ISSN 2395-1621

Home Automation Using Speech Processing

#1Ghorpade Sandhya, #2Nigade Pooja, #3Nikam Kiran,
#4Dr. Y. S. Angal

2poojanigade08@gmail.com

#123Department of E&TC,
#4Head, Department of E&TC

JSPM's BSIOTR, Pune, India.

ABSTRACT

This paper presents the development of home automation system based on voice command using Speech recognition. We are motivated to build a system which provide the support to elderly and disabled people at home and to create an automated and comfortable home environment. This system implemented using hardware and software that react to the human voice commands. Mel Frequency Cepstral Coefficient (MFCC) feature extraction technique along with the Vector Quantization (VQ) model formation technique is implemented to increase the accuracy of the system. Euclidean distance Classifier needs the minimum computation. The proposed system contemplate to control electrical appliances with relatively user-friendly interface and provide comfort for installation.

Keywords-- Mel-Frequency Cepstral Coefficient (MFCC), Vector Quantization (VQ).

ARTICLE INFO

Article History

Received: 3rd May 2017

Received in revised form :

3rd May 2017

Accepted: 8th May 2017

Published online :

13th May 2017

I. INTRODUCTION

Speech processing is one of the massive cultivate research areas in signal processing .Each year billions of pounds are being spent on supporting research in speech processing. The extreme aim of this research is to produce an interactive man machine communication.

The home automation system (HAS) concept has existed in the emerging market and becoming popular nowadays. Using Speech processing we can develop an automated system capable of controlling operation of various devices placed inside the home or office using voice command, so that respective user can handle, simply, such devices with secure authentication [1].

Speech Recognition is a technology which permitting the computer to identify and understand words spoken by a person using a microphone or telephone. Using a set of pre-programmed commands and instructions, user can talk with computer. Computer system that understands input speech enables user to have dialogue with the computer. User and the computer speaking as commands or in response to events, input, or other feedback would be included in these conversations. Speaking is easier and more sensitive than selecting buttons and menu items. Human speech has changed over many thousands of years to become an efficient method of sharing information and giving instructions [2].

II. METHODOLOGY

The home automation system using speech recognition system to make easy to elderly and physically challenged people with an easily operated home automation system that operates fully on voice commands. One of the major issue in now day society is wastage of energy, whereby energy consumption is continuously increasing year by year. The functional block diagram of the proposed system is shown in Figure.

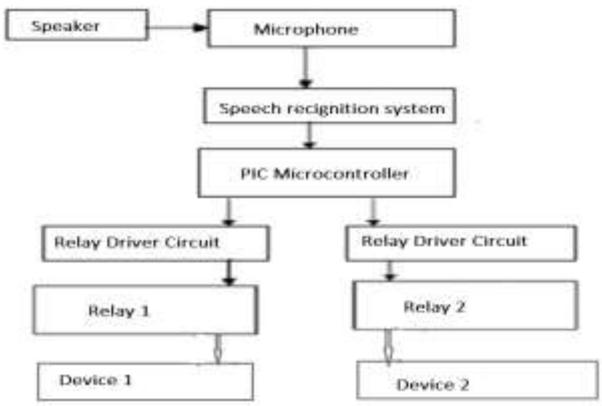

Fig.1 System block diagram

© 2017, IERJ All Rights Reserved

The speech input from microphone is given to the PC. Upon successful recognition of voice command the microcontroller actuates the corresponding electrical device like turning on lights, and TV and other electrical device using the relay module .In this system, we used two relays which are use for controlling switch on or switch off AC device and one relay to control the TV. Normally open circuit concept has been implemented in the voice control home appliances with application.

A. Main steps in speech recognition

Training phase:

1. Pre-processing of training speech samples.
2. Framing and windowing
3. Extracting appropriate speech features (feature vectors) from each windowed and frame.
4. Building a model for each word from feature vectors.

Testing phase:

1. Pre-processing of training speech samples.
2. Framing and windowing
3. Extracting appropriate speech features (feature vectors) from each windowed and frame.
4. Obtain Euclidean distance (or probabilities) of feature vectors with each word model
5. Decision based on minimum distance (or maximum probability).

