Voice & Speech Based Security System Using MATLAB

1. Silvy Achankunju
M Tech – II year
silvy.jan28@gmail.com

2. Chiranjeevi Mondikathi
Assistant Professor-EC Dept.
chirusgsits@gmail.com

Acropolis Institute of Technology & Research, Indore

Abstract: VSSS is designed for the users who intend to achieve high definition security system to secure office or home. The important features for a security system to be accessible are authentication as well as authorization. The human voice includes speech as well as voice. For the system to operate, speech as well as voice is mandatory to match the recorded sample. The design of system also involves a GSM system to provide a feedback message to the prime member of the unit with the information of user accessing the system. This project includes the hybrid of all approaches which are helpful in achieving more accurate result.

Keywords: Voice recognition, security system, speech recognition.

I. INTRODUCTION

Self security as well as security of the belongings is the critical matter of concern. In order to keep the surrounding safe, people assure the safety of the assets with biometric security systems. But these systems are also having limitations in their own way. Biometric systems (like iris scanning, thumb print, finger print, password etc) will be violated by making the authorized user unconscious.

VSSS is a system in which security is provided by user's voice as well as speech. Once the system agrees with the voice & speech of authorized user, then only the locked structure can become accessible.

As every human voice is unique in nature, so this characteristic is very helpful in maintaining the dignity of security. Also, VSSS is unlike biometric systems. It is difficult to break VSSS as human should be conscious while accessing the arrangement. Also various feeling states (anxiety, peaceful, delighted etc) of user, the input will also be unacceptable for the system, as the human pitch differs in different state. It is not necessary that the user will speak at same pitch in test duration as he/she spoke during train duration.

II. PROJECT DESCRIPTION

In the project, basically is hybrid of methods to acquire better result. The voice input is served through microphone of the system. The system is trained by the authorized person's voice sample. The analysis of the test duration voice piece is carried by MATLAB coding. The result of analysis will determine the accessing of system permitted or not

A GSM module is connected with the system to ensure the high definition security. If the system detects authorized user's voice, then the locked door will be opened as at the same time, a message, which contain the information of the user, will be sent to the principle member of the organization. Also, in case of intruder, the message is sent.

III. FLOW CHART

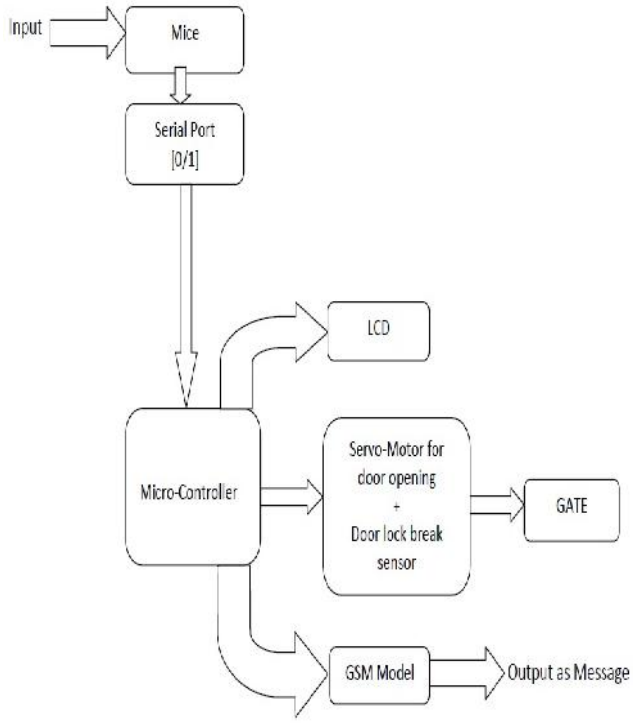

Fig 1: Flow chart of project

IV. Circuit Description

A) Hardware Model

The hardware of the project consists of following sections.

- Microcontroller
- LCD display
- Serial to parallel module
- GSM module

B) Software Model

The software section of the project is mainly divided in two categories. Train category and Test category. Both of these have more or less same segments.

- ✓ For train category
  - Record the voice sample
  - Sampling
  - Crop the above voice sample
  - Hamming window
  - FFT
  - MFCC
- ✓ For test category
  - Initialize the GSM port
  - Record the voice sample
  - Sampling
  - Crop the above voice sample
  - Hamming window
  - FFT
  - MFCC
  - MSE condition calculation
  - Euclidean Distance
  - Result

V. Explanation

A) Voice sample Recording –

Voice input is taken from microphone and is processed through MATLAB. The syntax to record the voice signal is

$$x = wavrecord\,(n\ , f_s)$$

where m = no. of samples of audio/voice signal

fs = sampling frequency

B) Sampling of audio signal –

Conventionally, a Nyquist criterion is helpful in sampling the audio signal.

