Comparative analysis of Dysarthric speech recognition: multiple features and robust templates

Arunachalam Revathi[1] · R. Nagakrishnan[1] · N. Sasikaladevi[2]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Research on recognizing the speeches of normal speakers is generally in practice for numerous years. Nevertheless, a complete system for recognizing the speeches of persons with a speech impairment is still under development. In this work, an isolated digit recognition system is developed to recognize the speeches of speech-impaired people affected with dysarthria. Since the speeches uttered by the dysarthric speakers are exhibiting erratic behavior, developing a robust speech recognition system would become more challenging. Even manual recognition of their speeches would become futile. This work analyzes the use of multiple features and speech enhancement techniques in implementing a cluster-based speech recognition system for dysarthric speakers. Speech enhancement techniques are used to improve speech intelligibility or reduce the distortion level of their speeches. The system is evaluated using Gamma-tone energy (GFE) features with filters calibrated in different non-linear frequency scales, stock well features, modified group delay cepstrum (MGDFC), speech enhancement techniques, and VQ based classifier. Decision level fusion of all features and speech enhancement techniques has yielded a 4% word error rate (WER) for the speaker with 6% speech intelligibility. Experimental evaluation has provided better results than the subjective assessment of the speeches uttered by dysarthric speakers. The system is also evaluated for the dysarthric speaker with 95% speech intelligibility. WER is 0% for all the digits for the decision level fusion of speech enhancement techniques and GFE features. This system can be utilized as an assistive tool by caretakers of people affected with dysarthria.

Keywords Gamma tone energy features · Modified group delay Cepstrum · Stockwell cepstral features · Dysarthric speech recognition · Speech enhancement · Clusters

✉ Arunachalam Revathi
    revathi@ece.sastra.edu

1   Department of ECE/SEEE, SASTRA Deemed University, Thanjavur, India

2   Department of CSE/SoC, SASTRA Deemed University, Thanjavur, India

1 Introduction

It is observed that dysarthric speakers are finding difficulty in producing speech because their muscles are weak. In addition, because of the neurological defects caused due to the degenerative motor disorder, dysarthric speakers exhibit speech impediments in speaking. Human communication occurs through speech, emotions, facial expressions, and sign language.

1.1 Related works

Speeches uttered by dysarthric speakers become distorted and disordered, making it difficult for others to understand their speeches. Dysarthric speakers take more time in uttering a single word, and the speaking rate is low. They do not have proper control over - the volume, loudness, articulatory movements in producing speech. The Recognition system [22, 26] recognizes the speeches of people affected with cerebral palsy by using acoustic and language models. The severity of the problem of dysarthric speakers is assessed to develop the system [27] to recognize their speeches. The voice conversion process [2] is used for dysarthric people. Kernel transformation and locality sensitivity Discriminant analysis transforms the spectral features into discriminative phoneme features. Transformation of normal speeches into dysarthric speeches to augment the data [11] and machine learning classifiers are used for classification. Speech recognition is implemented [25] using Convolutional networks. A combination of modeling techniques [7] is used for speech recognition to ensure better performance. Rhythmic knowledge [23] of dysarthric speeches improves the system's accuracy. The connectionist approach and speech recognition assess the problem associated with dysarthria using the Hidden Markov model (HMM). It is proposed to have a system [20] to transform the speech of a speaker with speech impairments into a more intelligible form. Voice conversion based on Non-negative matrix factorization [1] is better than Gaussian mixture modeling-based. Integration of acoustic models with articulatory information is used to improve the accuracy of the dysarthric speech recognition system. The problem of dysarthric speakers in speaking is analyzed [28] using deep learning networks. Modification is done on dysarthric speeches by temporal and spectral parameters [18, 19] to improve intelligibility. The severity of the problems associated with dysarthric speakers are analyzed [14], and HMM is used for modeling & classification. In this work on speaker-independent dysarthric speech recognition, Gamma tone energy features, Modified group delay cepstrum, stock well features, and cluster-based modeling techniques are used to assess the performance of the dysarthric speech recognition system. Performance is augmented by utilizing speech enhancement techniques before extracting features from speech. In the previous literature, speech enhancement techniques are not exploited in dysarthric speech recognition. So, our work emphasizes the use of existing speech enhancement techniques for improving the performance of the dysarthric speech recognition system.

2 Dysarthric speech database [12]

Evaluation of any speech-based recognition system is done using the metrics such as recognition rate and time taken for computation. Speech utterances of dysarthric speakers aged between 18 and 51 are taken. Speech utterances of two female speakers with 6% and 95% speech fluency are taken for evaluating the system. Four normal speakers assess the subjective

evaluation of their speeches, and the test results are taken for effective comparison between manual recognition and recognition by an automated system. Experimental evaluation outperforms manual recognition for the speaker with 3% speech intelligibility. Experimental and subjective evaluation performs equally for the speaker with 95% intelligibility. A table listing the dysarthric speakers considered in our work with their intelligibility scores [12] is indicated in Table 1. Speech intelligibility is assessed by performing a transcription of words by human listeners.
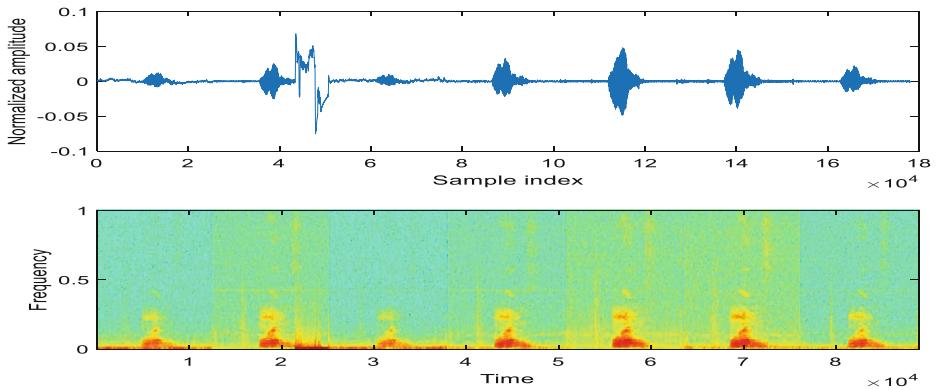
3 Analysis – Dysarthric speech

