iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4553

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Speaker Recognition Based on Feature Extraction in Clean and Noisy Environment

Ms. Manpreet Kaur[1], Prof. Puneet Mittal[2]
[1,2]*Department of Computer Science and Engineering*
[1]*Research Scholar,* [1]*Maharaja Ranjit Singh Punjab Technical University, Bathinda, India*
[2]*Assistant Professor (CSE),* [2]*Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, India*

Abstract: ***To improve the performance of speaker identification systems, an effective and robust method is proposed to extract features for speech processing, capable of operating in the clean and noisy environment. For capturing the characteristics of the signal, the Mel-frequency Cepstral Coefficient along with RASTA of the wavelet channels is calculated. Then the proposed feature extraction algorithm is evaluated on the speech database for text-dependent and text-independent speaker identification using the Gaussian Mixture Model (GMM) and Vector Quantization (VQ) identifier. Gaussian Mixture Models (GMMs) were used for the recognition stage as they give better recognition rate for the speaker's features than Vector Quantization. Some popular existing feature extraction methods MFCCs, LPC, LPC+DWT, MFCC+RASTA are also evaluated for comparison in this paper. Comparison of the proposed approach with the conventional feature extraction methods shows that the proposed method not only effectively reduces the influence of noise but also improves recognition accuracy. In addition, the performance of our method is very satisfactory in the noisy environment. A recognition rate of 98.63% was obtained using the proposed feature extraction technique.***

Keywords: ***GMM, VQ, RASTA, MFCC, LPC, DWT, Recognition Rate.***

I. INTRODUCTION

Speaker recognition has been an interesting research field for the last decades. Basically, speaker recognition of particular speaker is based upon the individual information stored in the speech waves. A lot of research has been carried out in the past years in order to create the ideal which is able to understand continuous speech in real time, from different speakers and in any environment. There is a lot of information about the gender, emotion, language being spoken and identity of the speaker which can be retrieved from the speech signal. Speech Signal can be used for many speech recognition, speech processing applications especially security and authentication. The significance of speech recognition lies in its simplicity. This simplicity and ease of operating a device using speech have many advantages like security devices, household appliances, cellular phones, ATM machines, computers etc [3]. There exists a number of difficulties which arises during speaker recognition, which are the existence of unwanted noise signals from the speaker's surrounding environment and speaker variability such as gender, speaking style, the speed of speech [5]. Speaker recognition is divided into two phases which are speaker identification and speaker verification. The registered speaker is found out on the basis of speech input in speaker identification phase, while verification is the task of automatically determining if a person really is the person he or she claims to be. In This article, we are primarily interested in speaker recognition in the text-dependent mode of isolated words and continuous speech applied to the English Language. Speaker recognition systems are classified as text dependent and text independent. In Text-dependent system same text is being spoken in the training phase as well as in the testing phase, whereas there is no restriction on text being spoken in text independent system. Text-dependent systems are easier to implement and less time consuming as compare to text independent systems. As a result, the Recognition rate is much better in the text-dependent system than in text independent system. For this purpose, a corpus of words was recorded from 30 speakers in the English Language. The first dataset of isolated words containing words of English language from 20 speakers and a second dataset of continuous speech from 10 speakers were used to develop our recognition system. The purpose of automatic speaker recognition is to identify the speaker by extracting features, characterizing and recognizing the information contained in the speech signal. It depends on characteristics of speakers which are affected by both the behavioral characteristics and the physical structure of the individual's vocal tract. There are three main phases in speaker recognition which are Front-end Processing, Speaker Modeling, and pattern matching. Front-end processing is used to highlight the relevant features and remove the irrelevant ones. After the first process, we will get the feature vectors of the speech signals. Pattern Matching is performed to verify the identity claim of the speaker.

Every speaker recognition systems contain two modules: feature extraction and feature matching. Feature extraction is the process of extracting a small amount of data from the speech signal which can be used to represent each speaker. For feature extraction, different techniques are used which are Linear Predictive Coding (LPC), MFCC, MFCCRASTA. Most of the speech recognition systems use Mel-Frequency Cepstral Coefficients (MFCCs) to better reflect dynamic changes. Feature recognition phase involves the actual procedure to identify the unknown speaker. For recognition step, VQ, GMM is used. Generally, Every Speaker recognition system has two phases which are training phase and the testing phase. In the training phase, a reference model is made with collected speech samples of all the registered speakers. In the testing phase, the input speech is matched with stored reference models and a decision is made on the basis of recognition [2]. In this article, an effective and robust feature extraction method for speech signals is proposed. A combination of DWT, MFCC, RASTA was developed to efficiently extract features from real speech signals and then recognize them. This hybrid approach will help to increase the percentage of the accurate speech recognition. After pre-processing of speech signals, the wavelet transform is applied to decompose the input signal into two different frequency channels which are lower frequency approximations and higher frequency details. The components of the low-frequency channel are known as approximations, whereas the high-frequency channel components are known as details. Successive approximations are performed in the decomposition process to decompose the signal to the desired level. Second, for capturing the characteristics of the individual speakers, the MFCCs of the approximations and detail channels are calculated. Based on this mechanism, the multi-resolution features of the speech signal can be easily calculated using the wavelet decomposition. The proposed technique is used in the feature extraction stage of a text-dependent and text-independent speaker identification system. GMM is used for identification stage and compared to the VQ technique. The rest of this paper is organized as follows: Section 2 gives an overview of Related Work that is being done in the field of Speaker Recognition. Section 3 describes the Feature Extraction, Feature Matching Techniques of Speaker Recognition System. Section 4 shows the Experimental Results. Section 5 describes the Conclusion.