According to the theorem, if the maximum frequency is fm, then information available in the audio signal can be reproduced by its samples but the necessary condition is

$$f_s = 2 * f_m$$

C) Crop audio signal –

As there is a silent portion in the audio signal, so it should be removed so that the filtering we will perform in future steps will involve the portions which contain the relevant information. All the unnecessary data to be cropped from the signal.

D) Hamming Windowing –

Hamming window is the combination of Hanning window & rectangular window. The selection of Hamming window is better for large signals. Hamming window equation is

$$h(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

where N = no. of samples in each frame.

Windowing is achieved by following formula

$$Y(n) = x(n) * w(n)$$

Where Y(n) = output signal

x(n) = input signal

w(n) = Hamming window

E) *FFT –*

Fast Fourier Transform is the efficient method to compute Discrete Fourier Transform. FFT is a digital processing tool. FFT formula is

$$X_k = \sum_{n=0}^{N-1} x_{ne^{-i2\pi k\frac{n}{N}}}$$

Where k = 0, 1 ….. N-1

F) *MFCC –*

Mel Frequency Cepstral Coefficient is the best for speaker recognition as it considers the human sensitivity of perception with respect to frequency.

Mel is the unit of pitch and it stands for Melody.

Mel filter will divide the wide spectrum to analysis the energy of signal at various points. Speech signal is partitioned into certain defined width window and set of parameters is derived from each window. Formula for mel filter is

$$Mel(f) = 2595 * log\left(1 + \frac{f}{700}\right)$$

G) MSE condition calculation –

It is a function that shows the average square of the difference between the value trained in the system & the value tested in that particular instant. The formula for MSE is

$$Mse = \frac{1}{l}\sum_{i=0}^{l}(\widehat{x_i} - x_i)^2$$

H) Euclidean Distance –

It is an effective method of determining the distance between two vectors. It is used to test the speech of the user. For instance, vector p and q, so the distance is calculated by following formula

$$Dis\,(p, q) = \sqrt{\sum_{l=1}^{n}(q_l - p_l)^2}$$

VI. RESULTS

(A) For train category

The system is trained by the USER 1 and word to train is – "**SILVY**"

- Original voice Signal of speaker Silvy for word "**SILVY**"