It is essential to analyze the speeches of dysarthric speakers and normal speakers. This is performed to assess the distortion level and degree of variation of dysarthric speeches with normal speeches. There is a variation in the pronunciation of utterances among dysarthric speakers. Depending on the speech intelligibility of dysarthric speakers, speeches uttered by them would have significant variations in style, obscurity, comprehension, and so on. So, manual recognition of their speeches would prove to be futile. Hence, it is necessary to develop an automated system to recognize their speeches and act as a translator facilitating others to help them provide necessary assistance. To show the characteristics of the speech uttered by a dysarthric speaker, spectrogram analysis is done on the speech utterances corresponding to the isolated digit "one." Figure 1 describes the speech along with its spectrogram for the digit "one" uttered by one of the dysarthric speakers considered in our work. Seven repeated trials of uttering the same word show more significant variations in volume, loudness, amplitude, and frequency components.

For effective comparison, utterances spoken by normal speakers are taken, and spectral analysis is done. Figure 2 gives the details about the seven trials of the utterance for the isolated digit "one" spoken by one of the normal speakers in the "TIMIT" database [8]. Speeches are pronounced uniformly with no variations in maximum amplitude, loudness, and frequency components. However, there is a pause between the utterances. So, the subjective and experimental assessments have yielded promising results for recognizing speeches of normal speakers.

Correlation analysis is done to understand the articulation of dysarthric and normal speakers uttering the words. Figure 3 indicates the results of the correlation analysis performed with speech utterances of dysarthric speakers. The amplitude of each sample in cross-correlated signal does not resemble the original utterance of dysarthric speakers, and it is revealed that speeches of the same utterance of the same dysarthric speaker are disjoint.

Table 1 Dysarthric speakers with speech intelligibility levels

| Speaker	Age	Speech intelligibility (%)
M01	⋗ 18	Very low (10%)
M04	⋗ 18	Very low (2%)
M07	58	Low (28%)
M09	18	High (86%)
F05	22	High (95%)
F03	51	Very low (6%)

Fig. 1 Illustration of variations in Speech signal with spectrogram – Dysarthric speaker (M07)

Figure 4 depicts the results of performing correlation analysis between speech utterances of normal persons [28]. When the samples of the cross-correlated signal are analyzed, it is revealed that utterances of the same normal speakers show a resemblance with a speaker's utterances.

4 Implementation of speaker independent speech recognition system

An automated system is developed to recognize isolated digits uttered by dysarthric speakers in this work. Speaker-dependent speech recognition systems are meant for persons with impairments. However, speaker-independent systems are developed; they would be universally used to recognize any dysarthric speaker's speeches. The training phase implemented in this work contains the feature extraction and template creation phases. The feature extraction phase encompasses the extraction of prospective features that are as consistent as possible within a speaker and vary as much as possible between speakers. Features should be uniquely or
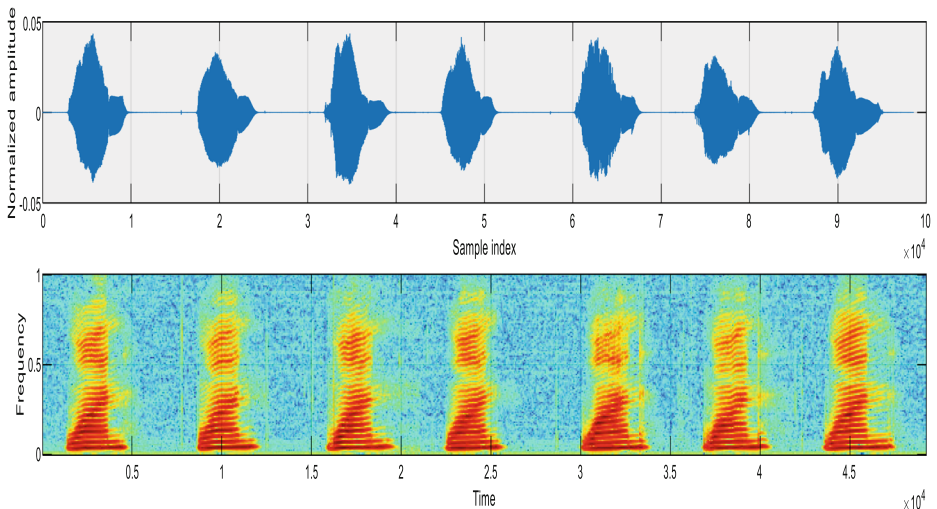
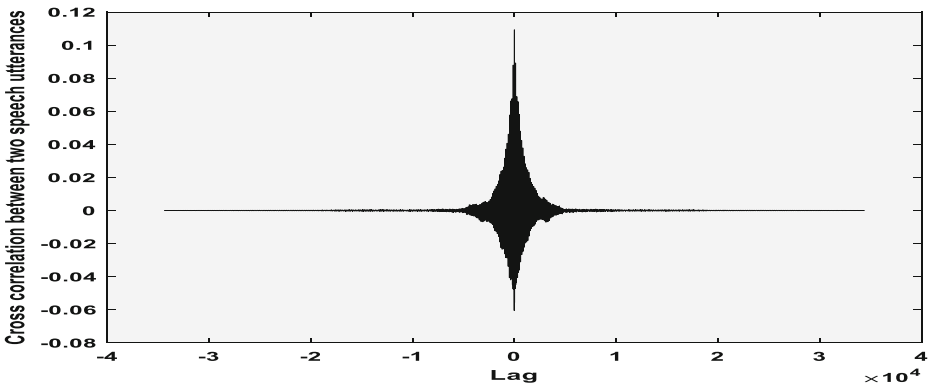**Fig. 2** Illustration - Speech signal and spectrogram – Normal speaker (TIMIT database) F5

