Dr.S.Selva Nidhyananthan et al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 4, Issue 4, Dec 2017, pp. 352-355

Speaker Recognition With Moment Features Using Lab View

Dr.S.Selva Nidhyananthan[1], K.Muthugeetha[2], V.Vallimayil[3], D.Suganya[4]

[1]Associate Professor, [2,3] UG Students, [4] Lecturer

[1-3]Mepco Schlenk Engineering College, Sivakasi, India, [4] VSVN Polytechnic College

nidhyan@mepcoeng.ac.in, kimgeetha1997@gmail.com, vvallimayil1995@gmail.com

Abstract*— Now a days, Biometrics is being used extensively for the purpose of security. Biometrics deals with identifying individuals with their physiological such as fingerprint, DNA,ECG etc or behavioral traits i.e. rhythm, voice etc. Voice is a most natural way of communication and non-intrusive as a biometric, voice biometric has characteristics of acceptability, cost ,easy to implement is required. Also voice based biometric system can be easily combined with other biometric systems to enhance the reliability and security of the system. This paper describes speaker recognition with moment features using Lab VIEW software. Speaker recognition consists of speaker verification and speaker identification. This project is to accumulate over a period of time few human being 's voice samples and check those voice samples with already stored data. In this project, silence removal, preprocessing, feature extraction has been done. For feature extraction, Mel Frequency Cepstral Co efficient (MFCC) is used. The moment features of speech is found for collecting database. Euclidean distance is found for the purpose of comparison.

***Index Terms*— Lab VIEW software; MFCC;**

I. INTRODUCTION

This paper introduces a speaker recognition using Lab VIEW software. Speaker recognition is the identification of a human being from their characteristics of sound. This is also called voice recognition. There is a basic dissimilarities between speaker recognition (recognizing who is speaking) and speech recognition (recognizing what is being said). And "voice recognition" is used for both speaker verification and identification. Additionally, there is difference between the act of authentication (called as speaker verification or speaker authentication) and identification. Recognize the speaker can simplify the task of convert speech in systems that have been trained on specific human being's voice or it can be used to prove or verify the identity of a speaker as a part of security process. In the modern world technology, speech based techniques are being widely using in voice recognition in ordinary PCs to biometric and forensic application[3].

Speaker identification system makes a one to many comparisons to establish the identity of an individual. The unknown speaker is identified when the known model of a speaker matches the input utterances[2]. In this paper, Mel frequency cepstral co-efficient has been used for formant detection. Data base of twenty persons having two samples per person including male and female has been created for analysis of result. The system (Speaker verification) is usually as a "gatekeeper" in order to provide access to a secure system. These system operate with the user's knowledge and typically require the user's co-operation. The developed system uses the Lab VIEW (Laboratory Virtual Instrument Engineering Workbench) 2014 platform.

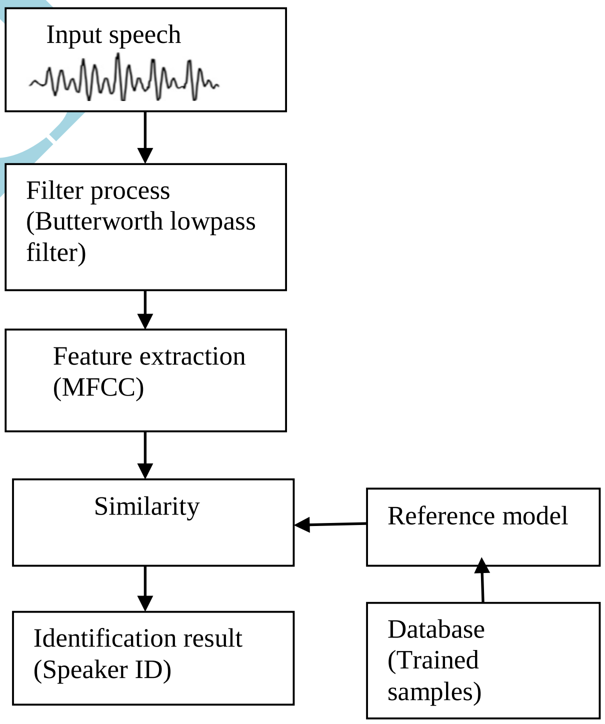

Figure 1.Speaker recognition Block diagram

II. PROPOSED METHODOLOGY

The first step is to record and collect the voice sample from a person. Twenty persons have been chosen and 2 samples of each person were taken in wav format. After recording and collection of voice, a low pass filter is used for the removal of noise and the voice sample is framed using framing technique and every signal is windowed using Hamming window technique. The hamming window has been extensively used in telephone communication signal processing. After applying Fast Fourier Transform (FFT), Mel filter bank is to be done. After finding the 20 coefficients for Mel filter bank, Discrete Cosine Transform (DCT) should be performed. The feature database contains attributes aspects such as Mean, Variance, Skewness, Kurtosis, 5th order moment and 6th order moment. The test speech signal was compared with the signals that are already stored in the database. Euclidean distance is found for the purpose of comparison.