II. RELATED WORK

In literature survey, various techniques have been proposed to improve the performance of Automatic Speaker recognition (ASR) systems in the presence of noise. DWT and RASTA are the Speech enhancement techniques which effectively reduces the effect of noise either using statistical information of noise or filtering the noise from a noisy speech before feature extraction. Techniques like perceptual linear prediction and relative spectra incorporate some of the features of the human auditory mechanism and give noise robust ASR. Ahmed et al. [7] discussed the effect of environmental noise in speaker verification system. Robust feature extraction plays an important role in improving the performance of speaker verification system. In this different features are combined to increase recognition rate. Coefficients are extracted by using Mel frequency cepstral coefficient (MFCC) feature extraction technique from the discrete wavelet transform (DWT) of the speech, with and without feature warping for improving the performance of speaker verification in the presence of noise. Performance of speaker verification technique was evaluated using different feature extraction techniques. Results indicate that the fusion of DWT-MFCC and MFCC is superior to other feature extraction techniques in presence of environmental noise. Mahmoud et al. [8] proposed a new method for speech recognition which is based on the hybrid approach of combining two feature extraction techniques which are discrete wavelet transform (DWT) and Mel frequency cepstral coefficients (MFCCs) to enhance the performance of recognition system. This method is implemented for 15 male speakers uttering 10 isolated words each which are the digits from zero to nine. Each digit is repeated 15 times. Performance of proposed system is compared to Mel frequency cepstral coefficient based method for feature extraction. Neural Networks (NN) is used for classification. MFCC is the most popular feature extraction technique used for feature extraction in speaker identification system. In a clean environment, MFCCs provides good performance for speaker identification but in noisy environments, its performance degrades. The noise was also removed from the speech signal with the help of Wavelet which can lead to a good representation of stationary and non-stationary segments of the speech signal. Maurya et al. [9] implemented speaker recognition for the speech samples in Hindi Language. For, this purpose they used combination of Mel frequency cepstral coefficient-vector quantization (MFCC-VQ) and Mel frequency cepstral coefficient-Gaussian mixture model (MFCC-GMM) for both text-dependent and text-independent systems. Recognition Accuracy of text independent system for MFCC-VQ and MFCC-GMM is 77.64% and 86.27% respectively. Accuracy of text-dependent system is better than text-independent system. Accuracy of speech samples of Hindi language is 85.49% and 94.12% using MFCC-VQ and MFCC-GMM approach. Nehe and Holambe [10] proposed a DWT and LPC based technique for isolated word recognition. The coefficients have been derived from the speech frames which are the result of the decomposing speech signal discrete wavelet transform feature extraction method. LPC coefficients which derive after decomposition of speech frame provides better representation than modeling the frame directly. The proposed method which is wavelet-based LPC is effective and efficient as compared to LPCC and MFCC because it takes the

combined advantage of LPC and DWT while estimating the features. This reduces the memory requirement and the computational time. M. S et al. [11] described a hybrid technique for speaker recognition. In this DWT based MFCC technique is employed for feature extraction. In DWT speech signal is divided completely into different frequency bands. Classification is done using SVM. Main steps involved in this speaker recognition system create a database (collection of voice samples in wav format), feature extraction, Training, Testing. After testing SVM is used to identify the speaker. Nidhyananthan et al. [12] described the use of Relative Spectra-Mel Frequency Cepstral Coefficient (RASTA-MFCC) feature extraction from the filter bank structure and Gaussian Mixture Model to improve the performance of text-independent speaker identification in the noisy environment. RASTA based MFCC feature extraction method is found to be more robust in the noisy environment compared with traditional MFCC method. MFCC is an efficient feature extraction method for identifying the speaker as it has the ability to capture speaker-specific information. In this, performance is improved by using RASTA processing of speech in the presence of convolution and additive noise. The proposed work combines these two processes to yield RASTA-MFCC feature which is robust to noise and to effectively capture the feature vectors. The proposed method with RASTA-MFCC feature and GMM modeling for speaker identification offers a speaker identification accuracy of 97.67% for noisy speech database of 50 speakers.

III. SPEAKER IDENTIFICATION SYSTEM

Speaker recognition is defined as the process in which unknown speaker is compared with a set of known speakers to find the best matching speaker in the database. The second component in speaker identification is testing, It is basically the task of comparing an unknown utterance to the training data and making an identification. Every Speaker Recognition system has two phases: Identification, Verification. In Identification, the aim is to match input voice samples with available voice samples. In speaker verification, the task is to identify the claimed person from available speech samples. The speaker identification is divided into two components: feature extraction and feature classification.

