www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN:2319-7242*
*Volume 4 Issue 3 March 2015, Page No. 10895-10898*

# Survey Paper on Automatic Speaker Recognition Systems

***Miss. Sarika S. Admuthe[1], Mrs. Shubhada Ghugardare [2]***

[1]ME Digital System Student, G. S. Moze COE, Balewadi, Savitribai Phule Pune University, India
sarika_admuthe@outlook.com

[2]Assistant Professor in Electronics & Telecommunication, G. S. Moze COE, Balewadi, Savitribai Phule Pune University, India
shubhada.ghugardare@gmail.com

**Abstract:** ***In this paper we present an overview of approaches for Speaker recognition. Speaker recognition is the process of automatically recognizing a person on the basis of individual information included in speech signals. Voiceprint Recognition System also known as a Speaker Recognition System (SRS) is the best-known commercialized forms of voice Biometric. The Speech is most essential & primary mode of Communication among of human being. In this paper we are focus on speaker identification and various feature extraction methods used for speaker identification system.***

**Keywords:** SRS, Speaker identification (SID) system, Feature Extraction method: MFCC, LPCC and GFCC.

## 1. Introduction:

Speech is one of the most important ways of human communication. Like fingerprints, it carries the identity of the speaker as voice print. The human speech is a signal containing mixed types of information; including words, feelings, language and identity of the speaker. There are a number of situations in which correct recognition of persons is required. The use of biometric-based (physiological and/or behavioral characteristics of a person) recognition is the most "natural" way of recognizing a person. This is also very safe as these characteristics cannot be stolen or forgotten. Biometric can be defined as study of life which includes humans, animals and plants. The word is taken from the Greek word where 'Bio' means life and 'Metric' means measure It is a system of identifying or recognizing the identity of a living person based on physiological or behavioral characteristics [1]. From the Table 1, it shows that the biometric patterns of voice are both high accordance with accessibility and acceptability.

**Table 1: Comparison between Biometric Identification**

| Biometric patterns | Iris | Face | Fingerprint | Voice |
|---|---|---|---|---|
| **Distinctiveness** | High | High | High | Moderate |
| **Robustness** | High | High | Moderate | Moderate |
| **Accessibility** | Low | Moderate | Moderate | High |
| **Acceptability** | Moderate | High | Moderate | High |

On the other hand, the real time conduct is one of the most important features of human speech interaction. Therefore effective, highly-accurate and time-efficient methods are necessary to deal with large amounts of speech information. Voiceprint Recognition System also known as a Speaker Recognition System (SRS) is the best-known commercialized forms of voice Biometric. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. The simplest to acquire, most used and pervasive in society and least obtrusive biometric measure is that of human speech. Voice is a most natural way of communication and non-intrusive as a biometric, Voice biometric has characteristic of acceptability, cost, easy to implement, no special equipment required. Also A voiceprint is a secure method for authenticating an individual's identity that unlike passwords or tokens cannot be stolen, duplicated or forgotten so, we have chosen to develop a (SID) Speaker identification System.

## 2. Literature Review:

Although in last six decades many techniques have been developed, we can achieve the ultimate goal of creating machines that can communicate naturally with people. The research in automatic speech and speaker recognition by machines has attracted a great deal of attention for six decades. In the 1950's, Bell Labs created a recognition system capable of identifying the spoken digits 0 through 9 was conducted by Biddulph and Balashek, while later systems were able to recognize vowels. Major attempts for automatic speaker recognition were made in the 1960. Table 2 Shows the work is already done in this field.

