

## An Approach to Extract Feature using MFCC

Parwinder Pal Singh, Pushpa Rani

Department of computer science engineering, Chandigarh University  
Assistant Professor<sup>1</sup>, Research Scholar<sup>2</sup>

**Abstract:** - Speech is the most natural and efficient way of communication between humans. Lots of efforts have been made to develop a human computer interface so that one can easily interact and communicate in an unskilled way. Speech recognition systems find their applications in our daily lives and have huge benefits for those who are suffering from some kind of disabilities. This paper presents an approach to extract features from speech signal of spoken words using the Mel-Scale Frequency Cepstral Coefficients. It is a nonparametric frequency domain approach which is based on human auditory perception system. Firstly, all the voice samples of isolated words are taken as the input and by using praat tool denoise all these samples. Then coefficients are extracted by using MFCC as these coefficients collectively represent the short term power spectrum of sound. All this implementation is build in Matlab.

**Keywords:** - Speech Recognition, Mel frequency cepstral coefficients (MFCC), cepstrum

### I. INTRODUCTION

Speech signals are naturally occurring signals and hence, are random signals. These information-carrying signals are functions of an independent variable called time. Speech recognition is the process of automatically recognizing certain word which is spoken by a particular speaker based on some information included in voice sample. It conveys information about words, expression, style of speech, accent, emotion, speaker identity, gender, age, the state of health of the speaker etc. There has been a lot of advancement in speech recognition technology, but still it has huge scope. Speech based devices find their applications in our daily lives and have huge benefits especially for those people who are suffering from some kind of disabilities [3] [4]. We can say that such people are restricted to show their hidden talent and creativity. We can also use these speech based devices for security measures to reduce cases of fraud and theft [7].

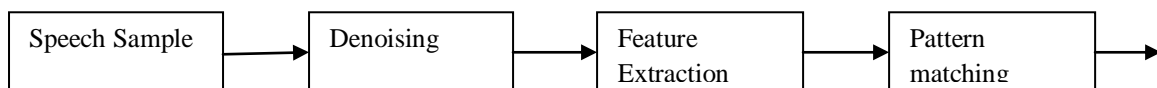


Fig. 1: Speech recognition system

Speech recognition mainly focuses on training the system to recognize an individual's unique voice characteristics. The most popular feature extraction technique is the Mel Frequency Cepstral Coefficients called MFCC as it is less complex in implementation and more effective and robust under various conditions [2]. MFCC is designed using the knowledge of human auditory system. It is a standard method for feature extraction in speech recognition. Steps involved in MFCC are Pre-emphasis, Framing, Windowing, FFT, Mel filter bank, computing DCT.

### II. FEATURE EXTRACTION



Fig. 2: MFCC 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.

