Comparative study of voice print Based acoustic features: MFCC and LPCC

Shivani Jain, Brij Kishore

Department of CSE, Apex Institute of Engineering & Technology, Jaipur, India

Abstract--- *Voice is the best biometric feature for investigation and authentication. It has both biological and behavioural features. The acoustic features are related to the voice. The Speaker Recognition System is designed for the automatic authentication of speaker's identity which is truly based on the human's voice. Mel Frequency Cepstrum coefficient (MFCC) and Linear Prediction Cepstrum coefficient (LPCC) are taken in use for feature extraction from the provided voice sample. This paper provides a comparative study of MFCC and LPCC based on the accuracy of results and their working methodology. The results are better if MFCC is used for feature extraction.*

Keywords— *authentication, human's voice, Speaker recognition system, text independent, feature extraction, LPCC, MFCC, pattern matching.*

I. INTRODUCTION

As the technology is varying day by day the advancement of technology is very much required for the user satisfaction. In the matter of security and investigation the system should be very powerful and advance. Speaker Recognition System is providing the best biometric authentication. The voice of human is a natural signal having unique features. It has both biological and behavioural features. The speech contains certain aspects of the speaker identity, emotion, gender etc. Speaker authentication finds its application in the speaker identity associated with the speech conveyed. This paper introduces two of the most popular Cepstral based feature extraction technique [1] MFCC, frequency mapped into mel-scale then converted to Cepstral domain and LPCC, linearly predicted frequency mapping converted to the Cepstral domain.

II. SPEAKER RECOGNITION SYSTEM

Speaker recognition, which requires two applications: speaker recognition and speaker verification, is the process of automatically identifying who is speaking on grounds of individual information included in voice waves.

2.1 Speaker verification—Speaker verification is the process of checking whether the speaker identity is who the individual claims to be.
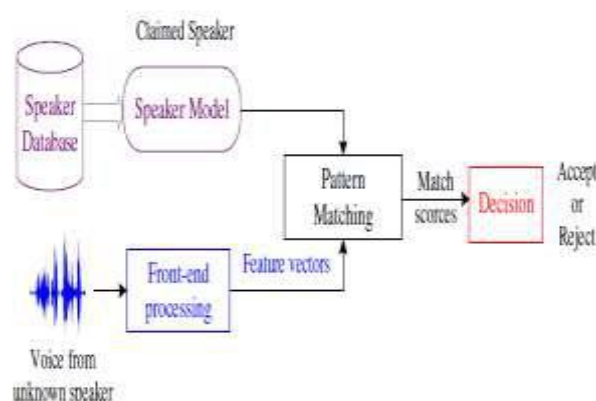

Fig.1: Basic Structure of Speaker Verification

2.2 Speaker identification--Speaker identification is the process of determining the identity of an unknown speaker by comparing his/her voice with sounds of registered speakers in the database. It's a one-to-many comparison.
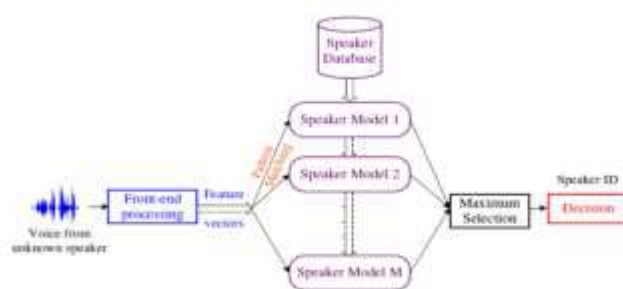

Fig.2: Basic Structure Speaker Identification

III. FRONT END PROCESSING (FEATURE EXTRACTION)

Feature extraction is to convert a voice signal to some type of parametric representation for further analysis and processing. Features obtained from spectrum of voice have shown to be the most efficient in automatic systems. Feature extraction is a process of reducing data while retaining the speaker discriminative information of the

speakers [3]. The selection of appropriate features along with methods to estimate (extract or measure) them known as feature selection and feature extraction.