**Table 2: Work already done**

| Sr. | Year | Approach | Characteristics |
|---|---|---|---|

| No. | | | |
|---|---|---|---|
| 01 | 2012[9] | Speaker identification | Speaker identification in noisy and reverberant condition separately not together |
| 02 | 2010[12] | Speaker Identification Using Admissible Wavelet Packet Based Decomposition | Proposed an admissible wavelet packet based filter structure for SID system, multi resolution capabilities of wavelet packet transform are used to derive the new features. |
| 03 | 2010 | Text dependent speaker verification systems | BFCC features perform well for SID system |
| 04 | 2009[10] | Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach | Iterative clustering approach for both speech and speaker recognition |
| 05 | 2007[11] | A Supervised Text-Independent Speaker Recognition Approach | Hausdorff-based metric, used in the speech feature vector classification process |
| 06 | [2007][7] | Significance of Formants from Difference Spectrum for Speaker Identification | Auto-associative neural network (AANN) and formant features |

## 3. Overview on Speaker Recognition System:

Speaker Recognition is a system that can recognize a person based on his/her voice. This is achieved by implementing complex signal processing algorithms that run on a digital computer or a processor. The speaker recognition system can be classified as speaker identification or speaker verification

### 3.1 Speaker Identification System:

Speaker Identification can be thought of as the task of finding who is talking from a set of known voices of speakers. It is the process of determining who has provided a given utterance based on the information contained in speech waves. The unknown voice comes from a fixed set of known speakers, thus the task is referred to as closed set identification. Speaker identification is a 1: N match where the voice is compared against N templates. Error that can occur in speaker identification is the false identification of speaker

### 3.2 Speaker Verification System:

Speaker Verification on the other hand is the process of accepting or rejecting the speaker claiming to be the actual one. Since it is assumed that imposters (those who fake as valid users) are not known to the system, this is referred to as the open set task. Speaker verification is a 1:1 match where one speaker's voice is matched to one template.
Table 3 shows the comparison of Speaker Identification system and Speaker Verification System [2]:

**Table 3: comparison between Speaker identification and Speaker verification**

| Sr. No. | Speaker Identification System | Speaker Verification system |
|---|---|---|
| 01 | Speaker is often reluctant. | Speaker is cooperative. |
| 02 | The system is required to test many patterns. | The system is required to test only one pattern. |
| 03 | The utterance may be anything. | The utterance may be a password. |
| 04 | System response may be slow. | System response is required to be fast. |
| 05 | SNR may be poor. | SNR is controllable. |

### 3.3 Architecture and operation of Speaker identification System [2]:

Speaker identification systems can be either text dependent or text independent. Text dependent systems will use the same test phrase whereas text independent systems have no restriction on the test phrase. Text independent system must use features which depend only on the speaker and are independent of the uttered word. The system operates in two modes –the training mode and the identification mode. Let us consider the system is designed for N speakers.

A training mode of identification system will be as shown in Figure 1. At least five different utterances from each speaker are available to the system. The system will finds the feature vectors corresponding to each utterance. The threshold is computed as the maximum possible variation in the feature vector.

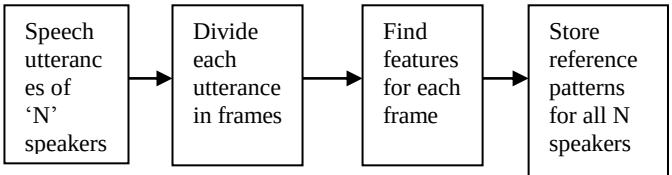


**Figure 1:** Block diagram of Training mode of Speaker Identification System

The identification mode will be as shown in Figure 2.The external utterance is used to find the speakers identity. The test utterance is divided into a number of frames. The feature vectors so obtained are compared with reference patterns stored for each speaker. Here the comparison is required with the patterns of all speakers.