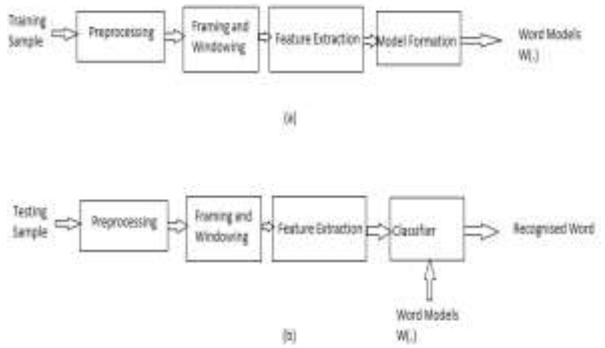

Fig.2 Main steps in Speech recognition a) Training phase b) Testing phase

PRE-PROCESSING

The objective of pre-processing step is to convert the sampled speech signal into a form suitable for feature extraction. This step consists basically of the end point detection (also called voice activation detection) and pre-emphasis.

a) End-Point Detection: An important problem in speech processing is detecting the presence of speech in a background of noise. This problem is often referred to as the end point location problem. The correct detection of a word's start and end point means that following processing of the data can be kept to a minimum. A major cause of error in isolated-word automatic speech recognition system is the inaccurate detection of the beginning and ending boundaries of test and reference pattern. It is essential for automatic speech recognition algorithm that speech segments be reliably separated from non-speech. The reason requiring an productive end pointing algorithm is that the computation for the processing the speech is minimum when the end points are accurately located [3 -6].

b) Pre- emphasis: The digitized speech signal is put through low order digital system (typically the first Order FIR filter) to spectrally flatten the signal to make it less susceptible to fine precision effects later in the signal processing. The digital system used in the pre-emphasizer is either fixed or slowly adaptive. Most widely used pre-emphasis network is the fixed order system.

In this case, the output of the pre-emphasis network is related to the input to the network by the difference equation. The most common value for is around 0.97.The pre-emphasis filter boost the signal spectrum approximately20dB/decade.

c)Enhancement: If the signal is corrupted with the some noise, it is necessary to remove the noise from the signal because the low Signal to Noise Ratio (SNR) degrades the system performance rapidly .There are several enhancement methods likes spectral subtraction ,spectral mean normalization and power spectrum difference etc. This enhancement step makes the final extracted feature more robust to the noisy condition.

FRAMING AND WINDOWING

In this step the pre-emphasized speech signal is blocked in to overlapping frames. Speech is quasi-stationary signal. If we consider a speech segment of length 10 msec. to 30 msec, it is almost stationary. By applying the frame blocking the signal is divided into frames of N samples and with adjacent frame being separated by M samples. The frames are generally overlapping in nature to avoid the blocking effect. This overlap is from 50% to 70 %. Next is the application of window on each frame to reduce signal discontinuities at either end of the block.

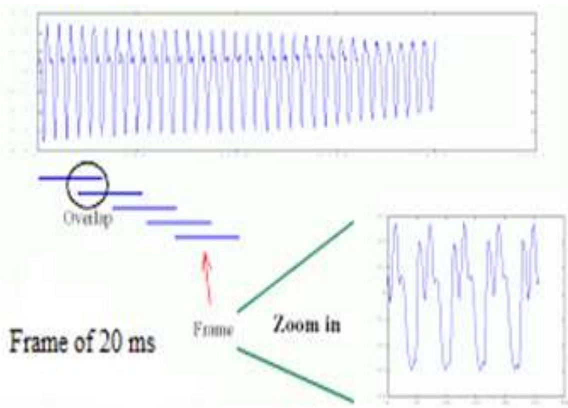

Fig. 3 The windowing process, showing the frame shift and frame size, assuming a frame shift of 10ms, a frame size of 20 ms, and a rectangular window.

The extraction of the signal takes place by multiplying the value of the signal at time n, s[n],with the value of the window at time n, w[n]

© 2017, IERJ All Rights Reserved

$$y[n]=w[n]s[n]$$

A typical window is the hamming window, which contract the values of the signal toward zero at the window boundaries, to keep away interruption.