Fig. 3 Correlation analysis – speech utterances of dysarthric speaker M07

specifically describing the characteristics of speakers. The success of the system also depends on the usage of modeling techniques. Modeling techniques are appropriately implemented to improve the success rate of the automated speech recognition system, and templates are created also should be unique to the speech/speaker. It is required that the test feature vectors are applied to the templates, and based on the type of classifier used, test speech is identified as associated with the pertinent template. So, the system's performance depends on the feature selection, robust template creation technique, and appropriate testing procedure that can be used to recognize the speeches of any speaker. It is also essential to use the techniques to improve the distorted speeches of dysarthric speakers comprehensively for better system performance.

4.1 Speech enhancement techniques

Various algorithms implemented using signal processing techniques aim to improve speech quality. These techniques improve the intelligibility or overall perceptual quality of the degraded speech. Speech enhancement implements noise removal techniques to improve the intelligibility of distorted speech. Speech enhancement finds applications in mobile phones, voice-over-internet protocol, audio teleconferencing, and digital hearing aids.
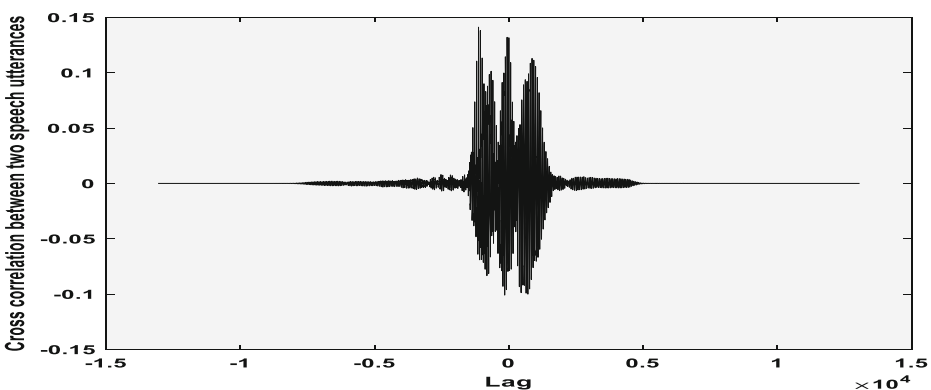

Fig. 4 Correlation analysis – speech utterances of normal speaker F5

The noisy speech is represented by the Eq. (1)

$$Y = X + N \tag{1}$$

Y Noisy speech vector.
X Clean speech vector.
N Noise vector.

Speech enhancement techniques used in our work are single-channel online speech enhancement [4], speech enhancement using minimum mean squared error log spectral amplitude and spectral amplitude extractor [5, 6], speech enhancement using wavelets [13], speech enhancement using probabilistic geometric approach & geometric approach [10, 15], and speech enhancement using phase spectrum compensation [24] for reducing distortion level of the speeches.

4.2 Feature extraction

Speech is considered a random signal defined as variations in acoustic pressure as a function of time. Features extracted from the speech are classified as the source and system features. Source features are extracted from the excitation signal. System features are normally cepstral features derived from the speech by analyzing the frequency domain. Filter banks in different frequency scales are designed to simulate the human ear mechanism of perceiving sounds. Features used for evaluating the work are Gamma-tone energy (GFE) features with filters scaled in BARK, MEL, and equivalent rectangular bandwidth (ERB) scales, MGDFC, and Stockwell features. The cochlea in-ear is a vital organ to receive sounds by using the basilar membrane present in it. Different frequency components are responded by the basilar membrane with vibration in different positions according to the intensity of sounds. Gamma tone filter bank used for Gamma tone energy feature extraction [17] would better simulate a mechanism of the basilar membrane to perceive speech as a representation of the sequence of sounds. Short-time phase spectra can also be used to analyze the characteristics of the speech, like short-time magnitude spectra. Transitions in the phase spectrum of the signal constitute formant frequencies as resonances of the vocal tract. Modified group delay cepstrum (MGDFC) [9, 16] is a feature extracted from speech. Discrete orthogonal stock well transform cepstrum (DOSTC) and discrete cosine Stockwell transform cepstrum (DCSTC) [16, 21] are obtained by performing transformation in DFT/DCT domains, bandwidth portioning, band selection, and IDFT/IDCT on selected frames of the pre-processed speech. The pre-processing techniques include voice activity detection, pre-emphasis, frame blocking, and windowing modules.

4.3 Implementation of template creation module

For any recognition system, it is needed to have models with a reduced set of parameters. Various modeling techniques create templates. Vector quantization (VQ) [3, 16, 17] based clustering technique is used to produce the reduced size of the cluster set from the larger size training data for each isolated digit uttered by a dysarthric speaker. During testing, feature vectors are extracted from the test speech after performing analysis based on pre-processing techniques and applied to the models. Euclidean distance is computed between the test vector and cluster centroids, and the minimum distance is derived. Finally, the digit is recognized as corresponding to the model based on a minimum of averages.

