IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727,
PP 55-57
www.iosrjournals.org

Text-Dependent Speaker Recognition System Using MFCC

Mr. Vijay K. Kale[1] , Dr.Prapti D. Deshmukh[2]
[1](Dept.of Computer Science & IT, Dr.Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India)
[2](MGM Dr.G.Y.Pathrikar College of Computer Science and IT, Aurangabad (MS), India)

Abstract: *In the past decade, interest in using biometric technologies for person authentication in security system. Voice is one of the most Promising and mature biometric modalities for access control.In this paper present text-dependent speaker recognition in noisy environment.Used low-pass filter to remove the background noise from noised speech signal.For extraction used speech parameterization method such as MFCC and completion of parameterization of speech, Vector quantization has been used in the learning and recognition purpose. We have achieved 90 % recognition rate for noisy environment as compare to existing system used in different technique.*

Keywords: *Feature, Recogination,* MFCC, Noise

I. Introduction

Recently the need for security has growing up. A biometrics is automatic mode of identification human being. Biometric is refers to identifying human being based on his physiological and behavioral characteristics. Physiological characteristics such finger print, Iris, retina, hand geometry, face. Behavioral characteristics such as signature, gait, voice, speaker, speech, keystroke etc.Speaker recognition has two major applications: a) Speaker Identification, and b) Speaker Verification.These are the technique for automatically recognizing who is speaking on the basis of individual informational included in speech waves. This technique make it possible to use the speakers voice to verify their identity and control accessing to service such voice dialing, banking by telephone, telephone shopping, database access service, information service, voice mail, security control for confidential areas, and remote access to computers. Speaker recognition is technique that can identify or verify of the pre known speakers [1]. The recognition process is based on the parameter of speech samples. Speaker recognition is classified in to two sections Text-dependent recognition and Text-independent recognition. Mel-frequency cepstral coefficients (MFCCs) are coefficients that are derived from a type of cepstral representation of the audio the difference between the cepstrum and the mel-frequency cepstrum is that in the MFCC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cestrum.This frequency warping can allow for better representation of sound, for example, in audio compression.

Pitch: Pitch or Fundamental frequency (F0) is a perceptual attribute of a periodic or nearly periodic sound, defined as the frequency of a sine wave that is matched to the target sound in a psychophysical experiment [2]. The pitch of the voice refers to how high or low the note produced by the vibrating vocal folds appears to be. The faster the vocal folds vibrate the higher the pitch. Conversely, slowly vibrating folds will produce a lower pitch.

Formants: Formants are spectral peaks of the sound spectrum which is defined here as P (f). They are basically resonances occurring in the vocal tract [3]. They generally occur every 1000 Hz and they are a concentration of acoustic energy at a particular frequency. There has been a lot of research done in this field to carry out speaker recognition using formants

Experimental Work

The database collected sample speech from various 50 speakers of 10 utterance of each. Input speech is recorded at sampling rate of 16000Hz.The example recorded speech sampled in digital form as following diagram.

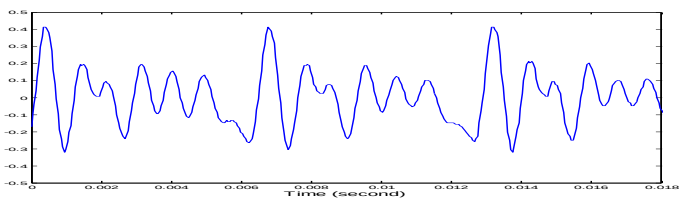

Figure 1: speech sample

A speaker recognition system has three important parts

- Preprocessing
- Feature extraction
- Pattern matching

Pre-processing: The first step of speech processing, the system needs to remove the silence and background noise from the speech signal. Silence detection strategy is based on the energy level in way that if variation of the signal sample in a known speech frame, compare to the frame mean are big enough, the frame is considered as frame, otherwise as a silence [4].

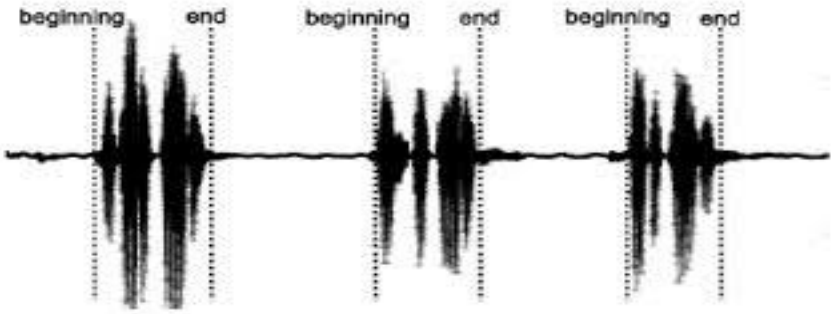

Figure 2: silence detection

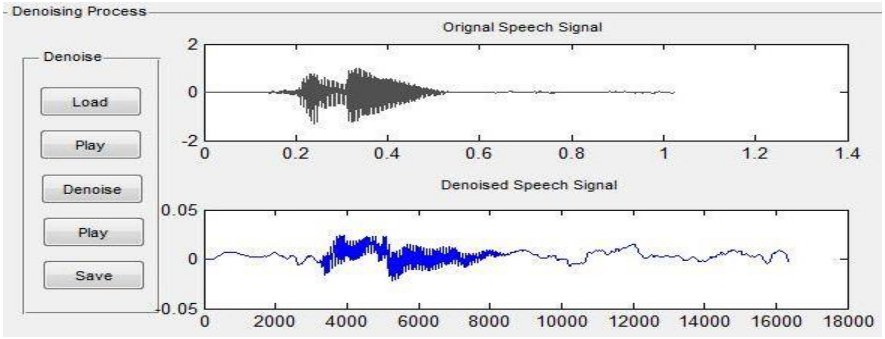

Figure 3: noise removes

After silence detection and noise remove from speech signal, the raw speech signal was prepared for feature extraction.

