"GUI Based Home Automation System Operated By Voice Command"

Ms. Rutuja Vijay Patil.
Department of Electronics and Telecommunication Engineering
Rajarambapu Institute of Technology
Rajaramnagar, Islampur- 415 409
E-mail: rutupatil.patil043@gmail.com

Abstract: This paper presents the design and implementation of a Home Automation system Operated by Voice Command. It uses human-computer interaction to control different devices. In this paper, ASR (Automatic Speech Recognition) has been used. The automatic speech recognition (ASR) system is used to accurately and efficiently convert a speech signal into a text message by using MATLAB. The speech signal is captured by using microphone. This signal is pre-processed at front-end for feature extraction using Mel Frequency Cepstrum Coefficient (MFCC). The obtained features are compared at back end for pattern matching using Dynamic Time Wrapping (DTW). After comparing features of recorded voice with database, controlling command is send to microcontroller to control devices. This system is helpful for quadriplegia and blind people.

Keywords: Home Automation, ASR, feature extraction, MFCC, pattern matching, GUI, VSPE.

I. INTRODUCTION

Home Automation becomes increasingly popular, people are in need of more home automation devices to upgrade their living spaces and enjoy a high-tech life. House is equipped with various consumer devices, that are generally controlled and managed manually or using different remote controls. But it is very difficult to control the multiple household appliances for elder people and people with disabilities such as quadriplegia patients and blind people.

Automatic Speech Recognition (ASR) is mostly used system for speech recognition. ASR system accurately and efficiently converts a speech signal into a text message independent of the device, speaker or the environment. In general the speech signal is captured and pre-processed at front-end for feature extraction and evaluated at back-end.

II. LITERATURE REVIEW

The reviews of literature and the technological aspects of human machine interaction through various speech recognition approaches are discussed. It also discusses the various techniques used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech recognition [1][2][7].

The non-parametric method for modeling the human auditory perception system, Mel Frequency Cepstral Coefficients (MFCCs) are utilize as extraction techniques. The non linear sequence alignment known as Dynamic Time Warping (DTW) is introduced. This paper presents the viability of MFCC to extract features and DTW to compare the test patterns [3].

This paper describes an approach of isolated speech recognition by using the Mel-Scale Frequency Cepstral Coefficients (MFCC) and Dynamic Time Warping (DTW). MFCC are extracted from speech signal of spoken words. To cope with different speaking speeds in speech recognition Dynamic Time Warping (DTW) is used [4].

Numeral recognition remains one of the most important problems in pattern recognition. This paper has discussed an effective method for recognition of isolated Marathi numerals. It presents a Marathi database and isolated numeral recognition system based on Mel-Frequency Cepstral Coefficient (MFCC) used for Feature Extraction and Distance Time Warping (DTW) used for Feature Matching or to compare the test patterns [5][6].

III. SPEECH DATABASE.

For accuracy in the key word recognition, we need a collection of utterances, which are required for training and testing. The collection of utterances in proper manner is called database. The total number of speakers was ten.

A. Acquisition setup

To achieve a high accuracy for key word recognition in home automation system we use 10×10 room for recording some key words without noisy sound. The Sampling frequency for all recordings was 11025 Hz. The speaker were Seating in front of the direction of the microphone. The distance of mouth to microphone is about 5-10 cm. The speech data is collected with the help voice recorder of PC.

B. Create isolated key words (database)

We have recorded isolated word commands such as Fan on, Fan off, Light on, Light off. The speech data is recorded with the help of wavrecord command and store in database using wavwrite command. The speech signals have been stored in form of wav file and given the labels of spoken words to the different files.

IV. PRINCIPLE OF VOICE RECOGNITION

A. Mel Frequency Cepstral Coefficient

There are several kinds of parametric representation of the acoustic signals. Among of them the Mel-Frequency Cepstral Coefficient (MFCC) is most widely used. MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore are best for speech recognition. The extraction of the features of acoustic signals is an important task to produce a better recognition performance. The correctness of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech.

Figure 1 shows Home Automation System Operated by Voice Command. In this block diagram we used MFCC and DTW technique for Key word recognition. Hardware is designed using these components microcontroller, Relay Driver Circuit, Relays, microphone, and RS232-Serial communication etc.