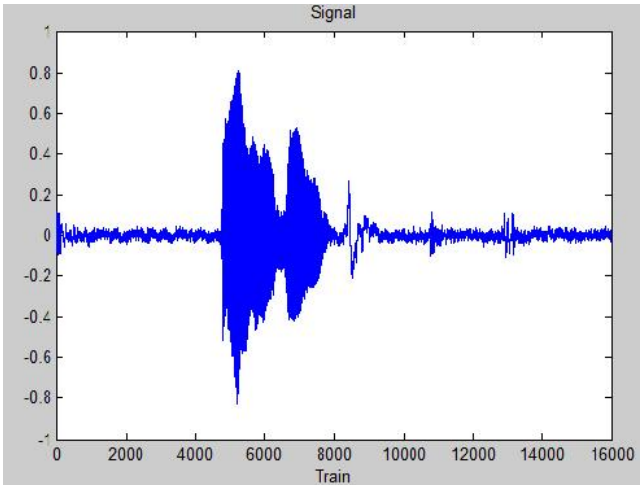

Fig 2: Voice Signal of USER 1

- Cropped signal

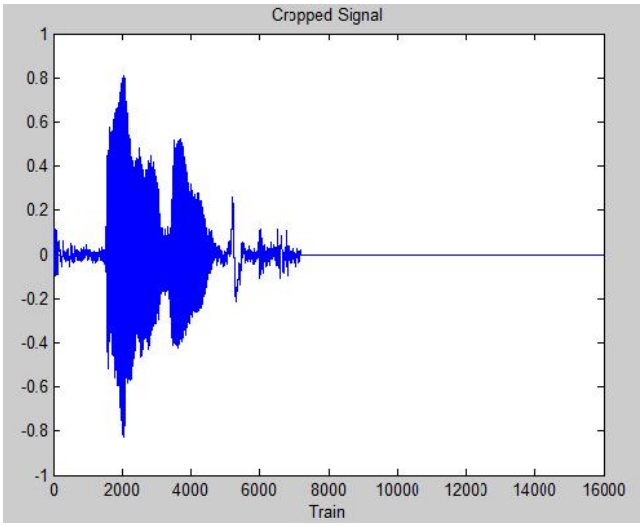

Fig 3: Cropped Signal of USER 1

- FFT of the above resultant signal

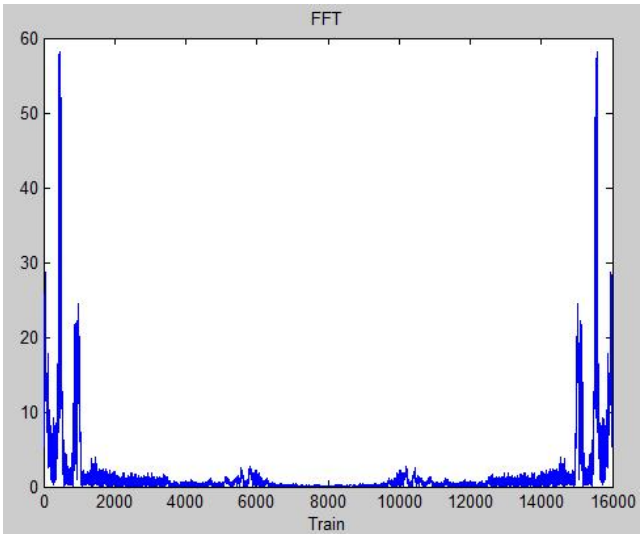

Fig 4: FFT of USER 1 cropped voice signal

- Mel graph of the voice of authorized person

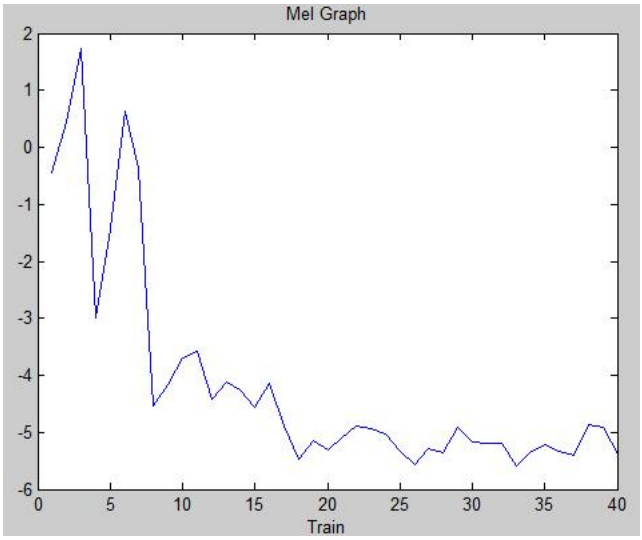

Fig 5: Mel Graph of voice signal of USER 1

(B) For test Category

All the essential steps will be carried for test category like cropping the original signal, taking FFT & Mel factor.

There exists two cases in this category

Case a) when the user is authorized

- Original voice signal of USER 1 for word "**SILVY**"

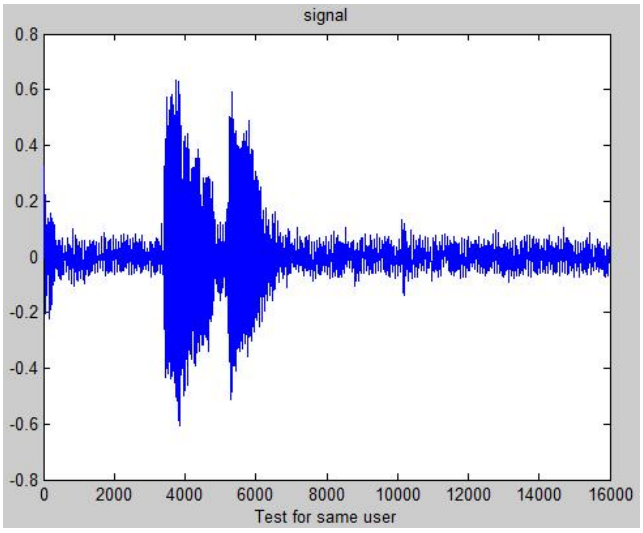

Fig 6: Voice Signal of USER 1

- After all process, Mel Graph for USER 1 voice signal

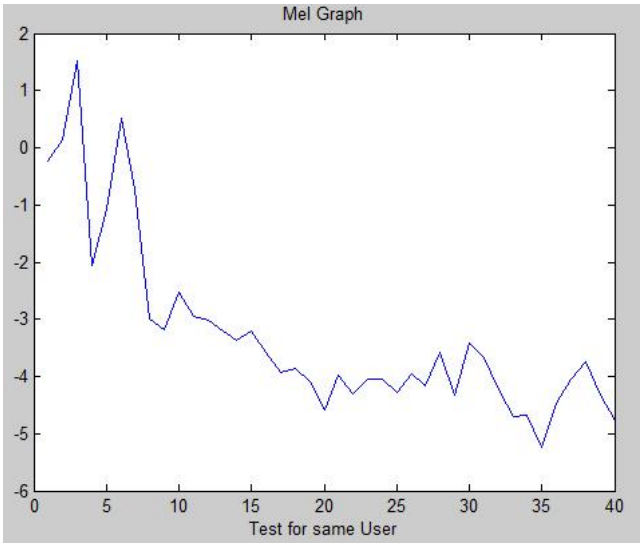

Fig 7: Mel Graph of voice signal of USER 1

It is clear from the Mel graph that USER 1 is authorized to access the system.