A. Description of Feature Extraction Techniques for Representation of Speech Signals

Speech is a complicated signal produced as a result of several transformations occurring at several different levels. An important problem in speech recognition systems is to find out a representation that is designed for extracting information content from speech signals. The goal of feature extraction is to represent any signal by a finite number of features of the signal. This is because a speech signal contains a lot of information and not all the information is relevant to specific tasks. The main feature extraction techniques are Linear Predictive Coding (LPC), Perceptual Linear Prediction (PLP), Mel-Frequency Cepstral Coefficient (MFCC).

1) *Discrete Wavelet Transform (DWT):* The wavelet transform (WT) is a technique for analyzing signals. The wavelet transform is used to decompose and reconstruct non-stationary signals efficiently [3]. The wavelet technique is considered a relatively new technique in the field of signal processing for feature extraction. Wavelet transforms have been used for automatic speech recognition by the several researchers for speech coding and compression, speech denoising, and enhancement and other processes. The wavelet transform is more suitable to deal with non-stationary signals like speech. Wavelets have the ability to analyze different parts of a signal at different scales. The wavelet transform is a transformation that provides a time-frequency representation of the signal [8]. In discrete wavelet transform the signal is divided into the high frequency and low-frequency content of the signal. The low-frequency content of the signal is called the approximations and the high-frequency components of the signal are called the details. The lower frequency contents provide a sufficient approximation of the signal while the finer details of the variation are contained in the higher frequency region. DWT is chosen due to its simplicity and ease of operation in handling complex signals such as the voice signal [15]
2) *Mel-Frequency Cepstral Coefficients (MFCCs):* MFCC is one of the most popular feature extraction techniques used in both speech and speaker recognition. It has the benefit that it is capable of capturing the phonetically important characteristics of speech. The main purpose of MFCC processor is to mimic the behavior of the human ears. MFCC extracts important characteristics from the speech signal, while at the same time deemphasizes all other information. MFCCs are considered as frequency domain features which are more accurate than time-domain features. The MFCCs are proved more efficient in the noisy environments as compare to other technique like LPC. MFCC is composed of five phases. The first step in MFCC feature extraction process is pre-processing in which the signals are pre-processed before extracting features from the speech signal. In framing, the speech signal is split into a number of frames. The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. Next step is to pass these windowed frames into the discrete Fourier transformer. This step is performed to convert the windowed frames into magnitude spectrum. The output of this step is known as spectrum, then the magnitude spectrum is passed to log and then to the inverse of discrete

Fourier transform which produces the final result as Mel-Cepstrum. The Mel-Cepstrum consists of the features which help to identify the speaker. For each speech frame, a set of MFCC is computed. This set of coefficients comprises the acoustic vector which shows the important characteristics of speech which are helpful for processing in speech recognition. A small drawback is that MFCC is not robust enough in noisy environments and combining it with other features improve its accuracy. The resultant matrices are referred to as Mel-Frequency Cepstrum Coefficients.

3) *Linear Predictive Coding (LPC):* LPC is the good analysis technique for extracting features, determining the basic parameter and computational model of speech. In LPC, speech samples can be approximated as a linear combination of past speech samples. The number of samples is referred to as the order of LPC. LPC has the capability for compression, synthesis, identification of the speech signal. LPC coefficients can be estimated using autocorrelation method. The LPC method provides a reliable, accurate and robust method to estimate the parameters. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. The assumption in LPC is that a prediction of the nth sample is made that can be shown by summing up the target signal's previous samples in the series of speech samples. So, with estimation of formants LPC analyzes the speech signal. It also removes the effects of formants from the speech signal and estimates the intensity and frequency. Inverse Filtering is used to remove the formants and remaining signal is called the residue. The drawback of LPC is its performance degradation in the presence of noise

4) *Relative Spectral Algorithm (RASTA):* RASTA is short term for Relative Spectral Algorithm. It is known as a technique that is developed as the initial stage of voice recognition. The main motive of the work is to improve the robustness of speech recognition system in the noisy environment. Speech enhancement in RASTA involves linear filtering process to find out short-term power spectrum of the noisy speech signal. The spectral values of input speech signal are compressed by a nonlinear compression rule before performing the filtering operation and expanded after filtering. Frequency sub-band filters are applied to the energy in each subband in order to remove the noise. In, voice signals stationary noises are often detected. Stationary noises are the noises that are present for the full period of a certain signal and does not have diminishing feature. This property does not change over time. Therefore RASTA is considered as an efficient tool to be included in the initial stages of voice signal filtering to remove stationary noises [15]

