A Survey on Speaker Recognition With Various Feature Extraction And Classification Techniques

Jyoti B. Ramgire[1], Prof. Sumati M.Jagdale[2]

[1]PG Student, Dept. Of Electronics and Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune 43, Maharashtra, India
[2] Associate Professor, Dept. Of Electronics and Telecommunication Engineering, Bharati Vidyapeeth's College of Engineering for Women, Pune 43, Maharashtra, India

---***---

Abstract - *Speech processing is more popular day by day for providing immense security. Authentication purpose speech is widely used. Speaker recognition is the process which can verify and identify the person who is speaking. Speech recognition system is different than speaker recognition system. Speaker recognition are widely used in industries, hospital, laboratory etc. Its advantages are more secure, easy implementation, more user friendly. For the area where security is very important, speaker recognition technique is one of the most widely used technique. It is also popular biometric technique. This paper describes an overview of different techniques that can be used in application of speaker recognition such as LPC, LPCC MFCC etc. Also discuss on different classifiers such as DTW, GMM, VQ, SVM. The main objective of this review paper is to summarize well known techniques for speaker recognition system.*

***Key Words*: Speaker recognition, Mel frequency cepstral coefficients(MFCC), Linear predictive coding (LPC), Linear Predictive Cepstral Coefficients (LPCC), Gaussian Mixture Model(GMM), Vector Quantization(VQ), Support Vector Machine(SVM), Dynamic Time Warping(DTW)**

1. INTRODUCTION

Speech signal contains different levels of information[14]. Speech signal can be used for speech recognition, speaker recognition or voice command recognition system[3]. Speaker recognition is used for many speech processing applications especially security and authentication. Today security is major requirement. Sometimes there may be confusion regarding speech and speaker recognition. Speaker recognition and speech recognition are very closely related systems but these two systems are different[14]. Speech recognition is the process of recognizing what is being said and speaker recognition is the process of recognizing who is speaking. Speech recognition has ability to automatically recognizing the spoken words of person based on information in speech signal[3].. Speaker recognition is classified as speaker identification and verification. The main aim of speaker recognition is to identify the speaker by extraction, characterization and recognition of the information contained in speech signal[14]. Speech recognition consist of speaker dependent and speaker independent.

The human speech is processed by machine depending on feature extraction and feature matching. Basic model of speaker recognition is shown in Figure 1[3].

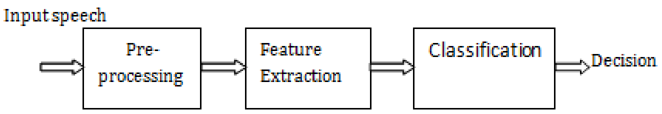

Fig -1: Basic model of Speaker Recognition system

Speaker recognition process is done in three steps. First is pre-processing is used to remove silent period from speech signal[3]. In speaker recognition, the feature is extracted using different techniques such as Linear predictive coding(LPC), Linear Predictive Cepstral Coefficients (LPCC), Mel frequency cepstral coefficients MFCC. For feature classification different classifiers are used such as Support Vector Machine (SVM), Vector Quantization(VQ), Gaussian Mixture Model(GMM), Dynamic Time Warping(DWT).

2. RELATED WORK

Table -1: Literature Survey

Author Name	Feature Extraction	Classifiers	Advantages
V. Tiwari et.al.[1]	LPC.LDB, MFCC	VQ	1. MFCC with hanning window using 32 filter has more efficiency. 2. Density matching property of vector quantization is powerful
K. Kaur, et.al.[2]	LPC, LPCC MFCC	VQ, GMM, SVM,DWT ,HMM	1.MFCC technique is more consistent with human hearing as compared to LPCC, MFCC. 2. GMM is best

			among classification models due to its good classification accuracy and less memory uses.
K. Dhameliya et.al.[3]	MFCC, LPC	GMM, ANN	For better speaker recognition it is possible to combine one or more techniques.

Literature survey reveals that, speaker recognition technique is one of the most widely used technique for the area where security is very important. It is also popular biometric technique[8]. Other biometric techniques are present but speech gives better result. Different feature extraction techniques are present such as LPC, LPCC, MFCC but MFCC better than other techniques for lower filter order.

3. SPEAKER RECOGNITION

The main objective of speaker recognition is to convert the acoustic audio signal into computer reliable form. Speaker recognition systems involve two phases such as training and testing. In training process take the input as speech signal and feature extraction is done using feature extraction technique. Feature vectors representing the voice characteristics of the speaker and are used for building reference model[1]. Actual recognition task is in testing phase. In testing phase speaker voice is matching with reference model using some matching technique. After level of matching decision is done.