Case b) when the user is unauthorized

- Original voice signal of USER 2 for word "**SILVY**"

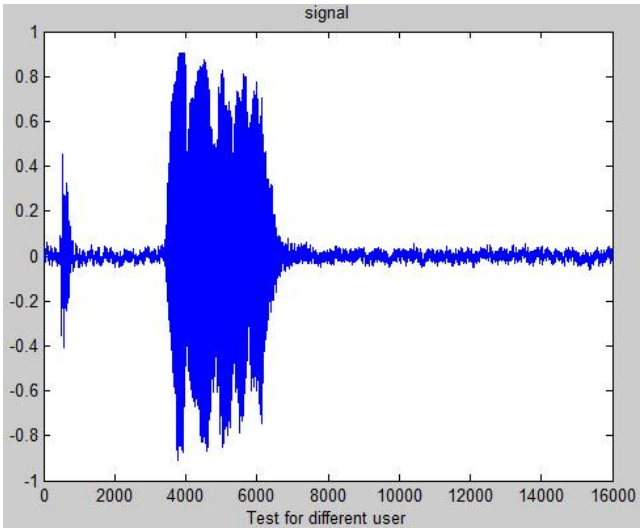

Fig 8: Voice Signal of USER 2

- After all process Mel Graph for USER 2 voice signal

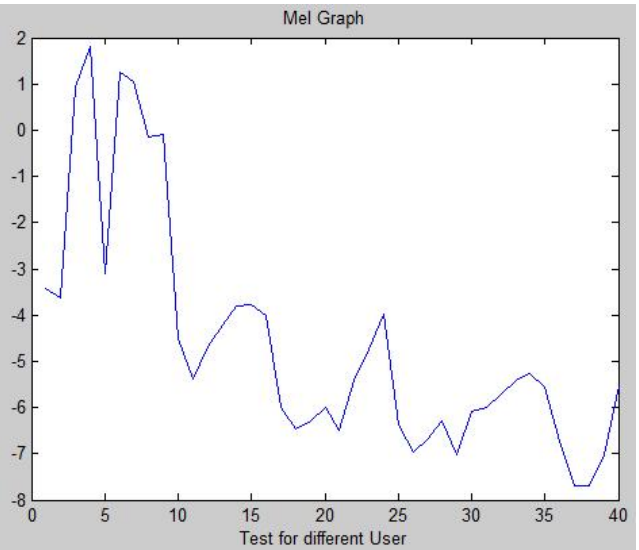

Fig 9: Mel Graph of voice signal of USER 2

Again it is evident from the Mel graph that the USER 2 is not authorized to access the system.

VII. CONCLUSION

The developed system is will check for the speech as well as the voice of the user of system. To gain the permission, both the conditions should be satisfied. Test for speech refers to check what is said. Test for voice refers to check who said. The developed system offers 70% of accuracy.

VIII. FUTURE SCOPES

The results can become more specifically accurate by using vector quantization and dynamic time wrapping.

IX. REFERENCE

1) Dr. Sayleegharge, Prof. Manisha Joshi, Bhaumikdoshi and Varun Sharma, "Gesture & Speech based appliance control," Int. Journal of Engineering Research and Applications, vol. 4, pp. 298-302, Jan. 2014.

2) Ibrahim Patel and Dr. Y. Srinivas Rao, "Speech recognition using HMM with MFCC-an analysis using Frequency spectral decomposition technique," Signal & Image processing: An International Journal, vol. 1, pp. 101-110, Dec. 2010.

3) Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithm using MFCC & DTW techniques," Journal of computing, vol. 2, pp. 138-143, Mar. 2010.

4) Koustav Chakraborty, Asmita Talele and Prof. Savitha Upadhya, "Voice Recognition using MFCCAlgorithm," International Journal of Innovative Research in Advanced Engineering, vol. 1, pp. 158-161

5) http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

About Authors:

Silvy Achankunju is pursuing M Tech – II year in VLSI & Embedded System from Acropolis Institute of Science & Technology, Indore. She has 1 year teaching experience. Her publication was on Voice based Security system using MATLAB & Embedded System in International Journal of Scientific Research, May-2015.

Chiranjeevi.Mondikathi is an Assistant Professor in Acropolis Institute of Technology and Research, Indore. He has completed his M. E. in Digital Communication from S.G.S.I.T.S College Indore. He has 3.5 years of teaching experience. He has presented paper in IACQER on "Face Recognition for Single and Different facial expressions, in IJERA on "Performance Evaluation of Two Reactive Routing Protocols, Embedded Systems (Automation of Car) and was awarded first prize in Gitam University, Nano Technology" in "Audisankara Engineering College.