Speaker Recognition Based Home Automation Using Matlab

Shilpa Khandade[1], Sucheta Khot[2]

[1]PG Student, Department of Electronics B.V.C.O.E.W. Pune-43, Maharashtra, India
[2]Department of Electronics B.V.C.O.E.W. Pune-43, Maharashtra, India

Abstract— *Due to decline in both physical and mental abilities, some elderly are not allowed to leave the bed without assistance. Some time they are unable to make the desirable bodily movements and repositioning. In this paper the home automation is obtained using MATLAB based speaker recognition. The feature extraction of speech signal is done by using MFCC and for selection of features of speech signal vector quantization is used. By using above two steps the speaker is recognized and then this is given to the microcontroller by using serial communication .Then the particular home appliance get operated.*

Keywords— Bodily movements ,Speech recognition ,Speaker recognition, FFT , DCT , MFCC ,Feature extraction.

I. INTRODUCTION

Underlying disease in older adults gives problems in mobility [1] .This disease is recognized by geriatricians[2].elderly person mostly suffer from unwanted falls while trying move from the bed without caregiver attendance .They are often not able to make the movements of their body so all time they need caregiver but some time it is not possible .The home automation system will help the patient and caregiver also .If care giver is not available at some time then the patient himself can manage the operation of home appliances .For human beings speech is the most common and vital way of communication. In speech recognition we are going to recognize certain words spoken by the patient by using individual information which is included in speech waves. Speech recognition is the analysis side in speech processing. The synthesis side might be called speech production. These two taken together allow computers to work with spoken language.

II. RELATED WORK

It is very helpful to use automated monitoring of rehabilitation exercises[3].It is also suitable for live captioning of speech live translation (speech-to-text and/or multilingual), replacement of Television remote control and other home automation systems, and in-vehicle applications [4].Automatic speech recognition is particularly suitable for pocket-size consumer devices like Smartphone[5].Hidden Markov Model (HMM) is a widely established method for classification . B. System Design and Implementation MFCC algorithm is used for speech recognition[5].The speech recognition has two steps which are Voice capture and Voice recognition .Voice capture step uses the microphone present in front of mouth of patient for capturing speech commands. Then the received commands are converted into binary codes according to its frequency .Then they are compared with predefined commands stored in the microcontroller. The Voice recognition step is used for comparing the converted binary codes with the one which is stored in the microcontroller[6].

III. MOTIVATION

If there is a bedridden patient in home it requires a caregiver who ideally observes the patient around the clock to prevent from bedsores. The caregiver has to provide a high degree of attendance to the elderly all the time. Also the knowledge and personality of caregiver affects the quality of nursing care. Lack of care taken by human caregivers causes unfortunate consequences. The patient and his family member both may suffer from these consequences.

IV. METHODOLOGY

1. Methods of Speech Recognition

1.1 Spectrum Normalization: In spectrum normalization the spectrum is normalized first by using linear normalization .Then the values of normalized spectrum are set to the interval[0,1] . After this observe the spectrum values of previously recorded signals and then find the difference between the spectrum values by comparing the test and target values[7].

1.2. Cross-Correlation: Assume that the signal and spectrum of recorded speech signal of same word are same .Find the cross-correlation of the signals then find maximum value position of the cross-correlation .Then find the difference of values right to maximum value and left to maximum value. Then take absolute of this difference find mean square error of the absolute value

.The mean square error of trained and target words should be minimum to match the speech signal[8].

1.3. Autocorrelation: Autocorrelation of the signal is nothing but computing cross-correlation of the signal with itself instead if computing with another signal .Autocorrelation finds the selfcorrelation of the signal .Thus training signals find the minimum difference between autocorelations[18].

1.4 Hidden markov model(HMM): In HMM the probabilities of state transition matrix and emission matrix are time independent.HMM is described by a vector and matrices[9].

