SPEECH RECOGNITION USING MATLAB

[1]ASEEM SAXENA, [2]AMIT KUMAR SINHA, [3]SHASHANK CHAKRAWARTI, [4]SURABHI CHARU

[1,2,3,4] Suresh Gyan Vihar University, Jaipur, Rajasthan, India
Email:- saxena.aseem11@gmail.com, amitsinha4@hotmail.com, moon.shashank87@gmail.com, surabhicharu10@gmail.com

Abstract-Speech Recognition is the process in which certain words of a particular speaker will automatically recognized that are based on the information included in individual speech waves. This paper enlightens upon the invention as well as technological advancement in the field of voice recognition and also focuses upon different steps involved for speaker identification using MATLAB Programming. In this paper firstly we will going to perform speech editing as well as degradation of signals by the application of Gaussian Noise. This background noise then will successfully removed by the application of Butterworth Filter. Moreover, the technique applied here is to develop a code using MATLAB Programming which will compare the pitch and formant vectors of a known speech signal which will then compare with the bunch of other unknown speech signals and prior to it choose the appropriate matches.

Keywords: Formants, Gaussian Noise, MATLAB Programming, Pitch Vector, Speech Editing, Speech Recognition.

I. INTRODUCTION

Speech is one of the most important medium by which a communication can take place. With the invention and widespread use of mobiles, telephones, data storage devices etc. has provided a major help in setting up of speech communication and its analysing. The term and the basic concept of speech identification was began in the early 1960`s with exploration into voiceprint analysis which was somewhat similar to fingerprint concept. It was in 1984 that a science fiction called "Star Trek to George Orwell`s," derived the concept that a machine can recognize the human voice. [1] Nowadays, with further growth & advancement in the field of speech recognition, the humans who are physically challenged such as blind and deaf can easily communicate with the machines. So in biological terms a voice that is being generated through trachea will be decoded by brain.

In this paper we will going to study about five basic principles that are involved in speech recognition as well as the related works that has been done on this technology till now [2].

These are:

- Speech Editing
- Speech Degradation
- Speech Enhancement
- Pitch Analysis
- Formant Analysis

Speech Editing

In this technique we will going to record a set of the speech signal in '.wav' (dot) wave format and taking a speech signal from the set of recorded speech waves on which we will going to perform

Speech Editing. Here the length of the vector representing this speech file must have a magnitude of 30,000. However, this vector is then divided into two separate vectors having equal length & in opposite order. Now with the help of MATLAB Programming & Tools we will going to develop a code by which the given wave file is read and then the same file is played in reverse order. The general representation of the speech editing waveform in forward mode as well as reverse mode is shown in fig.1.

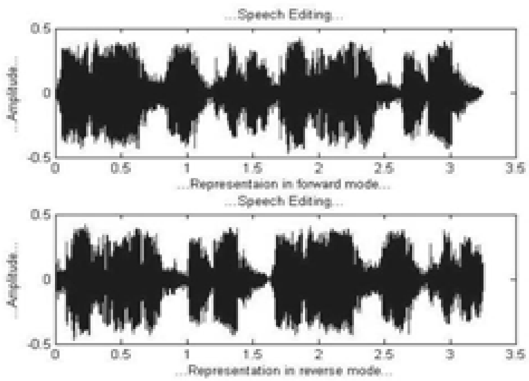

Fig.1: Representation of Speech Edited Waveform.

II. SPEECH DEGRADATION & ENHANCEMENT

Noise plays a vital role in speech enhancement as well as speech degradation. Thus noise estimation is one of the major part while performing the speech recognition task. Therefore, it is understood if the estimated noise is low it will not affect the speech signal but if the noise is high then speech will get distorted and loss intelligibility. So to remove the noise we have discussed two techniques i.e. speech degradation and speech enhancement. [3] The speech degradation technique involves the addition of gaussian noise to the original .wav format file with the help of MATLAB Function called randn(). Moreover, this process not only help us in making comparison between the clean file and the signal with the added gaussian noise, it also can be further viewed as which filter in DSP (Digital signal proceesing) such as Chebysev Filter, Butterworth Filter etc. can be worked better to remove this gaussian noise.The general representation of the degraded speech wave is shown in fig.2.

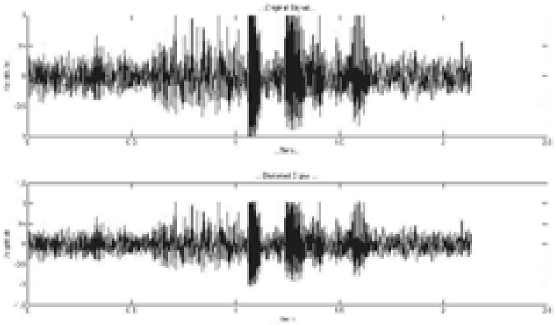

Fig.2: Representation of Speech Degraded Waveform.

