STM JOURNALS

Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720(online)
Volume 2, Issue 2
www.stmjournals.com

Voice Recognition Technique using Guassian Mixture Model

Anagha Girme, Nivedita Mahato*, Sucheta Manve, Abhijeet Banubakode
Department of Information Technology, JSPM'S Rajarshi Shahu College of Engineering, Tathwade, Pune, India

Abstract

This paper gives us some of the important information about the methods used for voice recognition system. Voice recognition is the identification of person with the help of characteristic of voice. Voice recognition has scope ranging from access control to forensics. In this system we are able to identify the speaker as well as verification of the speaker take place using feature extraction and feature matching technique using GMM. In this work the features of voice of the person are extracted with the help of Mel-Frequency Cepstral Coefficient (MFCC) and Subband based Cepstral Parameter (SBC). Hence the accuracy of speaker recognition increases giving speaker a more flexible system .In the experimental result SBC is more accurate than MFCC.

Keywords: *Mel-Frequency cepstral coefficient (MFCC), guassian mixture model (GMM), subband based cepstral parameter (SBC), feature extraction, speaker identification, speaker verification*

****Author for Correspondence*** E-mail: nivedita2293@gmail.com

INTRODUCTION

Today as the technology is evolving faster there is much increase in security issues. Bio informatics is such a technological evolution in which human characteristics such as eyes, fingerprints, voices, heartbeats etc are used for security purpose. Voice Recognition is the technology used to identify the person with the voice. Voice Recognition has two main terms in it - basically voice identification and voice verification. Voice Identification involves identifying the voice as well as the person to whom the voice belongs and tell the identity of the person whereas Voice Verification is to just verify the voice of the person [1].

Identification can be 1:1 and verification can be 1:N where N is number of voices of other person.The working system first takes the data set in the form of voices and stores it. The data set to store the data can be also given at runtime. The input file is given to the system. Then by selecting the algorithm to extract features, the features are extracted. After the feature extraction, classification is done. Thus the result of perfect matching is displayed only after the features of data set and input voice is matched [2].

Software Architecture

- Feature Extraction- The features of voice which are given as input are extracted using MFCC or SBC algorithm.
- Feature Matching and classification- The extracted features are the matched with the data set and the classification of voice whether the matching is perfect or poor is done [4].
- The whole architecture can be explained in the following diagram Figures 1 and 2.

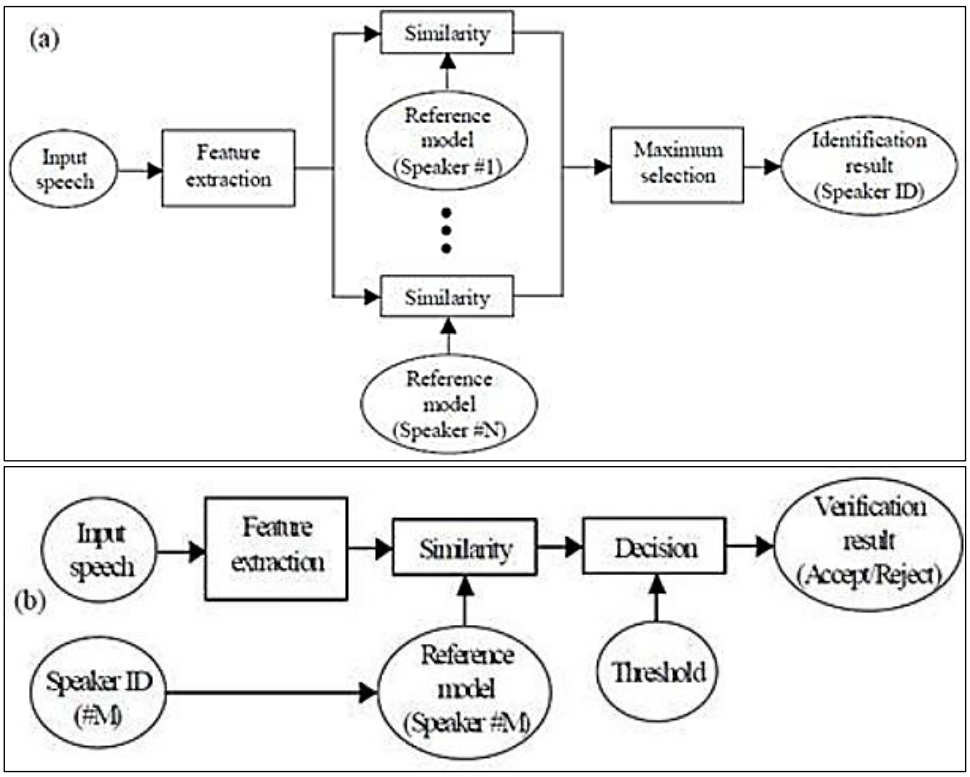

Fig. 1: *Voice Recognition Process (a) Speaker Identification, (b) Speaker Verification.*

