Feature Extraction Using Mel Frequency Cepstrum Coefficients for Automatic Speech Recognition

Dr. C.V.Narashimulu[1]

Mr. Touseef Sumer[2]

[1]Professor, [2]Assistant Professor, Dept. of ECE, Geethanjali College of engineering and Technology (Autonomous) Hyderabad.

Abstract—The most natural mode of communication for human being is Speech. The task of speech recognition is to convert speech into a sequence of words by a computer program. Automatic speech recognition (ASR) will plays a vital role in taking technology to the people. Find many applications of speech recognition such as direct voice input in aircraft, data entry, speech-to-text processing, voice user interfaces such as voice dialing. Generally ASR system can be divided into two different parts, namely feature extraction and feature recognition. In this paper we present MATLAB based feature extraction using Mel Frequency Cepstrum Coefficients (MFCC) for ASR. IT also describes the development of an efficient speech recognition system using different techniques such as Mel Frequency Cepstrum Coefficients (MFCC).

Keywords-- Automatic Speech Recognition, Mel frequency Cepstral Coefficient, Predictive Linear Coding

1. INTRODUCTION

Speech Recognition (is also known as Automatic Speech Recognition (ASR) or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program. Computer programs for speech recognition seem to deal with ambiguity, error, and non grammaticality of input in a graceful and effective manner that is uncommon to most other computer programs. Yet there is still a long way to go. We can handle relatively restricted task domains requiring simple grammatical structure and a few hundred words of vocabulary for single trained speakers in controlled environments, but we are very far from being able to handle relatively unrestricted dialogs from a large population of speakers in uncontrolled environments. Many more years of intensive research seem necessary to achieve such a goal. The idea of human machine interaction led to research in Speech recognition. Automatic speech recognition uses the process and related technology for converting speech signals into a sequence of words or other linguistic units by means of an algorithm implemented as a computer program. Speech understanding systems presently are capable of understanding speech input for vocabularies of thousands of words in operational environments.

Speech signal conveys two important types of information: (a) speech content and (b) The speaker identity. Speech recognisers aim to extract the lexical information from the speech signal independently of the speaker by reducing the inter-speaker variability. Speaker recognition is concerned with extracting the identity of the person.

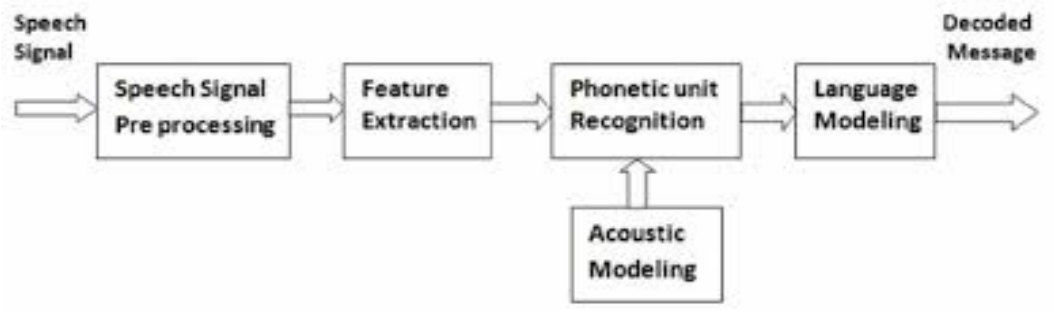


Figure 1 Speech Recognition System

2.The structure of proposed system consists of two modules

- Speaker Identification
- Speech Recognition

Speaker Identification Feature extraction is a process that extracts data from the voice signal that is unique for each speaker. Mel Frequency Cepstral Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech

These extracted features are Vector quantized using Vector Quantization algorithm. Vector Quantization (VQ) is used for feature extraction in both the training and testing phases. It is an extremely efficient representation of spectral information in the speech signal by mapping the vectors from large vector space to a finite number of regions in the space called clusters.