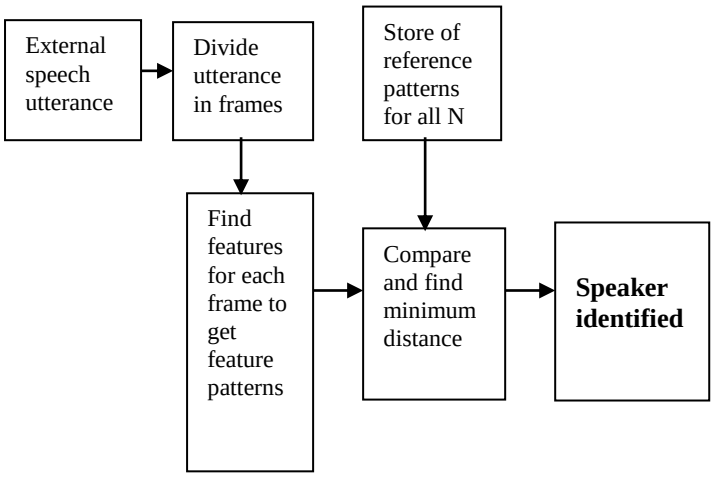


**Figure 2:** Block diagram of Identification mode of Speaker Identification System

## 4. Feature Extraction Methods:

Speech feature extraction is the signal processing frontend which has purpose to converts the speech waveform into some useful parametric representation. These parameters are then used for further analysis in speaker identification system. The list of widely used feature extraction techniques are as follows:

1. Linear Predictive Cepstral Coefficients (LPCC)
2. Perceptual Linear Predictive (PLP)
3. Mel-Frequency Cepstral Coefficients (MFCC)
4. Gammatone Frequency Cepstral Coefficients (GFCC)

### 4.1 LPCC (Linear Predictive Cepstral Coefficient):

Linear prediction analysis is an important method of characterizing the spectral properties of speech in the time domain. [5], in this analysis method, each sample of the speech waveform is predicted as a linear weighted sum of the past *p* samples. The weights which minimize the mean squared prediction error are called the predictor coefficients. The predictor coefficients vary as a function of time and it is usually sufficient to compute them once every 20 ms.

### 4.2 (PLP) Perceptual Linear Predictive Coefficients:

PLP feature extraction is similar to LPC analysis, is based on the short-term spectrum of speech. In contrast to pure linear predictive analysis of speech, LP modifies the short-term spectrum of the speech by several psychophysically based transformations.[6]. PLP performs spectral analysis on speech vector with frames of N samples with N band filters. Finally, LP analysis is done with FFT and the final observation vectors are extracted by taking the real values of inverse FFT.

### 4.3 (MFCC) MEL Frequency Cepstral Coefficients:

The MFCC is the most evident Cepstral analysis based feature extraction technique for speech and speaker recognition tasks. It is popularly used because it approximates the human system response more closely than any other system as the frequency bands are positioned logarithmically [4].

Feature extraction steps:

1. Pre-emphasize input signal
2. Perform short-time Fourier analysis to get magnitude spectrum
3. Wrap the magnitude spectrum into Mel-spectrum
4. Take the log operation on the power spectrum (i.e. square of Mel-spectrum)
5. Apply the discrete cosine transform (DCT) on the log-Mel power spectrum to derive Cepstral features and perform Cepstral.

### 4.4 GFCC: Gammatone Frequency Cepstral Coefficients:

Gammatone filters are realized purely in the time domain. Specifically, the filters are applied directly on time series of speech signals by simple operations such as delay, summation and multiplication. This is quite different from the widely adopted frequency-domain design, where signals are transformed to frequency spectra first and the gammatone filers then applied upon them. The time domain implementation avoids unnecessary approximation introduced by short-time spectral analysis, and saves a considerable proportion of computation involved in FFT [4].

Feature extraction steps:

1. Pass input signal through a 64-channel gammatone filter bank.
2. At each channel, fully rectify the filter response (i.e. take absolute value) and decimate it to 100 Hz as a way of time windowing.
3. Then take absolute value afterwards. This creates a time frequency (T-F) representation that is a variant of cochleagram.
4. Take cubic root on the T-F representation.
5. Apply DCT to derive Cepstral features.

