

Speech Recognized Automation System Using Speaker Identification through Wireless Communication

S. R. Suralkar¹, Amol C. Wani², Prabhakar V. Mhadse³

¹(Electronics & Telecomm Engg Dept, SSBT's COET, North Maharashtra University, Jalgaon, India)

²(Electronics & Telecomm Engg Dept, SSBT's COET, North Maharashtra University, Jalgaon, India)

³(Electronics & Telecomm Engg Dept, SSBT's COET, North Maharashtra University, Jalgaon, India)

Abstract : This paper discusses the methodology for a project named "Speech Recognized Automation System using Speaker Identification through wireless communication". This project gives the design of Automation system using wireless communication and speaker recognition using Matlab code. Straightforward programming interface of Matlab makes it an ideal tool for speech analysis in project. This automation system is useful for home appliances as well as in industry. This paper discusses the overall design of a wireless automation system which is built and implemented. The speech recognition centers on recognition of speech commands stored in data base of Matlab and it is matched with incoming voice command of speaker. Mel Frequency Cepstral Coefficient (MFCC) algorithm is used to recognize the speech of speaker and to extract features of speech. It uses low-power RF ZigBee transceiver wireless communication modules which are relatively cheap. This automation system is intended to control lights, fans and other electrical appliances in a home or office using speech commands like Light, Fan etc. Further, if security is not big issue then Speech processor is used to control the appliances without speaker identification.

Keywords — Automation system, MATLAB code, MFCC, speaker identification, ZigBee transceiver.

I. Introduction

Development of automation systems using speaker identification began in the 1960s with exploration into speech analysis using text matching, where characteristics of an individual's voice were thought to be able to characterize the uniqueness of an individual much like a fingerprint. The early systems had many flaws and research ensued to derive a more reliable method of predicting the correlation between two sets of speech utterances. The home evolutionary developments time from the era in which man became sedentary to stop living inside caves and start building their homes. These evolutionary trends of homes automation are focused on several main issues such as security, culture, leisure, comfort, energy savings, management and economic activities. Over the years much work has been done in the domain of automatic speech recognition for automation systems. The progress made is significant for small, medium and even large vocabulary systems. Speaker Identification based automation is one of the major growing industries that can change the way people live with security in operation. The aim of such Automation System is to provide those with special needs with a system that can respond to speech commands and control the on/off status of electrical devices, such as lamps, fans, television etc, in the home and office wirelessly. The system should be reasonably cheap and easy to install in any environment.

Generally, human voice conveys much information such as gender, emotion and identity of the speaker. The objective of voice recognition is to determine which speaker is present based on the individual's utterance [3]. Several techniques have been proposed for reducing the mismatch between the testing and training environments. Many of these methods operate either in spectral, or in cepstral domain. Firstly, human voice is converted into digital signal form to produce digital data representing each level of signal at every discrete time step like zero crossing. The digitized speech samples are then processed using MFCC to produce voice features. The performance of the speech recognition system is given in terms of error rate as measured for a specified technology. Vocal communication between person and computer includes synthesis of speech which is in Matlab database and matching of speech features with incoming speech commands. Speaker recognition involves speaker identification to output the identity of speaker among the population as switching on/off of the electrical equipment.

II. System Overview

Fig.1 show overall system confined with software and hardware with wireless units. A speaker identification based automation system is integrated system including separate transmitter and receiver module connected wirelessly and MATLAB software coding using MFCC algorithm to extract speech features for Speaker Identification.

