International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

"Speaker Verification with Lab VIEW"

Kanwar Sudeep Singh Sandhu[1], Sukhwinder Singh[2]

[1]Student, E&EC Department, PEC University of Technology, Chandigarh, India

[2]Mentor, E&EC Department, PEC University of Technology, Chandigarh, India

Abstract: ***The speaker verification is a process of verifying the identity of the claimants. It performs one to one comparison between a newly input voice print and the voice print for the claimed identity that is stored in the database. In this paper, linear predictive coding co-efficient has been used for formant detection. The peak frequencies in the frequency response of vocal tract are formants, which is being detected and compared for verification. Data base of twenty persons having five samples per person including male and female has been created for analysis of results. The System (Speaker verification) is usually employed as a "gatekeeper" in order to provide access to a secure system. These systems operate with the user"s knowledge and typically require the user"s cooperation. The developed system uses the LabVIEW (Laboratory Virtual Instrument Engineering Workbench) 2009 platform.***

Keywords: Automatic Speech Recognition, Biometry, Cross-Correlation, Fingerprint Recognition, Formant Detection, LabVIEW, Pre-Emphasis, Pre-Processing, Voice Recognition.

1. Introduction

In the present day of automated world, machine is replacing the human in every aspect of life. Due to this, the security concern regarding the authenticity of the user goes on increasing. Hence, it becomes necessary to include some constrains in order to reject impostors and allow only the authorized user to access the automated services. Traditional methods of PIN or password are not reliable enough to the security requirement of electronic transactions because some other unauthorized person (imposter) can steal the account number and password and the system will give the access to that unauthorized person. The most reliable method is to use the characteristics of the user such as speech, finger print, hand geometry etc. The user voice is the simplest, natural, easily acceptable way for the user recognition and can be transmitted easily.

1.1 Speaker Verification

Speaker Verification is a method of confirming the identity of an individual from his or her speech. There are several possible applications for speaker verification in Security and Identification systems. Now a day"s, biometrics is being used extensively for the purpose of security. It deals with identifying the individuals with their physiological factors such as finger print, face, DNA, ECG, etc. or behavioral traits i.e. rhythm, gait, voice etc. [1], [2]. Speaker Verification is a biometric system which provides positive verification of identity from individual"s voice characteristics [3]. Speech and music are the most basic means of adult human communication. As the technology is advancing and increasingly sophisticated tools become available to be used with speech and music signals, scientists can study these sounds more effectively, and invent new ways of applying them for the benefit of humankind. It includes coverage of the physiology and psychoacoustics of hearing as well as the results from research on pitch and speech perception, various methods and information on many aspects of automatic speech recognition (ASR) systems [4].

There are several possible implementations for systems that perform speaker verification. The implementation which is most frequently used makes use of a match filter. There are various characteristics of the speech which can be matched. These include magnitude, energy, frequency, linear predictive coding co-efficient (LPC) and Cepstrum Techniques [5]. In the present work, a speaker verification system has been developed using formant detection with LPC method. The speaker recognition is becoming very important in the biometry. Nowadays, only the fingerprint technology is fully accepted so that many real applications based on this technology have been created since now [6]. The speech technology is in the frame to be as successful as the fingerprint technology.

Speaker Verification

1.2 Biometrics

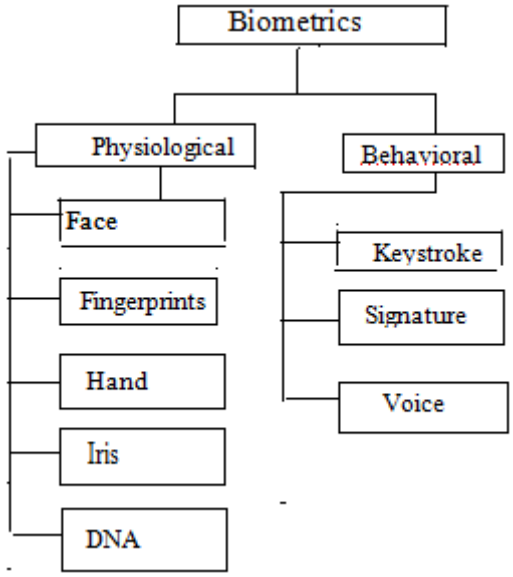
The term biometrics is now widely known as "the science of measuring physical characteristics, to verify the person"s identity and has been derived from the Greek words Bio (life) and Metric (to measure).It includes Voice recognition, Iris and face scans and fingerprint recognition [7]. Fig. 2 shows the use of Biometrics in various fields.