2.MFCC Based Speaker Recognition

Speaker recognition has two phases .First phase is training phase (Fig 1) and second phase is testing phase (Fig 2). In the training phase the MFCC features of the input speech are extracted and that are stored in the database as a reference .In the testing phase the MFCC features of testing speech signals are extracted and then that are compared with the features stored in database .By using the given decision logic the decision is given at the output.

Fig. 1 :Training Phase

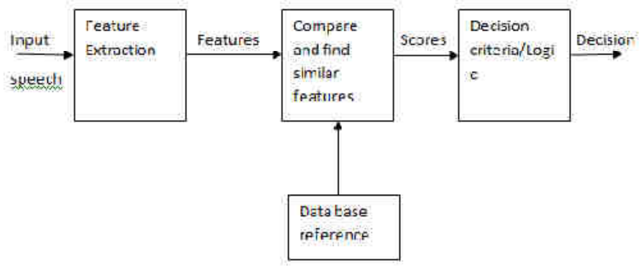

Fig. 2 :Testing Phase

3.MFCC feature extraction (Fig 3)

The first block in the MFCC feature extraction is take the input speech then it is given to pre-emphasis block to increase the energy of signal at higher frequency .In the framing the signal is segmented into small duration blocks. In the windowing each frame is multiplied with a hamming window to keep continuity of the signal.FFT converts signal from time to frequency domain to get the magnitude frequency response of each frame. The mel filter bank is used to get smooth magnitude spectrum. It normally uses triangular filter bank.DCT is the frequency to time domain conversion of a signal .When we apply DCT on log energy it gives different mel scale cepstral coefficients. Delta energy is used to find the velocity and acceleration of energy with MFCC.The mel scale is defined as (eqn.1)[11].

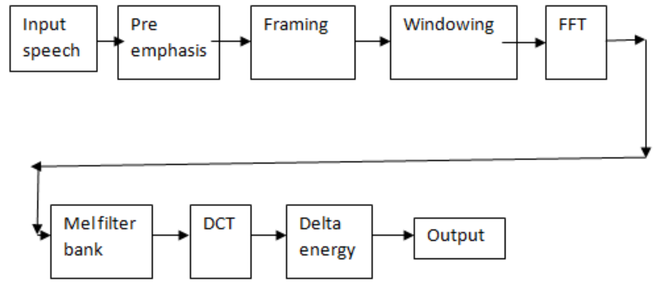

Fig.3: Flow of MFCC Feature Extraction

$$f_{mel} = 2595\log_{10}(1+f/700) \qquad (1)$$

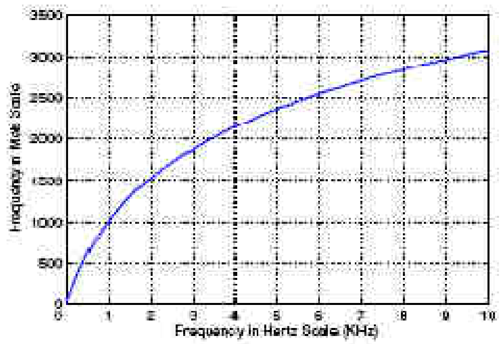

Fig.4: Relation Between Hertz and Mel Scales[11]

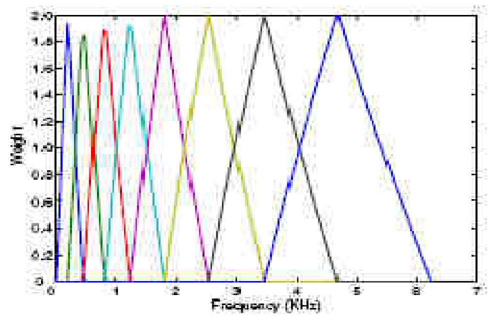

Fig.5: Mels Frequency Filter Bank The Vertical Axis Represents The Weights Of Filter Coefficients[11]