After feature extraction, feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features with the database

Speech Recognition System Hidden Markov Processes are the statistical models in which one tries to characterize the statistical properties of the signal with the underlying assumption that a signal can be characterized as a random parametric signal of which the parameters can be estimated in a precise and well-defined manner. In order to implement an isolated word recognition system using HMM, the following steps must be taken

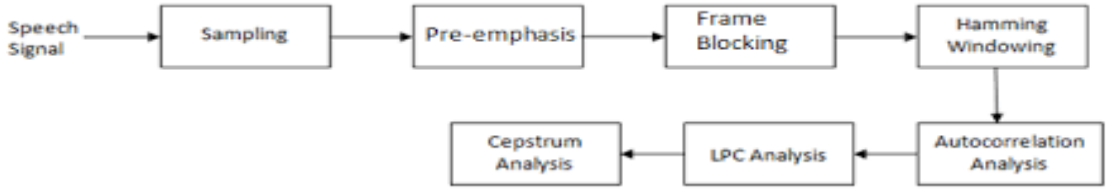


Figure 2 Mel Frequency Cepstrum Coefficient Block Diagram

The most commonly used acoustic features are mel-scale frequency Cepstral coefficients. Explanation of step by step computation of MFCC is given below:-

1. Pre-Emphasis- In this step isolated word sample is passed through a filter which emphasizes higher frequencies. It will increase the energy of signal at higher frequency.

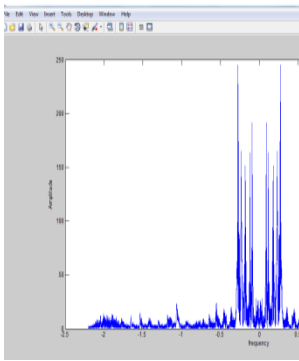
2. Frame blocking: The speech signal is segmented into small duration blocks of 20-30 ms known as frames. Voice signal is divided into N samples and adjacent frames are being separated by M (M
3. Y (n) = X (n) * W (n) Where W (n) is the window function
4. Fast Fourier Transform: FFT is a process of converting time domain into frequency domain. To obtain the magnitude frequency response of each frame we perform FFT. By applying FFT the output is a spectrum or periodogram.
5. Triangular band pass filters: We multiply magnitude frequency response by a set of 20 triangular band pass filters in order to get smooth magnitude spectrum. It also reduces the size of features involved. Mel (f) =1125* ln (1+f/700)
6. Discrete cosine transform: We apply DCT on the 20 log energy Ekobtained from the triangular band pass filters to have L mel-scale Cepstral coefficients. DCT formula is shown below Cm=∑k=1 N cos [m*(k-0.5)*π/N]* ,m=1,2……L Where N = number of triangular band pass filters, L = number of mel-scale Cepstral coefficients. Usually N=20 and L=12. DCT transforms the frequency domain into a time-like domain called frequency domain. These features are referred to as the mel-scale Cepstral coefficients .We can use MFCC alone for speech recognition but for better performance, we can add the log energy and can perform delta operation.
7. Log energy: We can also calculate energy within a frame. It can be another feature to MFCC. 2. Delta cestrum: We can add some other features by calculating time derivatives of (energy + MFCC) which give velocity and acceleration. Δ Cm (t) =[∑τ=−$MMCm$(t+τ)τ]/[∑τ=−$MM\tau$ 2] Value for M=2, if we add the velocity, feature dimension is 26.

2. SIMULATION

Pitch Features

	Sum	Maximum	Minimum	Variance	quartiles	Standard Deviation
F0	4.707	1	-1	0.0202	-0.7452	0.142
F1	5.873	0.561	-0.561	1.706	-0.0464	0.031
F2	2.8796	1	-1	0.0303	-1	0.1741
F3	1.4398	1	-1	1	-1	0.17
F4	9.7436	0.768	-0.76	0.05	0.711	0.241
F5	1.49	0.042	-0.042	1.03	-0.049	0.012
F6	1.4975	0.03	-0.03	1.1269	-0.0437	0.0106