4.4 Experimental evaluation – Results and discussion

An isolated digit recognition system for dysarthric speakers is evaluated by considering Gamma tone energy, Modified group delay cepstrum, and Stockwell cepstrum as features and VQ models and minimum distance classifier. Speech enhancement techniques are used to enhance the quality of distorted dysarthric speeches, and performance is analyzed for each technique for all features and models. The proposed speech recognition system consists of two phases for training and testing, respectively. Twenty-nine speech utterances for each digit spoken by five dysarthric speakers (M09, M07, M04, M01, F05) are taken for training, and seven speech utterances in each digit spoken by the speaker F03 are considered for testing. Speech utterances are concatenated and taken directly, and features are extracted. Enhanced speech is obtained by applying speech enhancement techniques on the concatenated speech, and features are extracted from the enhanced speech. Models are created for features of the original speech and enhanced speech. These models are stored as templates for each isolated digit. Test speech utterances are concatenated, features are extracted. Speech enhancement techniques are applied on the concatenated test speech, and features are extracted for enhanced speech. Test feature vectors are applied to the appropriate templates, and an isolated digit is recognized based on the minimum distance classifier. A decision level fusion classifier is used for Gamma tone energy features/MGDFC & Stockwell cepstrum and model by integrating results of all speech enhancement techniques. Table 2 indicates the type of speech enhancement technique used in our work. The spectrogram in Fig.5 depicts the characteristics of speeches in the time-frequency domain by applying speech enhancement techniques.

A decision-level fusion classifier is depicted in Fig.6. Table 3 details the experimental evaluation of speech enhancement techniques by using perceptual evaluation of speech quality (PESQ) as a metric. PESQ is computed between the speech utterance for the digit "one" uttered by the dysarthric

Table 2 Speech Enhancement Techniques Description

Enh-1	Single-channel online speech enhancement
Enh-2	Speech enhancement using mean squared error log spectral amplitude extractor
Enh-3	Speech enhancement using mean squared error spectral amplitude extractor
Enh-4	Speech enhancement using wavelet denoising
Enh-5	Speech enhancement using a probabilistic geometric approach
Enh-6	Speech enhancement using a geometric approach
Enh-7	Speech enhancement using phase spectrum compensation
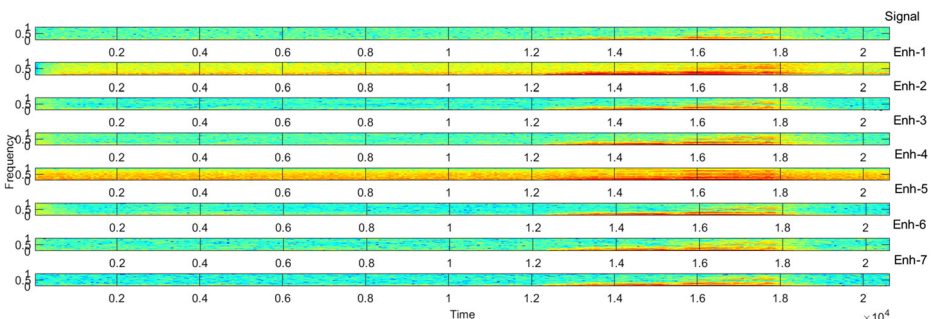
Fig. 5 Spectrogram plots – Speech enhancement techniques
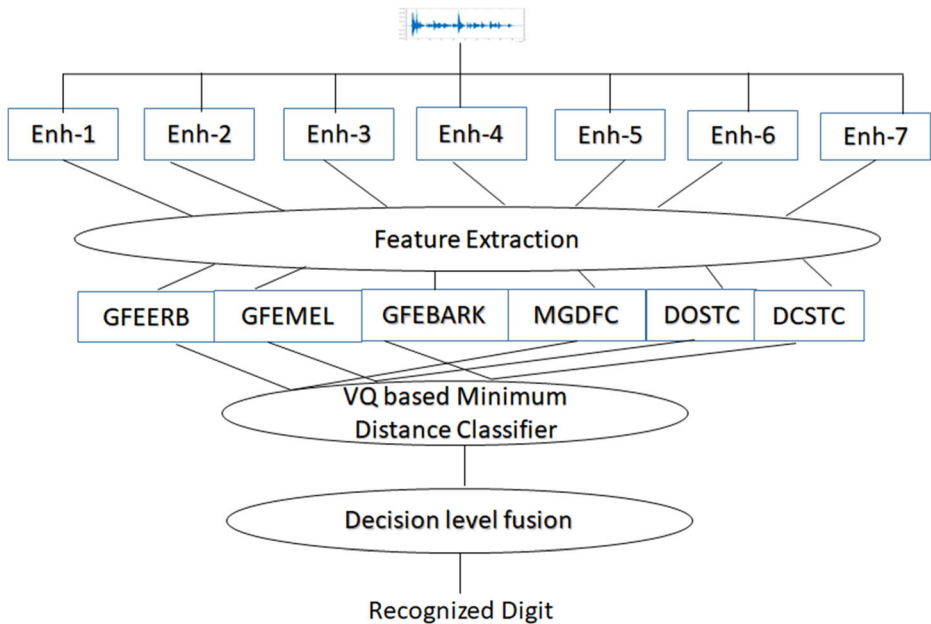
Fig. 6 Decision level fusion classifier

speaker with 6% and 95% speech intelligibility for all the speech enhancement mechanisms. Reference speech is the speech uttered by a dysarthric speaker with speech intelligibility, and degraded speech is the one uttered by a dysarthric speaker with 6% intelligibility.

This decision level fusion classifier classifies the pertinent speech based on the correct classification for any features, modeling technique, and speech enhancement techniques. Evaluation of the system is done for recognizing dysarthric speeches against the use of the feature with decision level fusion on speech enhancement techniques and the model. Individual word error rate (WER) for some isolated digits is 0%, with overall WER for decision level fusion of features, models, and speech enhancement techniques 4%. Individual WER for some isolated digits is relatively more because the testing is done with speech utterances of a dysarthric speaker with only 6% speech intelligibility. The system's accuracy can be enhanced by training the models with more data. Testing is also done for the speaker with 95% speech intelligibility. It is observed that overall WER is 0% for the speaker with 95% intelligibility. Hence, the system's accuracy depends on the use of test utterances from speakers with a level of intelligibility and the selection

