ISSN : 2348 - 9065 (Online)
ISSN : 2349 - 3143 (Print)

Controlling The Application Via Speech Processing Through Mel Frequency Cepstral Coefficients and Back Propagation Neural Method

[I]Sneha Sahu, [II]Neerja Dharmale
[I,II]Dept. of Electronics and Telecomm., Rungta College of Engineering and Technology, Bhilai, India

Abstract
Today, speech recognition is become additional and additional necessary. Varied speech applications area unit offered at intervals the market. Shopper electronic devices will operate through voice. It will use mobile phone as a controller and CELP (code excited linear prediction) parameters that unit area used for speech coding in mobile phones. Mel frequency Cepstral coefficients formula could be a technique that takes voice sample as inputs. When process, it calculates coefficients distinctive to a selected sample. During this project, simulation computer code referred to as MATLAB R2012a is employed to perform MFCC. The simplicity of the procedure for implementation of MFCC makes it most popular technique for voice recognition.

Keywords
Mobile phone, MFCC, CELP, Speaker verification, LSP

I. Introduction

The speech is that the most typical and first mode of communication among groups of people. Within the systems, some speech recognition systems use "training" that is additionally known as enrollment wherever every speaker reads text or isolated vocabulary. Accuracy is hyperbolic within the system by analyzing the person's original voice and uses it for fine-tune the popularity of that person's speech.

The term voice recognition refers distinctive the speaker, instead of what they are spoken language. For security method, recognizing the speaker will change the task simplify the task of translating speech in system that had been trained on a particular person's voice or to certify or verify the identity of a speaker it may be used. The essential diagram of the system is shown in Fig .1.

In the system, initial block is controlling device that is employed to regulate the system. Controlling device are often transportable or any microcontroller. Use mobile phone to regulate the system. Input to the current is speech command of one that desires to regulate the device at a distance. Over a point-to-point or point-to-multipoint line knowledge transmission, digital transmission or digital communications is that the transfer of a unceasingly variable analog signal over an analog channel, electronic communication is that the transfer of Whereas analog transmission is the transfer of a distinct messages over a digital or associate degree analog channel. The transmitted signal is received by the receiving network. Once this to acknowledge the speech, the speech recognition is gift within the system. In mobile phones CELP (code excited linear prediction) methodology is employed for speech coding.

For making minimally redundant illustration of a speech signal speech coding is employed. The aim of all speech coders is to attenuate the disturbance or noise at a given bit rate to succeed, or minimize the bit rate to obtain a given distortion with prime quality [1]. By speech coding process distortions are terribly less and system becomes much economical. Accuracy of the system is often a lot of and obtained smart result. In speech coding smart quality of speech are often obtained by analysis-by-synthesis methodology. To match the reconstructed speech wave shape the excitation signal is chosen by making an attempt as closely as doable to original speech wave shape within the time domain analysis-by-synthesis coder. After speech recognition, finally speech signal is transfer to any device. It are often any device like bulb, fan, machine, etc. And during this method final output is obtained.

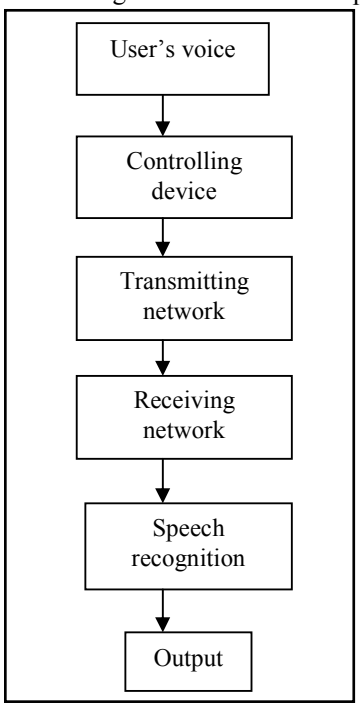

Fig 1: Block diagram of the system

II. Methodology

1. Classification of Speech

There are several parameters outline the potential of a speech recognition system [2].

(i) Isolated word

Isolated word contain sample windows, it receives single word or single utterances at a time. Isolated utterances are also a far better name with this work.