Spectral features are descriptors of the short-term voice spectrum. The spectral features express entirely or partially the physical characteristics of the vocal tract. Spectral features are MFCC, LPCC, and LFCC etc.

IV. MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

It is based on human hearing perception but it can't perceive frequencies above 1KHz [2]. Technique of computing MFCC is based on the short-term analysis, and thus from each frame a MFCC vector is computed. MFCC extraction is similar to the cepstrum calculation except that one special step is inserted, namely the frequency axis is warped according to the mel-scale. MFCC has two different types of filter, which are arranged linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. The complete functionality of MFCC is described below [8]:
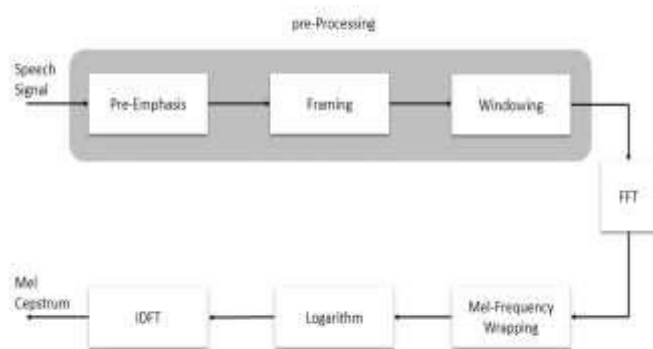

Fig.3: MFCC Methodology

V. LINEAR PREDICTIVE CEPSTRUM COEFFICIENT (LPCC)

Linear Predictive Coding (LPC) is a popular feature extraction technique for both speech recognition and speaker identification. The main idea behind LPC [2] is that a provided voice sample can be approximated as a linear combination of the past voice samples. LPC models signal as a linear combination of its past values and present input (vocal cords excitation) [6]. If the signal will be described only in terms of the linear combination of the past values then the difference between real and predicted output is called prediction error. LPC minimizes the prediction error to end out the coefficients. In speaker recognition task, LPC based on short-term analysis approach is used. This method is more suitable for real-time application. In speaker recognition area the set of prediction coefficients. In the context of speaker verification, LPCC are used to capture speaker specific

information manifested through vocal tract characteristics of the speaker is usually converted to the so-called Linear Predictive Cepstral Coefficients (LPCC), because the cepstrum is proved to be the most effective representation of speech signal for speaker recognition. The complete processing of LPCC is shown below [4],[10]:
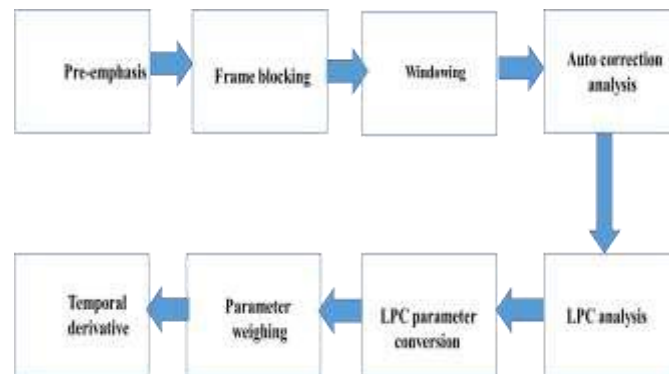

Fig.4: LPCC Methodology

VI. MFCC AND LPCC COMPARISON

1. MFCC are Cepstral coefficients computed on a warped frequency scale which is described on the basis of human auditory perception and LPCC are Cepstral coefficients that correspond to the human articulatory system based on linear prediction.

2. LPCC is used for feature Extraction at lower order and MFCC analysis is done with a fixed resolution along a Subjective frequency scale i.e. Mel-frequency Scale [6].

3. LPCC performs more accurately as compared to that of MFCC by 2.59% for authenticating a speaker. The study also suggests that on basis average time required for giving a decision, MFCC outperforms LPCC significantly by 3.73 sec [1].

