International Journal For Technological Research In Engineering
Volume 1, Issue 10, June-2014 *ISSN (Online): 2347 - 4718*

HOME SECURITY SYSTEM USING VOICE RECOGNITION

Ashwini Sanjay Garud[1], Rucha Arvind Dhamdhere[2], Dr. A.D.Rahulkar[3]
B.E. Instrumentation and Control Engineering
A.I.S.S.M.S's Institute of Information Technology
Pune, India.

Abstract: This paper presents the automatic operation of the door based on voice recognition. In this paper Mel Frequency Cepstral Coefficient (MFCC) are used to extract the discriminant features from the voice signal. Next, the Euclidean distance is used to compute the difference between two different voice signatures. Based on this distance the door is operated only if the distance is within specified range. The developed system can be useful for the physically handicapped person.
Keywords: Mel Frequency Cepstral Coefficient (MFCC), Euclidean Distance, Matrix Laboratory (MATLAB), Discrete Cosine Transform (DCT), Fast Fourier Transform (FFT), Hamming Window. Component.

I. INTRODUCTION

For this project we are using voice characteristic of a person that is unique to the individual. Hence to identify this uniqueness we are using Mel Frequency Cepstrum Coefficient (MFCC) algorithm. Earlier Linear Prediction Coefficients (LPC) algorithm was used to identify characteristics such as pitch, frequency. But now MFCC is used because the results that are obtained are more accurate, high efficiency, extraction of best parametric representation of acoustic signals, cannot perceive frequencies over 1kHz.

II. MEL FREQUENCY CEPSTRUM COEFFICIENT PROCESSOR

Since there is a need to extract the features of the speaker's voice and identify and differentiate between the pitch, gender, emotions, background noise etc. thus MFCC is used so as to obtain complete features and identify the authorized user.

Following steps are to be performed on the voice signal under MFCC algorithm:

1. Framing
2. Estimation of power spectrum.
3. Application of Mel filters
4. Summing the all energies in filter banks
5. Logarithm of all energies
6. Taking Discrete Cosine Transform
7. Generation of matrix containing Mel coefficients.

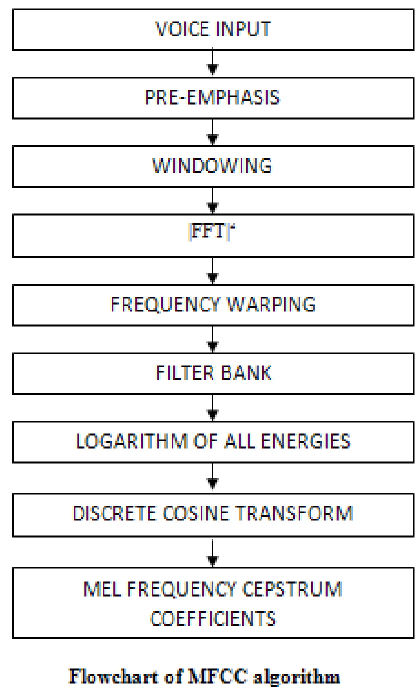

Fig. 1: Flowchart of MFCC algorithm

III. ALGORITHM

A. Speech Waveform: The speech waveform acts as an input signal to the MATLAB programmed personal computer or laptop.

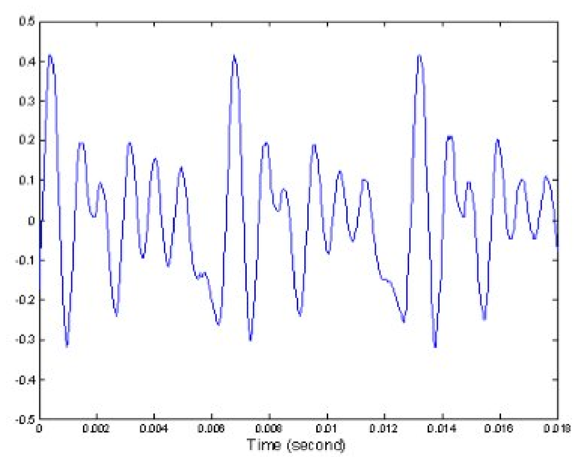

Fig. 2: Speech Waveform

B. Pre-emphasis stage: At this stage, the sampled speech waveform is passed through a filter which emphasizes higher frequencies. This process of emphasizing higher frequencies increases the energy of the signal at higher frequency.