5) *Proposed Hybrid Approach (DWT- MFCC-RASTA :*Hybrid methods try to reduce their limitations by combining the advantages of the combined techniques. The hybrid method is one of the emerging approaches that can improve speech recognition accuracy. In this article, a novel approach to develop a speaker identification system is presented. In this, speaker identification system is based on a hybrid set of speech features. This hybrid set consists of Discrete Wavelet Transform (DWT), Mel Frequency Cepstral Coefficient (MFCC), Relative Spectral Algorithm (RASTA). For the development of a robust and efficient speech or speaker recognition system pre-processing of speech signals is considered as a crucial step. Before feature extraction process discrete wavelet transform is applied to the speech signal. This is done to improve the accuracy of the recognition process and to make the system more robust to noise. The discrete wavelet transforms divide the signal into approximation and detail coefficients, we take only the approximation coefficients vector as input for feature extraction stage. Then, MFCC and RASTA feature extraction methods are used to find out coefficients. It was observed through extensive validation runs that the highest level of accuracy achieved was 98.63%. So, it was clear that combining features improves the accuracy. The percentage of verification is given by equation 1.1

$$\text{Accuracy (\%)} = \frac{\text{Test value}}{\text{Registered value}} \times 100 \qquad (1.1)$$

B. Speaker Modeling

Classification deals with recognition of an unknown speaker based on a similarity measure between an unknown speaker and the speakers that are enrolled. In classification, the speech produced by the speaker whose identity is to be recognized will be compared with all speaker's models in the database. Then, the identity of the speaker will be determined according to a specific algorithm. GMM and VQ are widely used and accepted techniques for pattern matching. GMM algorithm is computationally complex as compared to VQ but research proves it to be better than VQ. Generally, GMM method has high success rates as compare to Vector Quantization method. In order to accelerate the system and to reduce the execution time, we first make the system to identify the speaker. In training stage, coefficients are calculated. To reduce the speaker identification time, GMM is used. The classification of the speaker is a decision process which is based upon previously stored information for validating speaker. The feature matching techniques used in speaker recognition include Gaussian Mixture Models (GMM), Vector Quantization (VQ). Speaker modeling is carried out to find the best match in order to simply accept an unknown speaker or to reject it. Template model such as vector

quantization and stochastic model such as Gaussian Mixture model are two speaker models that are used for classification in text-dependent and text-independent speaker identification systems.

1) *Vector Quantization:* Vector Quantization (VQ) is a quantization technique to compress information in such a way that it maintains the most important or prominent characteristics. VQ is used in many applications such as for compressing data, recognizing voice etc. A smaller set of feature vectors representing the centroids of distribution is produced from a large set of feature vectors. VQ consists of extracting a small number of representative feature vectors as an efficient means of characterizing the speaker-specific features. This small set represents the centroids of the distribution. VQ is a process of mapping feature vectors from a vector space to a finite number of regions in that space. These regions are called clusters and they are represented by the centroids. A set of centroids which represents the whole vector space is called a codebook. In the testing phase, Euclidean distance between features of an unknown speaker and speaker models stored in the database can be calculated for identification purposes. The speaker is identified on the basis of minimum distance with the features of an unknown voice is selected as the identity of the unknown speaker [14]
2) *Gaussian Mixture Model (GMM):* Gaussian mixture model (GMM) is a conventional method for speech recognition, known for its effectiveness and scalability in speech modeling. GMM is simple, easy to evaluate and faster to compute method for speaker recognition. GMM is very competitive when compared to other pattern recognition techniques. A Gaussian mixture model is considered as a probabilistic model which assumes all the data points that are generated from a mixture of a finite number of Gaussian distribution with unknown parameters. GMM model is defined as the extension of vector quantization model. The limitation of GMM is that it requires a sufficient amount of training data to ensure good performance which increases the training time.

IV. EXPERIMENTAL RESULTS

A. Data Collection

In the data collection stage, 300 English words are recorded from 30 different speakers including both male and female with a sampling frequency. The database contains the speech data files of 30 speakers. These speech files consist of isolated words and continuous speech. Each speaker speaks 10 words out of which 5 are for training and 5 for testing. The data was recorded using a microphone, and all samples are stored in wav format files. Data is collected for two types of speech:

1) *Clean Speech:* For clean speech database, recordings of speakers are collected in a clean environment. Recordings of sound files in the wav format are collected from 30 speakers. Each speaker speaks 10 words. Then the collected database of 30 speakers is divided into training and testing database. Out of 10 words, 5 words are for training and 5 words are for testing
2) *Noisy Speech:* For noisy speech database, white Gaussian noise is added to the speech signal. Similarly, the collected database is divided into training and testing database.

B. Speaker Recognition Systems

Speaker recognition Systems are Classified into Text Dependent, Text Independent Systems

1) *Text-dependent System:* The dataset used in the test phase consists of speech files recorded from 30 speakers in a clean environment. Sounds are recorded using a microphone with a sampling frequency in wav format. In this system, the same text is being spoken both in training phase and in the testing phase. A database of 30 speakers is created which is divided into training and testing database. In this training and testing database includes similar speech utterances spoken by different speakers for speaker identification purpose
2) *Text-Independent System:* the text-independent system, there is no restriction on text being spoken. In this database of 30 speakers uttering different utterances is created. In this, there is no need to record similar word from one speaker a number of times. In this, each speaker can speak different words. Then the database of 30 speakers is created and divided into training and testing database.