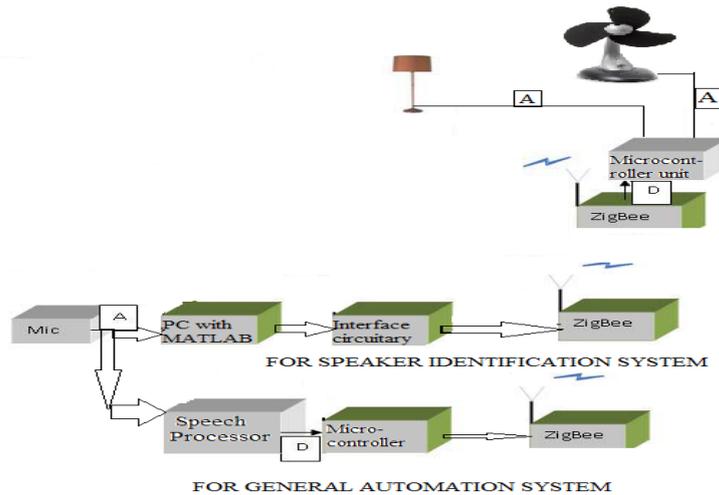


Fig.1: Functional Block Diagram of Automation system for Speaker Identification

This project aims to develop an automated system capable of controlling operation of various devices using voice command, so that respective user can handle, simply, such devices with authentication. Fig. 2 and Fig. 3 shows the architecture used for transmitter and receiver module, for speech interpretation an environmental microphone and communication interface, which captures and digitize voice more clearly in Matlab database and in speech processor. Microcontroller unit is used to control every operation of speech processor and speech commands coming from Matlab simulator.

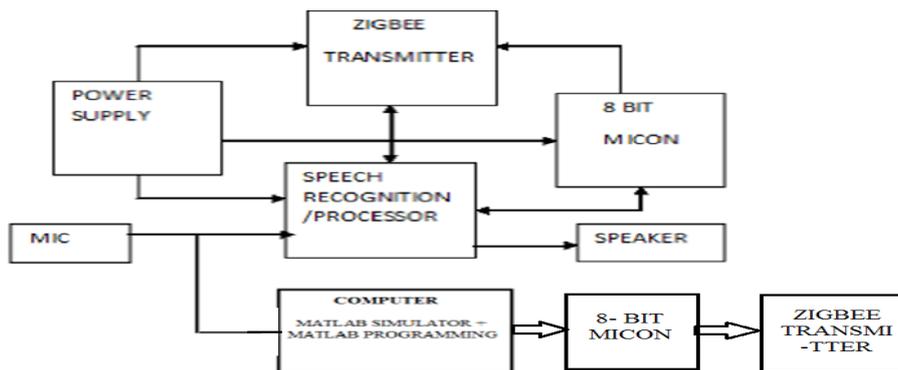


Fig. 2: Transmitter section.

Speaker identification based automation system is an integrated system that facilitates user with an easy-to-use home automation system that can be fully operated based on speech commands. This system is handled by particular user only, and hence more secure system. The system is constructed in a way that is easy to install, configure, run, and maintain. The functional blocks of the transmitter module is shown in Fig. 2.

Fig. 3 shows receiver section which converts digital signal from speech processor and Matlab simulator to analog signal to drive appliances using microcontroller unit.

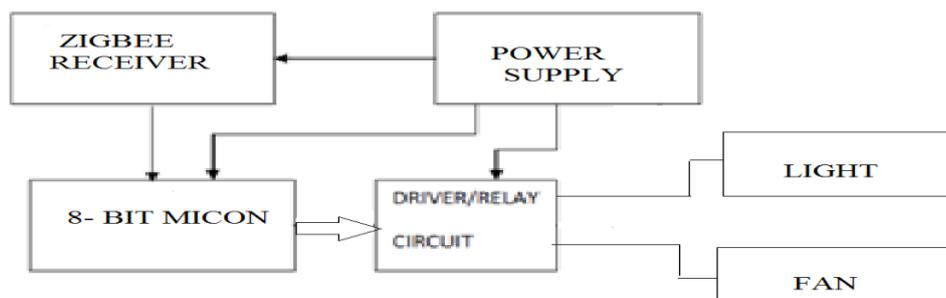


Fig. 3: Receiver section.

We have selected the most appropriate speech recognition method using Processor HM2007 which will provide instructions to activate the systems actuators. In this case the actuators will be switches or motors which control lighting, heating and other features inside the home and office. Here, speech signal is fed to speech processor for speech analysis and same signal is applied to MATLAB using speech detect and speech write commands speech is saved in Matlab database. After the above we have selected the signal bus and communication protocol between computer and actuators. The most common are: dedicated wiring, twisted pair, ZigBee, coaxial cable, power supply installed.

III. Speech Recognition

3.1 Speech Recognition Algorithm

A speech analysis is done after taking an input through microphone from a user. The design of the system involves processing of the input audio signal. At different levels, different operations are performed on the input signal such as Pre-emphasis, Framing, Windowing, Mel Cepstrum analysis and Recognition (Matching) of the spoken word [3]. The voice algorithms consist of two distinguished phases. The first one is training sessions, where recording of speech signal takes place associated with the application or uniqueness, while, the second one is referred to as operation session or testing phase, where particular users speech commands matching is tested, as described in Fig. 4.

