the intelligibility of dysarthric speech," *Speech Communication*, vol. 49, no. 9, pp. 743–759, Sep. 2007, doi: 10.1016/j.specom.2007.05.001.

[18] H. Tolba and A. S. El_Torgoman, "Towards the improvement of automatic recognition of dysarthric speech," in *2009 2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 277–281, doi: 10.1109/ICCSIT.2009.5234947.

[19] A. Sivanupandian and D. J. Jennifer, "Speech enhancement of dysarthric speech signal," *International Journal of Research in Engineering and Science (IJRES)*, vol. 10, no. 5, pp. 31–35, 2022.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Praat.org. https://www.fon.hum.uva.nl/praat/ (accessed: Jan. 23, 2022)

[21] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 3, pp. 972–980, May 2006, doi: 10.1109/TSA.2005.858051.

[22] D. Govind and S. R. M. Prasanna, "Expressive speech synthesis: a review," *International Journal of Speech Technology*, vol. 16, no. 2, pp. 237–260, Jun. 2013, doi: 10.1007/s10772-012-9180-2.

[23] H. Kim *et al.*, "Dysarthric speech database for universal access research," *9th Annual Conference of the International Speech Communication Association*, Brisbane, Australia, 2008.

[24] 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, Dec. 2012, doi: 10.1007/s10579-011-9145-0.

[25] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell, "The Nemours database of dysarthric speech," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, 1996, vol. 3, pp. 1962–1965, doi: 10.1109/ICSLP.1996.608020.

[26] J. Laures-Gore, S. Russell, R. Patel, and M. Frankel, "The Atlanta motor speech disorders corpus: motivation, development, and utility," *Folia Phoniatrica et Logopaedica*, vol. 68, no. 2, pp. 99–105, 2016, doi: 10.1159/000448891.

[27] E. Castillo Guerra and D. F. Lovey, "A modern approach to dysarthria classification," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, 2003, pp. 2257–2260, doi: 10.1109/IEMBS.2003.1280248.

[28] B. Akanksh, S. Vekkot, and S. Tripathi, "Interconversion of emotions in speech using TD-PSOLA," in *Advances in Intelligent Systems and Computing*, Springer International Publishing, 2016, pp. 367–378.

[29] W. Mattheyses, W. Verhelst, and P. Verhoeve, "Robust pitch marking for prosodic modification of speech using TD-PSOLA," *Proceedings of the IEEE Benelux/DSP Valley Signal Processing Symposium, SPS-DARTS*, pp. 43–46, 2006.

[30] M. Sondhi, "New methods of pitch extraction," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 2, pp. 262–266, Jun. 1968, doi: 10.1109/TAU.1968.1161986.

[31] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002, doi: 10.1121/1.1458024.

[32] R. Mochizuki and A. Kobayashi, "A low-band spectrum envelope modeling for high quality pitch modification," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 1, pp. I-645–8, doi: 10.1109/ICASSP.2004.1326068.

[33] R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek, "Toward phonetic intelligibility testing in dysarthria," *Journal of Speech and Hearing Disorders*, vol. 54, no. 4, pp. 482–499, Nov. 1989, doi: 10.1044/jshd.5404.482.

[34] S. Barreto and K. Ortiz, "Speech intelligibility in dysarthrias: influence of utterance length," *Folia Phoniatrica et Logopaedica*, vol. 72, no. 3, pp. 202–210, Apr. 2020, doi: 10.1159/000497178.

[35] K. M. Yorkston and D. R. Beukelman, "A comparison of techniques for measuring intelligibility of dysarthric speech," *Journal of Communication Disorders*, vol. 11, no. 6, pp. 499–512, Dec. 1978, doi: 10.1016/0021-9924(78)90024-2.

[36] P. S. Paithankar, "A comparative of perceptual and objective measures of speech intelligibility in children with hearing impairment," *International Journal of Scientific Development and Research (IJSDR)*, vol. 5, no. 11, pp. 71–75, 2020.

[37] N. Miller, "Measuring up to speech intelligibility," *International Journal of Language and Communication Disorders*, vol. 48, no. 6, pp. 601–612, Nov. 2013, doi: 10.1111/1460-6984.12061.

[38] H. M. Chandrashekar, V. Karjigi, and N. Sreedevi, "Spectro-temporal representation of speech for intelligibility assessment of dysarthria," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 390–399, Feb. 2020, doi: 10.1109/JSTSP.2019.2949912.

BIOGRAPHIES OF AUTHORS

Mahesh Giri received his B.E. degree in Electronics and Telecommunication from Savitribai Phule Pune University, Pune, India and his M.Tech. in Digital Electronics from the College of Engineering Pune, India. Currently, he is working as an assistant professor in the Department of Electronics and Telecommunication Engineering at METs Bhujbal Knowledge City, IOE, Nashik, India. In addition, he is a research scholar at Symbiosis International (Deemed University), Lavale, Pune, India. His research interests include signal processing, digital signal processing, and speech processing. He is working on an "Investigation on dysarthric speech using acoustic parameters of speech". His teaching experience is over eight years, and his industrial experience is one year. He can be contacted at Officeofanandi@gmail.com.

Neela Rayavarapu received her BE degree in Electrical Engineering from Bangalore University, Bangalore, India, in 1984, her M.S. degree in Electrical and Computer Engineering from Rutgers, The State University of New Jersey, USA, in 1987, and her Ph.D. degree in Electronics and Communication Engineering in 2012 from Panjab University, Chandigarh. She has been involved in teaching and research in electrical, electronics, and communication engineering since 1987. Her areas of interest are digital signal processing and its applications and control systems. She can be contacted at neela.raya27@gmail.com.