www.ijecs.in
*International Journal Of Engineering And Computer Science ISSN: 2319-7242*
*Volume 5 Issue 10 Oct. 2016, Page No. 18552-18556*

# A Review on Feature Extraction Techniques for Speech Processing

***Amandeep Singh Gill***
Assistant Professor, JMIETI,Radaur
Email id: Amanshergill33@gmail.com

ABSTRACT: Speech and language are considered uniquely human abilities Speech is a complex signal that is characterized by varying distributions of energy in time as well as in frequency, depending on the specific sound that is being produced. The aim of digital speech processing is to take advantage of digital computing techniques to process the speech signal for increased understanding, improved communication, and increased efficiency and Definition of various types of speech classes, feature extraction techniques, speech classifiers and performance evaluation are issues that requires attention in designing of speech processing system.

## I.INTRODUCTION

The main components of the human speech system are: The lungs, trachea, larynx, pharyngeal cavity, oral cavity, nasal cavity. Normally the pharyngeal and the oral cavity are grouped into one unit called the voral tract. The nasal cavity is normally called the nasal tract. The exact placement of the main organs is shown in figure 1.

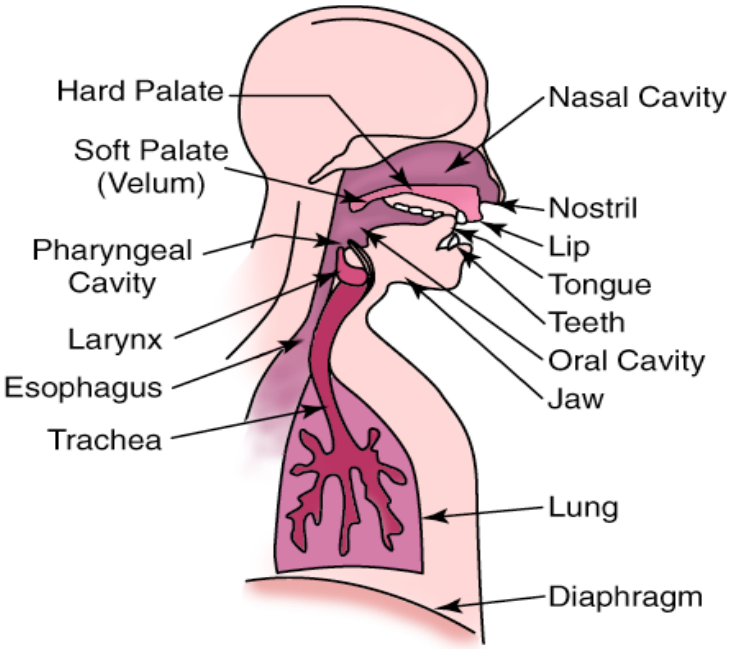


Fig 1. The Human Speech Production System

DOI: 10.18535/ijecs/v5i10.37

Muscle forces are used to press air from the lungs through the larynx. The vocal cords then vibrate, and interrupt the air and produce a quasi-periodic pressure wave. The pressure impulse are called pitch impulse. The frequency of the pressure signal is the pitch frequency or fundamental frequency. The frequency of the pressured signal is the part that define the speech melody[1]. The frequency of the vocal cord is determined by serval factors: The tension exerted by the muscles, it's mass and it's length. These factors vary between sexes and according to age. The pressure impulse are stimulating the air in the oral tract and for certain sounds also the nasal tract. When the cavities resonate, they radiate a sound wave which is the speech signal. Both tracts (Vocal and nasal) act as resonators with characteristic resonance frequencies.