3.1 Techniques Of Speaker Recognition

Classification of speaker recognition is as shown in Figure 1.[8]

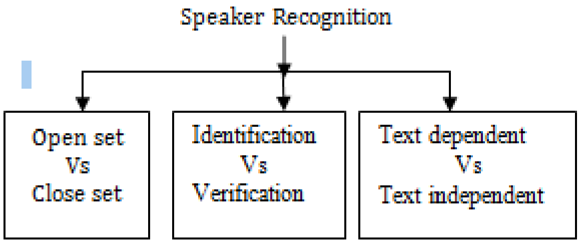
Speaker Recognition
Open set Vs Close set
Identification Vs Verification
Text dependent Vs Text independent

Fig-2: Techniques of Speaker Recognition

3.1.1 Open set Vs Close set

This classification is based on set of trained speaker available in system. Open set system contain any number of trained speaker. Close set system has only a specified users registered[8].

3.1.2 Speaker Identification Vs Verification

Speaker identification determines which registered speakers provides given utterance from a set of known speaker. Speaker verification is the process of accepting and rejecting identity claim of speaker[8]. Speaker verification is faster than speaker identification.

3.1.3 Text dependent Vs Text independent

Text dependent system involves same text being spoken both in training phase and in testing phase. In text independent system, there is no restriction on text being spoken[2]. Recognition rate is better is better in text in text dependent system than in text independent system.

4. FEATURE EXTRACTION TECHNIQUES

There are different techniques are used for feature extraction like Linear Prediction Coding (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Mel-Frequency Cepstrum Coefficients (MFCC).

4.1 Linear Predictive Coding (LPC)

This technique starts with the assumption that a speech signal is produced by a buzzer at the end of a tube. By estimating the formants LPC analyzes the speech signal. It removes the effects of formants from the speech signal, and estimates the intensity and frequency of the remaining buzz. The procedure used for removing the formants is called inverse filtering, and the remaining signal is called the residue[8]. Drawback of this technique is that performance degradation in presence of noise[2].

4.2 Linear Predictive Cepstral Coefficients (LPCC)

This technique is widely used to extract the features from speech signal. LPC parameters can effectively describe energy and frequency spectrum for sound frames[8]. LPCC gives smoother spectral envelop and stable representation as compare to LPC[2]. Drawback of this technique is that linearly spaced frequency band[2].

4.3 Mel-Frequency Cepstrum Coefficients (MFCC)

Mel-frequency Cepstral coefficients (MFCCs) were introduced by Davis and Mermelstein in the 1980s. MFCC is most popular technique and commonly used in most of application of speech signal for feature extraction. Feature extraction is first step in speech and speaker recognition system. Feature extraction means to identify the component of audio signal that are good for identifying the content and discarding all other stuff which carries information like background noise, emotion etc. The main principle of MFCC is filter bank coefficient[2]. The step by step process of MFCC is shown in Figure 2.

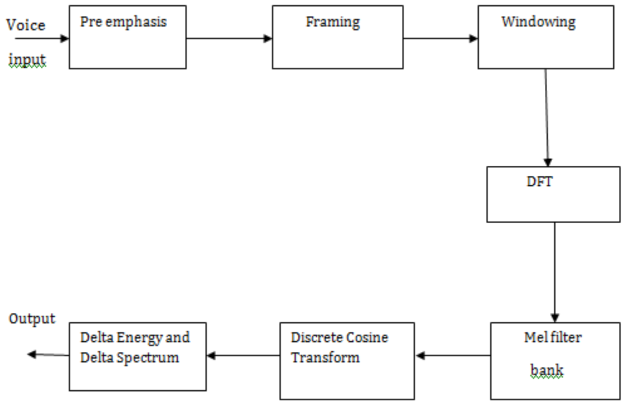

Fig-3: Block Diagram of MFCC

Following are different steps for MFCC:

I. Pre-emphasis

This is simple signal processing method. It increases the amplitude of higher frequency band and decrease the amplitude of lower frequency band[6]. Because higher frequencies are more important for signal disambiguation than lower frequencies.

II. Framing

The audio signals are constantly changing. For simplicity purpose it is necessary to take a constant signal for short time scale. If the frame is much shorter then there may not have enough sample goes to get reliable spectrum estimate. If the frame is longer then it changes throughout the frame. Signal is divided into frames of N samples and adjacent frame being separated by M[6].

III. Windowing

Windowing technique is used to minimize the signal discontinuity. In this process hamming windowing has to be multiplied with each frame for keeping the continuity of first and last point in the frame.

IV. Fast Fourier Transform

Fast Fourier Transform is used to convert each frame of N samples from time domain to frequency domain. FFT is perform to obtain magnitude of frequency response of each frame. When FFT is perform on each frame, assume that signal is periodic and continuous when warping around.

V. Mel Filter Bank Processing

The range of frequencies is very wide in FFT and voice signal does not follow the linear scale[6]. The output of FFT is multiplied by a set of 20 triangular bandpass filter to get log energy of each triangular bandpass filter.