Table 3 Assessment of speech enhancement techniques using PESQ

| Speech enhancement technique	PESQ
Enh-1	1.585303
Enh-2	1.542812
Enh-3	1.53715
Enh-4	1.442755
Enh-5	1.600728
Enh-6	1.470871
Enh-7	1.509319
Without speech enhancement	1.437014
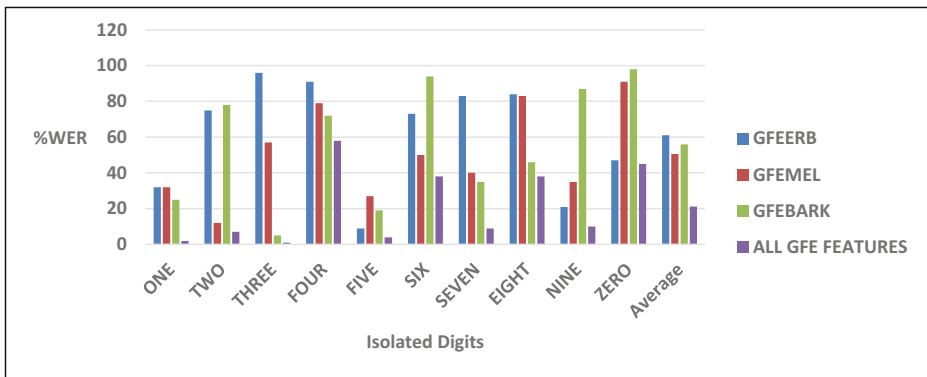
Fig. 7 Performance of the GAMMA TONE ENERGY features based dysarthric speech recognition system

of features, models, and speech enhancement techniques. Because of the variations in intelligibility levels of dysarthric speakers, the system poses difficulty in ensuring better accuracy for the speaker with low intelligibility for the individual consideration of features, model, and speech enhancement mechanisms But, the decision level fusion classification based on the integration of decision indices of features, model and speech enhancement techniques ensures the good WER of 4% besides 0% as individual WER for several isolated digits for the speaker with 6% intelligibility. Figure 7 indicates the system's performance for Gamma tone energy (GFE) features with filter bank analysis done in ERB, MEL, and BARK scales. Integration of features and speech enhancement techniques for the VQ model is 21% as overall word error rate (WER) for the speaker with 6% speech intelligibility.

Dysarthric speaker F03 (6% speech intelligibility).

Figure 8 indicates the system's performance for the integration of MGDFC, DCSTC, and DOSTC features and speech enhancement techniques for the speaker F03 with 6% speech intelligibility.
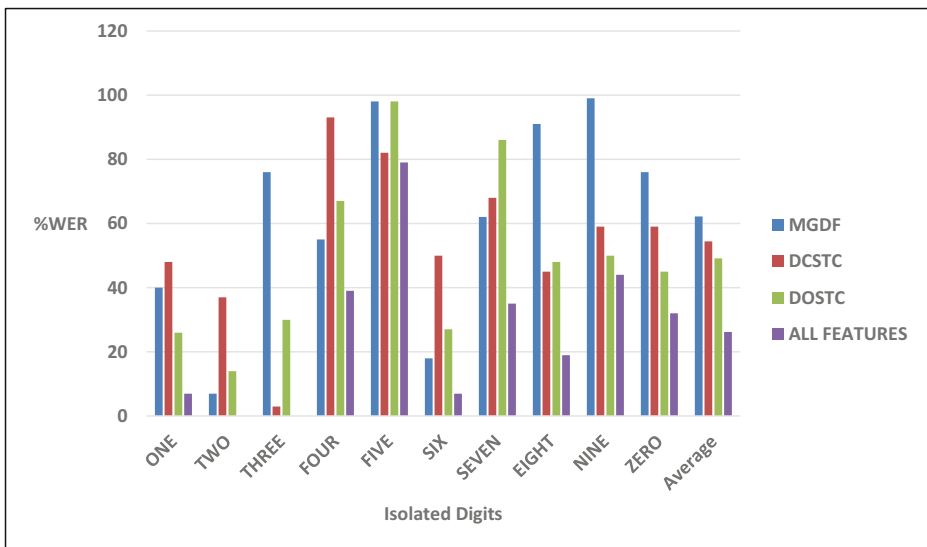
Fig. 8 Performance of the MGDFC, DCSTC, and DOSTC features based dysarthric speech recognition system Dysarthric speaker F03 (6% speech intelligibility)
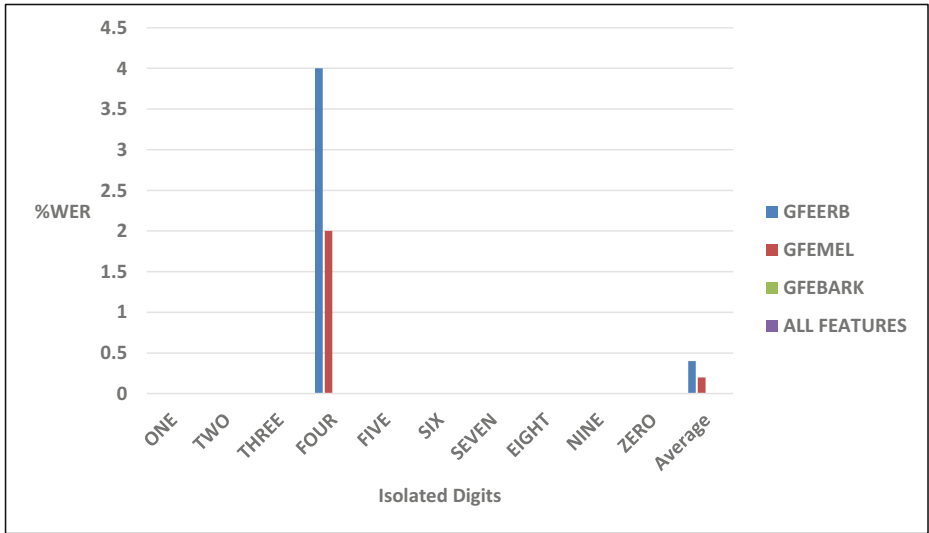
Fig. 9 Performance of the system (Gamma tone energy features) for the speaker F05 with 95% speech intelligibility

Figure 9 depicts the system's performance for Gamma tone energy features for the speaker with 95% speech intelligibility.