A. *To get input signal*

Input signal is taken from any audio signal or human being's voice. Input is in WAV format. Sampling frequency is calculated and sampling rate is calculated using audacity software.

B. *Preprocessing*

After getting input, silence and noise are removed using low pass filter. Low pass filter is used to allows signals with a frequency lower than certain cutoff frequency and losses the signal with frequencies higher than the cutoff frequency. Cutoff frequency is selected for our project is 1kHz.

C. *Feature extraction*

Mel Frequency Cepstral Co-efficient (MFCC) is used for feature extraction method. MFCC contains six steps they are framing, windowing, Fast Fourier transform(FFT),Mel Filter Bank, logarithm(LOG) and Discrete Cosine Transform(DCT)[2].

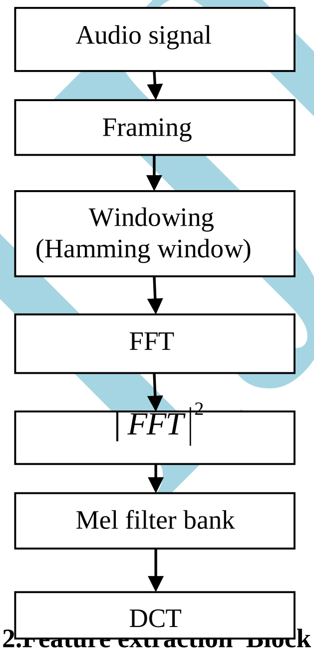

Figure 2.Feature extraction Block diagram

i. *Framing*

Framing is one of the important step in signal processing. The recorded discrete signal has a finite length but it is usually not processed whole. The preprocessed signals are framed. Speech or voice signals are blocked into frames of N sample. Adjacent frames are separated by M samples with the value M less than N. The first frame consists of first N samples and the second frame begins from M samples after the first frame and overlaps it by N-M samples and so on[9].

Let N=160samples, M=96samples, overlap (N-M)=64 samples(as per our speech signal). Why we are going to overlapping frames? The overlapping is used to increase precision of three recognition process. The length of a frame is increased normally to the power of two, that is in our case it would be 416 samples. The power of two is selected because the Fast Fourier Transform (FFT).

ii. *Windowing*

Before further processing, the individual frames are windowed. The frame itself is absolutely windowed by a hamming window. Windowing is the process of a small group of a larger dataset. There several types of windowing functions that can apply based upon the signal. An actual plot of a window shows that the frequency characteristics of a window are a continuous spectrum with a main lobe and several side lobes. The main lobe is around at each frequency component of the time domain signal, and the side approach zero. The side lobe response of a strong signal can overpower the main lobe response of a nearby weak sinusoidal signal. This paper uses a hamming window method. Hamming window method is better than other windowing method. The hamming window doesn't reach zero at both ends eliminating all discontinuity and it still has slight discontinuity in the signal. The hamming window works a good job of nullifying the nearest side lobe. To minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of the each frame.

$$y(n) = x(n)w(n), 0 \le n \le N-1$$

Where N=Frame length, w(n)=Window function.

By using above formula, Hamming window is measured.

iii. *Fast Fourier Transform*

Fast Fourier transform (FFT) is an algorithm that samples a signal over a period of time (or space) and divides it into its frequency components. Fast Fourier Transform are widely used for many application such as engineering, science, and mathematics.

$$X_k = \sum_{n=0}^{N-1} x_n . e^{-i2\pi kn/N}$$

$$= \sum_{n=0}^{N-1} x_n .[\cos(2\pi kn/N) - i.\sin(2\pi kn/N)]$$

After getting FFT value, absolute value of FFT is calculated.