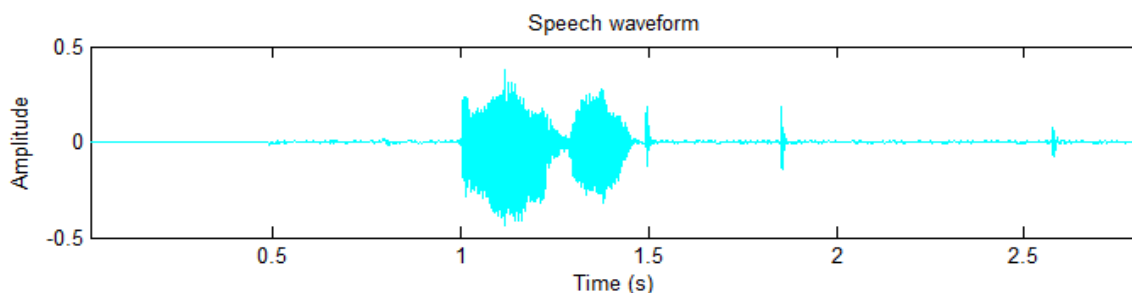


Fig.3. Speech waveform

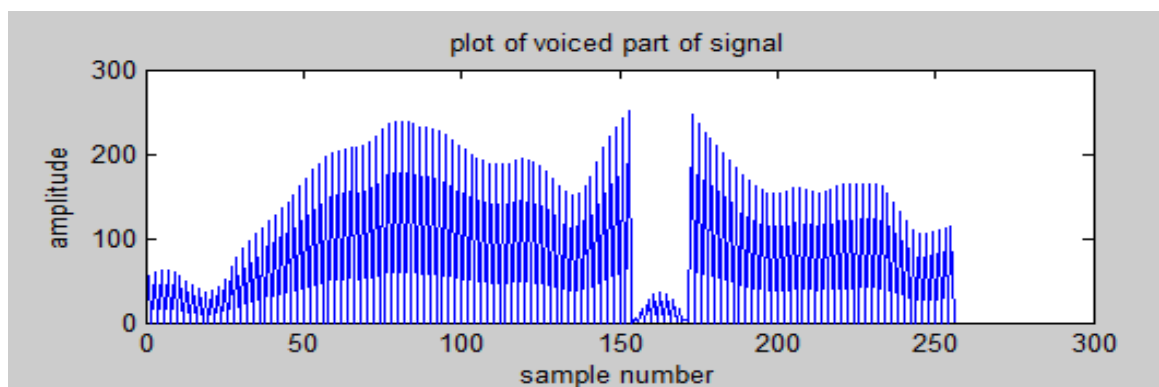


Fig.4. Plot of voiced part of signal

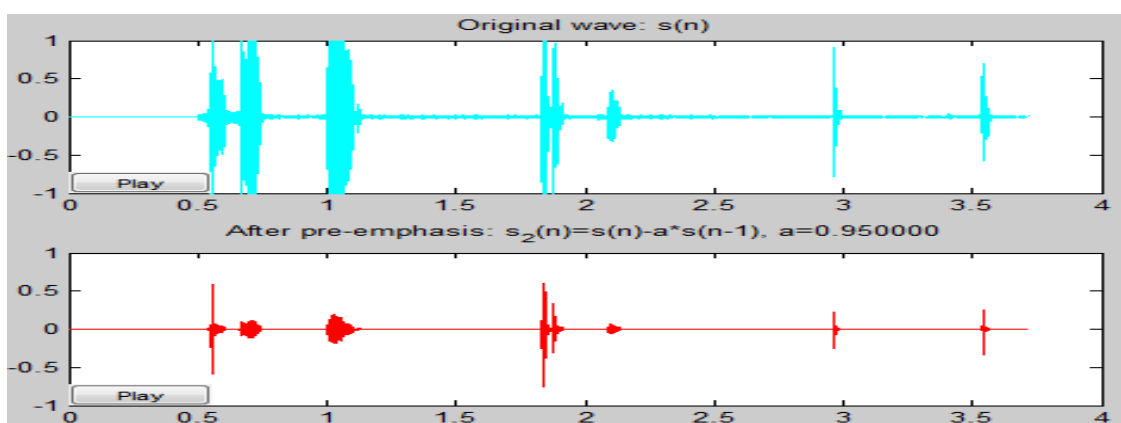


Fig. 5: Pre-Emphasis

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 ( $M < N$ ). Typical values for  $M=100$  and  $N=256$ . Framing is required as speech is a time varying signal but when it is examined over a sufficiently short period of time, its properties are fairly stationary. Therefore short time spectral analysis is done.
3. **Hamming Windowing:** Each of the above frames is multiplied with a hamming window in order to keep continuity of the signal. So to reduce this discontinuity we apply window function. Basically the spectral distortion is minimized by using window to taper the voice sample to zero at both beginning and end of each frame.