1.3 Speech in Biometry

Voice (or vocalization) is the sound produced by humans and other vertebrates using the lungs and the vocal folds in the larynx, or voice box. Voice is not always produced as speech. However, Infants babble and coo; animals bark, moo, whinny, growl and meow; and adult humans laugh,

sing, and cry. Voice is generated by airflow from the lungs as the vocal folds are brought closed together. When air is pushed past the vocal folds with sufficient pressure, the vocal folds vibrate [8]. If the vocal folds in the larynx did not vibrate normally, speech could only be produced as a whisper. Our voice is as unique as our fingerprint. It helps define our personality, mood, and health.

The range of use of the speech technology in the biometry is very wide. The most common one is the user authentication using his/her voice. In some cases, another speech technology is applicable, e.g. in the case of phone banking, the system can be fully automatic, which is realized by combination of the speech recognition, speaker recognition, and speech synthesis.



Biometric Characteristics

2. Methodology

The methods for speech recognition has been used in three phases :

a) Pre-Processing (Phase 1)
In Pre-Processing step, windowing and filtration of Voice Signal is done.

b) Feature Extraction (Phase II)
Using LPC Formant Detection method, the features of voice signal are extracted.

Key signal which is stored in the database is compared with the current data input by using Amplitude- Level measurement and Cross-Correlation tool.

2.1 System Implementation (Phase I)

2.1.1 Pre-Processing/ Pre-Emphasis
In this Research, in consideration with each available technique for speech recognition, an advanced method is presented that is able to classify the speech signals with high accuracy in a minimum time. In the presented method, first, the recorded signal is preprocessed.

The digitized sound signal contains relevant data and irrelevant information, such as white noise. Therefore, it requires a lot of storage space [9]. Most of the frequency components of speech signal are below 5KHz and upper ranges almost include white noise that have direct impact on system performance and training speed, because of its chromatic nature. So, speech data must be preprocessed. For pre-processing and post processing of voice signals, LabVIEW has been used on both the ends [10].

2.1.2 Audio Signal Processing
Sometimes referred to as audio processing, is the intentional alteration of auditory signals, or sound. As audio signals may be electronically represented in either digital or analog format, signal processing may occur in either domain. Analog processors operate directly on the electrical signal, while digital processors operate mathematically on the binary representation of that signal. Human hearing extends from approximately 20 Hz to 20 kHz, determined both by physiology of the human hearing system and by the human psychology. These properties are analyzed within the field of psychoacoustics [11].

2.1.3 VI Filters
The filter which is used here is basically an IIR filter. It records the sound from an external microphone with the help of the internal sound card and plays it simultaneously. It also filters the sound to reduce high frequency noise that may be added to the speech signal from the surrounding environment or the inherent noise added by the amplifiers that may be used before recording the sound. LabVIEW provides user control over the type of filter to be used, the order of the filter, the sampling frequency and the attenuation in the pass and stop bands. Although, made basically for sound recording and playing simultaneously in noisy environments, the VI can also be used as a sub VI for making other high level projects like equalizer, or any other application which may need any of the band-pass, band-stop, low-pass or high-pass filter, as it implements all these filters in a single VI.

2.1.4 Working of VI Filters
In the first phase, the VI was made to acquire sound from the sound card, filter it and output it simultaneously. The block diagram basically consists of the following parts:

1. Reading through the sound card.
2. Filtering the data.
3. Writing to the sound card.
4. Reporting the errors, if any, while executing the loop.