C. Windowing /Framing: We carry out the process of windowing to remove unwanted signal and divide the signal into over-lapping segments for accurate measurement. Framing is important stage in the algorithm. It is nothing but the sampling of signal. The frame of specified time slot is to be made from the voice signal since it is continuously changing. The time interval is restricted within 20ms to 40ms. Frames shorter than 20ms do not give enough reliable samples within the particular frame and frames longer than 40ms yield the signal which changes too much within the frame.

D. |FFT|²: We have used Hamming window to reduce the unwanted signals. Fast Fourier Transform is used to convert each frame from time domain to frequency domain.

E. Mel-Frequency Warping: Before we go to Mel-Frequency warping we need to understand what is Mel scale. As we know humans beings are good at differentiating changes at low frequencies than at high frequencies, so Mel scale is needed to differentiate between the perceived frequency, and pitch, of a pure tone to its actual measured frequency. The advantage of Mel scale is that it matches the frequency more closely than what humans hear.

F. Filter bank: At this stage spectrum is segmented into a number of critical bands which overlap the triangular filters.

G. Logarithm of all energies: Power spectrum is the distribution of the energy of a waveform among its different frequency components. Unlike linear scale logarithmic scale displays the value of quantity in terms of orders of magnitude. This operation is performed on all the energies to compress their ranges. The logarithm uses cepstral mean subtraction which is a channel normalization technique

H. Discrete Cosine Transform (DCT): We use DCT to convert the log Mel cepstrum to time domain.

IV. OBSERVATION S AND OBTAINED GRAPHS

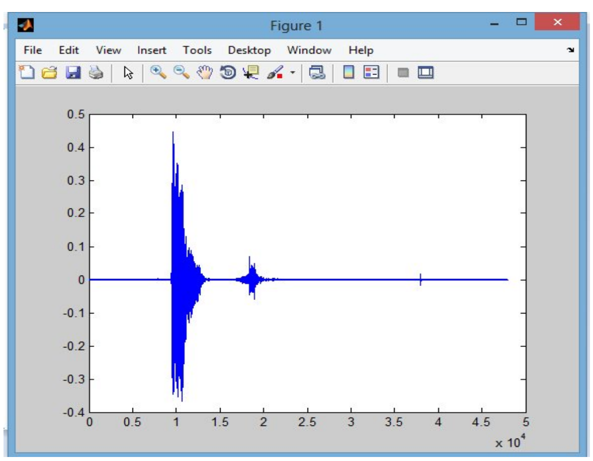

Fig. 3: This figure is achieved when the authorized user says the word “OPEN” in the main file.

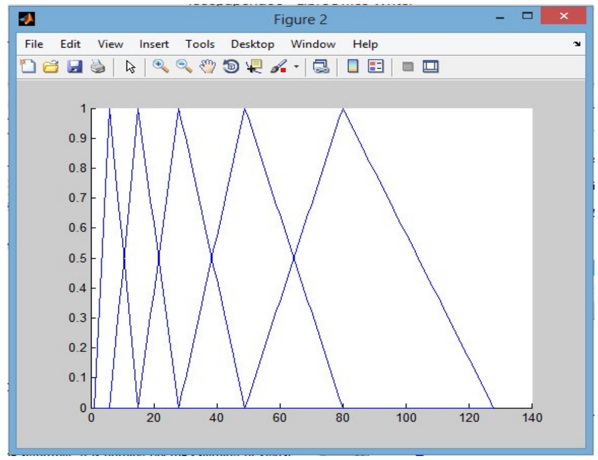

Fig. 4: This figure is achieved when the user says “OPEN” and filter banks are generated.