The speech enhancement technique enlightens upon the major use Speech Degradation technique i.e. removal of gaussian noise from the original speech wave. In this technique firstly the degraded signal i.e original signal mixed with gaussian noise is first converted to the frequency domain with the help of FFT tool in MATLAB Programming. Then higher frequency noise components are then removed with the help of 3rd order Butterworth low pass filter, according to the equation,

$$HB(u,v) \cong 1/(1+(\sqrt{2}-1)\left(\frac{D(u,v)}{Do}\right)^{2n})$$

Where

D (u,v) is the rms value of u and v, Do determines the cut-off frequency, n is the filter order. The reason to choose butterworth filter here because it has the capability to filter the Gaussian noise more closely & approximates an ideal low pass filter as the order, n, is increased. The resulting filtered signal was then scaled and plotted with the original noisy signal to compare the filtering result and the general representation of speech enhancement type waveform is shown in fig.3

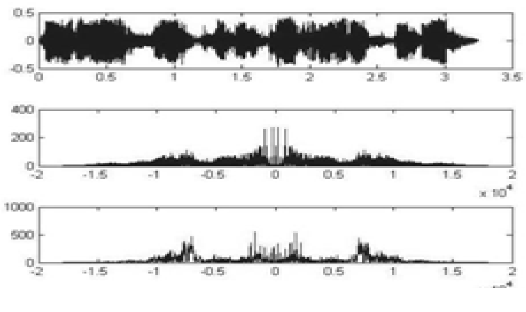

Fig.3: Representation of Speech Enhancement Waveform

Pitch Analysis

Pitch in terms of speech analysis can be defined as a technique which allows the ordering of sounds on a frequency-related scale. Pitch analysis helps us in identifying the state of speech of a person. The considered states are neutral, happy, sad. Therefore it is very important to understand the concept of pitch pitch analysis.[4] The paper describes a technique that involves the extraction of basic parameters of Pitch Analysis. Now, the calculation of the average pitch of the entire .wav format speech file that are recorded in our data base of different speakers was done and found to have a certain value which can be used in voice recognition, where the differences in average pitch can be used to characterize a voice file. The general representation of pitch analysis with respect to time is shown infig.4.

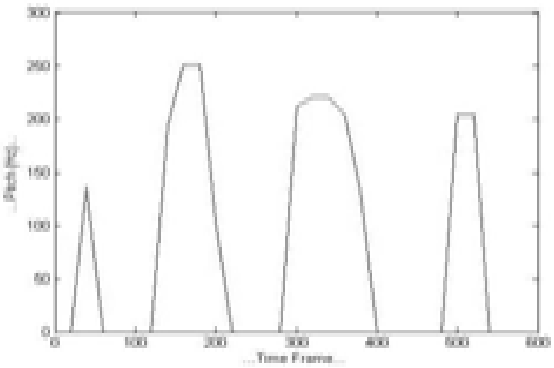

Fig.4: Representation of Pitch Analysis Waveform

Formant Analysis

In Formant Analysis technique, we will going to perform on any of the .wav format speech file taken from the set of recorded .wav speech signal. Further with the help of MATLAB Programming we had prepared a code for Formant Analysis. With the help of this code we can easily calculate the first five formants that are present in .wav speech file, calculation of difference between the vector peak positions of these five formants, vector position of the peaks in the power spectral density were easy calculated and can be used to determine the speech file. The general waveform of formant analysis can be shown in fig.5

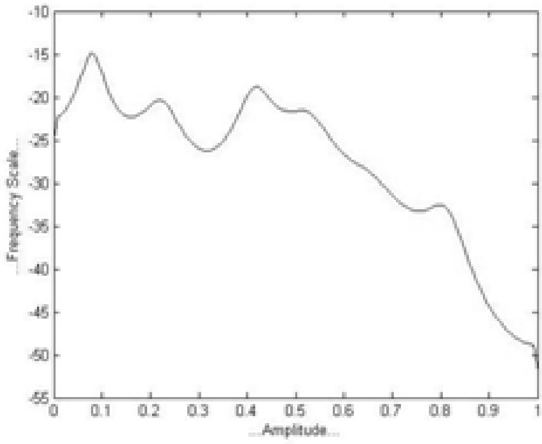

Fig.5: Representation of Formant Analysis Waveform

III. WAVEFORM COMPARISON

Formants in normal language can be defined as the spectral peaks of the sound spectrum.With the help of above discussed Pitch and Formant Analysis, a waveform comparison code was written with the help of MATLAB Programming. Thus, based on this code we can easily characterized Speech waveform files. In this process a reference .wav file was used which is then compared with the remaining .wav files. Moreover, a sorting routine is performed in which sorting and comparison of the average pitch of the reference file with all the other 5 .wav files. The technique further includes the comparison of formant vector of the reference file to all .wav files, and thus sorting for the top 3 average pitch correlations and then again sort these files by formant vectors correlations and then sort these by average pitch. In this way, we can easily recognise the speaker.