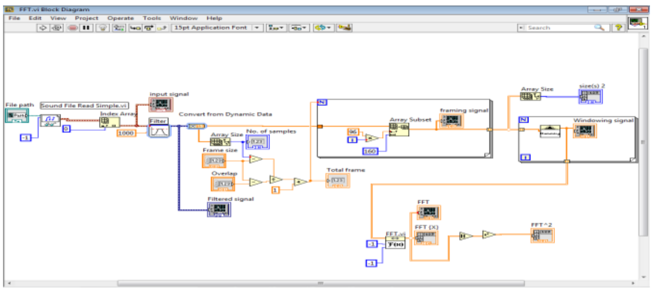

Figure 3.Block diagram of Feature Extraction

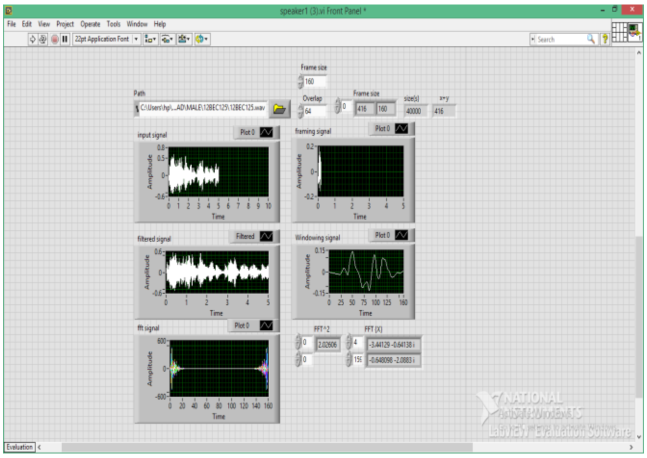

Figure 4.Front panel of Feature Extraction

iv. Mel filter bank

The human ear decides frequencies non-linearly across the audio spectrum and factual evidence propose that designing a front-end to operate in a corresponding manner develop recognition performance.

$$Mel(f) = 2595 * \log_{10}(1 + f/700)$$

To implement this filter bank, the window of speech data is metamorphosed using a Fourier transform and the magnitude is taken. The magnitude coefficients are then accumulated by correlating them with each triangular filter. Here accumulated means that each FFT magnitude coefficient is multiplied by the equivalent filter gain and the results binned. Thus, each accumulate holds a weighted sum reproducing the spectral magnitude in that filter bank channel. Normally the triangular filters are spread over the whole frequency range from zero to nyquist frequency.

However, band limiting is often useful to neglect unwanted frequencies or avoid allocating filters to frequency regions in which there is no useful signal energy. If Mel-scale filter bank parameters are required directly, then the target kind should be set to MELSPECTRUM. By preference, log filter bank parameters can be generated by fixture the target kind to FILTERBANK.

Figure 5 and Figure 6 describes Mel filter bank for 20 coefficients and its values used for finding DCT coefficients.

v. DCT

A discrete cosine transform expresses a finite order of data points in terms of a sum of cosine functions oscillating at different frequencies. The discrete cosine transform is widely used for the speech processing. It is often used because of its energy compaction, which results in its coefficients being more concentrated at low indices than the coefficient of the DFT. This allows us to approximate a signal fewer coefficient. There are several coefficient of the DCT.

$$C(k) = \sum_{n=0}^{N=1} s(n)\cos\left(\pi k \frac{n-1}{2N}\right), \text{k=0,1,2…N-1.}$$

Where, s(n) is a real signal and its inverse is,

$$s(n) = \frac{1}{N}\left[C(0) + 2\sum_{k=1}^{N-1} C(k)\cos\left(\pi k \frac{n+1}{2N}\right)\right]$$

Where, N=0,1,2,…N-1.

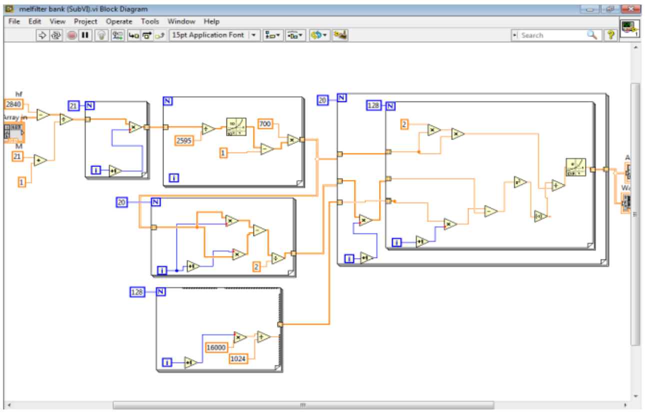

Figure 5.Block diagram of Mel filter bank

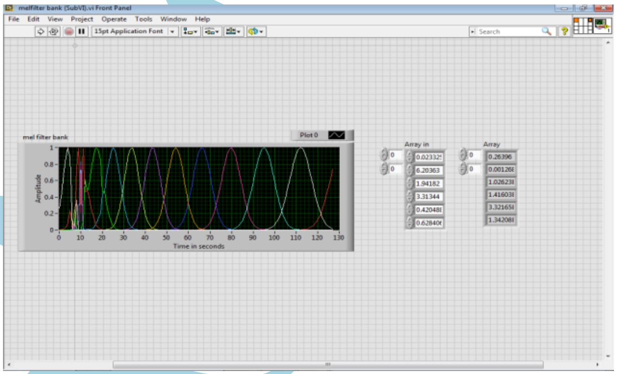

Figure 6.Front panel of Mel filter bank

III. FEATURES FOR DATABASE

After finding the DCT co efficient values for 20 filters and collecting database for some features. To find Mean,Variance,Skewness,Kurtosis,5th order moment and 6th order moment .

nth Order moment formula has been used,

$$\mu_n = \mathrm{E}\left[(X - \mathrm{E}[X])^n\right]$$

$$= \int_{-\infty}^{+\infty} (x-\mu)^n f(x)dx.$$

In Lab VIEW ,we use mathscript node to calculate features for collection of database.