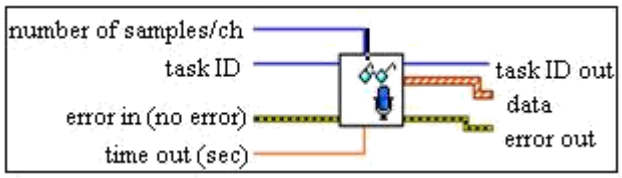
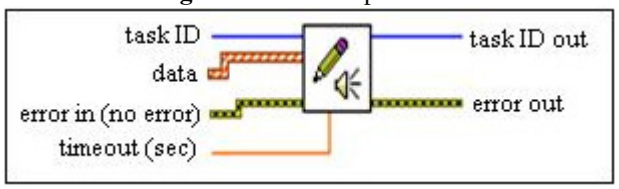
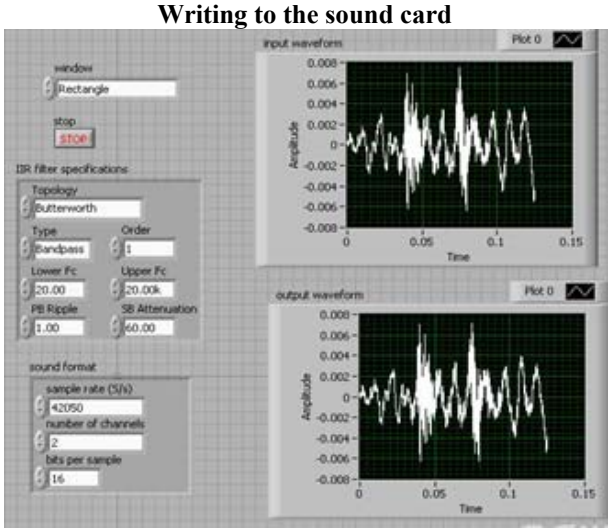


Figure 3: Sound Input Read



Writing to the sound card



2.1.5 IIR Filter Front End

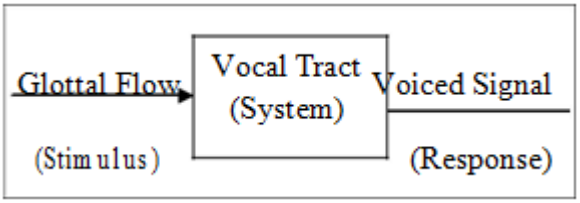
The VI filters use the input voice coming from the microphone (in accordance with the controls set by the user) and simultaneously outputs it to the speaker. The recorded file is stored in a user specified location with a user specified name. It also displays the time domain graphs of the data on the front panel, which is shown in Fig.5.

2.2 Feature Extraction (Phase II)

2.2.1 Voice Signal Analysis

Output of the Pre-Processed signal enters the Second Phase for further signal processing and feature extraction using LPC Formant Detection Method. When we pronounce a vowel or a voiced consonant, the vocal cords periodically vibrate to generate glottal flow. The glottal flow is composed of glottal pulses. The period of a glottal pulse is the pitch period. The reciprocal of the pitch period is the pitch, also known as the fundamental frequency. The vocal tract acts as a time-varying filter to the glottal flow. The characteristics of the vocal tract include the frequency response, which depends on the position of organs, such as the pharynx and tongue. The peak frequencies in the frequency response of the vocal tract are formants, also known as formant frequencies. In signal processing, a voice signal is a convolution of a time-varying stimulus and a time-varying filter. The time-varying stimulus is the glottal flow. The characteristics of the vocal tract include the frequency response, which depends on the position of organs, such as the pharynx and tongue. The peak frequencies in the frequency response of the vocal tract are formants, also known as formant frequencies. In signal processing, a voice signal is a convolution of a time-varying stimulus and a time-varying filter. The time-varying stimulus is the glottal flow. The time-varying filter is the vocal tract [13].

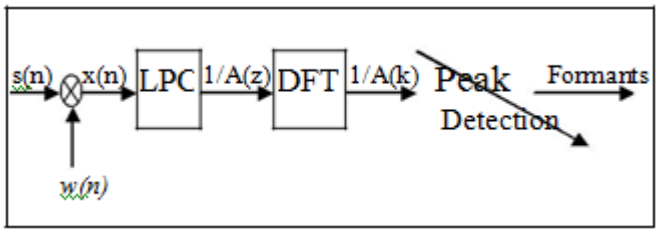
Researchers study formant tracks and pitch contour to understand how formants and pitch evolve over time. The second attempt describes how to detect formant tracks and pitch contours from voice signals by using the National Instruments LabVIEW graphical development environment. This section includes the Voice Signal Analysis, which is built with LabVIEW. Fig 6 shows the Vocal Tract System.