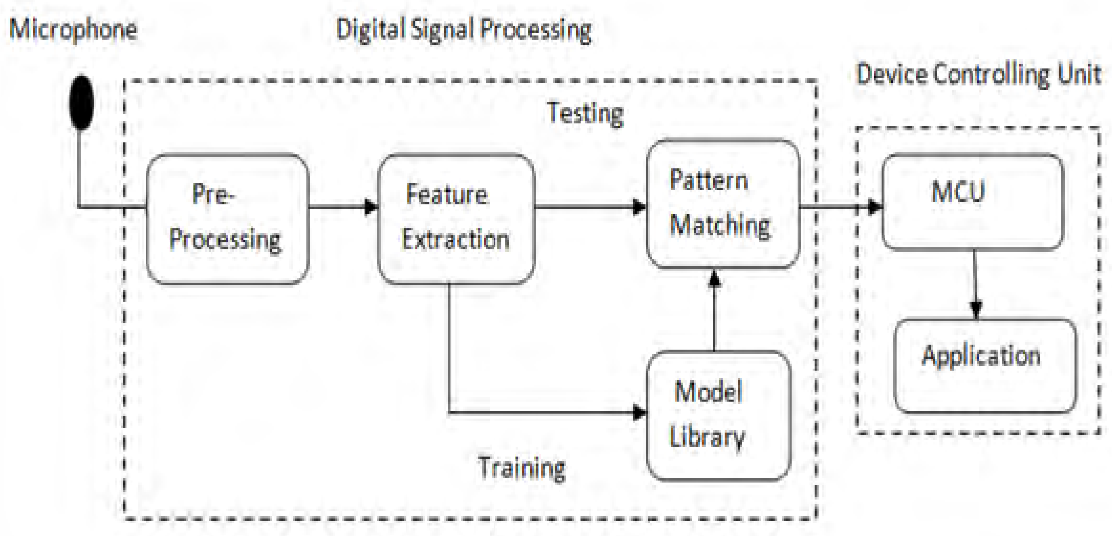

Figure 1 Proposed system of Home Automation System Operated by Voice Command.

1) Pre-emphasis:

This step processes the passing of signal through a filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency in order to improve SNR.

$$y(n) = x(n) - 0.9x(n-1)$$

The goal of pre-emphasis is to compensate the high frequency part that was suppressed during the sound production mechanism of humans.

2) Framing and Windowing:

The process of segmenting the speech samples into a small frame with the length within the range of 20 to 40 msec. In this step the continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M ($M < N$). The first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by N - M samples. Overlapping rate is 50% to avoid the risk of losing the information from the speech signal. After dividing the signal into frames that contains nearly stationary signal blocks.

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. The concept here is to minimize the spectral distortion by using the window to taper the signal to zero at the beginning and end of each frame. Hamming window is used here.

3) Fourier Transform:

To obtain a good frequency resolution, a 512 point Fast Fourier Transform (FFT) is used, which converts each frame of *N* samples from the time domain into the frequency domain.

4) Mel-Frequency Filter Bank:

The frequencies range in FFT spectrum is very wide and voice signal does not follow the linear scale. Thus for each tone with an actual frequency, *f*, measured in Hz, a subjective pitch is measured on a scale called the 'Mel' scale. The Mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels. Therefore we can use the formula in Equation 4 to compute the Mels for a given frequency f in Hz.

$$Mel(f) = log * (1 + f/700)$$

5) Discrete Cosine Transform:

Mel-cepstrum coefficients contain only real part. We can convert them to the time domain using the Discrete Cosine Transform (DCT). The set of coefficient is called acoustic vectors.