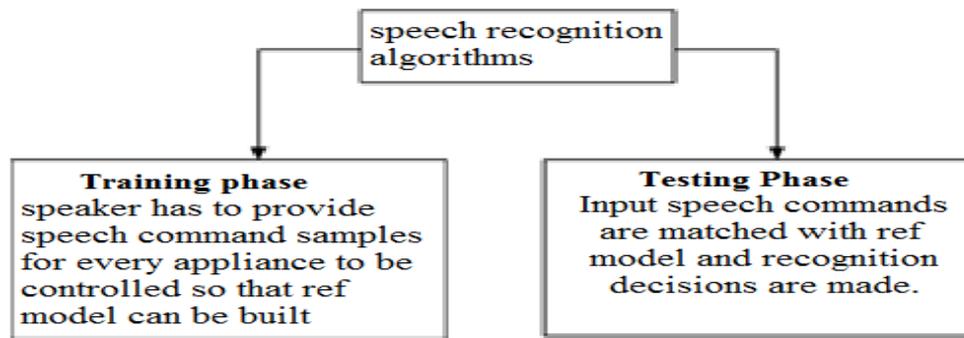


Fig. 4: Speech Recognition Algorithm.

3.2 MFCC Algorithm

The extraction of the best parametric representation of speech signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions which cannot perceive frequency over 1Khz. MFCC is based on known variation of the human ear's critical bandwidth with frequency. 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 [3]. It checks zero crossing in speech signal.

The cepstrum is a common transform used to gain information from a person's speech signal. It can be used to separate the excitation signal (which contains the words and the pitch) and the transfer function (which contains the voice quality). It is the result of taking Fourier transform of decibel spectrum as if it were a signal. We use cepstral analysis in speaker identification because the speech signal is of the particular form above, and the "cepstral transform" of it makes analysis incredibly simple. Mathematically, cepstrum of signal = FT[log{FT(the windowed signal)}].

MFCCs are commonly calculated by first taking the Fourier transform of a windowed function of a signal and mapping the powers of the spectrum obtained onto the mel scale, using triangular overlapping windows. Next the logs of the powers at each of the mel frequencies are taken, Direct Cosine Transform is applied to it. The MFCCs are the amplitudes of the resulting spectrum.

Step-I: 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.

Step-II: Framing

The short time analysis is fundamental to speech analysis techniques. It is assumed that over a long interval of time speech waveform is not stationary but over a sufficiently short time interval about 10-30msec, it can be considered stationary. This is due to that fact that the rate at which the spectrums of the speech signal changes is directly dependant on the rate of movement of the speech articulators. Most speech analysis systems

operate at uniformly spaced time intervals or frames of typical duration 10-30 msec. In frame blocking, 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 and The second frame begins M samples after the first frame, and overlaps it by N - M samples. Similarly, the third frame begins 2M samples after the first frame (or M samples after the second frame) and overlaps it by N - 2M samples. This process continues until all the speech is accounted for within one or more frames. Typical values for N and M are N = 256 (which is equivalent to ~ 30msec windowing and facilitate the fast radix-2 FFT) and M = 100 [2].

Step-III: Windowing

The next step in the speech processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. When we perform Fourier Transform, it assumes that the signal repeats, and the end of one frame does not connect smoothly with the beginning of the next one. This adds some glitches at regular intervals. So we have to make the ends of each frame smooth enough to connect with each other. This is possible by a processing called Windowing. This is done by Hamming windowing method.

$$w(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi n}{N-1}\right)$$

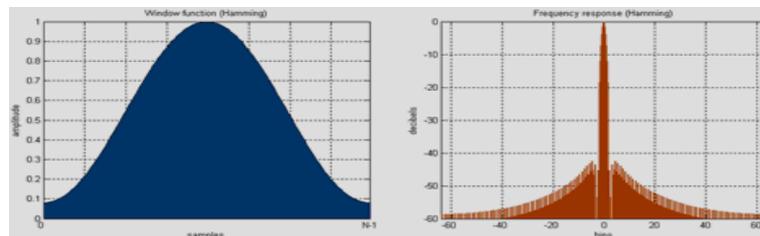


Fig. 5: Hamming Window.