Figure 10 gives the performance analysis of the system for MDGFC, DCSTC, and DOSTC features and integrated speech enhancement for the female speaker with 95% speech intelligibility.

Figure 11 describes the evaluation of the system for decision level fusion of features and speech enhancement techniques for the VQ-based minimum distance classifier.
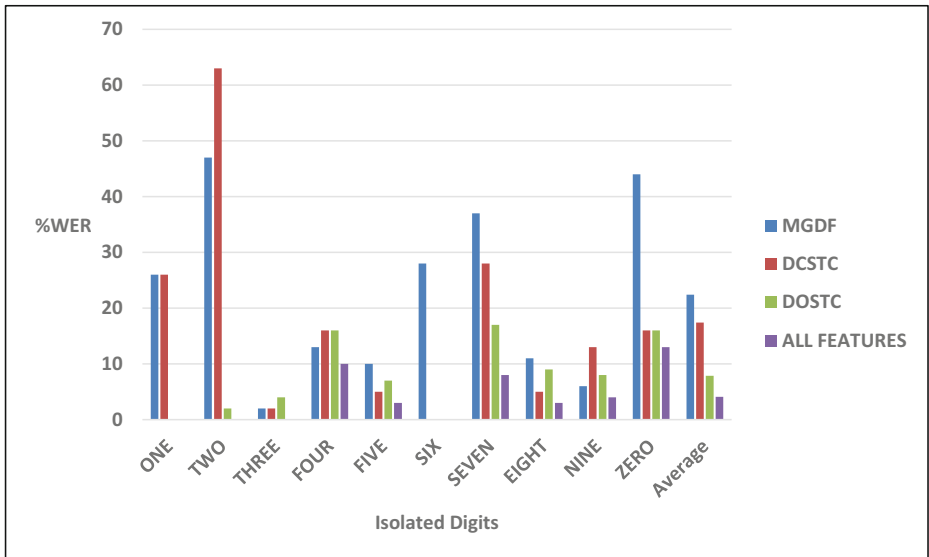
Fig. 10 Performance of the system (MGDFC, DCSTC, and DOSTC features) for speaker F05 with 95% speech intelligibility
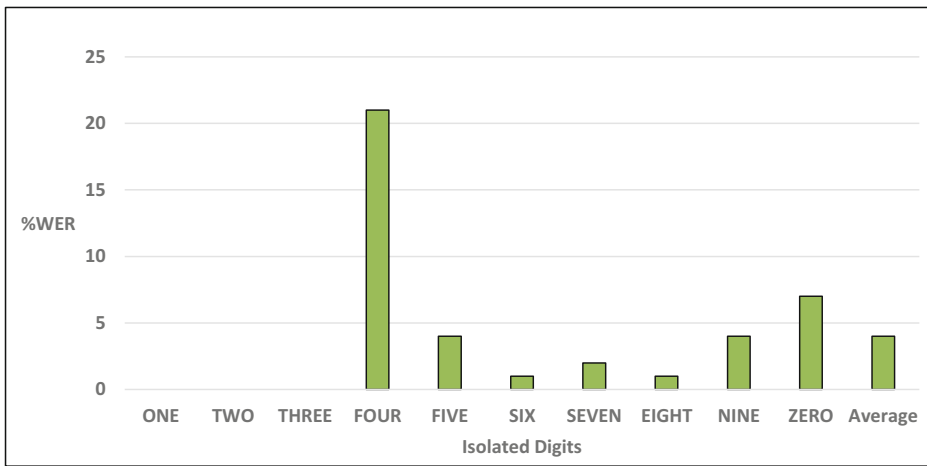
Fig. 11 Performance of the decision level fusion system for the speaker F03 with 6% speech intelligibility

The decision level fusion classifier performs well for a dysarthric speech recognition system ensuring that WER is less than 10% for all the isolated digits except the digit 'Four.' Fig.12 gives comparative analysis between experimental and subjective assessment of dysarthric isolated digit recognition system for the speaker F03 with 6% speech intelligibility. For many isolated digits, experimental evaluation outperforms the manual recognition of isolated digits. However, experimental evaluation and subjective assessment have provided 0% WER for the speaker F05 with 95% speech intelligibility for all the isolated digits. WER is 100% for manually recognizing the isolated digits' two' and 'five.' WER is less than 2% for experimental assessment of these digits. It is ensured that experimental evaluation using the features, models, and speech enhancement techniques has provided better accuracy than manual recognition. Four normal speakers are instructed to spontaneously listen to the utterances in each isolated digit spoken by the F03 speaker, and subjective evaluation is done accordingly.
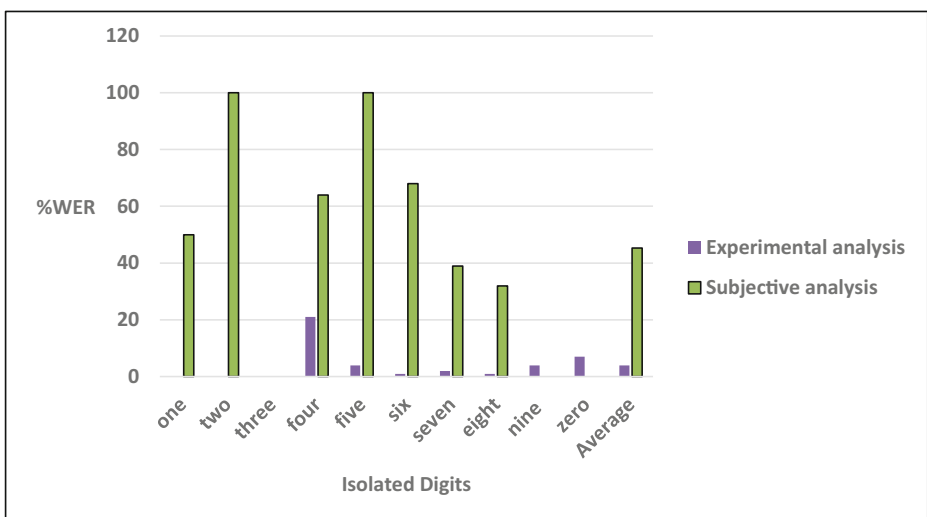