(ii) Connected word

Connected word and isolated word each are of comparable kind however enable separate vocalization to be run along minimum pause between them.

(iii) Continuous speech

In this computer can examine the content and permits the user to talk naturally. Numerous ways are there to see determine

utterances boundaries and difficulties occurred in it.
(iv) Spontaneous speech
Spontaneous speech means that a system ought to be capable to handle a spread of natural speech feature like words being run along.

2. Speech Recognition Techniques

Nowadays totally different techniques area unit are obtained for speech recognition. The target of speech recognition is to characterize, analyze, extract, and acknowledge info regarding the speaker identity. Speech recognition techniques are employed in alternative ways. For determinant the speech characteristics numerous techniques area unit used. During this technique speech is analyzed in numerous manners and speaker identity is verified. The speech information contains totally different variety of information that shows the speaker identity. This involves speaker specific information because of excitation supply, vocal tract, and behavior feature. Speech analysis techniques are employed in numerous functions. It is terribly essential to use and also the speech analysis stages area unit of the subsequent three varieties.
(i)Segmentation analysis
In segmentation analysis, speech is analyzed victimization the frame size and shift within vary of 10-30 ms to extract speaker information. Vocal tract information of speaker recognition is extract by this technique.
(ii)Sub segmental analysis
Sub segmental analysis is outlined as speech analyzed victimization the frame size and shift in vary range 3-5 ms. For the excitation state this method is employed to principally analyze and extract the characteristic.
(iii) Supra segmental analysis
In this work, victimization the frame size speech is analyzed. This method is principally accustomed mainly analyze the behavior and characteristic of the speaker.

3. CELP based speaker verification

In this paper, a CELP primarily based speaker verification methodology is planned for physical science devices together with mobile phones. Within the CELP primarily based speaker verification, the CELP coding methodology used for transportable language is applied to the encoded voice information to perform speaker verification. A system is planned here for in operation client electronic devices by voice mistreatment CELP (code excited linear prediction) parameters that area unit normally employed in speech coding in mobile phones. There are two characteristics of the CELP primarily based speaker verification. First is speaker verification used solely the encoded voice information and it doesn't want the decoding method. So that is will perform on any electronic devices and also the second is once the mobile phone is employed to control the patron physical science devices. It will use the voice coding operate that's in-built to the transportable itself. No extra information is needed to control the system [3]. Speaker verification has two styles of phases' enrollment and verification.
In methodology there are two phases first is training and other is testing. In the training phase first load data, data should be loaded to recognize speech. Data should be loaded with the help of mobile phone. After data PSD (power spectral density) is done. Power spectral density could be alive of a signal's power intensity within the frequency domain. In follow, the PSD could be a computed from the FFT (fast Fourier transform) spectrum of a symptom. The PSD characterized the amplitude versus frequency content of a random signal. After that silent is removed. Then apply Mel frequency cepstral coefficient feature extraction technique.

4. Mel frequency cepstral coefficients

MFCC is that the most generally used feature extraction technique. These coefficients represent audio supported perception and square measure derived from the Mel frequency cepstrum. This method is taken in to account to be the most effective on the market approximation of human ear. The diagram of the structure of the associate MFCC processor is shown in Figure 2.
Like LPCC, the input signal is initial knowledgeable the 1st order digital all- pole filter for pre- emphasis therefore on spectrally flatten the signal and so this resultant signal is passed for windowing wherever it's divided into frames exploitation overacting windows. When windowing initial FFT and then Mel scale filter banks are applied so on acquire the Mel-spectrum. FFT is essentially used for the conversion of the speech signal from time domain to frequency domain. Mel scale filter bank consists of a series of triangular band pass filter banks which are arranged in such the simplest way in order that the lower boundary of one filter is found at the centre frequency of the future filter and the upper boundary of the same filter is situated at the centre frequency of the next filter. The Mel scale is graduated that resembles the manner during which human ear perceives sound. Mel scale filter bank maps the powers of the spectrum obtained on top of onto the Mel scale by exploitation triangular overlapping windows [4].
After the signal is knowledgeable the filter banks, log energy at the output of every filter bank is calculated. The log is taken to rework into cepstral domain. Finally, DCT is applied to every Mel spectrum (filter output) to convert the values back to real values in time domain. This transformation ornamentation relates the options and initial few coefficients are joined along as a feature vector of a selected speech frame. Since DCT accumulates most of the data contained within the signal to its lower order coefficients by discarding the upper order coefficients, thus a substantial reduction in machine value is accomplished. After MFCC, finally information is tested and output is obtained.