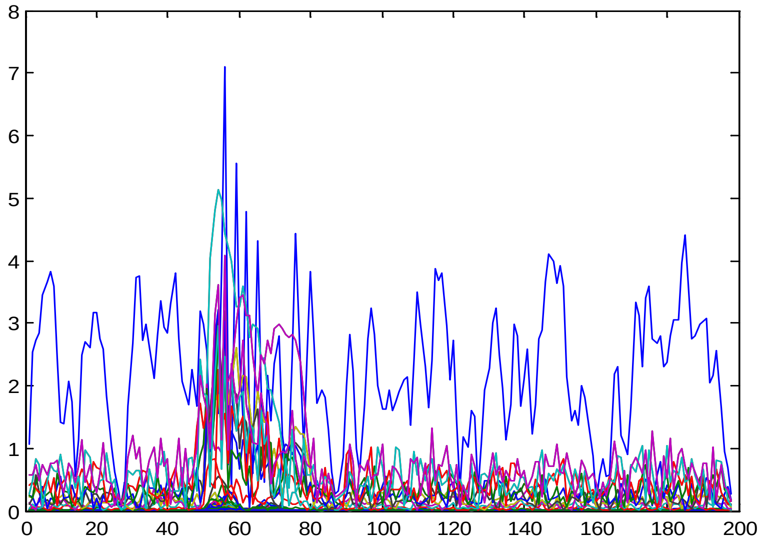

Figure 2 MFCC coefficients for 'Fan On' command.

B. Dynamic Time Warping

The features extracted from each command word using MFCC are stored as a data base. When a valid command word is taken for testing, the MFCC of that command is computed. This is compared with the MFCCs stored in the data base and matched using the technique of dynamic time warping. Dynamic programming is an extensively studied and widely used tool in operation research for solving sequential decision problems. Dynamic time warping (DTW) is an algorithm for measuring similarity between two sequences which may vary in time or speed. The feature vector obtained from the incoming input speech has to be compared with the feature vectors in the data base. One of the popular methods of comparing two feature vectors is to calculate the Euclidian distance between the two feature vectors. The formula used to calculate the Euclidean distance can be defined as following:

The Euclidean distance between two points

$$P(p1, p2, p3, \dots \dots, pn), Q(q1, q2, q3, \dots \dots, qn)$$

is given by the formula

$$\text{ED} = ((p1 - q1)^2 + (p2 - q2)^2 + \cdots (pn - qn)^2)^{½}$$

Word which gives the least distance is identified as the command word.

The problem with this approach is that different command word has different length. Hence the feature vectors will be of different lengths. The matching process needs to compensate for length differences and take account of the non-linear nature of the length differences within the words. The Dynamic Time Warping algorithm achieves this goal; it finds an optimal match between two sequences of feature vectors which allows for stretched and compressed sections of the sequence.

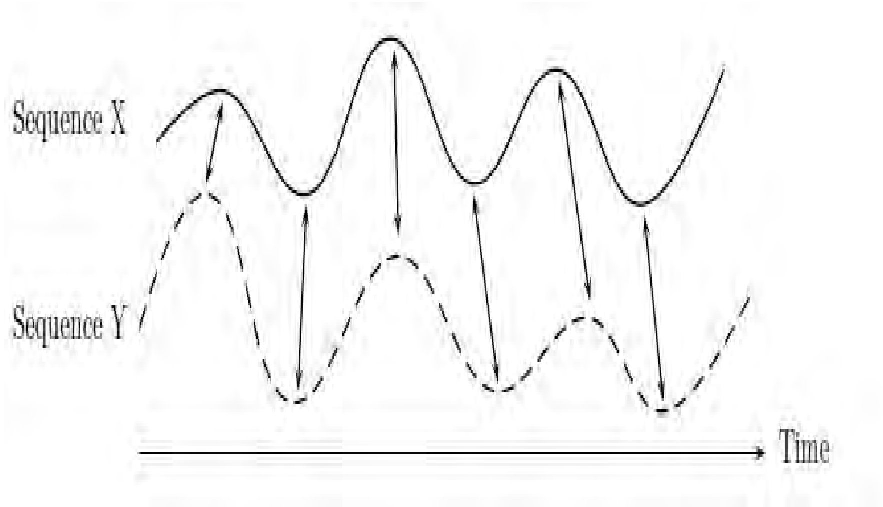

Figure 3 Time alignment of two time-dependent sequences

Consider two time series Q and C of length n and m respectively.

$$C = (c1, c2, c3, \ldots, cm),\ Q(q1, q2, q3, \ldots, qn)$$

To align two sequences using DTW we construct an *n*-by-*m* matrix where the (i^{th}, j^{th}) element of the matrix contains the distance *d* between the two points. Warping path *W* is a contiguous set of matrix elements that defines a mapping between *Q* and *C*.

