AUDIO FEATURE EXTRACTION USING MEL-FREQUENCY CEPSTRAL COEFFICIENTS

Phyo Thu Zar Tun[1]

Mandalay Technological University, Mandalay and 095, Myanmar

Abstract

The speech signals, random signals are naturally occurred and it has the independent carrying signal according to the time. To extract the characteristics or features of speech signal is important for the analysis of speech audio signal because it is very useful in classifying these signals based on this features. In this proposed system, the feature extraction is analyzed upon the spoken digit word wav file by using Mel-frequency cepstral coefficients (MFCC) technique and it is implemented with the Matlab Programming. The MFCC feature extraction technique is widely used in speech recognition because it is robust, effective and simple to implement.

Keyword: Audio, features, spectrogram, MFCC

1. INTRODUCTION

The extraction feature of audio signal is to reduce the amount of data and to choose the various features from the extracted features. It is a kind of measuring with numerical representation that can use to characterize in audio segmentation. The extracted features from an audio signal are different from each other and these features can optimize using various algorithms. When the audio input file carrying many features can analyze by using feature extraction techniques and then select the best features for classification. The feature extraction can do at time domain or the frequency domain and spectrograms can use to extract audio features that carry the spectral. A spectrogram is the representation of an audio signal by showing the frequency spectrums in time. It can calculate with Fast Fourier Transform (FFT) over the series of overlapping windows that are extracted from the audio signal. This proposed system is analyzed the feature extraction of speech digit wav file by using MFCC feature extraction technique.

2. LITERATURE REVIEW

The author extracted the features from the input speech for identification of speech. The extraction procedures: linear predictive coding (LPC), Linear predictive cepstral coefficient (LPCC), Perceptual linear prediction (PLP) and Mel frequency cepstal coefficient (MFCC) were analyzed [1]. In that paper, the four techniques for feature extraction were used and they all have their own advantages and disadvantages. The Mel frequency cepstrum can give a better performance rate and it is widely used to mimic the human auditory system [1].

The feature extraction technique is a method of filter in audio signal and the authors were analyzed the voice classification and feature extraction for speech recognition by using different type of Mel-frequency cepstral coefficients (MFCC). In this paper, the different MFCC techniques were discussed among them delta-delta MFCC feature extraction technique is better than the other feature extraction techniques [2]. The authors were presented the three feature extraction techniques: MFCC, LPC and PLP and among them the MFCC is repeatedly used for feature extraction because it is the most real individual acoustic speech [3].

3. METHODOLOGY

The purpose of feature extraction is to know the characteristics of speech signals and it can perform by using feature extraction techniques such as MFCC, Log-mel Spectrogram and LPC etc. The input speech digit wav file is read and then extract features with MFCC technique.

3.1.1. Mel-Frequency Cepstral Coefficients (MFCC)

The MFCC computes the mel frequency cepstral coefficients of the speech signal. The process of MFCC is in entire speech data in a batch and it is partitioned the speech signal into frames and computed the cepstral features for each frame. The MFCC's features can convert to statistics for use in the task of classification. Its distributions can observe from the probability density functions of each of the mel-frequency cepstral coefficients. The MFCC is split the entire data into overlapping segments and the windowlength is determined the length of each segment. The overlaplength is the determination of length of overlap between segments [5].

The process of MFCC is pre-emphasis, frame blocking, windowing, fast fourier transform, mel-frequency filter bank and discrete cosine transform. The pre-emphasis is the sample to pass through the filter for emphasizing the higher frequencies. In framing, the input speech signal is segmented as small duration block known as frames and this process is essential in speech because of speech is time varying signals. The sample of framing for input speech signal is illustrated in Figure 1.

In order to perform the continuity of speech signal, these frames is multiplied with windowing methods.

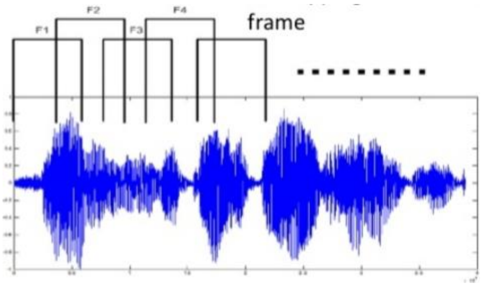

Figure 1. Sample of Framing for Speech Signal [4]

The windowing function is performed to smooth the signal for computation. The output of windowing is

$$Y(m)=X(m)W_n(m),\ 0<=m<=N_m-1$$

Where, N_m is the quantity of samples in every frame and X(m) is input speech signal and $W_n(m)$ is the hamming window. Many window functions are existed among them hamming window is mainly used for speech signal analysis because its resulting frequency resolution is better than other windowing methods. The equation for the hamming window is

$$W_n(m)=0.54-0.46\ \mathrm{Cos}\ (2\pi/(N_m-1)),\ 0<=m<=N_m-1$$

The function of fast fourier transform is to convert the time domain into the frequency domain.

$$Y(w)=F[h(t)*X(t)]=H(w)*X(w)$$

The input for the fast fourier transform is the windowed signal and its output is the discrete frequency bands. The fast fourier transform sizes are 512, 1024 or 2048. The discrete fourier transform formula is

$$X[k]=\sum_{n=0}^{N-1} x[n]\, e^{\frac{j2\pi kn}{N}}$$

The Mel is a unit of pitch and the mel-frequency scale is the approximation of linear frequency (1KHz) and then close to logarithmic for higher frequencies. The mel scale is applied the filter banks according to the spectrum and the output is the sum of filtered spectral components. The MFCC's log energy computation is performed by the logarithm of the square magnitude of the mel filter bank's output. And it compresses the dynamic range of values.

$$\mathrm{Mel}(f)=2595*\log10(1+f/700)$$

The discrete cosine transform is processed the convertion from the log mel spectrum into time domain and it is known as mel frequency cepstrum coefficient. The step by step process of MFCC for feature extraction is shown in Figure 2.