Vocal Tract System

2.2.2 Detection of Formants and Pitch

Formant tracks and pitch can be calculated using several methods. The most popular method is the Linear Predictive Coding (LPC) method. This method applies an All-pole model to simulate the vocal tract. Fig. 7 shows the flow chart of formant detection with the LPC method. In this Figure, applying the window w(n) breaks the source signal s(n) into signal blocks x(n). Each signal block x(n) estimates the coefficients of an all-pole vocal tract model by using the LPC method. After calculating the discrete Fourier transform (DFT) on the coefficients A(z), the peak detection of 1/A(k) produces the formants.



Formant Detection with LPC Method

2.3 Comparison of Signals (Phase III)

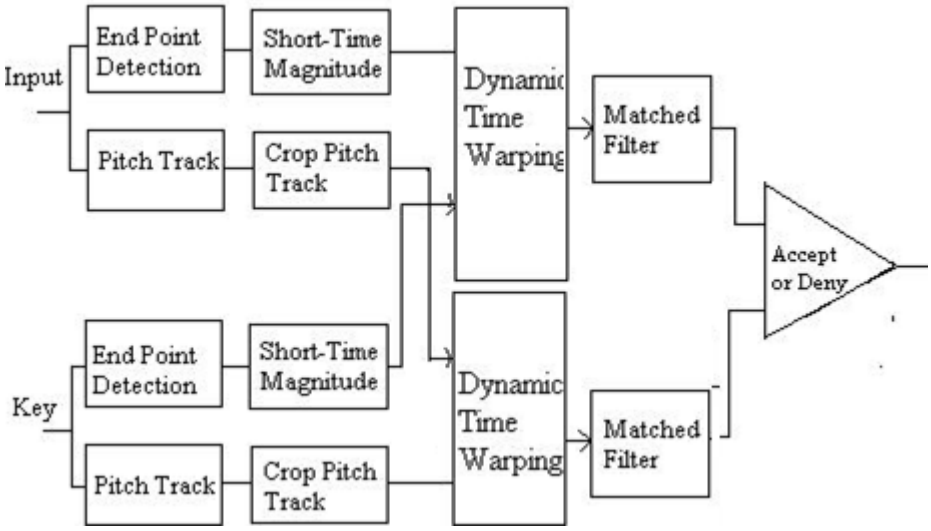
2.3.1 Peak Detection and Signal Correlation

Output of 2nd attempt is configured and measured for formants of each person output for verification and identity of Individuals output of a person or group of persons, which has been configured below from the Amplitude and Level Measurement Tool. Further, the output voice of each can me compared with itself and with voice of another person with convolution and cross- correlation tool.

2.3.2 Registration Process to add new user to the database

Once a person has been entered and a key has been established, an authorization attempt breaks down into the following steps:

1. The person speaks a password into the microphone.
2. Signal is pre-processed with windowing and filtering.
3. LPC formant extraction is used for the short-time magnitude of the signal as is the pitch tract.
4. Each is cropped and dynamically time warped so that corresponding points on the signals are aligned.
5. Now that the signals rest on top of each other, matched filter (for both magnitude and pitch) is used to determine a numerical value for their correlation.
6. These numbers are compared to the thresholds that were set when the key was first created. If both the magnitude and pitch correlations are above this threshold, the speaker has been verified.



Block Diagram of Speaker Verification System

3. Results and Discussion

3.1 Registration

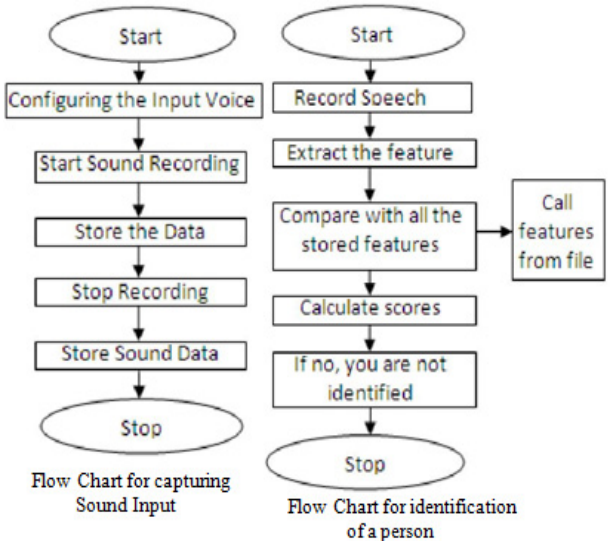
LabVIEW software of Speaker Verification includes two steps:

1. Registration of the user.
2. Testing for Authentication.

Registration is the necessary step in the Automatic Speaker Verification System which makes database storage in the system. Figures 9 and 10 show the complete flowcharts for the registration of a new user and capturing his voice along with the identification of the speaker.

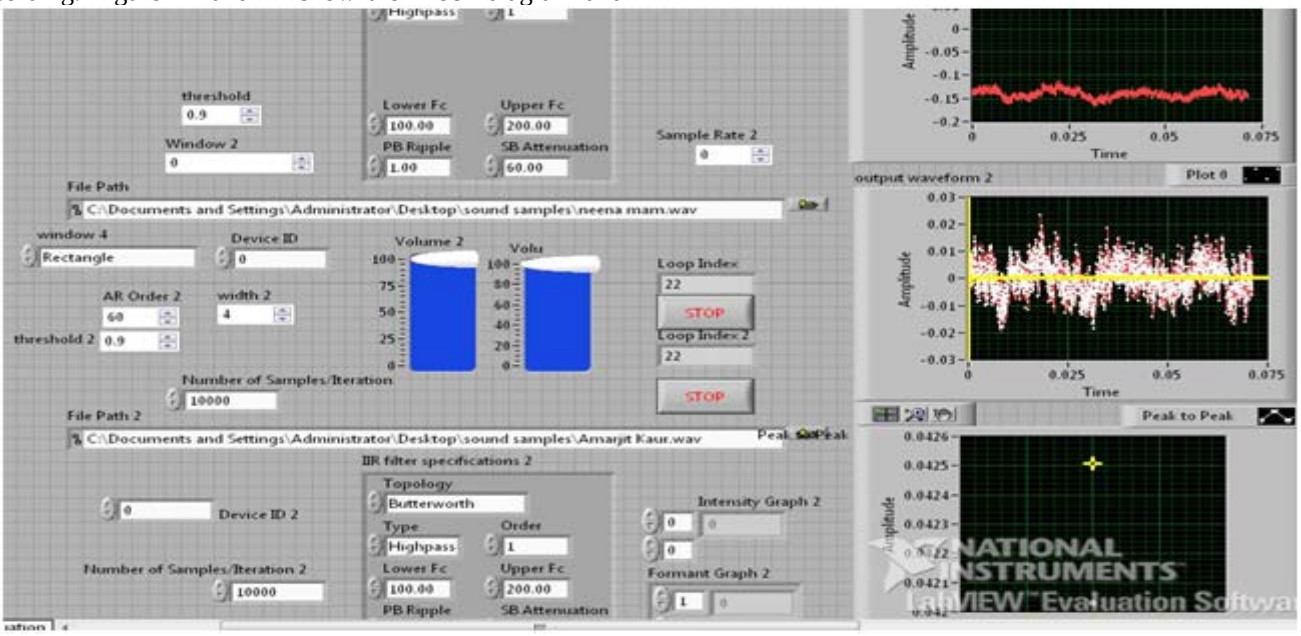
3.2 Database

The Database is made in LabVIEW, which includes both male and female. Twenty persons have been chosen and 3 sample of each were taken. In this way a total of 60 voice sample were used for training and testing. In which Recording and Stop button is used to start and stop recording. Figure 11 and 12 show the Block diagram and front panel of the VI sound recording and comparison system.