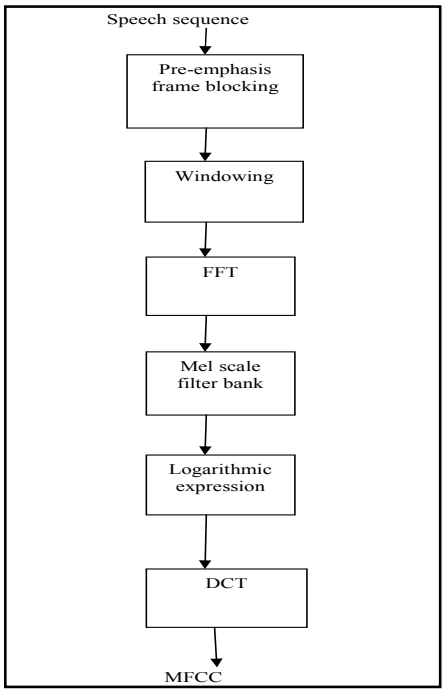

Fig.2 : Basic layout of MFCC

ISSN : 2348 - 9065 (Online)
ISSN : 2349 - 3143 (Print)

III. Result and Discussion

Voice samples of two speakers saying different sentences, first person speak "Open Microsoft PowerPoint" and the other person speak "Open VLC" at two completely different instants were knowledgeable through the MFCC formula and their several MFCC Coefficients were extracted.