Feature Extraction: The most important part of speaker recognition is the extraction of features from the speech signal. This step helps us to bring the important part of data from set of samples. This make easier to find the different types of features extraction techniques. Linear Predictive Cepstral Coefficients (LPCC), Mel-frequency Cepstral Coefficients (MFCC) [5].To extract the feature from the speech signal mel-frequency cepstral coefficients are frequently used speech parameterization in speaker recognition. The example of MFCC feature extraction from speech sample in digital form as following figure

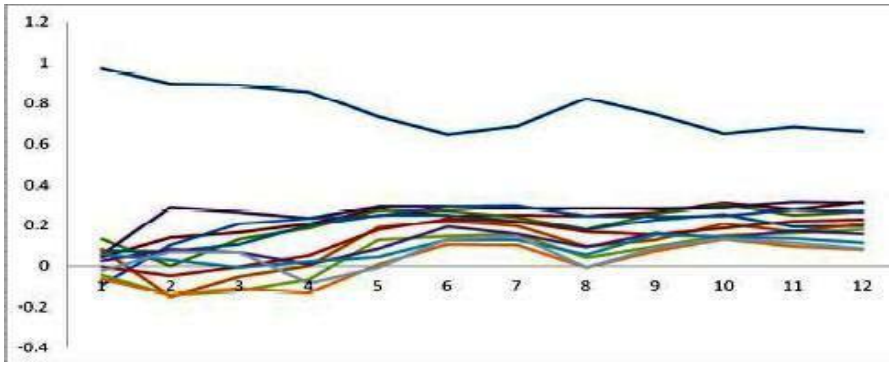

Figure 4: Plot of MFCC 13 features for 12 frames

The extracted feature speech samples of a speaker are quantized to number of centroids using quantization algorithm MFCC are calculated in training phase and testing phase. The speaker uttered same word in a training session and once in testing session later[6].Then the Euclidean distance the MFCC of each speaker in training phase to centroids of individual speaker in testing phase.The main goal is that of recognize the speaker from set of N known speaker.

Feature Matching:

The feature matching techniques used in speaker recognition include Dynamic Time Wrapping (DTW), Hidden Markov Model (HMM) and Vector Quantization (VQ). We have used VQ approach to ease of implementation and high accuracy. The vector quantization also called pattern matching quantization is often used in lossy data compression[7].VQ divides a large set of points into groups having approximately the same number of points close set to them.Each group is represented centroid point. After the enrolment session, the acoustic vectors extracted from input speech of a speaker provide a set of training vectors. The next imported step is to build a speaker specific VQ codebook for this speaker using those training vectors

System Result:

The database consists of 50 distinct speakers including both male and female. It also contains 50 speech samples used for training and testing the speaker recognition system. Speech samples recorded. Recognition rate of the system defined as follows.

$$RR \frac{Total\ Number\ of\ Speakers}{Total\ Number\ Reconised\ speakers} \times 10$$

Sr.No	Feature Maching Technique	Feature Maching Technique	Recognition Rate
1	MFCC	VQ	90 %

Table1: Recognition Rate of our System

III. Conclusion

The system performance of measured on the basis of accuracy, time taken to compute the feature recognition, it was observed that the speaker recognition system performs well in noisy environment. The entire research process was carried out using MATLAB. It noticed that achive the 90 % recognition rate. Using MF CC and VQ Technique.

References

[1]. Campbell, J.P., Jr.; "Speaker recognition: a tutorial" Proceedings of the IEEE Volume 85, Issue 9, Sept. 1997 Page(s):1437 – 1462.
[2]. PRADEEP. CH," Text Depedent Speaker Recognition Using MFCC And LPC VQ", National Institute of Technology, Rourkela, 2007.
[3]. Fu Zhonghua; Zhao Rongchun; "An overview of modeling technology of speaker recognition", IEEE Proceedings of the International Conference on Neural Networks and Signal Processing Volume 2, Page(s):887 – 891, Dec. 2003.
[4]. Nakai, M.; Shimodaira, H.; Kimura, M.; "A fast VQ codebook design algorithm for a large number of data", IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume 1, Page(s):109 – 112, March 1992.
[5]. D. Reynolds and B. Carlson, "Text-Dependent Speaker Verification Using Decoupled and Integrated Speaker and Speech Recognizers," In EUROSPEECH in Madrid, ESCA, pp.647-650, 1995.
[6]. Mangesh S. Deshpande and Raghunath S. Holambe; Speaker Identification Using Admissible Wavelet Packet Based Decomposition International Journal of Signal Processing 6:1, 2010
[7]. S. Furui., "Speaker independent isolated word recognition using dynamic features of the speech spectrum", IEEETransactions on Acoustics, Speech and Signal Processing, Vol. 3, pp. 52–59, Feb. 1986.B. S. Atal,"Text-independent speaker recognition," J.A.S.A., 52, 181 (A), 83th ASA, 1972.