4. LPCC provide a comparatively better performance as compared to MFCC for a robust fixed phrase speaker verification and on basis time taken for each trial, MFCC provide a better real based Fixed phrase Speaker Verification [1].

5. It may also be concluded that LPCC serve as a better acoustic feature as compared to MFCC for higher accuracy in designing an Fixed phrase Speaker Verification [1].

6. The performance of the Mel-Frequency Cepstrum Coefficients (MFCC) may be affected by the number of filters and type of window.

7. It has been observed that the conventional features extraction technique like LPCC and MFCC are sensitive to noise.

8. LPCC consistently outperforms MFCC, mainly due to its better performance in the female trials. This can be explained by the relatively shorter vocal tract

in females and the resulting higher formant frequencies in speech.

9. The features like MFCC and LPCC carry both the speech and speaker dependent information.

10. MFCC shows better recognition rate with 99.78% for Isolated, 99.88% for paired and 99.82% for Hybrid words. LPCC shows the recognition rates of 95.82%, 97.02%, 96.62% for Isolated, Paired and Hybrid words respectively [5].

11. The principle behind the use of LPCC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. But in MFCC the Mel scale is logarithmic scale that resembles the way in which human ear perceives sound. Mel scale filter bank maps the powers of the spectrum obtained above onto the Mel scale by using triangular overlapping windows. The Mel scale is represented by the following formula [2]:

a. Mel $(f) = 2595*\log 10(1 + f/700)$

12. MFCC and LPCC methods are applied to the overlapping frames of speech signal, the dimension of feature vector depends on dimension of frames [10].

13. From the simulation results we conclude that MFCC algorithm, which require more computation but perform better than LPCC in terms of efficiency and accuracy.

14. MFCC gives consistent results and is robust to noise due to the fact that it is based on human perception of speech [9].

VII.   CONCLUSION

MFCC and LPCC both are the important features obtained from speaker's voice. There are different measures for comparing the performance of these features. According to number of filters, different algorithms, dimension of frames etc. , the features vary in their performance.

REFERENCES

[1] Songitha Mishra and Rabul Laskar "*Comparison of MFCC and LPCC for a fixed phrase speaker recognition system, time complexity and failure analysis*" 2015 International Conference on Circuit, Power and Computing Technologies [ICCPCT]

[2] Vibha Tiwari"*MFCC and it applications in speaker recognition*" 2010 International Journal on Emerging Technologies

[3] Xinhui Zhou, Daniel Garcia-Romero, Ramani Duraiswami, Carol Espy-Wilson, Shihab Shamma "*Linear versus Mel Frequency Cepstral Coefficients for Speaker Recognition*" Department of Computer Science, University of Maryland, College Park, USA

[4] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi "*Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*" Journal of Computing

[5] Taabish Gulzar, Anand Singh and Sandeep Sharma "*Comparative Analysis of LPCC, MFCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks*" International Journal of Computer Applications

[6] Tsang-Long Pao, Chun-Hsiang Wang and Yu-Ji Li " *A Study on the Search of the Most Discriminative Speech Features in the Speaker Dependent Speech Emotion Recognition*" 2012 International Symposium on Parallel Architectures, Algorithms and Programming

[7] Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar *"A Review on Speech Recognition Technique"* 2010 International Journal of Computer Applications

[8] Dr.Mukesh Rana, Saloni Miglani "*Performance Analysis of MFCC and LPCC Techniques in Automatic Speech Recognition*"2014 International Journal Of Engineering And Computer Science

[9] P. Prithvi, Dr. T. Kishore Kumar2 "*Comparative Analysis of MFCC, LFCC, RASTA -PLP* " 2015 International Journal of Scientific Engineering and Research (IJSER)

[10] Urmila Shrawankar, Dr. Vilas Thakare "*Techniques for feature extraction in speech recognition system: A comparative study*".