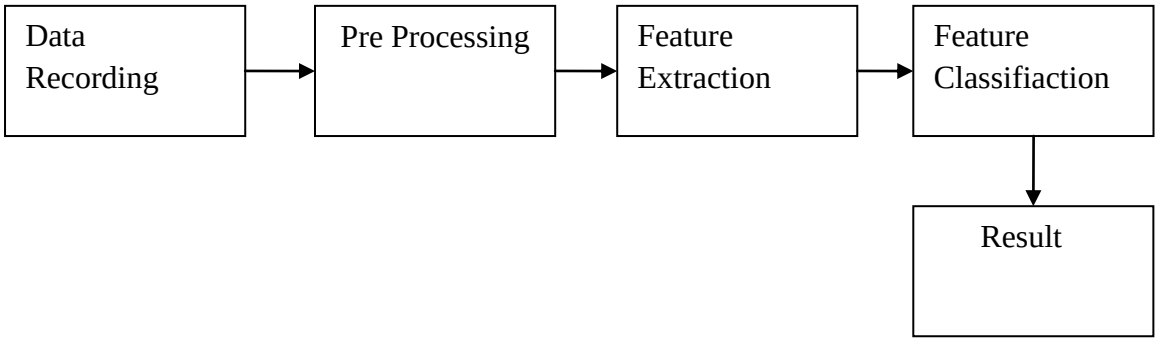


Fig 2 Speech Processing Block diagram

The general block diagram for speech processing is shown in figure 2.

## II. TYPES OF RECORDED SPEECH DATA

### A. Isolated speech

It requires single utterance at a time. Often, these types of speech have "Listen/Not-Listen states", where they require the speaker to have pause between utterances. Isolated word might be better name for this type[2].

### B. Connected word

Connected word require minimum pause between utterances to make speech flow smoothly. They are almost similar to isolated words.

### C. Continuous speech

Continuous speech is basically computer's dictation. It is normal human speech, without silent pauses between words. This kind of speech makes machine understanding much more difficult.

### D. Spontaneous speech

Spontaneous speech can be thought of as speech that is natural sounding and no tried out before.

## III. FEATURE EXTRACTION FOR SPEECH PROCESSING

Speech feature extraction is responsible for transformation of the speech signals into stream of feature vectors coefficients which contains only that information which is required for the identification of a given

utterance. As every speech has different unique attributes contained in spoken words these attributes can be extracted from a wide range of feature extraction techniques and can be employed for speech recognition task. But extracted feature should meet certain criteria while dealing with the speech signal such as: extracted speech features should be measured easily, extracted features should be consistent with time, and features should be robust to noise and environment [9]. The feature vector of speech signals are typically extracted using spectral analysis techniques such as Mel- frequency cepstral coefficients, linear predictive coding wavelet transforms. The most widely used feature extraction techniques are discussed below:

3.1 LPC (Linear Predictive Coding): It is desirable to compress signal for efficient transmission and storage. Digital signal is compressed before transmission for efficient utilization of channels on wireless media. For medium or low bit rate coder, LPC is most widely used. The LPC calculates a power spectrum of the signal. It is used for formant analysis. LPC is one of the most powerful speech analysis techniques and it has gained popularity as a formant estimation technique. While we pass the speech signal from speech analysis filter to remove the redundancy in signal, residual error is generated as an output. It can be quantized by smaller number of bits compare to original signal. So now, instead of transferring entire signal we can transfer this residual error and speech parameters to generate the original signal. A parametric model is computed based on least mean squared error theory, this technique being known as linear prediction (LP). By this method, the speech signal is approximated as a linear combination of its p previous samples. In this technique, the obtained LPC coefficients describe the formants. The frequencies at which the resonant peaks occur are called the formant frequencies. Thus, with this method, the locations of the formants in a speech signal are estimated by computing the linear predictive coefficients over a sliding window and finding the peaks in the spectrum of the resulting LP filter. We have excluded 0th coefficient and used next ten LPC Coefficients. In speech generation, during vowel sound vocal cords vibrate harmonically and so quasi periodic signals are produced. While in case of consonant, excitation source can be considered as random noise. Vocal tract works as a filter, which is responsible for speech response. Biological phenomenon of speech generation can be easily converted in to equivalent mechanical model. Periodic impulse train and random noise can be considered as excitation source and digital filter as vocal tract[4].

3.2 MFC (Mel-frequency cepstral coefficients)

They are captured from cepstral representation of the audio clip .MFCC is most popular Feature Extraction method. MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. For phonetically important characteristics of speech signal is expressed in Mel Frequency Scale.

$$Mel(f) = 2595 * \log 10 (1 + f/700)1$$

Mel Frequency scale has linear frequency spacing below 1000 HZ and logarithmic speak above 1000 HZ.

MFCC Block Diagram is shown in figure 3.