IV. RELATED WORKS

[5]Controlling Of Device Through Voice Recognition Using MATLAB

In this paper a technique is described in which firstly a speech command can be determined by power of speech signal which can be taken by the help of microphones being connected to the computer itself. Using MATLAB Programming the sampling of the speech signal takes place with sampling rate of 8000 samples/sec according to nyquist criteria i.e.

$F=2*fm$

Where

F=Sampling Frequency, fm = maximum component of frequency being present in speech signal. The sampled signal is then filtered off by using band pass filter lying in the range of 300 Hz-4000 Hz, which filters all the speech signal lying below 300 Hz of frequency range. Moreover, it includes the algorithm for the creation of speech templates which can be achieved by calculating the power of each sampled signals respectively. Now, a dictionary is created in which same command is taken many times and thus the average of these will represent a speech template which can be then stored in a library. Now the command templates that are being received by microphone are than compared with the templates being stored in library according to Euclidean`s Distance i.e. Euclidian Distance=$\Sigma_{i=1}((dic[i]-com[i])^2)$ Where i denotes the number of sample points. Thus, the command will be detected by a particular device and it will performs the operation accordingly & if it does not match, the device should not follow that command. The general block diagram of the interfacing is shown in fig.6.

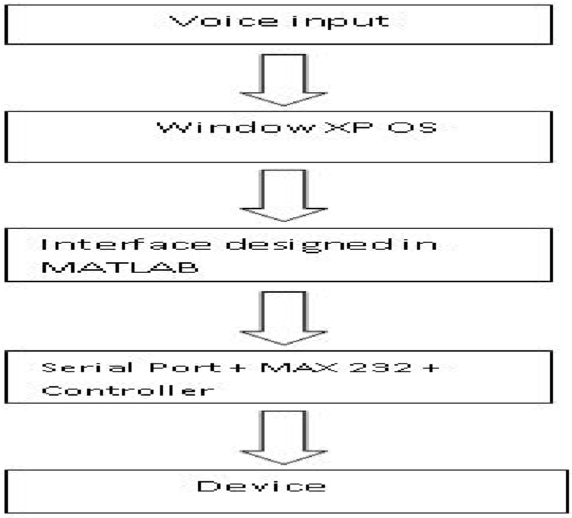

Fig.6: Block Diagram- Interfacing

[6]Speech Recognition using Digital Signal Processing

This paper truly based on independent speech recognition which can be achieved by the use of Mel Frequency Cepstrum Coefficients which will process the input speech signal and further will recognise the speaker. This above task can be performed by using MATLAB Programming use of Digital Signal Processing (DSP) as a hardware platform This phenomena is broadly classified into three categories in order to understand the concept of speech recognition.

- Speaker Identification
- Speaker Verification
- MFCC

Speaker Identification

The general block diagram of Speaker Identification is shown in fig.7.

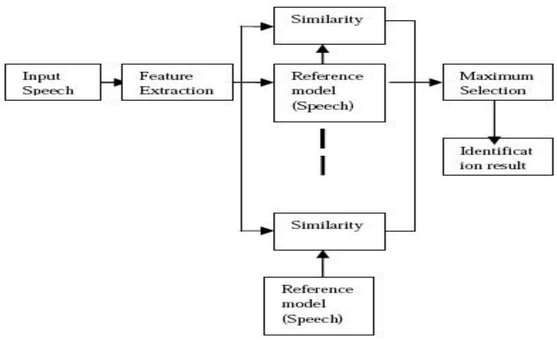

Figure a: (speaker identification/recognition)

Fig.7: Block Diagram of Speaker Identification

In the above figure, when the input signal i.e. the speech signal is provided the Speech Extraction process simply extracts all the necessary phenomena such as pitch, frequency etc. which will be needed at the end for the recognition of speaker. Further, the Speech Matching Block will simply matches all the properties that are extracted before with the set of speech signals being stored. The speech recognition system consist of two separate phases. The first one is referred to the enrolment sessions or training phase while the second one is referred to as the operation sessions or testing phase. Due to this the system can construct an efficient model for that speaker.

Speaker Verification

The general block diagram of Speaker Verification is shown in fig.8.