Mel spectrums are obtained by applying the triangular filterbank to the speech signal.Then the mel spectrums are transformed back into the time domain using DCT which gives MFCC coefficients . if we denote the Mel power spectrum coefficients by *Sk* (where *k* is the index of the Mel-spaced filters and *k* = 1, 2, . . .,*K*) then the MFCC coefficients (*Cn*) are calculated as[11]

$$C_n = \sum_{K=1}^{k} (\log S_k) \cos[n(k-1/2)\prod/k] \;; n=1\ldots..L \qquad (2)$$

MFCC coefficients are represented in the form of MFCC feature vector as[11]

$$C=[C_1,C_2,C_3,\ldots\ldots\ldots\ldots,C_L] \qquad (3)$$

4.Vector Quantization

Vector quantization is used for the selection of features. It takes large no of feature vectors as input and gives small no of feature vectors as output. Storing each single feature vector generating from the training phase is not possible. So the probability distribution of feature vectors is done and it is possible by quantizing each feature vector to relatively small numbers of template vectors.

5.Algorithm of Speech Recognition

5.1 Create database using MATLAB

1. Select the current directory path
2. Create train database folder in the database path
3. Loop the voice samples From1: n
4. For i=1

   Sampling frequency -16000Hz
   Recording time is 3 seconds
   Use wave record function to record the sample
   Playback the wav file
   i+1

5.2 Speech recognition using MATLAB

1. Select the current directory
2. Open the train database folder
3. Set sampling frequency=16KHz, and data bits-8
4. Load the wav samples one by one
5. Apply MFCC algorithm to the samples
6. Store the melcepts in cc1 variable

5.3 Read wav file function

1. Set Sampling frequency -16000Hz
2. Recording time 3 seconds
3. Use wave record function to record the sample
4. Playback the wav file
5. Store the live wav file
6. Apply MFCC function to the live samples
7. Store the melcepts in CC variable
8. Read sample1
9. Use distance measure function for CC1 with CC
10. Find distance for all samples
11. Compare the distance between CC1 to CC
12. Display the sample with minimum distance.

V. RESULTS

1.Fig.4 to Fig.7 are of training phase and Fig.8 to Fig.13 are of testing phase.