C. Training and a Testing Phase

1) *Training :* after applying a combination of three feature extraction techniques which are MFCC, RASTA, and discrete wavelet transform in order to improve the accuracy of the system. Each registered speaker provides samples of their speech in the clean and noisy environment. There are different techniques for extraction of features from a voice signal. The voice is trained using MFCC, LPC, LPCDWT, MFRASTA, Wavelet-based MFRASTA.
2) *Testing Phase :* The voice of the speaker to be identified or verified is given as input speech. The input speech is matched with the stored reference model using Vector Quantization (VQ), Gaussian Mixture Model (GMM). GMM is used for the

recognition process as it is more efficient as compared to VQ. In the recognition stage, the extracted features are compared with the reference model and the word that has the best match will be the output. The results showed that a recognition rate of 98.63% using the wavelet-based MFCCRASTA. The results are obtained using Matlab. In this section, experimental results have been shown for various test signals with proposed algorithms. All the experiments have been done in MATLAB 2013a version with 4 GB RAM and i5 processor for speed specifications. The results show that for clean speech, an identification rate of 98.63% is achieved by the proposed feature extraction while the identification rate for noisy speech is 93.41%. When the test patterns are corrupted with Gaussian noise, the performance of the system using different feature extraction techniques affected significantly. With noise, identification rate reduces to 93.41% while the proposed technique demonstrates much better noise robustness with a satisfactory identification rate of 98.63%.

D. Figures and Tables

Results were shown in the form of tables, figures. Recognition Accuracy for text dependent systems for 30 speakers in clean and noisy Environment are shown as below. Tables show Recognition Rate obtained with GMM, VQ. GMM shows better recognition rate as compare to VQ. Then Comparison of Text Dependent and Text Independent Systems is performed.

TABLE 1
TEXT DEPENDENT CLEAN SPEECH

Features	Matching Techniques	
	VQ	GMM
MFCC	84.21	88.32
LPC	77.54	83.36
LPC-DWT	85.15	90.12
MFRASTA	87.32	93.21
DWMFRASTA	92.64	98.63

TABLE 2
TEXT DEPENDENT NOISY SPEECH

Features	Matching Techniques	
	VQ	GMM
MFCC	71.23	80.43
LPC	59.42	73.13
LPC-DWT	75.24	83.62
MFRASTA	81.32	82.16
DWMFRASTA	86.24	93.41

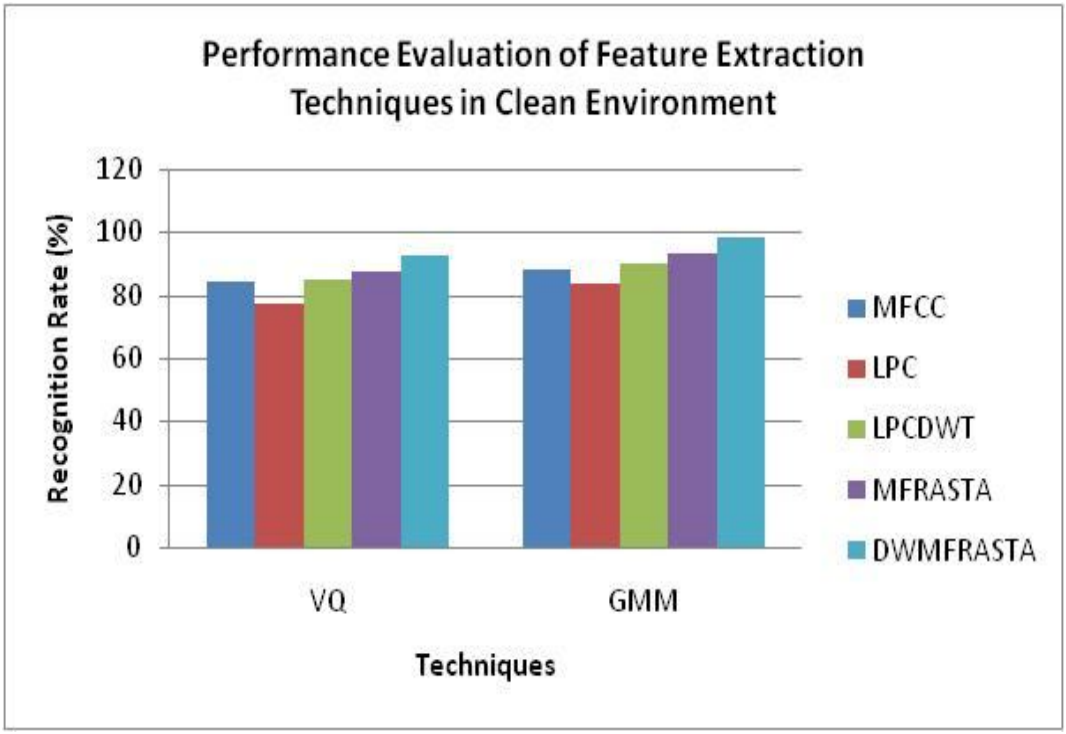


Fig 1 Performance Evaluation of Feature Extraction Techniques in Clean Environment