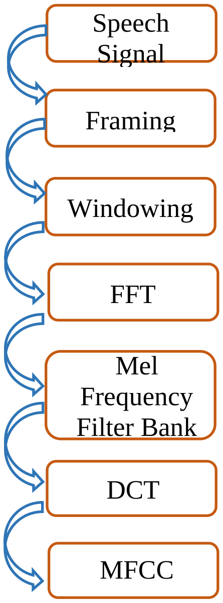

Figure 2. Process of MFCC Feature Extraction

4. RESULTS

This proposed system is the analysis of feature extraction for speech digit wav which is implemented by recording and downloaded from the https://zenodo.org/badge/61622039.svg. All of the implementation for feature extraction using MFCC technique is experiment in Matlab programming. The input speech digit signal is shown with waveform in Figure 3.

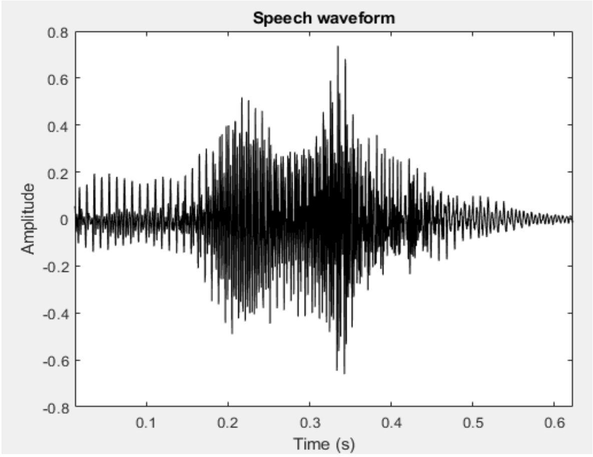

Figure 3. Input Speech Signal Waveform

The total energy of MFCC features in each critical band is calculated using the following equation

$$Y(i)= \sum_{k=0}^{N/2} \log|s(n)|\, H_i(k.\frac{2\pi}{N})$$

Where, N is framelength, s(n) is DFT signal which is calculated by FFT, H is the critical band and N is the number of points used in DFT. The total energy of input speech digit wav file is shown in Figure 4.

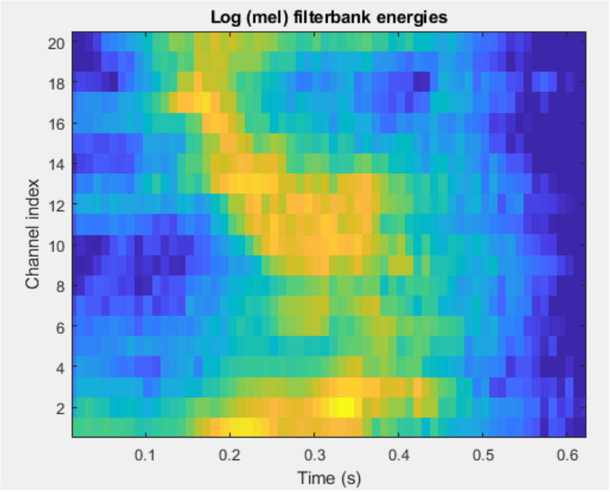

Figure 4. Filterbank energy by log(mel)

The short term power spectrum of an input speech signal is represented as the mel-frequency cepstrum [6] and it is shown in Figure 5.

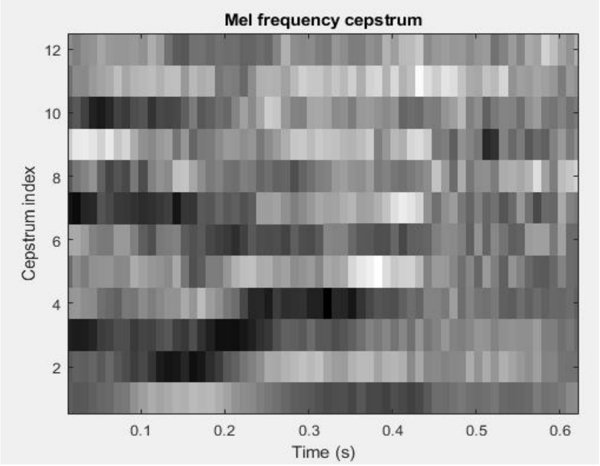

Figure 5. Mel frequency cepstrum

5. CONCLUSION

The process of feature extraction is the transformation of the raw signal into feature vectors. The MFCC is composed of the advantages from the cepstrum analysis with perceptual frequency scale. The MFCC feature extraction technique is more robust and effective in speech signal processing. Moreover the feature selection can improve to get better accuracy for speech recognition or classification from feature extraction by using feature extraction techniques.

6. ACKNOWLEDGEMENT

This is my first step of research paper writing for audio signal processing therefore I have a lot of weak points to be a better research paper in this field.

REFERENCES

[1] Athira menon.G, Anjusha.V.K, “Analysis of Feature Extraction Methods for Speech Recognition”, International Journal of Innovative Science, Engineering & Technology, Vol. 4 Issue 4, April 2017.

[2] Rajeev Ranjan, Abhishek Thakur, “Analysis of Feature Extraction Techniques for Speech Recognition System”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), May 2019.

[3] Bhuvaneshwari Jolad & Dr. Rajashri Khanai, “Different Feature Extraction Techniques for Automatic Speech Recognition: A Review”, International Journal of Engineering Sciences & Research Technology, February 2018.

[4] www.slideshare.net , Text-independent speaker recognition system

[5] www.mathworks.com

[6] www.wikipedia.com