$$Y(n) = X(n) * W(n)$$

Where  $W(n)$  is the window function

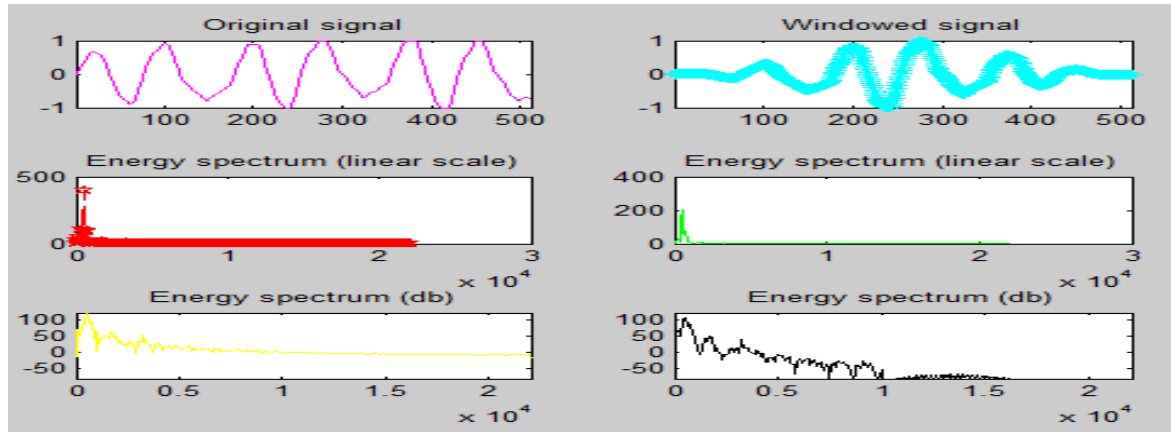


Fig. 6: Edges become sharp by using Hamming Window

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.

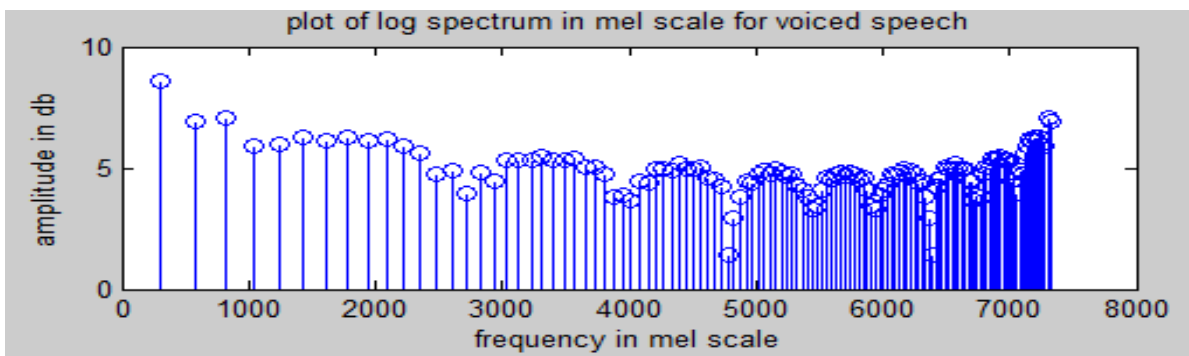


Fig.7. Spectrum of voiced speech

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.