Fig. 12 Comparative analysis – Experimental and subjective assessment – F03 dysarthric speaker

Table 4 Performance evaluation – integration of GFE features for speech enhancement techniques

Speech enhancement technique	one	two	three	four	five	six	seven	eight	nine	zero	Average
No enhancement	63.6	67.9	14	91	41	85	63	79	83	99	69
Single-channel online speech enhancement	39.8	67.9	53	90	62	77	52	99	39	85	66
Speech enhancement using mean squared error log spectral amplitude extractor	76.1	56.8	49	97	32	96	78	92	69	95	74
Speech enhancement using mean squared error spectral amplitude extractor	47.7	35.8	13	83	27	84	87	87	76	##	64
Speech enhancement using wavelet denoising	31.8	79	19	95	26	82	34	88	55	97	61
Speech enhancement using a probabilistic geometric approach	44.3	81.5	85	97	14	89	72	86	60	66	69
Speech enhancement using a geometric approach	55.7	85.2	11	72	6.7	85	77	99	31	88	61
Speech enhancement using phase spectrum compensation	47.7	37	60	95	23	72	69	47	40	99	59

Table 4 describes the efficiency of the dysarthric speech recognition system with the integration of GFE features for the speech enhancement techniques considered separately. It indicates the performance of the isolated digit recognition system in terms of WER. Average WER is less for the system with phase spectrum compensation speech enhancement technique.

Table 5 indicates the comparative analysis between our proposed method and the paper [26].

Table 5 Comparative analysis – Proposed work and the paper [26]

Paper	Parameters considered
[26]	Task – Character recognition for dysarthric speeches (speaker-dependent) Dataset – ATR Japanese speech database and number of words – 216,432,648,864 Features used – MFCC and its velocity and acceleration coefficients Modeling technique used - RNN. Performance metric – CER- Dysarthric speaker 1 – Top- 1- 23.2% and Top-3 -19.2% CER-Dysarthric speaker 2 – Top-1-30% and Top-3 -24.9%
Our proposed method	Task – Isolated digit recognition for dysarthric speeches (Speaker independent) Dataset – TORGO database and number of isolated digits - 10 Features used – GFEERB,GFEMEL,GFEBARK, MGDFC,DCSTC and DOSTC Modeling technique used – VQ Speech enhancement techniques used – 7 The performance metric used – WER F03 speaker with 6% speech intelligibility Decision level fusion of speech enhancement techniques for GFE features – 21% WER Decision level fusion of speech enhancement techniques for MGDFC, DOSTC, and DCSTC features −26% WER F05 speaker with 95% speech intelligibility Decision level fusion of speech enhancement techniques for GFE features – 0% WER Decision level fusion of speech enhancement techniques for MGDFC, DOSTC, and DCSTC features −4% WER Experimental and Subjective assessment F03 speaker – Integration of features and speech enhancement techniques – 4% WER - Experimental assessment F03 speaker – 45.3% WER - Subjective evaluation

5 Conclusion

This paper discusses the use of speech enhancement techniques to improve the quality of distorted dysarthric speeches to augment the recognition system's performance to identify the isolated digits uttered by dysarthric speakers. This automated system uses feature extraction, VQ-based modeling technique, and speech enhancement techniques to assess the performance of the dysarthric speech recognition system by using test utterances of speakers with 6% and 95% speech intelligibility. The feature extraction process imbibes the extraction of Gamma tone energy features with filters in filter bank calibrated in ERB, MEL, and BARK scale, modified group delay cepstrum, and Stockwell Cepstral features. Templates are created using the VQ-based modeling technique for the training set of feature vectors. Test utterance is classified as corresponding to the pertinent digit by using the minimum distance classifier. Test results are compared with manual recognition of isolated digits uttered by the speakers considered for testing. The decision level fusion classifier implemented using the integration of features and speech enhancement techniques yielded better results than the subjective assessment of test utterances corresponding to the isolated digits. Experimental evaluation outperforms the subjective assessment for testing the utterances of a dysarthric speaker with 6% intelligibility. However, the manual recognition and experimental evaluation have yielded the same results for the dysarthric speaker with 95% speech intelligibility. This automated system can be a translator to convert the speeches of dysarthric speakers into intelligible speeches so that caretakers can provide necessary assistance to the dysarthric speakers. It helps improve the social status of the dysarthric speakers and assists them in leading a respectful life in society.

Data availability All relevant data are within the paper and its supporting information files.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain any studies with human participants.

Competing interests The authors have declared that no competing interest exists.

Conflict of interests All the authors declare that they have no conflict of interest.