Fig. 4 :GUI

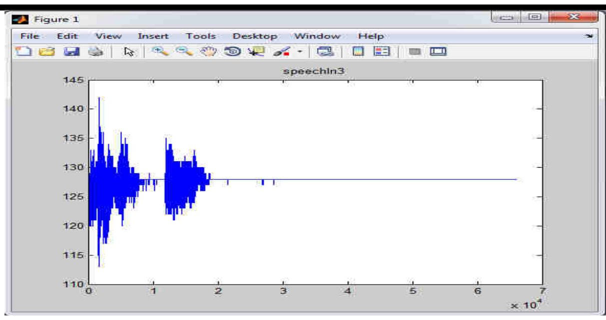

Fig. 5: Input Speech Signal

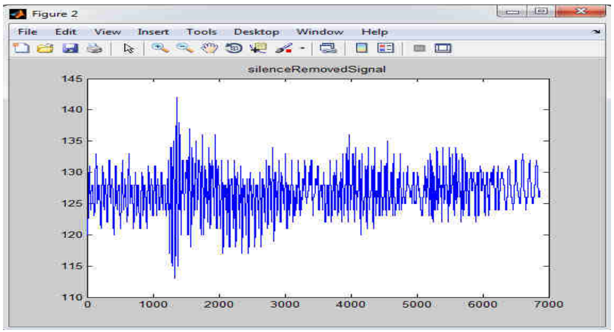

Fig. 6: Silence Removed from Input Speech Signal

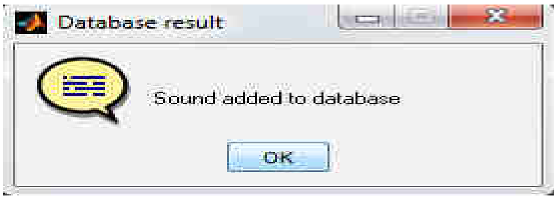

Fig. 7 :Input Signal Added to Database

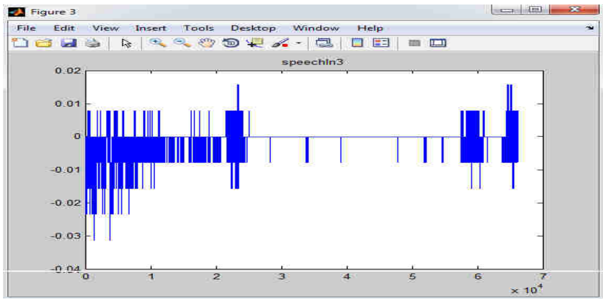

Fig. 8 :Input Speech Signal from Same User

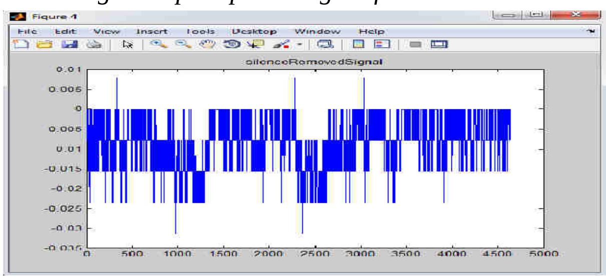

Fig. 9 : Silence Removed from Input Speech Signal

Fig. 10: Recognized User ID

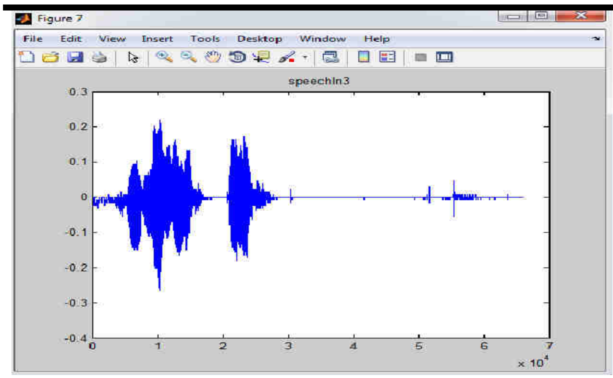

Fig. 11: Speech Input for Recognition from Different User

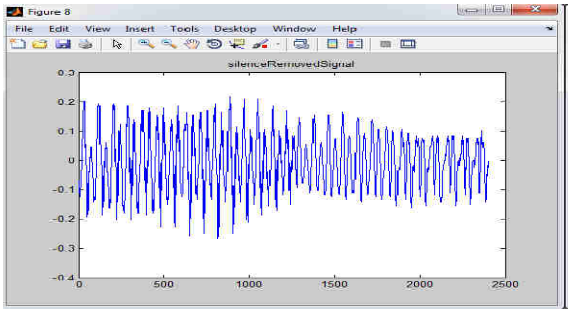

Fig. 12: Silence Removed from Input Speech Signal

Fig. 13: Gives Warning as a Wrong User

Fig. 14: Whole System

In the above figure fan and bulb these two home appliances are connected at the output .If we say 'fan on' through the microphone then the MATLAB will recognize speech and speaker then according to that specific commands are given to the microcontroller and the fan will get on .The above steps will be repeated for fan off , bulb on and for bulb off .