## 5. Current Status of Speaker Recognition System in India:

**1. Indian Institute of Technology, Guwahati:** Study of Source Features for Speech Synthesis and Speaker Recognition and Development of Person Authentication System based on Speaker Verification in Uncontrolled Environment, Development of Speech based multi-level authentication system.

**2. Indian Institute of Technology, Kharaghpur:** Development of speaker verification software for single to three registered user(s) and speaker verification system to increase security in limited user Environment, FPGA based Automatic Speaker Recognition, Development of (SRS) Speaker Recognition Software for Telephone Speech, and Development of speech database for speaker recognition application

**3. Indian Institute of Technology, Madras:** TATA Power Project involves Speaker Identification in Radio Communications Channel.

**4. CFSL, Chandigarh:** CFSL is the first Forensic Laboratory in the Country to develop text independent speaker identification system.

**5 National Institute of Technology (NIT), Silcher and NEHU Shillong:** Development of Speech based multi-level authentication system.

## 6. Conclusion:

In this paper, we have presented an extensive survey of automatic speaker recognition systems. We have presented a study of the current status of research being carried out in the field of speaker recognition in India. Totally, four feature extraction algorithms namely MFCC, LPC, PLP and GFCC are implemented and its performances were observed. By investigating these feature vectors along with the speaker identification techniques it was found that the GFCC features gave better results and accuracy for training and testing for all noise speech data.

## 7. Acknowledgement:

We would like to thank the publishers, researchers and teachers for their guidance. We would also thank the college authority for providing the required infrastructure and support. Last but not the least we would like to extend a heartfelt gratitude to my family members for their support.

## References:

[1] Tiwalade O. Majekodunmi, Francis E. Idachaba, "A Review of the Fingerprint, Speaker Recognition, Face Recognition and Iris Recognition Based Biometric Identification Technologies", of the World Congress on Engineering 2011 Vol II WCE 2011, July 6 - 8, 2011, London, U.K. ISBN: 978-988-19251-4-5 ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[2] Dr. Shaila D. Apte, "Speech and audio processing", Wiley India, 2012.

[3] Anand Vardhan Bhalla, Shailesh Khaparkar, Mudit Ratana Bhalla "Performance Improvement of Speaker Recognition System", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 3, March 2012.

[4] Vimala. C., Radha. V, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal (WCSIT), ISSN: 2221-0741,Vol.2, No. 1, pp. 1-7,2012.

[5] R. W.Schafer and L. R. Rabiner, "Digital representations of speech signals", IEEE, vol. 63, pp. 662-677, Apr. 1975.

[6] Urmila Shrawankar, "Techniques for Feature Extraction in Speech Recognition System: A Comparative Study", SGB Amravati University.

[7] Kishore Prahallad, Sudhakar Varanasi, Ranganatham Veluru, Bharat Krishna M, Debashish S Roy "Significance of Formants from Difference Spectrum for Speaker Identification", INTERSPEECH-2006, paper 1583-Tue1CaP.1

[8] L.R.Rabiner and R.W. Schafer, "Digital Processing of speech signals" (sixth impression), Pearson Education and Dorling Kindersley Pvt. Ltd, 2011.

[9] D. Garcia-Romero, X. Zhou, and C. Y. Espy-Wilson, "Multicondition training of Gaussian PLDA models in i-vector space for noise and reverberation robust speaker recognition," in *Proc. ICASSP*, 2012, pp.4257–4260

[10] A.Revathi1, R. Ganapathy and Y. Venkataramani , "Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach", International Journal of Computer science & Information Technology (IJCSIT), Vol 1, No 2, November 2009

[11] Tudor Barbu "A Supervised Text-Independent Speaker Recognition Approach", World Academy of Science, Engineering and Technology 33, 2007.

[12] Mangesh S. Deshpande and Raghunath S. Holambe; Speaker Identification Using Admissible Wavelet Packet Based Decomposition International Journal of Signal Processing 6:1, 2010