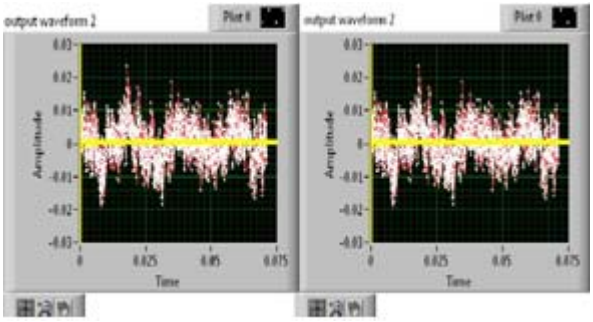


Flow Chart for capturing Sound Input

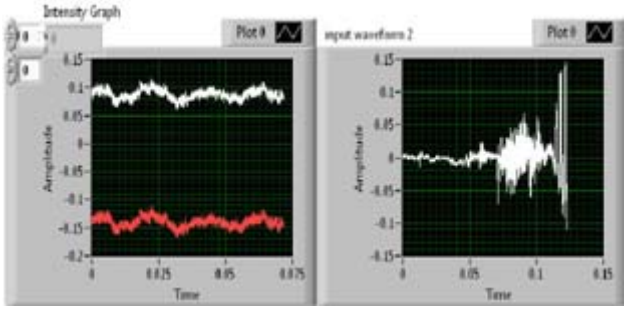
Flow Chart for identification of a person



Front Panel of Signal Verification

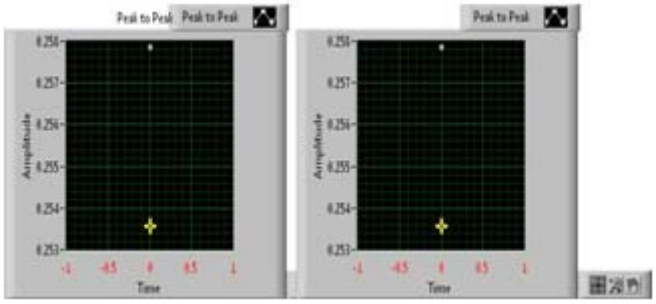


Output Voice Signal Waveforms and Peak Values of same person

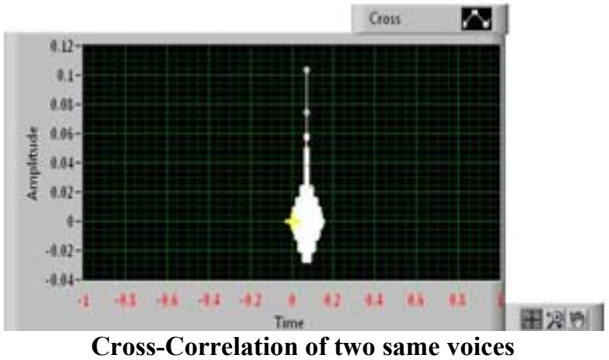


Input Voice signal waveforms of different persons

After the first step of Registration, the user enters in the data base. For this, the registration program has been employed which enrolls the user by: Entering the Name e.g. entering the biometric signal so that speech signal of the user can be processed as shown in the above figure.



Peak to Peak values of the same person

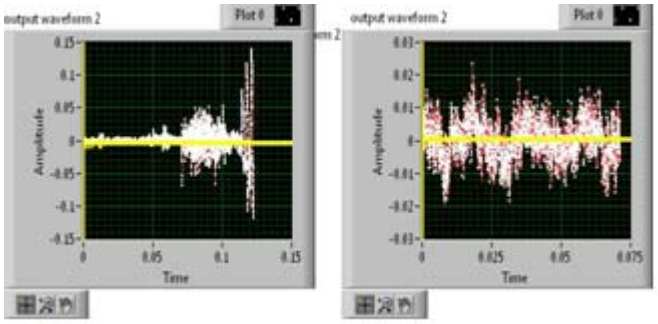


Cross-Correlation of two same voices

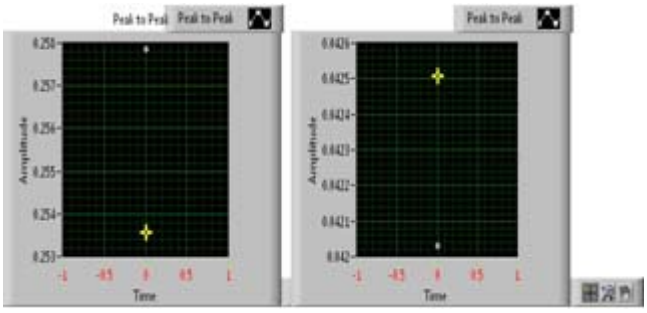
3.3 Verification

Verification has been done by calculating the pitch and magnitude of the new input speech signal from a Speaker with the estimated speakers that are already registered. The process is:

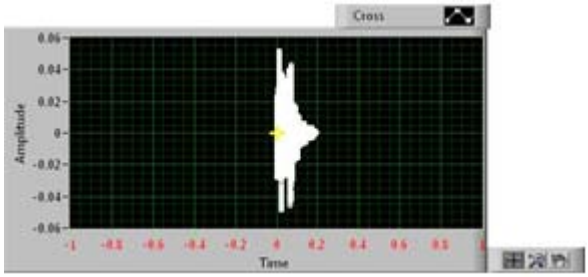
1. Enter Voice Print (Speech Signal)
2. Extract LPC co-efficient
3. Calculate the distance D and compare it with the given distance.



Output Voice signal waveforms of different persons



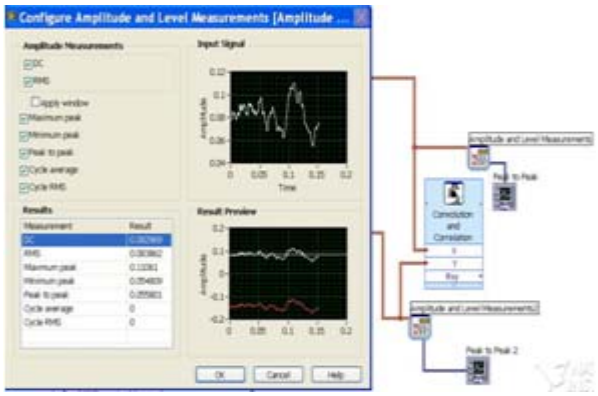
Output Voice signal peak values of different persons



Cross correlation output of two different Voices

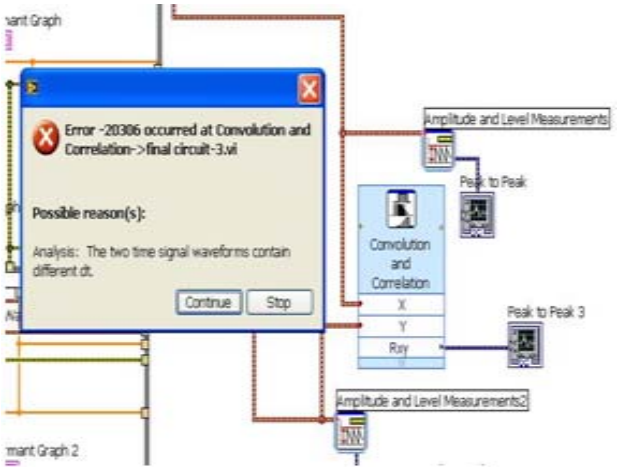
3.4 Extracted Feature Measurement (LPC of Speech Signal)

LPC Cepstrum Coefficient has been extracted from the speech Signal, compared and aligned for starting phase of both the signals. Finally the peak to peak value has been measured for each and matched with the key signal for identification.



Comparison of both the signals

During testing of the system, the system checks the identity of the person. The experiments have been conducted on the database stored in the LabVIEW which is the key data and is compared with the new input voice signal. The results obtained are shown in Table 1 and 2.



Error due to comparison of two different Signal Waveforms

Table 1: Result Verification with same voice signals (Where key is the stored data and PP is the Peak to Peak value of the Voice Signal)

Sr. No	*Key*	*Person*	*Pp value of key*	*Pp value of person*	*Cross correlation*	*Cycle RMS*	*Result /Remarks*
1	Am-2	Am-2	0.0163	0.0163	0.04	0.00513	Match
2	Am-2	Am-3	0.0163	0.0211	0	0.00322	No Match
3	Am-3	Am-4	0.211	Low signal Length (sample didn"t reference Value and less than threshold point)			
4	Ksb-1	Ksb-1	0.0442	0.0442	0.4	0.006405	Match
5	Ksb-1	Ksb-2	0.0442	Less loop Index or low ZCR			
6	Me	Me	0.01166	0.1166	0.013	0.002003	Match
7	Me	Me-1	0.01166	0.01368	0	0.002091	Match
8	Me-1	Me-1	0.01368	0.01368	0.04	0.002091	Match
9	Me-4	Me-4	0.01195	0.01195	0	0.002152	No Match
10	Me-2	Me-3	0.0147	0.0108	0.04	0.002152	Match