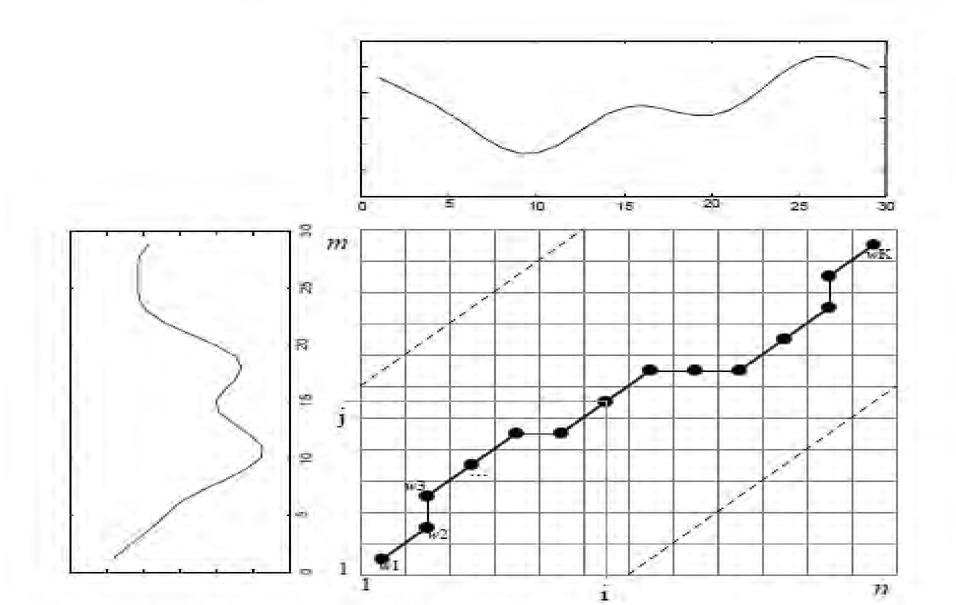

Figure 4 Minimum path between stored and reference keyword.

C. Device Controlling Unit

1) Using Simulation in PROTEUS

The devices like bulb are controlled by using Proteus software and Keil software. According to recognized word command, the specific data is send to controller through serial communication. For interfacing of Matlab with Proteus through serial communication, we use Virtual Serial Port Emulator Software (VSPE). In Proteus data is received by COMPIN and send to microcontroller through Max232. The received data send to port pins for controlling devices.

2) Using Hardware

In actual implementation of device controlling unit we use 8051 development board to control devices. In this Matlab send specific data to P89V51RD2 controller via serial communication and MAX 232 IC. This received data send to port pins and control the device using relay and relay driver IC.

V. RESULT

The recorded database of voice commands is as shown below in Figure 5.

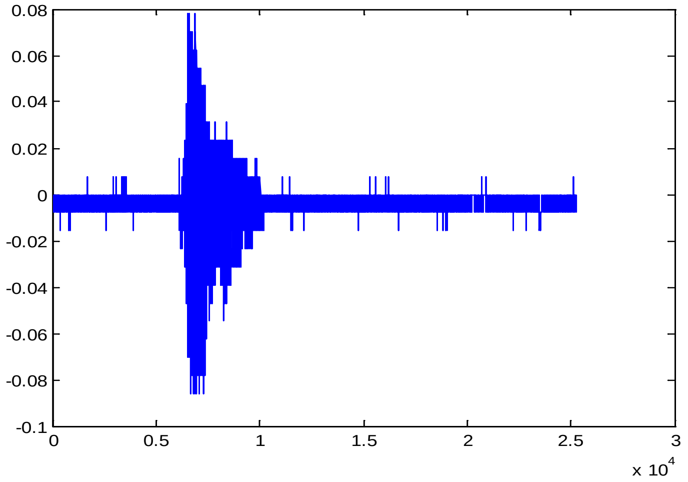

Figure 5 Database of voice command ‘Fan on’.

The recognized voice commands converted into text is displayed on GUI window along with FFT Spectrum and MFCC coefficients. Also their calculated parameters are displayed on it, as shown in figure 6.