6.While Performing Tests Of The Systems Various Parameters Are Calculated As

TABLE I.PARAMETERS OF THE SYSTEM

Parameters	Time
Average enrollment time required	8 sec
Average time required to store data	4 sec
Average time required to recognize input data	4 sec
Efficiency	84%

TABLE II. FA,FR Values

Sr.No.	Person No.	No.of Attempts	FA	FR
1	P1	5	0	0
2	P2	5	0	2
3	P3	5	0	2
4	P4	5	0	2
5	P5	5	0	1
6	P6	5	0	1
7	P7	5	0	1
8	P8	5	0	0
9	P9	5	0	0
10	P10	5	0	0
11	P11	5	0	0
12	P12	5	0	1
13	P13	5	0	0
14	P14	5	0	2
15	P15	5	0	0
16	P16	5	0	0
17	P17	5	0	2
18	P18	5	0	0
19	P19	5	0	0
20	P20	5	0	0
21	P21	5	0	0
22	P22	5	0	1
23	P23	5	0	1
24	P24	5	0	3
25	P25	5	0	1
Total	-	125	-	20

FA : FAULT ACCEPTANCE , FR : FAULT REJECTION

VI. CONCLUSION

Speech and speaker recognition can be used for home automation. It helps the elderly patients who are unable to do bodily movements .They can easily operate home appliances without leaving their seats .Speech signal of human is nonlinear in nature. So MFCC are derived using the logarithmically spaced filter bank and it gives the better results.

ACKNOWLEDGEMENTS

I would like to thank my all the staff members of BVCOEW, Pune for being moral support through the period of my project study whose help and shared knowledge is the main part of my project.

REFERENCES

[1] S. Bennett, Z. Ren, R. Rockwood, and F. Knoefel ,"In-Bed mobility monitoring using pressure sensors", IEEE Trans. Instrumentation And

Measurement, Vol. 64, No. 8,pp.2110-2120, Aug. 2015.

[2] C. MacKnight and K. Rockwood,"Research analysis of the hierarchical assessment of balance and mobility (HABAM)", J. Clin. Epidemiol., vol. 53, no. 12, pp. 12421247, 2000.

[3] M. Huang, J. J. Liu, W. Xu,N. Alshurafa, X. Zhang, and M. Sarrafzadeh,"Uusing Pressure Map Sequences for Recognition of On Bed Rehabilitation Exercises *",IEEE Journal of biomedical and health informatics*, Vol.18, No. 2, pp.411-418 March 2014 .

[4] S. Ahn and H. Ko,"Background noise reduction via dual- channel scheme for speech recognition in vehicular environment ", IEEE Trans. Consumer Electron., vol. 51, no . 1, pp. 22 - 27, 2005.

[5] Steven J. Anderson, A. C. M. Fong, Jie Tang,"Robust tri-modal automatic speech recognition for consumer applications", IEEE Trans. on Consumer Electronics, Vol. 59, No. 2, pp.352-360, May 2013.

[6] M.Senthil Sivakumar , Jaykishan Murji, Lightness D Jacob, Frank Nyange, M.Banupriya,"Speech Controlled Automatic Wheel Chair", Pan African International Conference on Information Science, Computing and Telecommunications,pp.70-73,2013

[7] Buera, Luis, et al. ,"Cepstral vector normalization based on stereo data for robust speech recognition", IEEE Trans. on Audio, Speech, and Language Processing,pp, 1098-1113,2007.

[8] Chen, Jingdong, J. Benesty, and Y. Huang," Robust time delay estimation exploiting redundancy among multiple microphones", IEEE Trans. On Speech and Audio Processing,pp.549-557,2003.

[9] Varshney N & Singh.S, " Embedded Speech Recognition ystem", International Journal of Advanced Resear in Electrical, Electronics and Instrumentation Engineering ,pp.9218-9227 2014.

[10] D. Vijayasenan, F. Valente, and H. Bourlard,"An Information Theoretic Combination of MFCC and TDOA Features for Speaker Diarization" ,IEEE Trans. on Speech and Audio Processing,VOL. 19, NO. 2,pp.431-438, Feb.2011.

[11] Ishwar S. Jadhav et al, " Human Identification using Face and Voice Recognition," International Journal of Computer Science and Information Technologies, Vol. 2 (3) , pp. 1248-1252,2011.