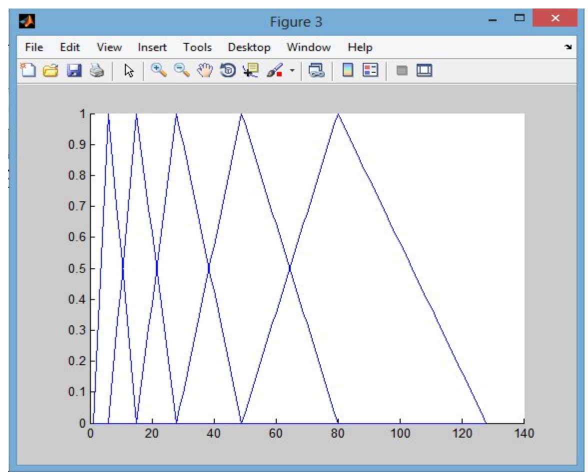

Fig. 5: This figure is achieved when user says “OPEN” and thus filter banks are generated.

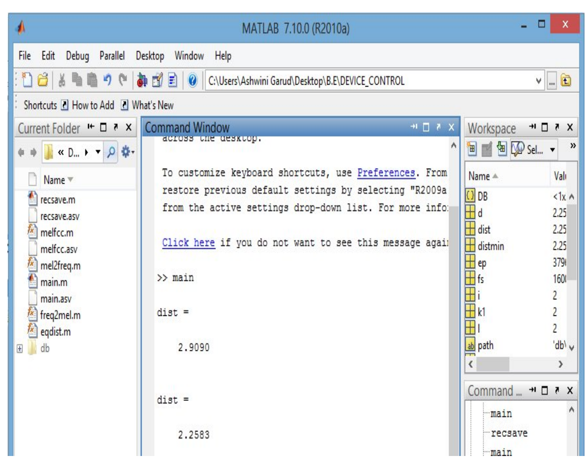

Fig. 6: The following MFCC coefficients are obtained when the word “OPEN” is uttered.

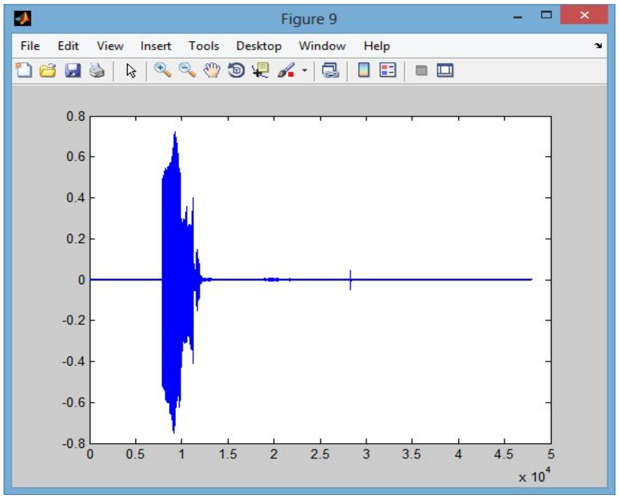

Fig. 7: The following graph is obtained when the user says "CLOSE".

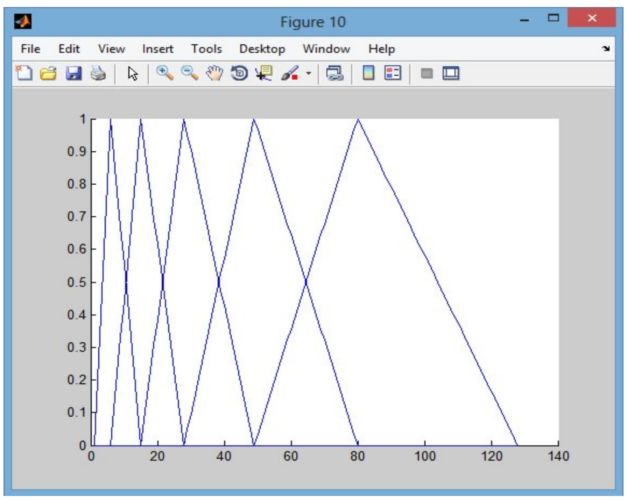

Fig. 8: This figure represents the filter banks which are generated when the user utters the word "CLOSE".