Fig 3(a): Power spectral density of 1st person

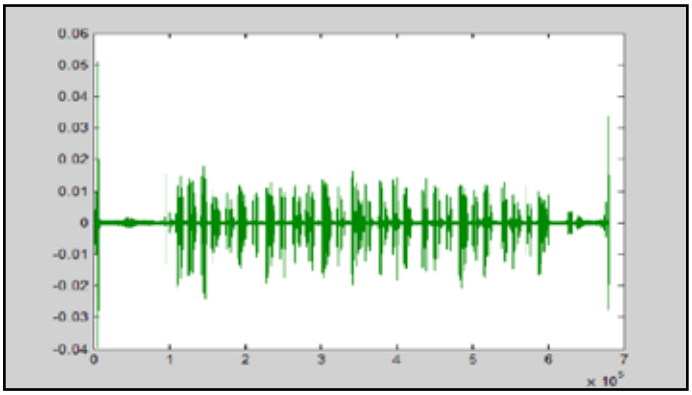

Fig 3(b): Remove silent of the speech signal of 1st person

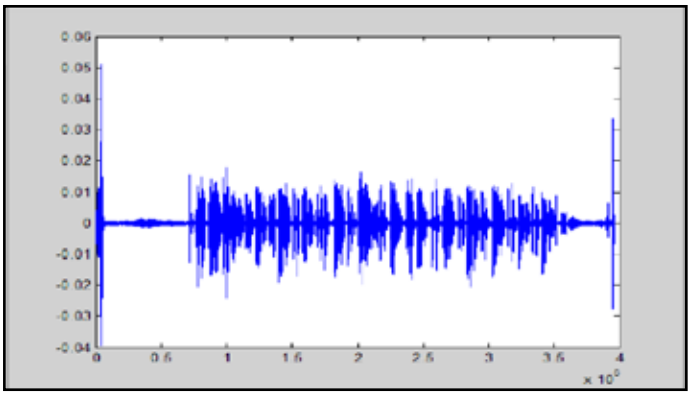

Fig 3(c): Remove silent of the speech signal of the 1st person

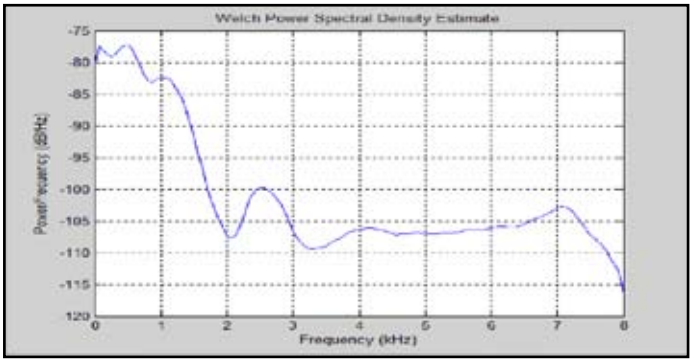

Fig 4(a): Power spectral density of 2nd person

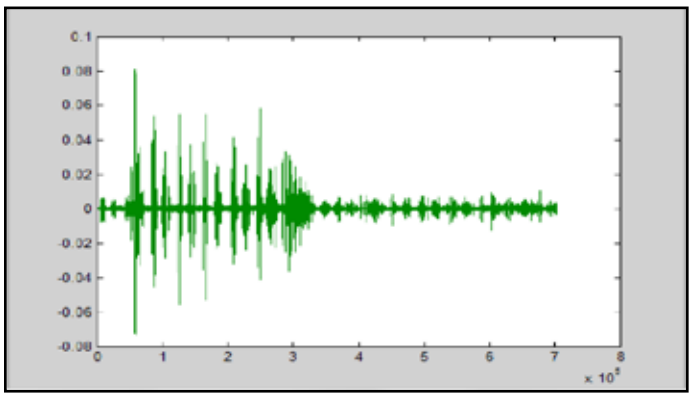

Fig 4(b): Remove silent of the 2nd person

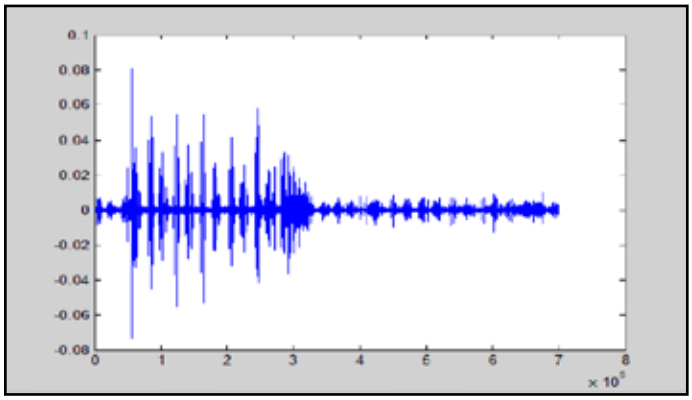

Fig 4(c): remove silent of the 2nd person

IV. Conclusion and Future Scope

Experiment has been conduct successfully. Data set can be collected from mobile phone of different persons. It absolutely was discovered that MFCCs for each individual user was distinctive. Certain variations were discovered as a result of distinction within the vicinity of the recording space. It may be further improve by the utilization of better feature extraction.

References

[1] *Nimisha Susan Jacob, Ancy S. Anselam, Sankuntala S. Pillai "Performance analysis of CS-ACELP speech coder" IJEAT, pp. 191-195,2015.*

[2] *Shikha Gupta, Amit Pathak, Achal Saraf. "A study on speech recognition system". IJSETR, pp. 2192-2196, 2014.*

[3] *Masatsugu Ichino, Yasushi Yamazaki, Hiroshi Yoshiura. "Speaker verification method for operation system of the consumer electronics devices". IEEE, pp. 96-102, 2015.*

[4] *Taabish Gulzar, Anand Singh, Sandeep Sharma (2014). "Comparative analysis of LPCC, MFCC and BFCC for the recognition of hindi words using artificial neural networks". IJCA, pp. 22-27, 2014.*

[5] *Koustav Chakraborty, Asmita Talele, Prof. Savitha Upadhya. "Voice recognition through MFCC Algorithm". IJIRAE, pp. 158-161, 2014.*