DOI: 10.18535/ijecs/v5i10.37

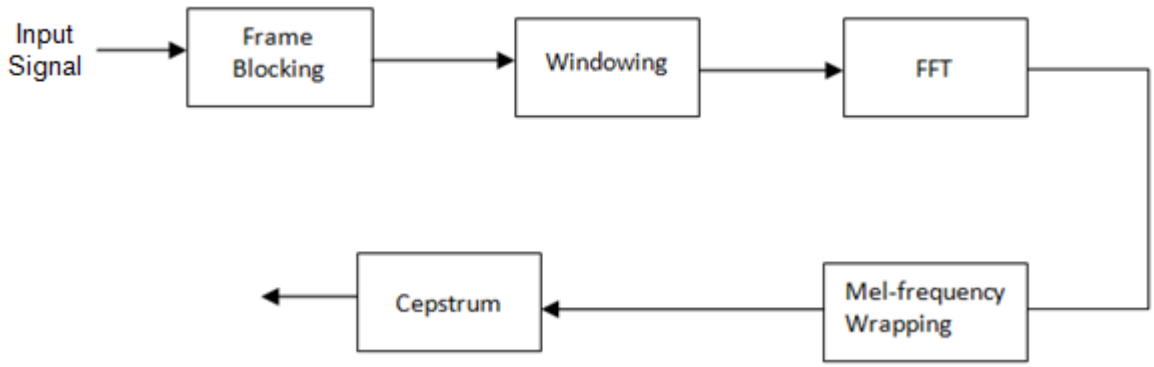


Figure 3. Block Diagram of MFCC

*A. Frame Blocking*

It removes the acoustical interface present in the beginning and ending of sound file.

*B. Windowing*

· Improves the sharpness of harmonics

· Removes the discontinuous of signal by tapering beginning and ending of the frame zero

· Decrease spectral distortion created by the overlap

· Decrease the error provide by FFT.

*C. FFT*

· Convert each signal from time domain to frequency domain

· Its calculation time is about ten times lower than classic DFT

*D. Mel Frequency wrapping*

Before this stage mel filter bank is used .Each filter gives ceptral coefficient .Signal is plotted against the Mel Spectrum to mimic human Hearing.

*E. Ceptrum*

*M*el Ceptrum is converted back to standard frequency scale. This is the key for speech preconisation[5,6].

## DISCRETE WAVELET TRANSFORM (DWT)

The basic idea of DWT in which a one dimensional signal is divided in two parts one is high frequency part and another is low frequency part. Then the low frequency part is split into two parts and the similar process will continue until the desired level. The high frequency part of the signal is contained by the edge components of the signal. In the DWT decomposition input signal must be multiple of 2n. Where, n represents the number of level. To analysis and synthesis of the original signal DWT provides the sufficient information and requires less computation time[7,8].

## CONCLUSION

Different feature extraction techniques and recognition techniques are discussed in this paper and it can be concluded that performance of MFCC technique is superior to LPCC performance. This paper attempts to provide a comprehensive survey on speech feature extraction techniques. Speech processing has attracted scientist as an important regulation and has created a technological influence on society.

## REFERENCES

[1] M.A.Anusuya, “Speech Recognition by Machine,” International Journal of Computer Science and Information security, Vol.6, No.3, 2009

[2] S.J.Arora and R.Singh, "Automatic Speech Recognition: A Review, "International Journal of Computer Applications, vol60-No.9, December 2012

[3] Santosh K.Gaikward and Bharti W.Gawali, "A Review on Speech Recognition Technique," International Journal of Computer Applications, vol 10, No.3, November 2010

[4] Lindasalwa Muda, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques ", Journal Of Computing, Volume 2, Issue 3, March 2010

5] Nidhi Srivastava and Dr.Harsh Dev"Speech Recognition using MFCC and Neural Networks", International Journal of Modern Engineering Research (IJMER), march 2007

[6] Dr.R.L.K.Venkateswarlu, Dr.R.Vasantha Kumari and A.K.V.Nagavya, "Efficient Speech Recognition by Using Modular Neural Network", Int. J. Comp. Tech. Appl., Vol 2 (3)

[7] Bishnu Prasad Das and Ranjan Parekh, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", International Journal of Modern Engineering Research (IJMER) , Vol.2, Issue.3, May-June 2012

[8] Om Prakash Prabhakar and Navneet Kumar Sahu, "A Survey On: Voice Command Recognition Technique," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May 2013

[9] Milind U. Nemade and Prof. Satish K. Shah, "Survey of Soft Computing based Speech Recognition Techniques for Speech Enhancement in Multimedia Applications", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2013