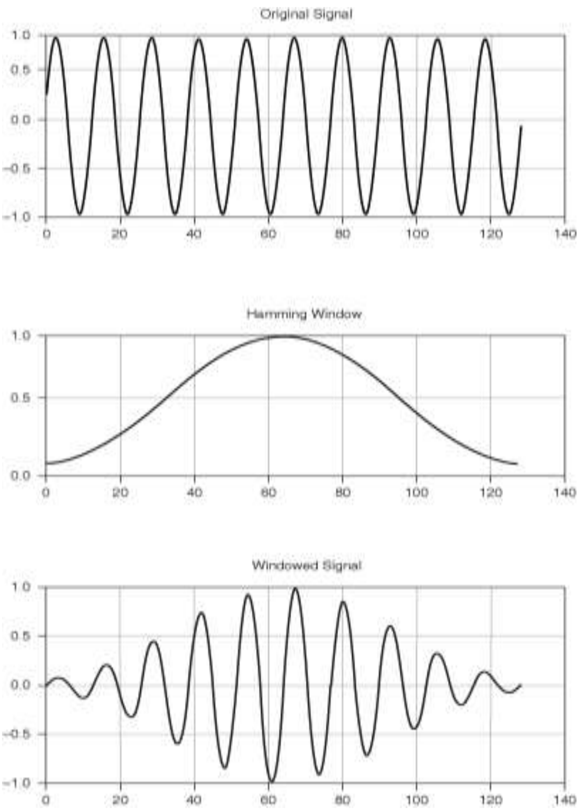

Fig. 4 Windowing technique

The equation is as follows (assuming a window that is L frames long):

$$w[n] = 0.54 - 0.46\cos(2\pi n/L) \qquad 0 \leqslant n \leqslant L-1$$
$$0 \text{ otherwise}$$

FEATURE EXTRACTION

In speech recognition systems, feature extraction and recognition are two important modules. The primary goal of feature extraction is to find robust and particular features in the acoustic data. The recognition module uses the speech features and the acoustic models to decode the speech input and produces text results with high accuracy. A number of speech feature extraction methods have been proposed, such as linear predictive cepstral coefficients (LPCCs), Mel-frequency Cepstral coefficients (MFCCs) and perceptual linear predictive coefficients (PLPs) [2]. In temporal analysis the speech waveform itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis [3]. The main goal of feature extraction step is to compute a saving sequence of feature vectors providing a compact representation of the given i/p signal.

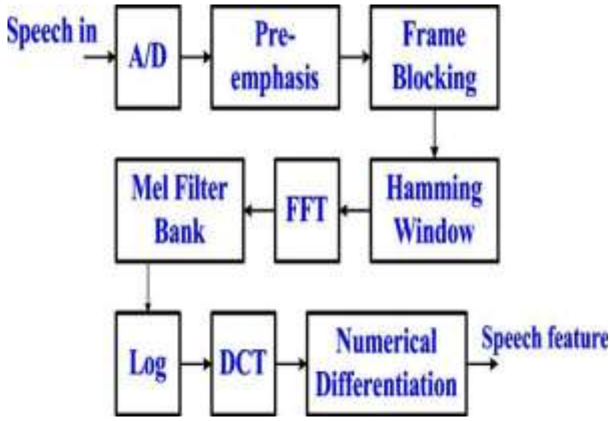

Fig.5 MFCC processor

Mel Frequency Cepstral coefficient: It is a perception based method. MFCC feature extraction approach gives a good discrimination and a small correlation between components. Characteristics of the slow varying part concentrated in the low cepstral coefficients. The results of the FFT will be information about the amount of energy at each frequency band. During MFCC computation the intuitions is implemented by creating a bank of filters which collect energy from each frequency band. Finally, we take the log of each of the mel spectrum values.

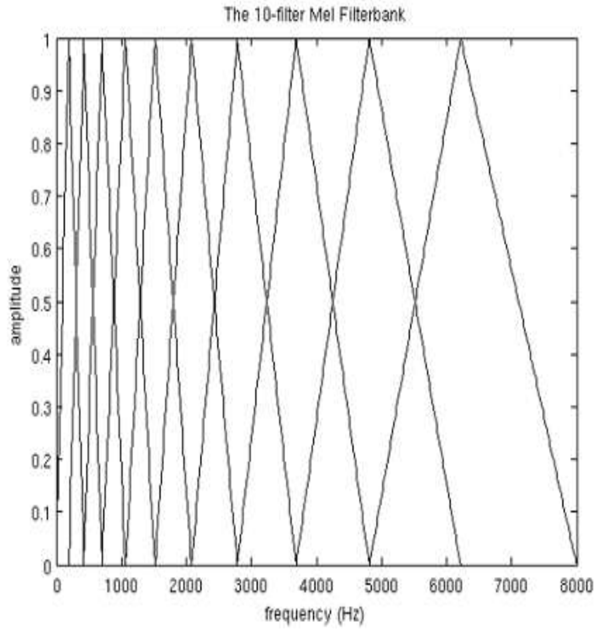

Fig.6 Mel filter bank

MFCC are also increasingly finding uses in music information such as genre classification, audio similarity measure etc.