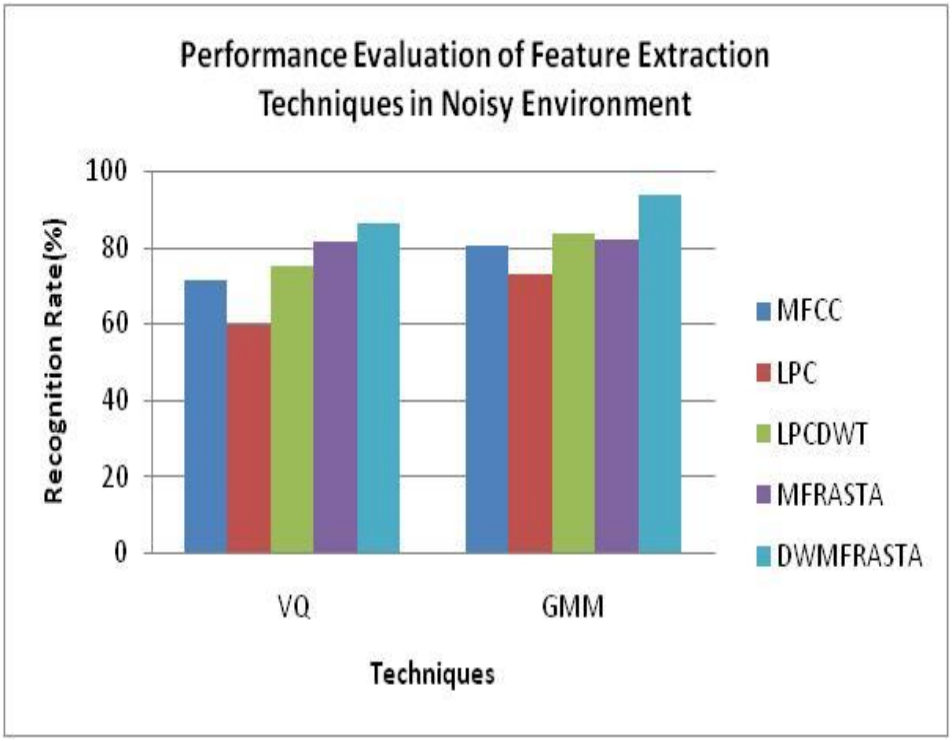


Fig 2 Performance Evaluation of Feature Extraction Techniques in Noisy Environment

In the Following Table Performance of Proposed Wavelet-based MFRASTA Hybrid Approach for Text Dependent and Text Independent System is shown. Performance of Text Dependent system is better than Text Independent System. From Results it is shown that DWMFRASTA-GMM Performs better than DWMFRASTA-VQ. GMM shows good Recognition Accuracy as compare to VQ.

TABLE 3
PERFORMANCE OF DWMFRASTA FOR TEXT DEPENDENT AND TEXT INDEPENDENT

System	Speakers	Techniques	
		DWMFRASTA-VQ	DWMFRASTA-GMM
Text-Dependent System	Male 1	85.74	87.93
	Male 2	2.47	89.36
	Male 3	100	97.36
	Female 1	98.26	96.33
	Female 2	77.23	100
	Female 3	80.36	99.21
Text-Independent System	Male 1	79	83.12
	Male 2	82	82.41
	Male 3	77.02	81.03
	Female 1	78.84	90.36
	Female 2	77.49	92.45
	Female 3	80.01	89.61