References

1. Aihara R, Takashima R, Takiguchi T et al (2014) A preliminary demonstration of exemplar-based voice conversion for articulation disorders using an individuality-preserving dictionary. J Audio Speech Music Proc 5(2014):1–10. https://doi.org/10.1186/1687-4722-2014-5
2. Aihara R, Takiguchi T, Ariki Y (2017) Phoneme-discriminative features for Dysarthric speech conversion. Proc Interspeech 2017:3374–3378 https://doi.org/10.21437/Interspeech.2017-664

3. Arunachalam R (2019) A strategic approach to recognizing the children's speech with hearing impairment: different sets of features and models. Multimed Tools Appl 78:20787–20808. https://doi.org/10.1007/s11042-019-7329-6

4. Doire CSJ, Brookes M, Naylor PA, Hicks CM, Betts D, Dmour MA, Jensen SH (2017) Single-Channel online enhancement of speech corrupted by reverberation and noise. IEEE/ACM Trans Audio Speech Lang Proc 25(3):572–587. https://doi.org/10.1109/TASLP.2016.2641904

5. Ephraim Y, Malah D (1984) Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process 32(6):1109–1121. https://doi.org/10.1109/TASSP.1984.1164453

6. Ephraim Y, Malah D (1985) Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process 33(2):443–445. https://doi.org/10.1109/TASSP.1985.1164550

7. España-Bonet C, Fonollosa JA (2016) Automatic speech recognition with deep neuralnetworks for impaired speech. In: International Conference on Advances in Speech and Language Technologies forIberian Languages. Springer, Cham, pp 97–107. https://doi.org/10.1007/978-3-319-49169-1_10

8. Selouani SA, Dahmani H, Amami R, Hamam H (2012) Using speech rhythm knowledge to improve dysarthric speech recognition. Int J Speech Technol 15(1):57–64

9. Hegde RM, Murthy HA, Gadde VRR (2007) 'Significance of the modified group delay feature in speech recognition. IEEE Trans Audio Speech Lang Process 15(1):190–202 https://ieeexplore.ieee.org/document/4032772/

10. Aihara R, Takashima R, Takiguchi T, Ariki Y (2014) A preliminarydemonstration of exemplar-based voice conversion for articulation disorders using an individuality-preserving dictionary. Eurasip J Audio Speech Music Process 2014(1):1–10

11. Jiao Y, Tu M, Berisha V, Liss J (2018) Simulating Dysarthric Speech for Training Data Augmentation in Clinical Speech Applications. 2018 IEEE international conference on acoustics, speech, and signal processing (ICASSP), Calgary, pp 6009–6013. https://doi.org/10.1109/ICASSP.2018.8462290

12. Tu M, Berisha V, Liss J (2017) Interpretable objective assessment of dysarthric speech based on deep neural networks. In Interspeech, pp 1849–31853

13. Lallouani A, Gabrea M, Gargour CS (2004) Wavelet-based speech enhancement using two different threshold-based denoising algorithms, 1st edn. Canadian Conference on Electrical and Computer Engineering 2004 (IEEE Cat. No.04CH37513), Niagara Falls, Ontario, pp 315–318. https://doi.org/10.1109/CCECE.2004.1345019

14. Lee SH, Kim M, Seo HG, Oh BM, Lee G, Leigh JH (2019) Assessment of Dysarthria Using One-Word Speech Recognition with Hidden Markov Models. J Korean Med Sci 34(13):e108. Published 2019 April 8. https://doi.org/10.3346/jkms.2019.34.e108

15. Lu Y, Loizou PC (2008) A geometric approach to spectral subtraction. Int J Speech Commun 50(6):453–466. https://doi.org/10.1016/j.specom.2008.01.003

16. Revathi A, Sasikaladevi N (2019) Hearing impaired speech recognition: Stockwell features and models. Int J Speech Technol 22:979–991. https://doi.org/10.1007/s10772-019-09644-3

17. Revathi A, Sasikaladevi N, Nagakrishnan R, Jeyalakshmi C (2018) Robust emotion recognition from speech: Gamma tone features and models. Int J Speech Technol 21:723–739. https://doi.org/10.1007/s10772-018-9546-1

18. Rudzicz F (2011) Articulatory knowledge in recognition of Dysarthric speech. IEEE Trans Audio Speech Lang Process 19(4):947–960. https://doi.org/10.1109/TASL.2010.2072499

19. Islam MT, Shahnaz C, Zhu WP, Ahmad MO (2018) Enhancement of noisy speech with low speech distortion based on probabilistic geometric spectral subtraction. arXiv preprint arXiv:1802.05125

20. Rudzicz F (2013) Adjusting dysarthric speech signals to be more intelligible. J Comp Speech Lang 27(6):1163–1177. https://doi.org/10.1016/j.csl.2012.11.001

21. Stark AP, Wójcicki KK, Lyons JG, Paliwal KK (2008) Noise driven short-time phase spectrum compensation procedure for speech enhancement. In: Ninth annual conference of the international speech communication association

22. Kim H, Hasegawa-Johnson M, Perlman A, Gunderson J, Huang TS, Watkin K, Frame S (2008) Dysarthric speech database for universal access research. In: Ninth Annual Conference of the International Speech Communication Association

23. Sloane S, Dahmani H, Amami R et al (2012) Using speech rhythm knowledge to improve dysarthric speech recognition. Int J Speech Technol 15:57–64. https://doi.org/10.1007/s10772-011-9104-6

24. Revathi A, Sasikaladevi N, Nagakrishnan R, Jeyalakshmi C (2018) Robust emotion recognition from speech: Gamma tone features and models. Int J Speech Technol 21(3):723–739

25. Takashima Y, Nakashima T, Takiguchi T, Ariki Y (2015) Feature extraction using pre-trained convolutive bottleneck nets for dysarthric speech recognition. 2015 23rd European Signal Processing Conference (EUSIPCO), Nice, pp 1411–1415. https://doi.org/10.1109/EUSIPCO.2015.7362616

26. Takashima Y, Takiguchi T, Ariki Y (2019) End-to-end Dysarthric Speech Recognition Using Multiple Databases. ICASSP 2019–2019 IEEE international conference on acoustics, speech, and signal processing (ICASSP), Brighton, pp 6395–6399. https://doi.org/10.1109/ICASSP.2019.8683803

27. Thoppil MG, Kumar CS, Kumar A, Amos J (2017) Speech signal analysis and pattern recognition in diagnosing dysarthria. Ann Indian Acad Neurol 20:352–357 http://www.annalsofian.org/text.asp?2017/20/4/352/217159

28. Garofolo JS (1993) Timit acoustic phonetic continuous speech corpus. Linguistic Data Consortium, 1993