$$\text{Mel}(f) = 1125 * \ln(1 + f/700)$$

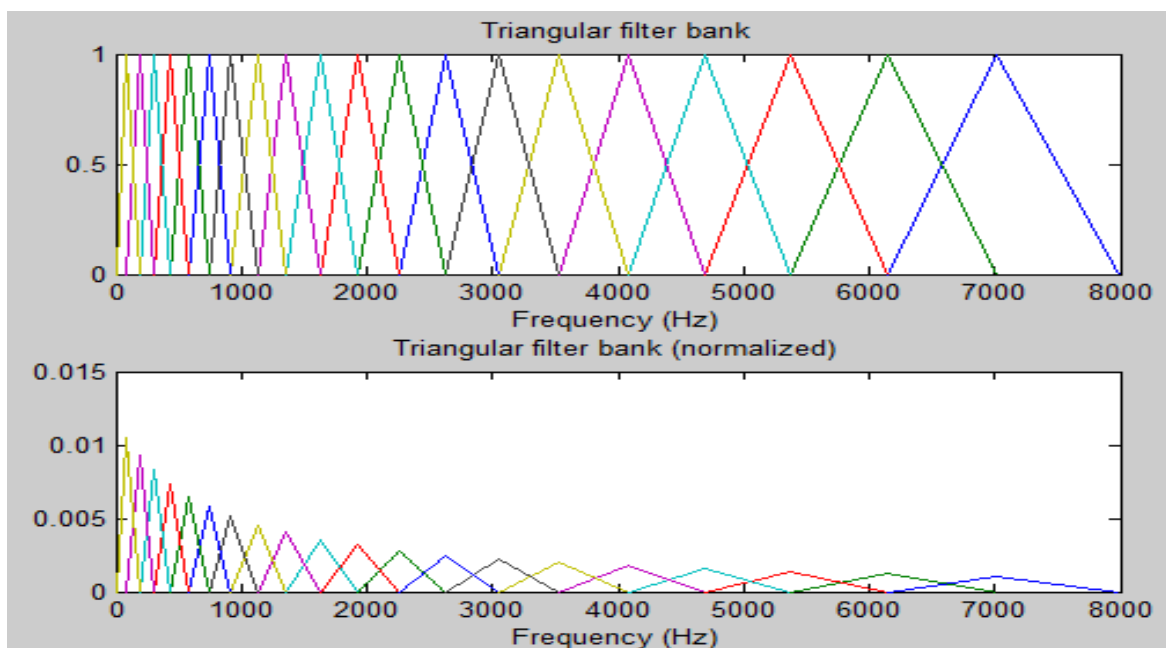


Fig. 8: Normalizing features

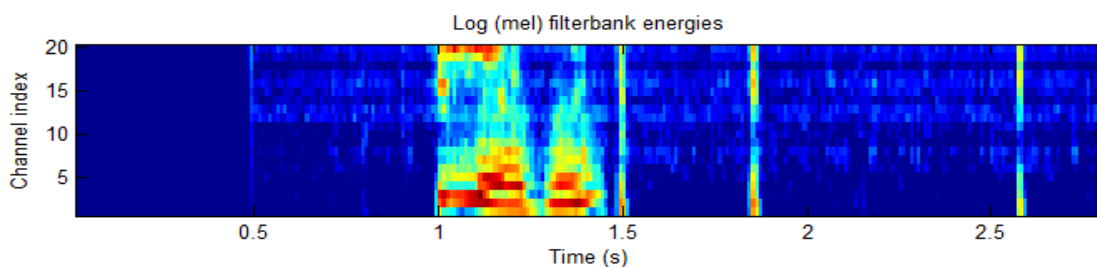


Fig.9. Plot of filter bank energies

6. **Discrete cosine transform:** We apply DCT on the 20 log energy  $E_k$  obtained from the triangular band pass filters to have L mel-scale cepstral coefficients. DCT formula is shown below

$$C_m = \sum_{k=1}^N \cos [m*(k-0.5)*\pi/N] * E_k, m=1,2,\dots,L$$

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 quefrequency 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.