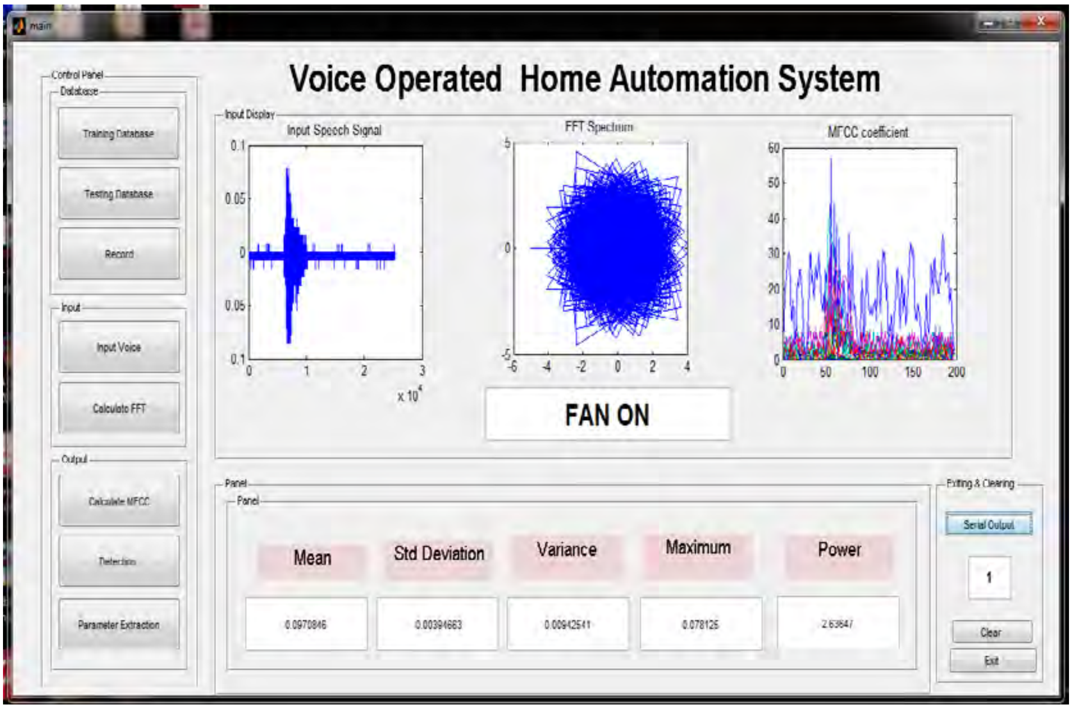

Figure 6 GUI for “Fan On “command

The assignment of COM pins with Virtual Serial Port Emulator (VSPE) software is as shown below in Figure 7.