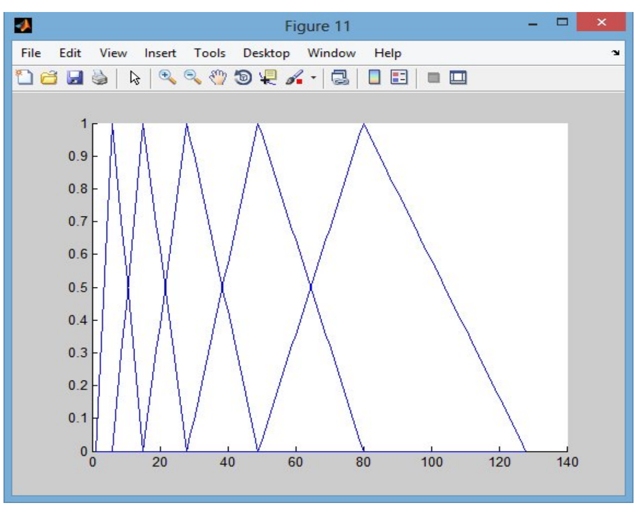

Fig. 9: This figure is generated when the user utters the word "CLOSE"

The figure confirms that the input voice is matched with the reference voice which was stored in the database. The finding of this study is consistent in the sound proof environment. This process affects greatly because of the external noise. The accuracy of this automatic door operating system on voice recognition using MFCC algorithm is 75%.

V. ACKNOWLEDGMENT

It is our great pleasure in expressing sincere and deep gratitude towards our guide Dr. A.D. Rahulkar Professor in department of Instrumentation and Control for all his valuable guidance and constant support throughout the course of our project and help for trying out different ways. We have been privileged to work under his supervision, and we truly appreciate his help. We take this opportunity to thank our Head of Department, Mr. H. P. Chaudhari and our principal Dr. P. B. Mane for their valuable support. Last but not the least we thank Mr S. C. Rajgade, Mrs D. U. Shinde, Mrs A. A. Shinde, Mrs V. V. Kale, Mrs G. S. Ingle, Mrs S. V. Kulkarni, Ms D. P. Inamdar, Ms R.D. Valsangkar, Mrs N. B. Mane, Mrs M. M. Tare, all my friends, our family and other non-teaching staff who helped us directly or indirectly with our project. take this pleasure to acknowledge with gratitude to our project guide Dr. A. D. Rahulkar, Department of Instrumentation And Control Engineering, for his expert opinions and innovative suggestions, and encouraging us to work har towards achieving the aim and objective of the project throughout the period. I would like to extend my gratitude to thank Head of the Department Mr. H. P. Chaudhari and all staff members for genuine support and encouragement.

REFERENCES

[1] L.R. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, N.J., 1993.

[2] L.R Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, N.J., 1978.

[3] S.B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences", IEEE Transactions on Acoustics, Speech, Signal Processing, Vol. ASSP-28, No. 4, August 1980.

[4] Y. Linde, A. Buzo & R. Gray, "An algorithm for vector quantizer design", IEEE Transactions on Communications, Vol. 28, pp.84-95, 1980.

[5] S. Furui, "Speaker independent isolated word recognition using dynamic features of speech spectrum", IEEE Transactions on Acoustic, Speech, Signal Processing, Vol. ASSP-34, No. 1, pp. 52-59, February 1986.

[6] S. Furui, "An overview of speaker recognition technology", ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 1-9, 1994