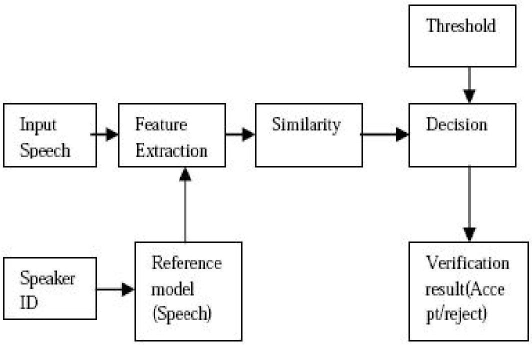

Fig.8: Block Diagram of Speaker Verification

The working of this block is similar to speaker identification process with a difference that here we have included a threshold block above the decision block as shown in above figure.

Mel-Frequency Cepstrum Coefficients Processor (MFCC):

MFCC's is a type of algorithm i.e. basically used to define relationship between human ear's critical bandwidths with frequency. This method is basically used for analysing and extraction of pitch vectors. The general diagrammatic representation of the block diagram of MFCC is shown in fig.9.

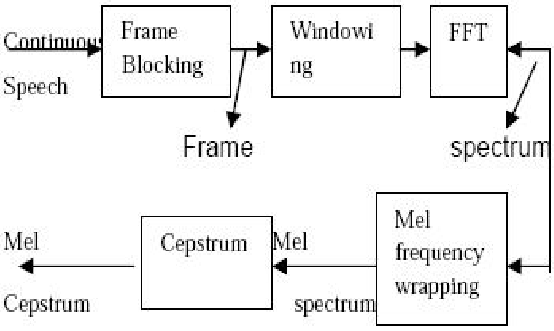

Fig.9: Block Diagram MFCC

V. LIMITATIONS OF RELATED WORKS

- The above discussed papers were failed to discuss general features and characteristic behaviour of speech signals were not described properly.
- The above discussed task of speech recognition can be easily challenged by highly variant input speech signals.
- The above discuss process of handling devices can be overruled in presence of some noise.
- The above discussed papers, were failed to discuss about the use of filters for the noise removal.
-

System Proposed

Speech recognition process is one of the most important step that has brought the humans and machines more closely to each other. The main aim of this paper is to describe the basic features as well as characteristic behaviours of speech waves. It also enlightens upon speech editing as well as the basic filtering techniques. Moreover, the use of generation of the codes using MATLAB Programming makes it more advanced in the study of speech recognition process.

System Architecture

As discussed earlier about the basic limitations and drawbacks, here in this section, we will going to present an architecture by which we can easily overcome the above problems.The basic block diagram of our system architecture is shown in fig.10.

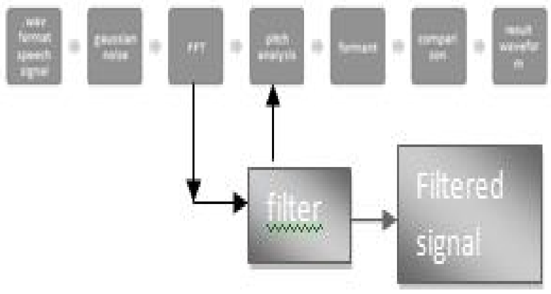

Fig.10: System Architecture Block Diagram

In the given block diagram a given .wav format file was taken in which we added the Gaussian noise and its FFT was taken which is then used for pitch analysis. Moreover, we have also generated programming codes with help of MATLAB Programming, for the formant analysis. Furthermore, we have used the filter as shown in block diagram by which the Gaussian noise can be removed. Therefore through our paper we have successfully described about the waveform comparison with the help of above discussed principles and have successfully defined how the speech recognition is carried out.

CONCLUSION

This paper successfully defines about various characteristics and behaviour of speech signals and also entails upon the setting up of communication between human speech signals with the machines. In this paper we have generated codes with the help of MATLAB Programming which requires .wav format speech signals. Thus, in order to remove this limitation there is a requirement for the study of various formats of speech signals which can be used for the communication with the machines.

REFERENCES

[1] [1].**"Speaker Recognition Using MFCC and Vector Quantization Model".**

[2] E. Darren Ellis Department of Computer and Electrical Engineering – University of Tennessee, Knoxville Tennessee 37996 topic on "Design of a Speaker Recognition Code using MATLAB"

[3] Topic on "Noise estimation Algorithms for Speech Enhancement in highly non-stationary Environments"ijcsi.org/papers/IJCSI-8-2-39-44.pdf

[4] Topic on "Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person" ijcsi.org/papers/IJCSI-9-4-1-296-299.pdf

[5] Topic on "Controlling of Device through Voice Recognition Using Matlab1" ISSN NO: 2250-3536 VOLUME 2, ISSUE 2, MARCH 2012 http://www.ijater.com/Files/fd54f0ee-1aae-4e88-ab81-b89f32ac57c6_IJATER_03_35.pdf

[6] Topic on "Speech Recognition using Digital Signal Processing" ISSN:2277-9477, Volume2,Issue 6.

★★★