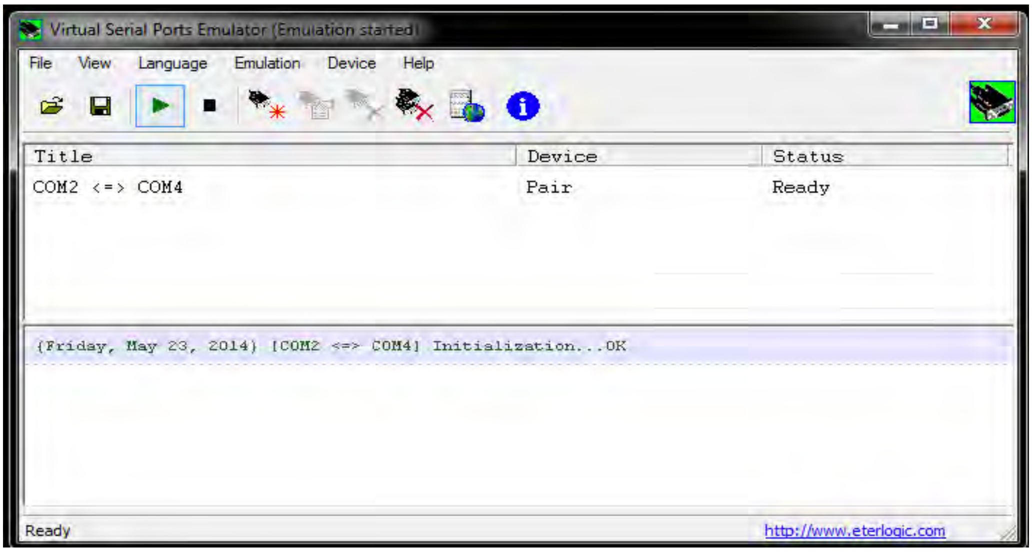

Figure7 Assigning of COM port using VSPE

1) Simulation Result of device controlling unit

The devices are controlled in Proteus as shown in Figure 8

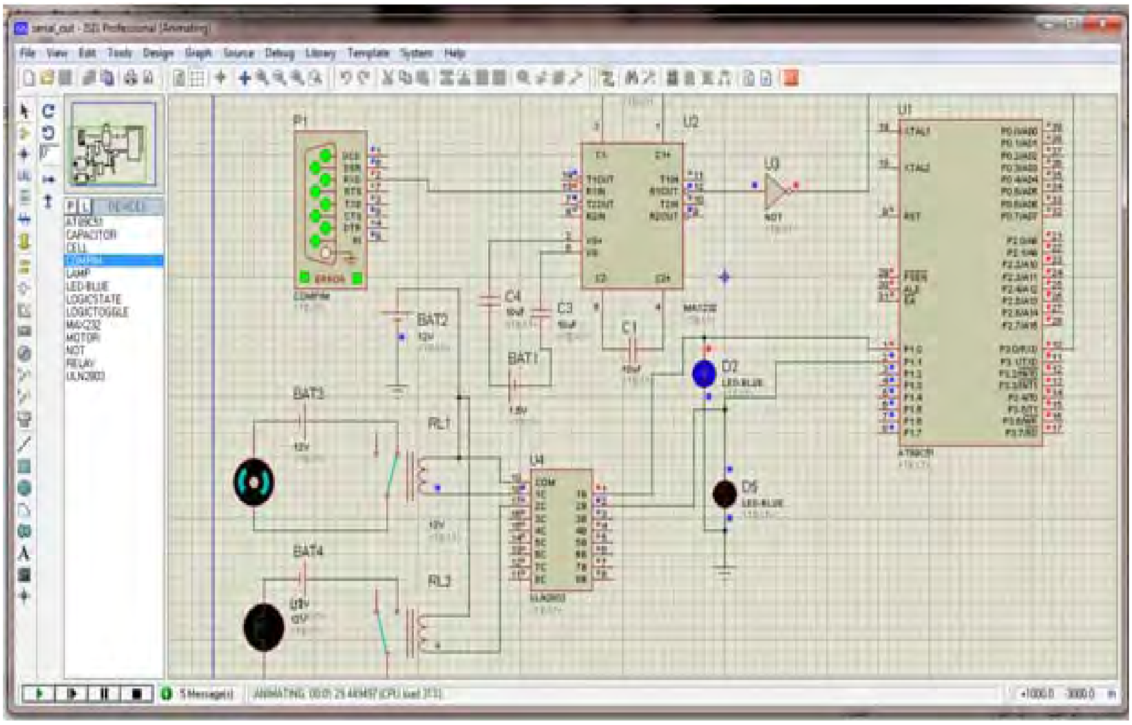

Figure 8 controlling of motor for command “Fan on”