Fundamental Frequency



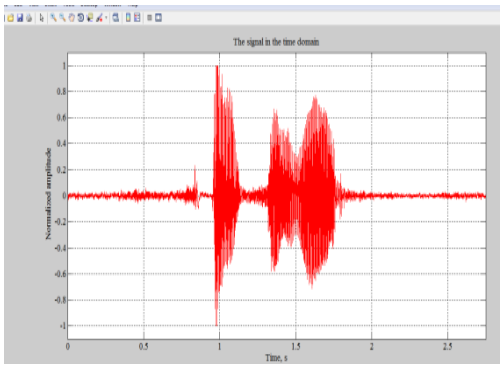


Figure 4 Fundamental Frequency

Figure 5 The Signal in Time Domain

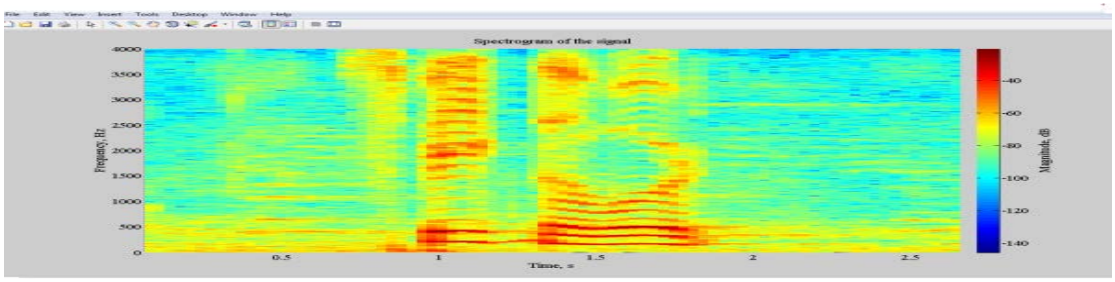


Figure 3 Spectrogram of the Signal

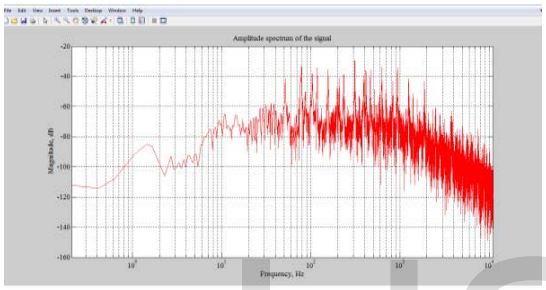


Figure 4 Amplitude Spectrum of the signal

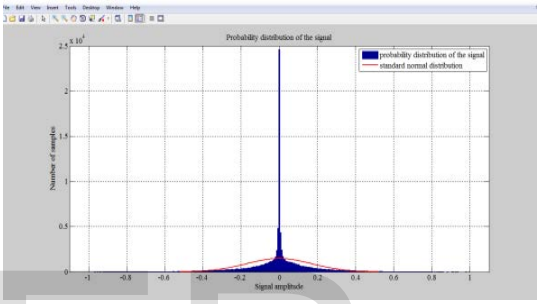
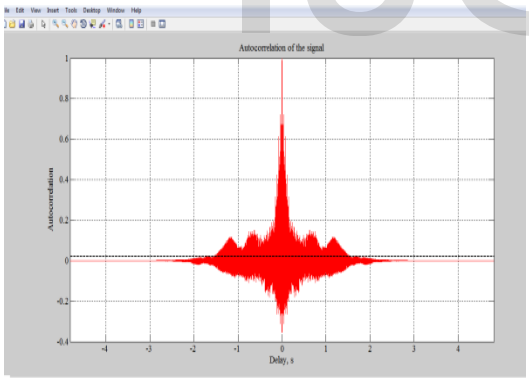