VI. Discrete Cosine Transform

This is the process to convert the log Mel spectrum into time domain[6]. The result is called Mel Frequency Cepstrum Coefficients[3].

5. FEATURE CLASSIFICATION

In speaker recognition system another important part is classifications[12]. In classification stage the patterns are classified into different classes. There are many classifiers are used such as DWT, GMM, SVM, VQ, etc. From that selection of classifier is an important task. But there is no fix criteria for the selection of classifier. Many pattern classifiers are explored for developing speech systems like, emotion classification, speech recognition, speaker verification, speaker recognition.

5.1 Dynamic Time Warping

Dynamic Time Warping algorithm calculates the distance between two sequences which may vary in time or speed. Time warping is done non-linearly to normalize the timing differences between test utterance and the reference template. Then time normalization distance is calculated between patterns. Authentic speaker is identified with minimum time normalized distance. It is beneficial for variable length input feature and require less storage space[2].

5.2 Vector Quantization

Vector Quantization is classical quantization technique from signal processing. It allows modelling of probability density functions by distribution of prototype vector. By using vector quantization the extracted speech feature of speaker are quantized to a number of centroids. These centroids compose the codebook of that speaker[1]. It is used for data compression and require less storage. VQ is computationally less complex. Memory requirement is achievable for real time applications[2].

5.3 Gaussian mixture model

Gaussian Mixture Model is a density estimator. GMM is an unsupervised learning algorithm. It require less training and test data. Expectation maximization algorithm is used to estimate GMM parameter from training data. A sequence of feature extracted from input signal. By computing log likelihood the distance of the given sequence from the model is obtained. GMM performs better as it requires less amount of data to train the classifier hence memory requirement is less[2].

5.4 Support Vector Machine

For classification of speech or speaker recognition support vector machine is simple and effective algorithm. SVM is supervised learning algorithm[2]. It is more useful for binary classification. But it has poor performance in speaker recognition due to its fixed length vector.

6. CONCLUSION

In this review paper, there is discussion on the speaker recognition that can be used for many speech processing applications specially for security and authentication. There are most commonly used feature extraction techniques are discussed from that MFCC are widely used. Also discuss different feature classification techniques for speaker recognition.

ACKNOWLEDGMENT

I would like to thank Prof. S.M. Jagdale for their valuable assistance to complete this work.

REFERENCES

[1] V. Tiwari, "MFCC and its applications in speaker recognition", *IEEE International Jouranal on Emerging Technologies*, Volume-1, Issue-7, May 2013, pp 33-37.

[2] K. Kau and N. Jain, "Feature Extraction and Classification for Automatic Speaker Recognition System – A Review", *International Journal of Advanced Research in Computrt Science and Software Engineering*, Volume 5, Issue 1, January 2015, pp. 1-6.

[3] K. Dhameliya, "Feature Extraction And Classification Techniques for Speaker Recognition: A Review", *IEEE International Conference on Electrical, Electronics, Signal, Communication and Optimization*, January 2015, pp. 1-4.

[4] S. Nakagawa, L. Wang and S. Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information," *IEEE Transaction on Audio, Speech and Language Processing*, Vol. 20, No. 4, May 2012, pp. 1085-1095.

[5] M. AdaMmsk, B. VonSolms, " An Open Speaker Recognition Enabled Identification And Authentication System", *IST-Africa 2014 Conference Proceedings* 2014,pp 1-8.

[6] Lindasalwa M. , M. Begam and I. Elamvazuthi, "Voice Recognition Algorithm using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", *Journal Of Computing*, Volume 2, Issue 3, March 2010, pp 138-143.

[7] E. Chandra, K. Manikandan, M. Kalaivani, "A Study on Speaker Recognition System and Pattern Classification Techniques", *International Journal Of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, Vol 2, Issue 2, February 2014, pp. 963-967.

[8] P. Chaudhary and M. Vagadia, "A Review Article on Speaker Recognition with Feature Extraction", *International Journal of Emerging Technology and Advanced Engineering*, Volume 5, Issue 2, February 2015,pp. 94-97.

[9] R. Bharti and P. Bansal, "Real Time Speaker Recognition System using MFCC and Vector Quantization Technique", *International Journal of Computer Applications*, Volume 117 No. 1, May 2015, pp, 25-31.

[10] S. Suuny, D. Peter, K. Jacob, "Performance Of Different Classifiers In Speech Recognition", *International Journal of Research in Engineering and Technology*, Volume: 02 Issue: 04 Apr-2013, pp. 590-597.

[11] S. Madikeri and H. Murthy, "Mel Filter Bank Energy Based Slope Feature and Its Application to Speaker Recognition," *IEEE National Conference on communication* (NCC), Bangalore, January 2011, pp. 1-4.

[12] S. K. Singh. " Features And Techniques For Speaker Recognition", M. Tech. Credit Seminar Report, Electronic Systems Group, EE Dept, IIT Bombay submitted Nov 03.