2) Hardware implementation of device controlling unit

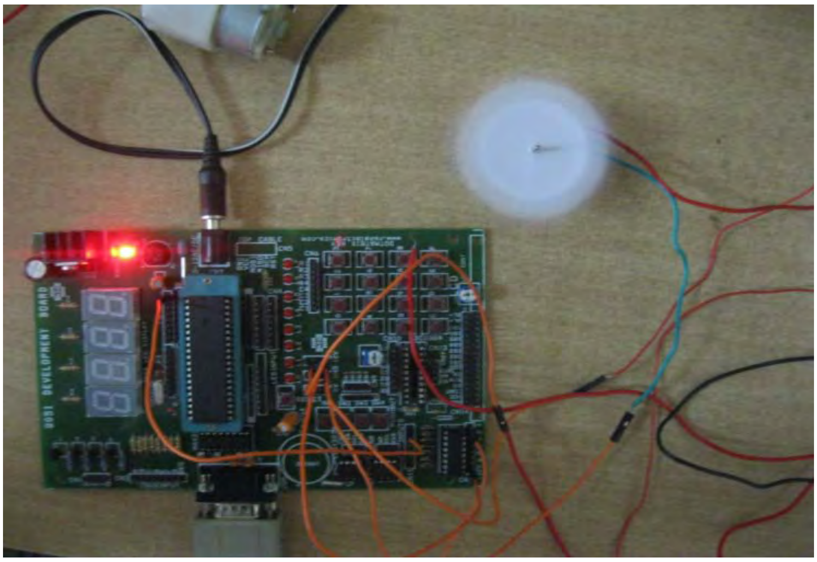

Figure 9 controlling of 12V DC motor for command “Fan on”

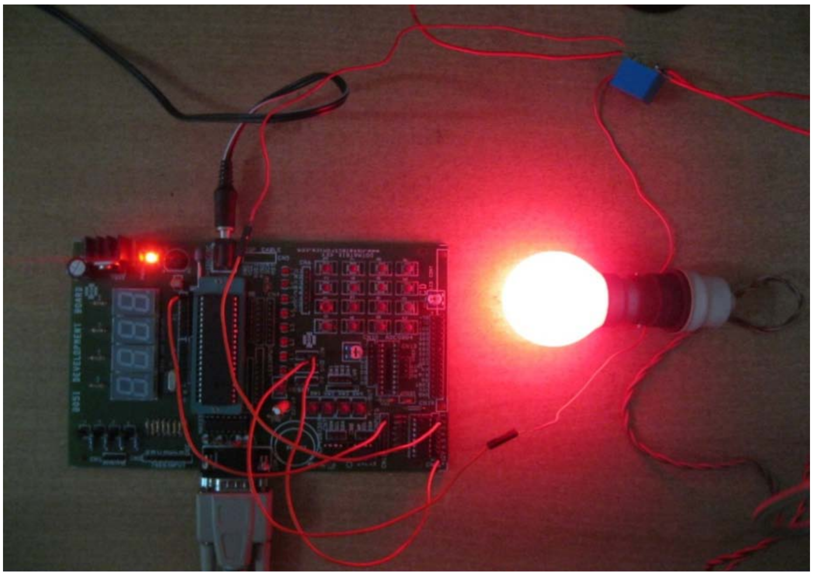

Figure10 controlling of 230V Bulb for command “Light on”

VI. CONCLUSION

We have successfully implemented voice recognition system using MATLAB. The recognized voice commands are displayed on GUI in text form. As voice command is recognized by MATLAB, the specific data is send to microcontroller. The controlling of devices using microcontroller is implemented according to received data using Proteus. This system is targeted at elder, blind and quadriplegia people.

This model for the speech recognition was tested in noise free and noisy environment with varying speakers.

ACKNOLOGEMENT

The authors would like to thank the RIT Institute for the timely and important help in the field of signal processing. The authors also wish to thank the reviewers for their very detailed review of the manuscript and for their help in clarifying some aspects of the reported results.

REFERENCES

[1] Krishan Kant Lavania, Shachi Sharma and Krishna Kumar Sharma," Reviewing Human-Machine Interaction through Speech Recognition approaches and Analyzing an approach for Designing an Efficient System" International Journal of Computer Applications, Volume 38, 2012.

[2] Aarti V. Jadhav and Rupali V. Pawar, "Review of Various Approaches towards Speech Recognition", International Conference on Biomedical Engineering, pp: 99 – 103, IEEE, 2011.

[3] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol.2, issue 3, March 2010.

[4] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, "Isolated speech recognition using MFCC and DTW", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, August 2013.

[5] "Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 National Conference on Emerging Trends in Engineering & Technology (VNCET-30 Mar'12.

[6] Bharti W. Gawali, Santosh Gaikwad, Pravin Yannawar, Suresh C.Mehrotra, "Marathi Isolated Word Recognition System using MFCC and DTW Features" ACEEE Int. J. on Information Technology, Vol. 01, No. 01, Mar 2011

[7] Katti and Anusuya," Speech Recognition by Machine: A Review", International Journal of Computer Science and Information Security, Volume: 64, pp: 501 - 531 IEEE, 2009.