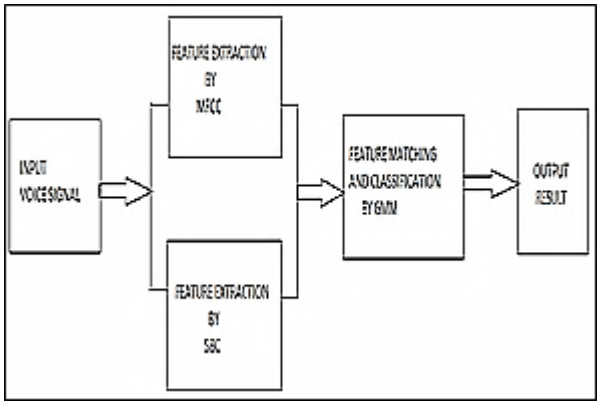

Fig. 2: *Block Diagram of Voice Recognition System.*

Feature Extraction

Input data to the system will be taken as voice. Voice can't be directly taken by any algorithm for further processing. Its characteristic or features have to be considered for further processing. So we take features of voice such as pitch and its frequency wavelength and extract them. Feature extraction is done with the help of two algorithms viz, MFCC or SBC [5].

- **MFCC Algorithm**

This algorithm has been proven to be a powerful tool in field of voice recognition. Speech frames are used to compute MFCC algorithm. MFCC in Figure 3 has following procedure first we apply the Fourier transform on the input signal. Then we map the power of spectrum obtained in above step to the Mel scale. Then we take logs of the power of Mel frequencies at Mel signal. Then we take discrete cosine transform of Mel log powers. Here we convert the Mel log spectrum back to time. Thus we get the result as mel frequency cepstrum coefficient [6]. As the Mel coefficient is real numbers we are able to convert them into time domain using discrete cosine transform. Finally the score is calculated. If we get the score which is greater than 6.8 then the input is a perfect match or the input falls in the poor quality [7].

Each voice sample is divided into frames and its MFCC is computed and stored in form of matrix. In frame blocking step, the voice signal is blocked into frames of N samples and adjacent frames are separated by M(M<N). Thus the first frame is of N samples and the other begins with M n so on. The next step is window step in which the frames are processed to remove the discontinuities in beginning as well as the end of each signal to reduce signal distortion [8]. In next step we

convert the N samples from domain of time to its frequency domain with the help of Fast Fourier Transformation. FFT is fast algorithm which is defined on set of N samples *{x}* as,

$X_n = \sum_{k=0}^{N-1} x_k \, e^{-2\pi jn/N}$ here 0, 1, 2……N-1

finally the log Mel spectrum is converted back to time and thus we obtain the MFCC.

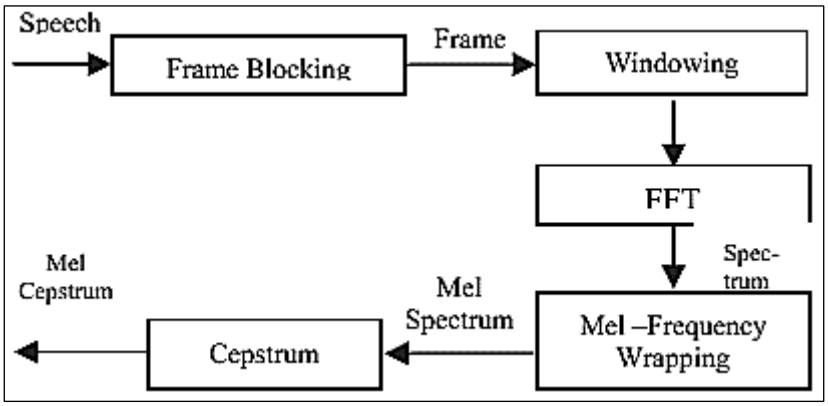

Fig. 3: *Block Diagram of MFCC.*

- **SBC Algorithm**

In MFCC the calculation is done in two stages. In first the Mel signals log and in second we apply DCT to convert it into time. In SBC the first stage is same as MFCC but in derivation it uses Wavelet packet transform instead of Fourier transform. SBC proves to be a better technique than MFCC. Discrete cosine transformation is applied to derive sbc parameters from subband energies. It uses wavelet packet tree for comparison [9].