IV. MINIMUM DISTANCE CLASSIFIER

Euclidean distance is used for minimum distance classifier method.

$$|x - y| = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Where , x is the trained sample,

y is the testing sample,

k=No. of features.

The above distance formula is used for finding minimum distance between trained samples and testing sample.

Table 1 describes that moment features of speakers and it contains Mean,Variance,Skewness,Kurtosis,5th order moment (Hyperskewness) and 6th order moment (Hyperflatness).

Moment features	Speaker1	Speaker2	Speaker3	Speaker4
Mean	0.00318	0.00608	0.00364	0.00234
Variance	0.20453	0.20005	0.20456	0.20653
Skewness	0.10632	0..02226	0.58008	0.02507
Kurtosis	8.01769	5.00562	9.44689	4.34228
5^{th} order	300.564	239.626	254.6	307.129
6^{th} order	894924	720329	657032	820714

Table 1.Moment features for Database collection

V. RESULT

The moment features is found in the Lab VIEW software using mathscript node to access matlab functions. The feature database is used for one to one comparison between trained samples and test speech signal. To finding the minimum distance between features database and test speech with the help of Euclidean distance. Table 2 shows the Identification results speakers.

Sl.No.	List of Speakers	Identification (Correct/Incorrect)
1	Speaker 1	CORRECT
2	Speaker 2	INCORRECT
3	Speaker 3	INCORRECT
4	Speaker 4	INCORRECT
5	Speaker 5	CORRECT
6	Speaker 6	CORRECT
7	Speaker 7	CORRECT
8	Speaker 8	CORRECT
9	Speaker 9	CORRECT
10	Speaker 10	CORRECT
11	Speaker 11	CORRECT
12	Speaker 12	CORRECT
13	Speaker 13	INCORRECT
14	Speaker 14	CORRECT
15	Speaker 15	CORRECT
16	Speaker 16	CORRECT
17	Speaker 17	CORRECT
18	Speaker 18	INCORRECT
19	Speaker 19	CORRECT
20	Speaker 20	CORRECT

Table 2.Speaker Identification Result

The speaker recognition system in LabVIEW environment is an efficient programming platform which gives 75% accuracy.

VI. CONCLUSION

This work describes speaker recognition systems as a part of the biometric security system. The speaker identification using MFCC method was implemented on Lab VIEW 2014 Platform. The feature have been extracted and stored in a database to be compared with a testing speech. In testing session ,Euclidean distance to finding minimum distance between trained and test signal.

REFERENCES

[1] Douglas A. Reynolds, James R. Glass, Timothy J. Hazen *and* Ji Ming.*"* Robust Speaker Recognition in Noisy Conditions*".* IEEE transactions on audio, speech, and language processing, vol. 15, no. 5, July 2007.

[2] Saurabh bhardwaj, smriti srivastava, madasu hanmandlu and J. R. P. Gupta." GFM based methods for speaker identification". IEEE transactions on cybernetics, vol. 43, no. 3, June 2013.

[3] Mohammed algabri, hassan mathkour, mohamed a. Bencherif, Mansour alsulaiman, and mohamed A. Mekhtiche." Automatic speaker recognition for mobile forensic applications". Hindawi ,March2017.

[4] Douglas A. Reynolds *and Richard C. Rose,"* Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models" IEEE transactions on audio, speech, and language processing, vol. 3, no. 1, January 1995.

[5] Simarpreet Kaur and Purnima," Speaker Verification using LabVIEW" International Journal of Computer Applications Volume 21– No.4, May 2011.

[6] Kumar Rakesh, Subhangi Dutta and Kumara Shama" Gender recognition using speech processing techniques in labview". International Journal of Advances in Engineering & Technology, May 2011.

[7] Vibha Tiwari," MFCC and its applications in speaker recognition". International Journal on Emerging Technologies ,February 2010.

[8] Athira Aroon and S.B. Dhonde ,"Speaker Recognition System using Gaussian Mixture Model". International Journal of Computer Applications Volume 130 – No.14, November 2015.

[9] Ashish kumar panda and Amit kumar sahoo, "Study of speaker recognition system",2011.

[10] Kunjithapatham Meena,Kulumani Subramanian, and Muthusamy Gomathy,"Gender classification in speech recognition using fuzzy logic and neural network".International Arab Journal of Information Technology volume 10,No.5,September 2013.

[11] National instruments visit www.ni.com