MODEL FORMATION

Signal modeling represents the process of converting sequences of speech samples to observation vectors representing events in a probability space. Vector quantization (VQ) is a classical quantization technique from signal processing which allows the modelling of probability density function by the distribution of

© 2017, IERJ All Rights Reserved

prototype vector. It was originally used by Dudley for data compression and speech coding in 1960's and smith in the 1960's.VQ algorithm proposed by Linde et al. was used for IWR by Bush et. Al. and Burton et al.It works by dividing a large set of points (vectors) into groups having approximately the same no. of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms with minimum distortion to make a speech specific VQ codebook. The need to reduce to competition and memory at nearest neighbour is the cheap motivation behind VQ modeling. This approach also avoids the problems of segmenting speech into meaningful subunits .Research in VQ focused on the methods for generating the codebook, the type of distortion measure and efficient structure achieve high rate low distortion VQ [9, 10].

III.DATABASE

Voice samples are used as the database .We have collected voice samples of 10 speakers (5 boys and 5 girls) and 200 samples of each speaker. These samples are taken in two phases; first is Training phase and another is Testing Phase. For example for single user we have taken 10 samples of each word for training phase and 10 samples of same words for testing phase that means we require 100 samples for training and 100 for testing phase in total. In model formation training samples are compared with testing samples. So it is needed to take large number of samples of different speakers in order to increase the accuracy of a system. Normally, human voice signal frequency range is 3 KHz to 20 KHz. But in our project application area is home environment so we select the sampling rate 8000 KHz and 32 bits resolution for high accuracy.

Record And Save

Speech sample recording can be done by using two methods First is by simulating the program for recording the sample in MATLAB simulation tool.

Secondly by using the speech recording softwares. We used Cool Edit V5 and Sound Forge V11.0.24for taking samples. This software provides recording of voice samples in various sampling rate, resolution and noise reduction tools.

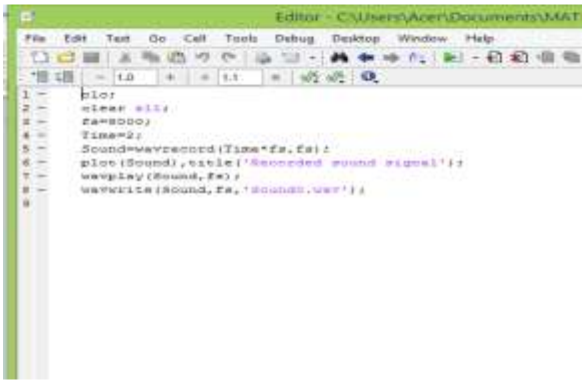

Fig. 7 Matlab code to record speech sample

Above figure shows the matlab program for recording a speech sample. After simulation we get the recorded sound signal in the output window. This speech signal is saved in the database. Following figure shows the recorded sound signal.