Figure 5 Probability Distribution of the Signal



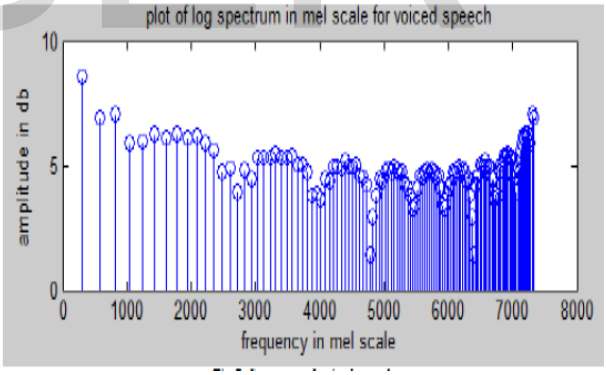


Figure 6 b Log cepstrum plot

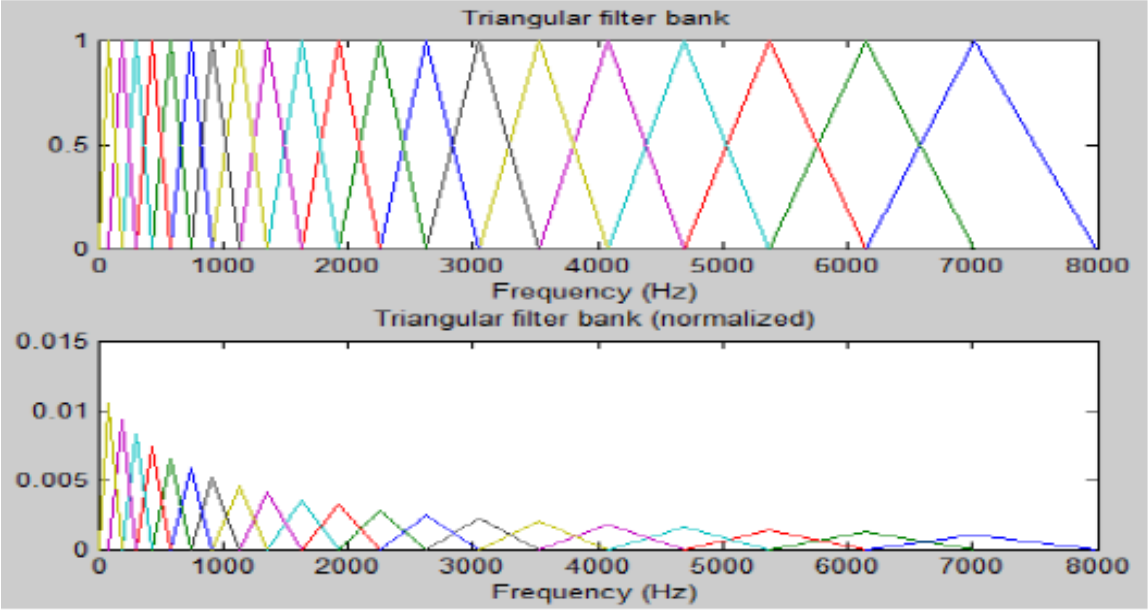


Figure 7 Tringular filter bank

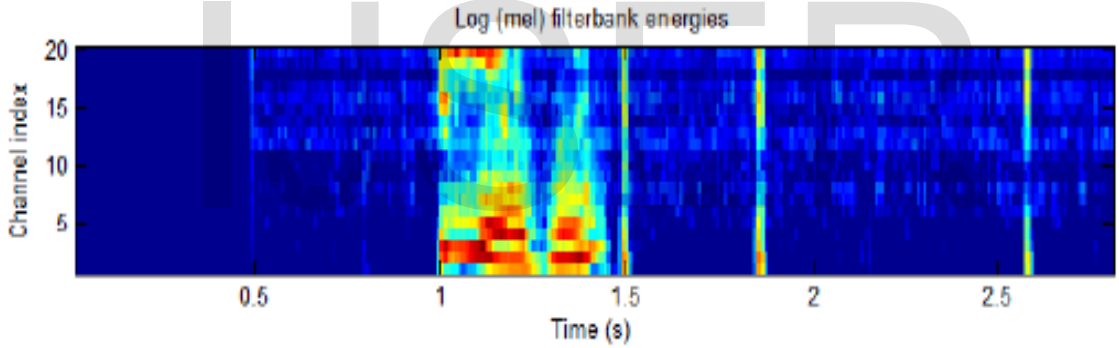


Figure 8 Filter bank energies

CONCLUSION

This paper we have successfully denoise the input sample and while extracting the MFCC coefficients A real-time speaker recognition system using MFCC has been achieved and the experimental result has been analyzed using MATLAB we also taken into the consideration of Delta energy function and draw a conclusion that we can increase the MFCC coefficient according to our requirement. Features are extracted based on information that was included in the speech signal. Extracted features were stored in a .wav file. In our future work MFCC coefficients for designing a speaker independent system type

REFERENCES

[1] Rishiraj Mukherjee, Tanmoy Islam, and Ravi Sankar “text dependent speaker recognition using shifted mfcc” IEEE, 2013,

[2] Santosh k. Gaikwad, Bharti W. Gawali and Pravin Yannawar, "A Review on Speech Recognition Technique", international journal of computer applications, November 2010**.**

[3] A. Shafik, S. M. Elhalafawy, S. M. Diab, B. M. Sallam and F. E. Abd El-samie, "A Wavelet based Approach for Speaker Identification from Degraded Speech", International Journal of Communication Networks and Information Security (IJCNIS), December 2009.

[4] Yousef Ajami Alotaibi, "Comparative Study of ANN and HMM to Arabic Digits Recognition .

Systems", JKAU: Eng. Sci., Vol. 19 No. 1, pp: 43-60 (2008 A.D. / 1429 A.H.)

[5] N.N. Lokhande, N.S. Nehe and P.S. Vikhe, MFCC based Robust features for English word Recognition, *IEEE*, 2012.

[6] L. Muda, M. Begam and I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping(DTW) Techniques, *Journal of Computing, 3(2)*,2010.

[7] Anjali, A. Kumar and N. Birla, Voice Command Recognition System based on MFCC and DTW, *International Journal of Engineering Science and Technology, 2(12)*,2010