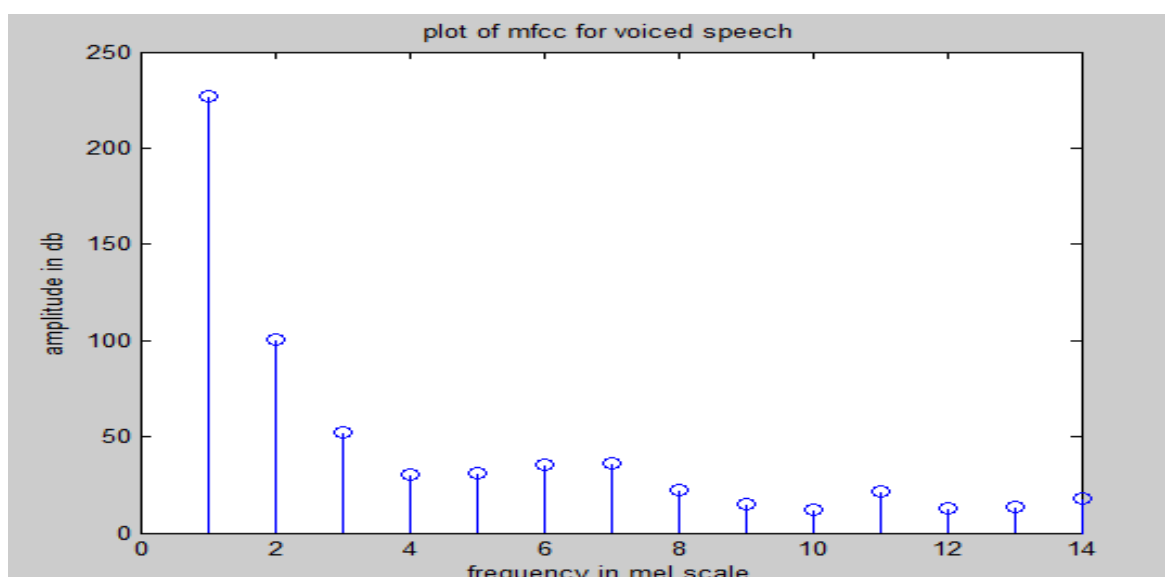


Fig.10. Plot of MFCC coefficients

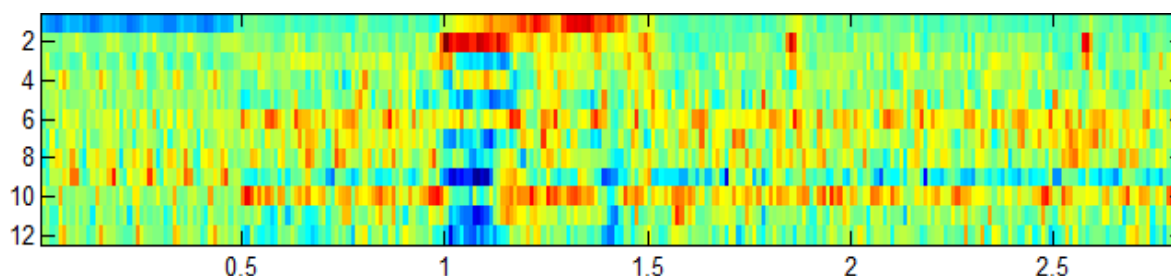


Fig.11. Plot of Mel frequency Cepstrum

1. **Log energy:** We can also calculate energy within a frame. It can be another feature to MFCC.
2. **Delta cepstrum:** We can add some other features by calculating time derivatives of (energy + MFCC) which give velocity and acceleration.

$$\Delta C_m(t) = \frac{[\sum_{\tau=-M}^M C_m(t+\tau)\tau]}{[\sum_{\tau=-M}^M \tau^2]}$$

Value for M=2, if we add the velocity, feature dimension is 26. If we add both acceleration and velocity, the feature dimension is 39.

### III. CONCLUSION

In this research we have successfully denoise the input sample and while extracting the MFCC coefficients 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. We can add velocity and acceleration to extract 39 MFCC coefficients. The MFCC feature extraction technique is more effective and robust, and with the help of this technique we can normalizes the features as well, and it is quite popular technique for isolated word recognition in English language. Features are extracted based on information that was included in the speech signal. Extracted features were stored in a .mat file. In our future work we will do another breakthrough in the field of research, and will use these extracted MFCC coefficients for designing a speaker independent system type.

### REFERENCES

- [1] S. Dhingra, G. Nijhawan and P. Pandit, Isolated Speech Recognition using MFCC and DTW, *International journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 8(2), 2013.
- [2] C. Poonkuzhali, R. Karthiprakash, S. Valarmathy and M. Kalamani, An Approach to feature selection algorithm based on Ant Colony Optimization for Automatic Speech Recognition, *International journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 11(2), and 2013.
- [3] V. Sharma and P. Sharma, Discrete and continuous Mouse Motion using Vocal and Non-Vocal Characteristics of Human Voice, *International journal of Computer Science and Engineering Technology*, 4, 2013.
- [4] C. Ittichaichareon, S. Suksri and T. Yingthawornsuk, speech Recognition using MFCC, *International Conference on Computer Graphics Simulation and Modeling*, 2012.
- [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.