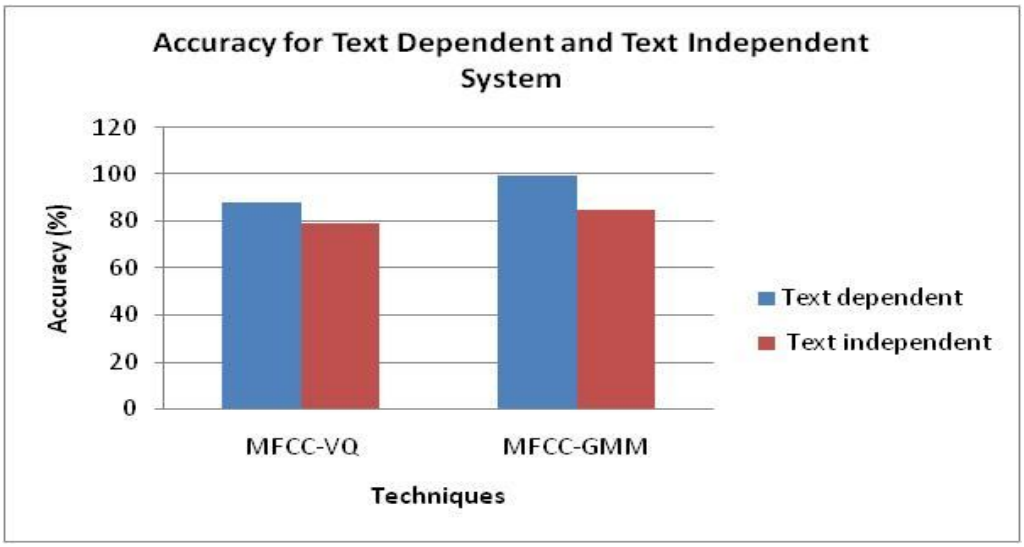


Fig 3 Accuracy for Text Dependent and Text Independent System

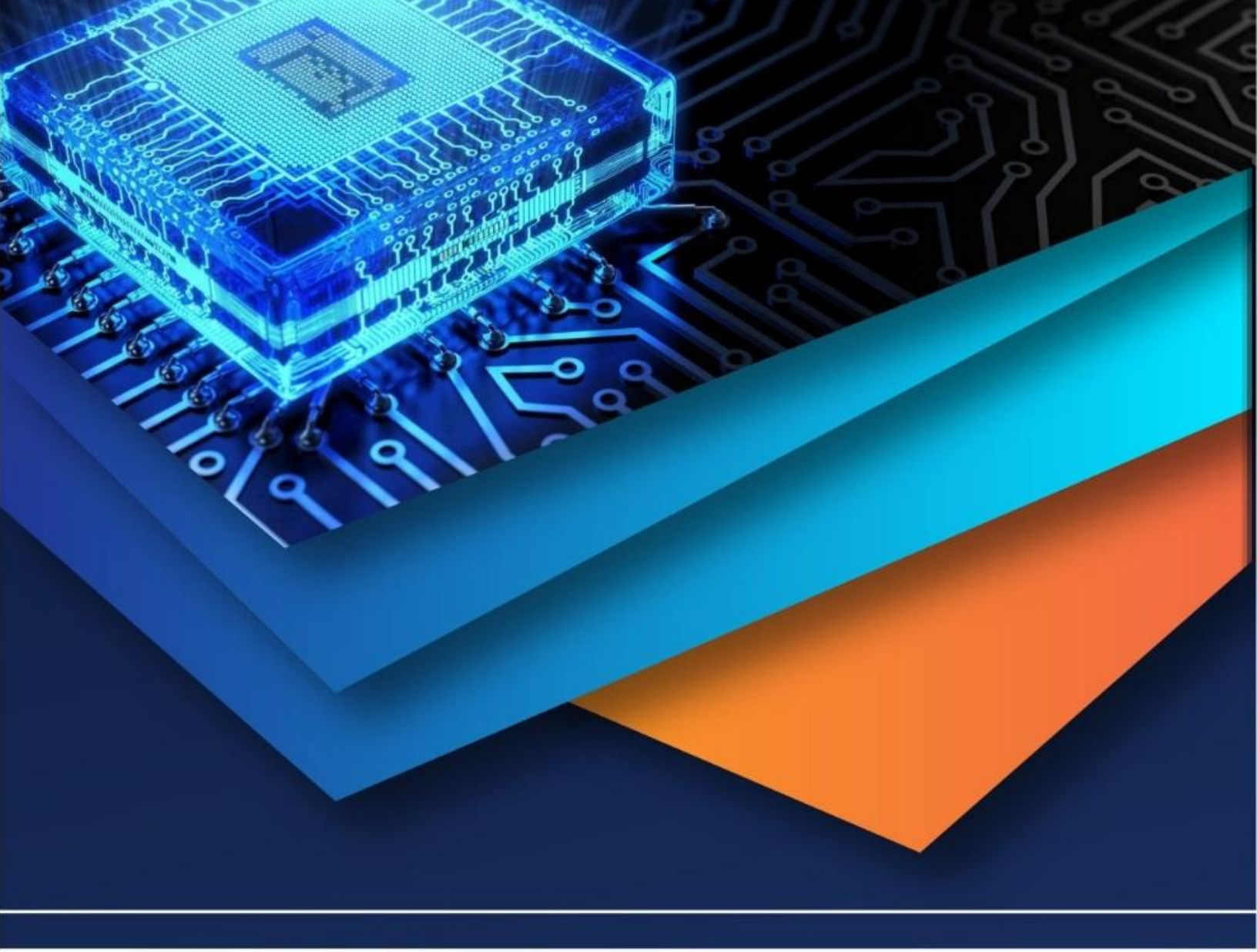
V. CONCLUSION

In this paper, we presented an effective and robust new feature extraction technique based on a combination of three feature extraction techniques which are MFCC, DWT, RASTA. The discrete wavelet transform is performed on the speech signal before extracting the features to improve the accuracy of recognition and to make the system more robust to noise. Based on the time-frequency analysis of the wavelet transform, approximations and details resolutions channels are obtained. After discrete wavelet transforms the MFCC, RASTA is applied on the input signal for capturing the characteristics of the speech signals. Results showed that the proposed technique gives better performance than the MFCCs features. In addition, the technique reduces the problem of noise and improves efficiently the recognition rate when dealing with noisy speech signals compared to the MFCCs which operate well in the clean environment. In this thesis, 5 different methods which are MFCC, LPC, LPC+DWT, MFCC+RASTA, MFCC+DWT+RASTA are compared. In order to improve the accuracy of the system, different combinations of these 5 methods have been tested. It is found that the best combination is MFCC+DWT+RASTA who has the best performance with 98.63%

recognition rate which is better than MFCC which is one of the best feature extraction techniques based on the human hearing system. GMMs were used for the recognition stage as they give more advanced representation than VQ. On the basis of experimental results, the proposed method can make an effectual analysis. The average identification rate of the system was 98.63%. The hybrid system is one of the emerging approaches that can improve speech recognition accuracy and will take an important role in speech technology research. The stated results show that the proposed method can make an accurate and robust classifier.