Here the score is also computed same like MFCC but feature extraction method is completely different than MFCC. Here the score range is 21.5. If the score is greater than 21.5 than the input is perfect match otherwise it goes in poor quality.

The subband energies derive the SBC parameter by following equation,

$$\sum_{i=1}^{l} log\, S_i \, Cos\left(\frac{n(i-0.5)}{L} * \pi\right)$$

Where, n=1.....n which is number of SBC parameter and L is total number of frequency bands.

FEATURE MATCHING AND CLASSIFICATION

After the features are extracted we need to match them with the features of already stored voice data set which follows the classification process. The classification process includes to identify to whether the voice matches the given input voice. This is done with help of GMM algorithm. The Gaussian mixture model is the mixture of M Gaussian and is given by following equation,

$$p = \left(\frac{\vec{x}}{\lambda}\right) = \sum_{i=1}^{M} p_i b_{i(\vec{x})}$$

where *x*= D dimensional random vector, *bi(x)* and *i=1,2,3.....M* are component densities, and *pi* and *i=1,2,3.....M* are mixture weights. Each component density is D dimensional function in the form,

$$b_{i(\vec{x})} = \frac{1}{(2\pi)^{D/2}|\sum i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x}-\mu_i)^T \sum_i^{-1}(\vec{x}-\mu_i)\right\}$$

where denotes the mean vector and denotes the covariance vector matrix [10].

CONCLUSION

Voice Recognition System can be efficiently implemented using Mel-Frequency Cepstral Coefficient (MFCC) and Subband based Cepstral Parameter (SBC). Hence, we also conclude that SBC is more preferable than MFCC as accuracy level given by SBC algorithm is more as compared to MFCC algorithm.

The input data set undergoes feature extraction and feature matching methods. The proposed model could go a long way in increasing the security level. On the basis of these methods much research has been done [8].

REFERENCES

1. Zhizeng Luo , Zhao. Speech Recognition and Its Application in Voice-based Robot Control System Jinghing. *Robot research laboratory. Zhejiang Province, China: Hangzhou Institute of Electronic Engineering Hangzhou*. 2010.
2. Romero F, Caballero-Morales SO. Speaker identification using Neural Networks on an FPGA. *División de Estudios de Posgrado.*
3. Srikanth Ronanki , Bajibabu B, Prahallad Kishore. *Duration Modelling In Voice Conversion Using Artificial Neural Networks. International Institute information Technology, Hyderabad, India*. http://ravi.iiit.ac.in/~speech/publications/2012_Conf_P003.pdf
4. Hamza A AlAbri, Ahmed M AlWesti, Mohammed A. AlMaawali, *et al. NavEye: Smart Guide for Blind Students Qaboos University*.
5. Fredrickson, Tarassenk. Text-Independent Speaker Recognition Using Neural Network 'Techniques. *Oxford University, UK*.
6. Sankar Ravi , Sethi Netoo Singh . Robust Speech Recognition Techniques Using a Radial Basis function Neural Network for Mobile Applications. *Department of Electrical Engineering. University of South Florida Tampa, Florida*.
7. Yang Hai, Yunfei Xu, Huang Houjun, *et al*. Voice biometrics using linear Gaussian model. 2014
8. *Laboratory of Speech Acoustics and Content Understanding*, Chinese Academy of Sciences, 21 Beisihuan XiLu, Beijing.
9. Dey Subhadeep, Barman Sujit, Bhukya Ramesh K., et al. Speech Biometric Based Attendance System. *Department of Electronics and Electrical Engineering Indian Institute of Technology, Guwahati, India.*
10. Trujillo-Romero F, Caballero- Morales SO. Speaker identification using Neural Networks on an FPGA. *División de Estudios de Posgrado. Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, 2012.

Cite this Article

Anagha Girme, Nivedita Mahato, Sucheta Manve, Abhijeet Banubakode. Voice Recognition Technique Using Guassian Mixture Model. *Journal of Artificial Intelligence Research & Advances*. 2015; 2(2): 12–15p.