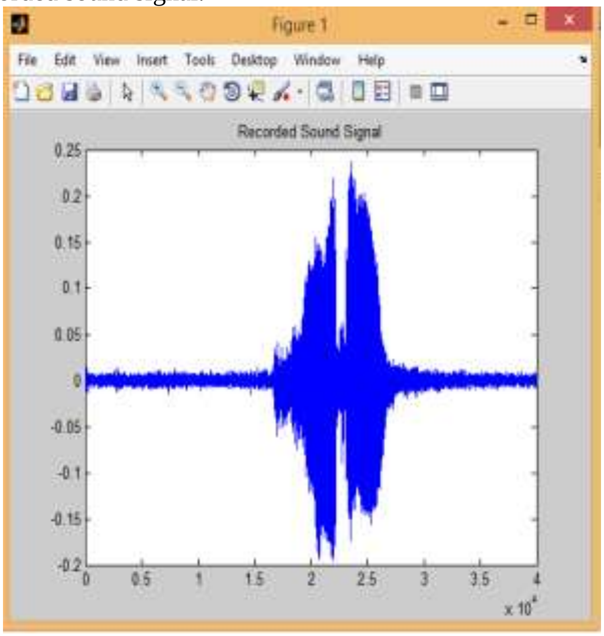

Fig. 8 Recorded Sound Signal

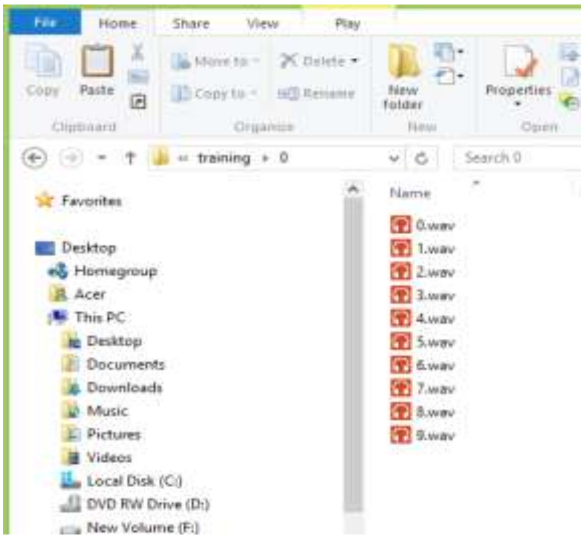

Fig. 9 Recorded sample saved in database

IV.SUMMARY AND DISCUSSIONS

Voice Controlled Home Automation is a very different concept than what is presently available in h market. This would make automation more easy and intuitive. The people will be able to interact with the system. It also is an important aspect in the present world where people are so busy, this would help them in easing the basic functionality of their life. The world around us is going digital in every aspect we can imagine and it is happening fast, we also need to move forward with it. Our system is a great initiative step in automation, it would also provide with security. As it is based on voice recognition we can assign particular password to each

© 2017, IERJ All Rights Reserved

user and the automation will respond to the correct passwords only. Thus by implementing voice operated system for automation and security purpose we can develop the man-machine interaction.

ADVANTAGES

1) Increases the level of home Security.
2) Provide high living standards to human being.
3) As the voice recognition is used reliability is more than other systems.
4) Saves energy and time for controlling home appliances.

APPLICATIONS

1) The System is an integrated system to facilitate elderly and disabled people with an easy to use home appliance.
2) For office use.
3) For Industrial use.
4) In military applications for security purpose.

REFERENCES

[1] C. Schmandt, Voice Communication with computers conversational system, MIT opens Courseware, Van Nastrand Reinhold, New York, 2008.

[2] V .M. Velichko and N. G. Zagoruyko," Automatic recognition of 200 words," Int .J.Man-Machine studies, vol. 2, pp.223, June .1970.

[3] L. R. Rabiner and M.R. Sambur," An Algorithm for determining the end points of isolated utterance," The BELL Systems Tech .Journal, pp.297-315, Feb.1975.

[4] M. H. Savoji, "Endpoint of speech signals," Speech Communication, vol. 8, no. 1,pp. 46-60, March 1989.

[5] J. C. Junqua, B. Mak and B. Reaves, "A Robust algorithm for word boundary detection in the presence of noise," IEEE Trans. on speech and audio processing, vol.2, no.3, pp.406-412, July 1997.

[6] Y. S. Angal, R.H. Chile and R.S. Holambe," Comparison of end point detection algorithms in isolated words speech signal processing," Journal of Instrument Society of India, vol.40,no. 4 ,pp.305-307,Dec -2010.

[7] Y. S. Angal, R.H. Chile and R.S. Holambe," Speech recognition for isolated words using MFCC and VQ," Journal of Instrument Society of India, vol.41, no. 3, pp.201-204, Sep -2011.

[8] Y. S. Angal ,R.H. Chile and R. S. Holambe, "Speech recognition of isolated words based on feature extracted from Mel filter and Bark filter like wavelet packet structure derived from Triplet Half band Filter bank," Computer and electrical engineering ,Elsevier press.

[9] B.H .Juang and A.H. Gray,"Multiple stage vectors quantization for speech coding," in proc. IEEE Inter. Conf. on Acoustic, Speech, Signal Processing, Paris, pp.597-600, May 1982.

[10] J.W. Picone," Signal modelling techniques in speech recognition," IEEE Trans. on Speech and Audio Processing, vol.81, no.9, Sep 1993.

[11] J. Markhoul, "Linear prediction: A tutorial review," proceedings of the IEEE, vol. No.4, vp.561-580, 1975.

[12] S. Furui,"Cepstral analysis techniques for automatic speaker verification," IEEE Control System Magazine, pp. 53-68, 2002.

[13] Q. Zhu and A. Alwan," The effect of additive noise on speech amplitude spectra: A quantity analysis," IEEE signal processing letters, vol. 9, no 9, pp .275-277, 2002

© 2017, IERJ All Rights Reserved