REFERENCES

[1] N. P. K, D. Mathew, A. Thomas, "Comparison of Text-Independent Speaker Identification Systems using GMM and me- Vector Methods", In International Conference on Advances in Computing & Communications (ICACC), Cochin, India, 2017, pp. 47-54.
[2] M. I, A. S. Ali, H. S. Ali, "Wavelet-based Mel-Frequency Cepstral Coefficients for Speaker Identification using Hidden Markov Models", Journal of Telecommunications, vol. 1, no. 2, pp. 16-21, 2010.
[3] E. Rady, A. Yahia, E. Dahshan, H. Borey, " Speech Recognition System Based on Wavelet Transform and Artificial Neural Network", Egyptian Computer Science Journal, vol. 37, no. 3, pp. 85-96, 2013.
[4] P. K. Kurzekar, R. R. Deshmukh, V. B. Waghmare, P.P. Shrishrimal, " A Comparative Study of Feature Extraction Techniques for Speech Recognition System", vol. 3, no. 12, pp. 18006- 18016, 2014
[5] M. Cutajar, E. Gatt, I. Grech, O. Casha, J. Micallef, "Comparative study of automatic speech recognition techniques", IET Signal Processing, vol. 7, no. 1, pp. 25-46, 2013.
[6] C. Hsieh, E. Lai, Y. Wang, "Robust Speaker Identification System Based on Wavelet Transform and Gaussian Mixture Model", Journal of Information Science and Engineering, pp. 267-282, 2003.
[7] A. Ali, D. Dean, B. Senadji, V. Chandran, G. Naik, "Enhanced Forensic Speaker Verification using a combination of DWT and MFCC feature warping in the presence of Noise and Reverberation Conditions", IEEE Access, vol. 5, pp. 15400-15413, 2017.
[8] M. Abdalla, H. Abobakr, T. Gaafar, " DWT and MFCCs based Feature Extraction Methods for Isolated Word Recognition", International Journal of Computer Applications, vol. 69, no. 20, pp. 21-26, 2013.
[9] A. Maurya, D. Kumar, R.K. Agarwal, " Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach", In International Conference on Smart Computing and Communications, Kurukshetra, India, pp. 880-887, 2017.
[10] N. Nehe, R. Holambe, " DWT and LPC based feature extraction methods for isolated word recognition", EURASIP Journal on Audio, Speech and Music Processing, pp. 1-7, 2012.
[11] S. M. S, M. K. R, " Speaker Recognition using DWT-MFCC with multi- SVM Classifier", International Journal of Electronics and Communication Engineering (ICEHS), pp. 47-51, 2017.
[12] S. S. Nidhyananthan, R. S. S. Kumari, "Noise Robust Identification using RASTA-MFCC feature with quadrilateral Filter Bank Structure", Springer, pp. 1321-1333, 2016.
[13] R. K. Prasad, A. N. Mishra, S. N. Sharan, " Text-Dependent Speaker Identification Using VQ and DTW", VSRD International Journal of Electrical, Electronics & Communication Engineering, vol. 1, no. 8, pp. 453-459, 2011.
[14] S. B. Dhonde, S. M. Jagade, " Comparison of Vector Quantization and Gaussian Mixture Model using Effective MFCC features for Text-Independent Speaker Identification", International Journal of Computer Applications, vol. 134, no. 15, pp. 11-13, 2016.
[15] S. Mungamuri, P. P. S. Subhashini, "Text-Dependent Speaker Recognition using RASTA, LPC and Discrete Wavelet Transform", International Journal & Magazine of Engineering, Technology, Management and Research, vol. 2, no. 7, pp. 240-247, 2015.
[16] K. Sarmah, "Comparison Studies of Speaker Modeling Techniques in speaker Verification system", International Journal of Science and Engineering, vol. 5, no. 5, pp. 75-82, 2017.
[17] S. M. S, M. K. R, "Review on Feature Extraction and Classification Techniques in Speaker Recognition", International Journal of Engineering Research and General Science, vol. 5, no. 2, pp. 78-83, 2017.
[18] V. Sharma, P. K. Bansal, "A Review on Speaker Recognition Approaches and Challenges", International Journal of Engineering Research & Technology, vol. 2, no. 5, pp. 1581-1588, 2013.
[19] J. B. Ramgire, S. M. Jagdale, "A Survey on Speaker Recognition with various Feature Extraction and Classification Techniques", International Research Journal of Engineering and Technology, vol. 3, no. 4, pp. 709-712, 2016.
[20] V. C, R. V, "Suitable Feature Extraction and Speech Recognition Technique for Isolated Tamil Spoken Words", International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp. 378-383, 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY