

Human Recognition using Voice Print in LabVIEW

Dr. S.Selva Nidhyananthan¹, K.Muthugeetha², V.Vallimayil³

¹Associate Professor, ^{2,3} UG Students

¹⁻³Mepco Schlenk Engineering College, Sivakasi, India

Abstract

This paper describes speaker recognition 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, pre-processing, feature extraction has been done. For feature extraction, Mel Frequency Cepstral Co efficient (MFCC) is used. The moment features of speech are found for speaker identification.

Keyword: Lab VIEW software, computer, MFCC feature extraction

INTRODUCTION

This paper introduces a speaker recognition using Lab VIEW software. Speaker recognition is an identity of authentication process which automatically identifies individuals with the intrinsic characteristic conveyed by their voice. This is also called voice recognition. "Voice recognition" is used for both speaker verification and speaker 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 and verify the identity of a speaker as a part security process. In this paper, Mel frequency cepstral coefficient has been used for formant detection. Data base of fifty persons having two samples per person including male and female has been created for analysis of result. These systems operate with the user's knowledge and typically require the user's co-operation. The developed system uses the LabVIEW (Laboratory Virtual Instrument Engineering Workbench) 2014 platform.

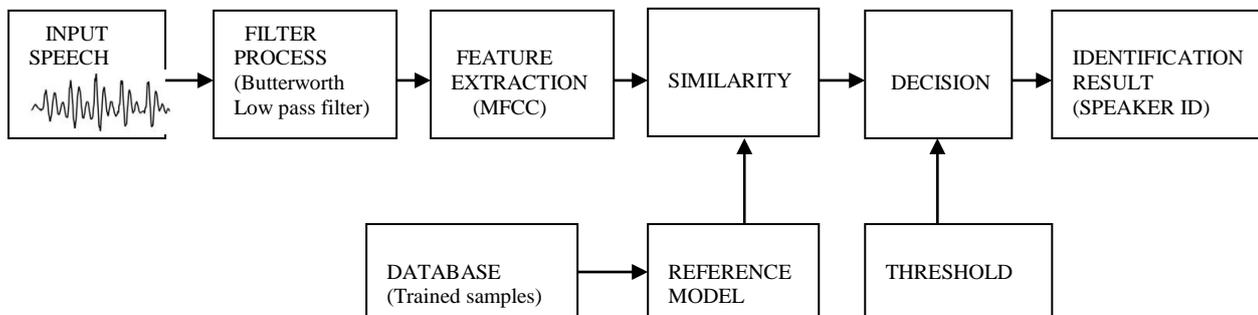


Figure 2. Block diagram of speaker recognition

METHODOLOGY

Initially, record and collect the voice from a person. Fifty persons have been chosen and 2 sample of each person were taken in .wav format. After recording and collection of voice, a low pass filter is used for silence and noise removal and voice sample is framed using framing technique and every signal is windowed using Hamming window technique voice samples are extracted using MFCC method.

A. To get input signal

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

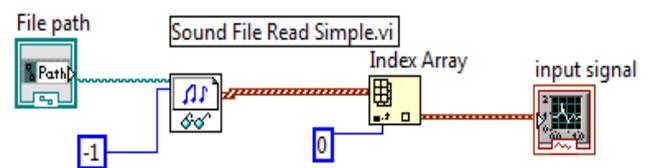


Figure 1. Reading the input signal in labVIEW

Figure 1 describes that sound file reading in LABVIEW software. File path is taken from already recorded voice samples.

B. Pre-processing

After getting input, silence and noise are removed using low pass filter. Low pass filter is used to allow the signals with a frequency lower than certain cut off frequency and block the signal with frequencies higher than the cut off frequency. We assume the cut off frequency is 1 KHZ.

FEATURE EXTRACTION

Mel Frequency Cepstral Co-efficient (MFCC) is used for feature extraction method. MFCC method contains six steps

involves there are Framing, windowing, Fast Fourier transform (FFT), Mel Filter Bank and Discrete Cosine Transform (DCT).

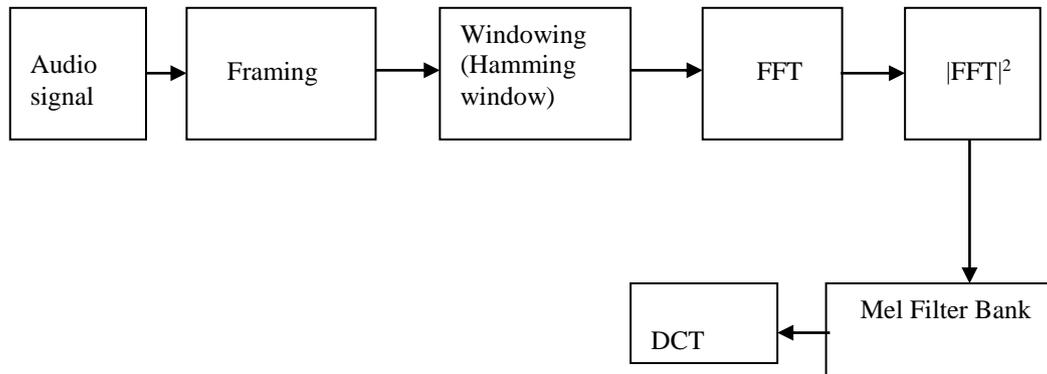


Figure 3. Feature Extraction method

A. Audio signal

The filtered signal is given as an input signal by following the block diagram. The input signal may be any audio signal or human voice.

B. 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 pre-processed 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. Let N=160samples, M=96samples, overlap (N-M)=64 samples(as per our speech signal). The purpose of overlapping is 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. Voice signals change their characteristics from time to time.

C. Windowing

Hamming window is used for our project. Hamming window is derived from following equation.

$$w(n) = 0.54 - 0.46 \cos(2\pi n / N - 1), 0 \leq n \leq N - 1$$

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

1) *Level-1 why windowing method:* A level-1 when frequency content of a signal is computed errors can arise when we take a limited duration snapshot of a signal that actually lasting for a longer time. windowing is a way to reduce these errors.

2) *Level-2 Advantages of windowing :* A level-2 Windowing is used to suppress the discontinuity and resulting spurious high frequencies in the frequency analysis by 'tapering' the recorded signal smoothly to zero at the start and end of the recording period.

3) *Level-3 why we are using hamming window:* A level-3 Hamming window is used to reduce the side lobe levels in signal. Also provides smaller main lobe width and sharp transition band. Hamming window is having better selectivity for large signal.

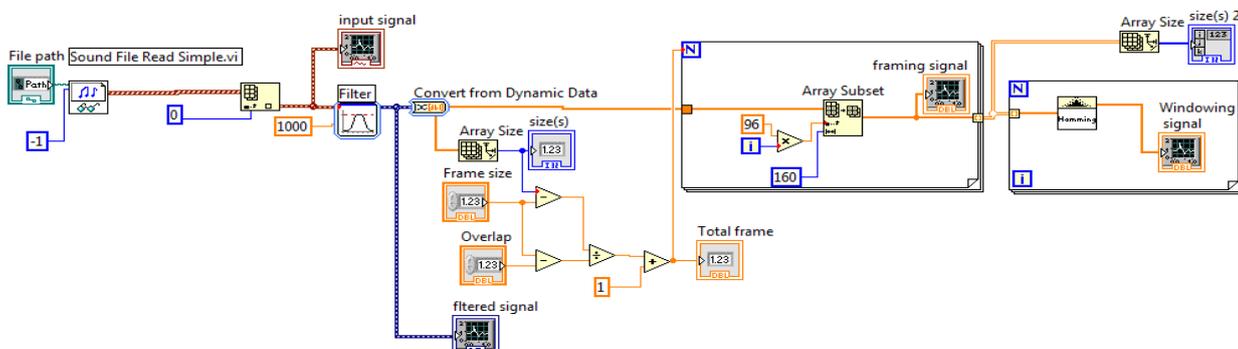


Figure 4. Framing and Windowing Process in Lab VIEW

D. FFT(Fast Fourier Transform)

FFT is used to convert a signal from its original domain to a representation in the frequency domain. And also used to simply characterize the magnitude and phase of a signal. And also absolute value of the FFT can be taken.

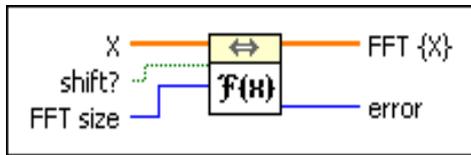


Figure 5. FFT in LabVIEW

Figure 5 shows that X is the input after completing the windowing signal. To manually select the FFT size .FFT it will converts time domain to frequency domain.

E. Mel Filter Bank

Mel filter is applied for 20 filters coming from the Fast Fourier Transform signal. 20 filters are assumed by us. Assuming no of filter is our choice. A Mel scale to the power spectrum to extract the frequency bands. The aim of the mel scale is to mimic the non-linear human ear perception of sound.. It is time-frequency distributor. Figure 6 describes that Mel filter bank for 20 filters. In that we create subVI for Mel filter bank.

F. DCT(Discrete Cosine Transform)

The discrete cosine (DCT) is used to decorrelate the filter bank co efficient a process is referred as whitening. In our project is applied for 20 co-efficient value because of applying DCT function the time domain signal is converted into frequency domain signal. Figure 7 describes that DCT for LabVIEW software. DCT will gives sum of cosine function.

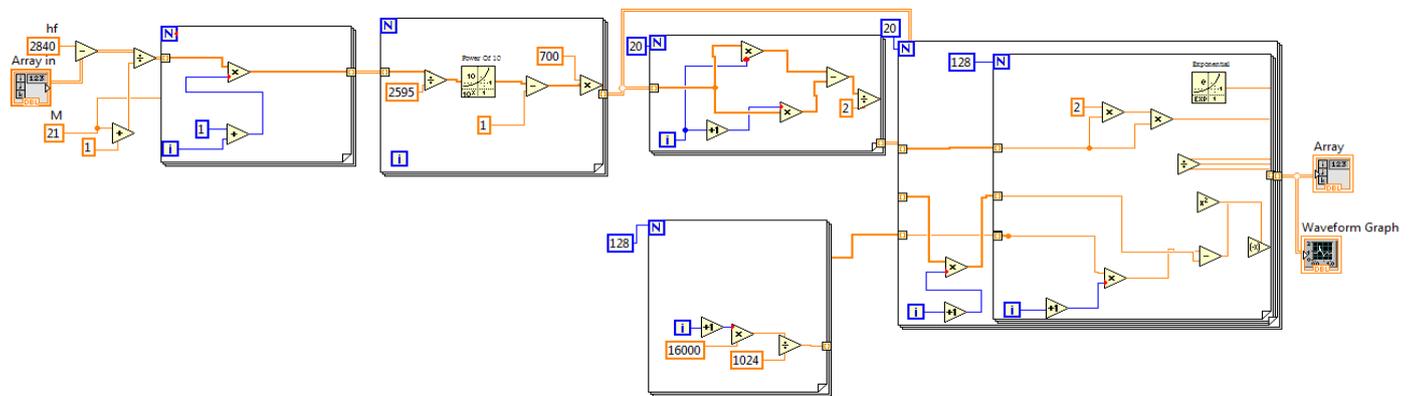


Figure 6. Block diagram of Mel Filter bank

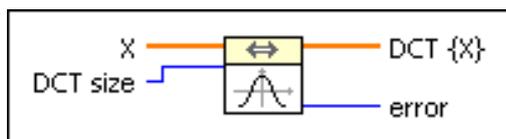


Figure 7. DCT in LabVIEW

FEATURES FOR DATABASE

Getting 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 for Database.

nth Order moment formula has been used,

$$\mu_n = E[(X - E[X])^n]$$

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

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

Table 1. Moment features for Database.

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 describes the speech features contains Mean, Variance, Skewness, Kurtosis for various speakers. The moment features are used to find the minimum distance

between trained samples and test sample for identifying speaker. To find out which speaker we have used MFCC coefficients .In MFCC coefficients we have compared the mean values of the speakers but in the mean there is minimum variation in the values .so that we have moved to higher order for collecting database.

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

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.

MINIMUM DISTANCE CLASSIFIER

Euclidean distance is used for minimum distance classifier method.

RESULT AND DISCUSSION

Table 2. Identified speakers

SPEAKERS	S1	S2	S3	S4	S5
S1	631.855	1.49E+10	1.93E+10	2.74E+09	1.16E+09
S2	1.49E+10	7929.11	4.40E+09	1.22E+10	1.37E+10
S3	1.93E+10	4.40E+09	21868.7	1.66E+10	1.81E+10
S4	2.74E+09	1.22E+10	1.66E+10	1828.6	1.58E+09
S5	1.16E+09	1.37E+10	1.81E+10	1.58E+09	370.264

Table 3 shows that corresponding speakers will give minimum value. For example, S4 as testing speech sample and its features compared with the all speakers in the database. Table 3 describes that identify the speakers by using threshold value. We have chosen threshold value as 8000. If difference between the speakers below threshold means it will gives ACCEPT and also identify the speaker id and its difference above threshold means it will gives REJECT to display in the front panel .

In the above table, Yellow color shaded box indicates correctly identified speakers and Red color shaded box indicates incorrectly identified because the value is greater than threshold value.

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, Euclidian distance to finding minimum distance between trained and testing signal. The experiments have been conducted on the database stored in the .wav files and it has

been observed that the system is accurate up-to a value of 85%.

REFERENCES

- [1] Simarpreet Kaur and Purnima,,” Speaker Verification using LabVIEW” International Journal of Computer Applications Volume 21– No.4, May 2011
- [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] 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
- [4] Athira Aroon and S.B. Dhonde, Speaker Recognition System using Gaussian Mixture Model”. International Journal of Computer Applications Volume 130 – No.14, November 2015.
- [5] 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.
- [6] Vibha Tiwari, ”MFCC and its applications in speaker recognition”. International Journal on Emerging Technologies, February 2010./
- [7] Kumar Rakesh, Subhangi Dutta and Kumara Shama” Gender recognition using speech processing techniques in labVIEW”. International Journal of Advances in Engineering & Technology, May 2011.
- [8] Mohammed algabri, hassan mathkour, mohamed a. Bencherif, Mansour alsulaiman, and mohamed A. Mekhtiche.” Automatic speaker recognition for mobile forensic applications”. Hindawi , March2017.
- [9] A.K. Jain, A. Ross and S. Prabhakar, “An Introduction to Biometric Recognition,” IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video Based Biometrics, vol.14, no.1, pp.4-20, Jan.2004.
- [10] F Orság, “Some Basic Techniques of the Speech Recognition”, In: Proceedings of 8th Conference STUDENT EEICT 2002, Brno, CZ, FEKT VUT, pp. 5, ISBN 80- 214-2116, 2002.
- [11] L.R. Rabiner, B.-H. Juang, “Fundamentals of Speech Recognition” (Prentice-Hall, Englewood Cliffs, 1993 .
- [12] Ashish kumar panda and Amit kumar saho, ”Study of speaker recognition system”,2011.
- [13] 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.
- [14] National instruments visit www.ni.com