Table 2: Result Verification with different voice signals (Where key is the stored data and PP is the Peak to Peak value of the Voice Signal)

S. No	Key	person	Pp value of key	Pp value of Person	Cross correlation	Cycle RMS	Result /Remarks
1	Am-2	Ksb-1	0.0163	0.0442	0	0.00513	Not Verified
2	Ksb-1	Me	0.0442	0.1166	0	0.002003	Not Verified
3	An	Rl	0.09	0.131	0	0.0136	Not Verified
4	Rk	Rl	0.00862	0.131	0	0.004367	Not Verified
5	Me-1	Rk	0.01368	0.00862	0	0.002091	Not Verified
6	Za	Rk	0.258	0.00862	0	0.071566	Not Verified
7	Za	Am-2	0.258	0.0163	0	0.00513	Not Verified
8	An	Me-3	0.09	0.0108	0	0.001607	Not Verified
9	RT	ND	0.0554	0.0552	0	0.01381	Match/error
10	RT	Am-3	0.0554	0.774	0	0.01381	Not Verified

4. Conclusion and Future Scope

4.1 Conclusion

This work describes Speaker Recognition System as a part of the Biometric Security System. Mainly, this work aims at the Speaker Verification and the research in this field. The Speaker Verification system using formant detection with LPC is implemented on LabVIEW2009 platform. There are two session of system. First is Registration and second is Testing. In Registration session, firstly pre-emphasis of the signal is performed. Features are extracted and stored in a file to be compared with the query. In the testing session, voice print of unknown speaker is taken. 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 94%.

4.2 Future Scope

Automatic speaker recognition system using LabVIEW is an efficient program giving almost 94% accuracy but still there are chances for improvement.

1. The main problem of the system is the external noise. By using some other noise elimination methods, the performance of the system can be improved.
2. Different method of Silence Remove can be used.
3. High quality microphone can be used to improve the system accuracy.
4. The system can be tested on the larger database.

References

[1] Federal Bureau of Investigation Educational Internet Publication, 1997, DNA testing, "http://www.Jbi.gov/kids/dna/dna.htm.
[2] F. J. Prokoski, R. B. Riedel, and J. S. Coffin, "Identification of individuals by means of facial thermography," in Proceedings of The IEEE 1992 International GA, USA 14-16 Oct., pp. 120-125, 1992,IEEE.
[3] L.R. Rabiner, B.-H. Juang, 1993 "Fundamentals of Speech Recognition" (Prentice-Hall, Englewood Cliffs.
[4] F Orsag, "Some Basic Techniques of the Speech Recognition", In: Proceedings of 8th Conference student EEICT 2002, Brno, CZ, FEKT VUT, pp. 90-94, 2002,ISBN 80- 214-2116.
[5] Noll, A.M. 1967. Cepstrum Pitch Determination Journal of the Acoustical Society of America 41 (February): 293-309.
[6] The project: Speaker Verification, from ni.com.
[7] A.K. Jain, A. Ross and S. Prabhakar, Jan.2004 "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.
[8] A.Barney, C.H Shadle, and P.O.A.L. Davies, "Fluid Flow in a Dynamical Machanical Model of the Vocal
[9] Folds and Tract. I: Measurements and Theory, "J. Acoustical society of America, Vol. 105, no, 1, pp 444-445, Jan. 1999.
[10] D.R Rodman, "Computer Speech Technology, Boston", Mass.: Artech House, 1999. Carnahan Conference on Security Technology: Crime Countermeasures, Atlanta.
[11] Using LabVIEW for Voice Signal Analysis, ni.com. B. Gold, N Morgan, "Speech and Audio Signal Processing", New York, USA, John Wiley & sons, inc, 2000.
[12] A.V Oppenheim, R.W Schafer, Buck, J.R.: Discrete-Time Signal Processing, 2nd ed, Upper Saddle River, NJ, Prentice Hall, 1999.
[13] J.R Deller, J.H.L Hansen, Proakis, J.G. Discrete-Time Processing of Speech Signals, New York, USA, IEEE Press, 2000, ISBN 0-7803-5386-2.