Step-IV: Fast Fourier Transform

The Fast Fourier Transform is used to convert each frame of N samples from the time domain into the frequency domain. The FFT is a fast algorithm to implement the Discrete Fourier Transform (DFT) which is defined on the set of N samples {xn}, as follow:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, \quad n = 0,1,2,\dots,N-1$$

Note that we use j here to denote the imaginary unit, i.e. j= √-1. In general Xn 's are complex numbers.

Step- V: Mel Filter Bank Processing

The range of frequencies in FFT is very wide and voice signal does not follow the linear scale. The bank of filters according to Mel scale as shown in figure 6 is used [3] to smoothen scale.

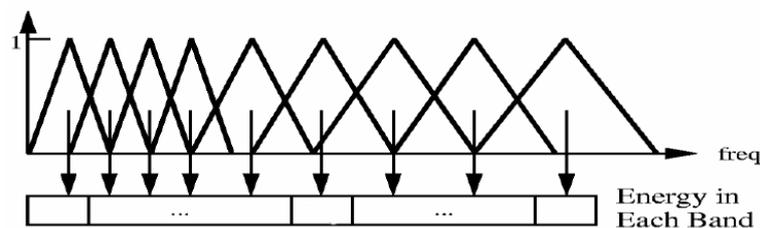


Fig. 6: Mel scale filter bank.

The Mel for given frequency F in Hz:

$$F(\text{mel}) = [2595 * \log_{10} [1+f] 700]$$

Step-VI: Discrete Cosine Transform

This is the process just reverse to FFT to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstrum Coefficient.

Step-VII: Delta Energy and Delta Spectrum

The levels of speech signal and the frames changes, therefore, there is a need to add features related to the change in cepstral features over time. The energy in a frame for a signal X(t) in a window from time sample t1 to time sample t2, is represented at the equation below:

$$\text{Energy} = \sum X^2(t)$$

IV. Hardware Design

In this section we present the hardware description of the modules used for adaptation of speech processor HM 2007 and wireless communication like Zigbee Module.

4.1. Speech Processor Module

The main element of the design consist of speech synthesizer/recognition ASIC and ZIGBEE transceiver module. The speech recognition IC needs to be programmed for various speech commands. This is done by using a MIC, analog speech signal from MIC are stored in the internal memory of the IC after being digitized using ADC blocks which is internal process in processor. Once the learning/programming is completed. The IC is ready to accept the commands. A command issued by the user through MIC will be digitized and compared with the digitized commands already stored in the internal memory of IC. When a match is received, microcontroller status will be updated accordingly. The microcontroller in turn will generate a specific data which will be transmitted through RF channel using ZIGBEE transmitter.

4.2 Microcontroller Module

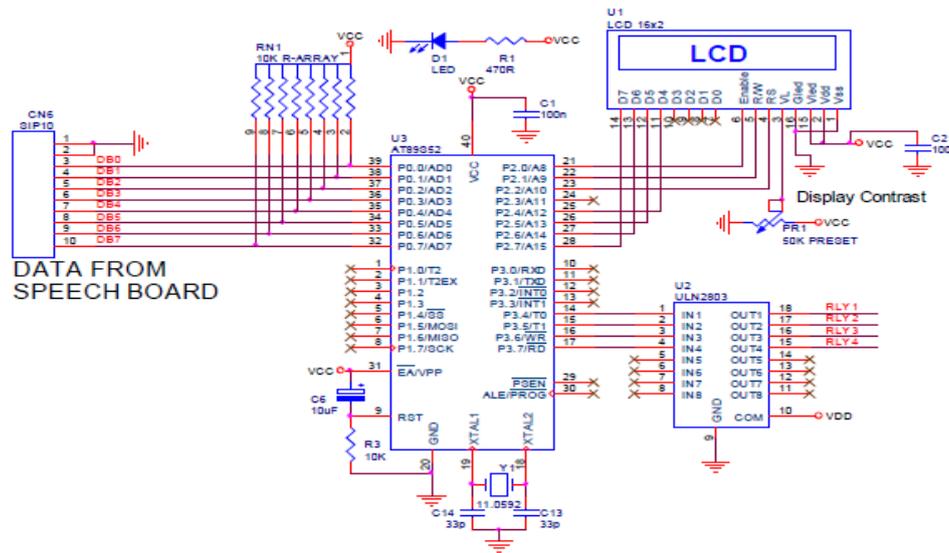


Fig. 7: Microcontroller Module.

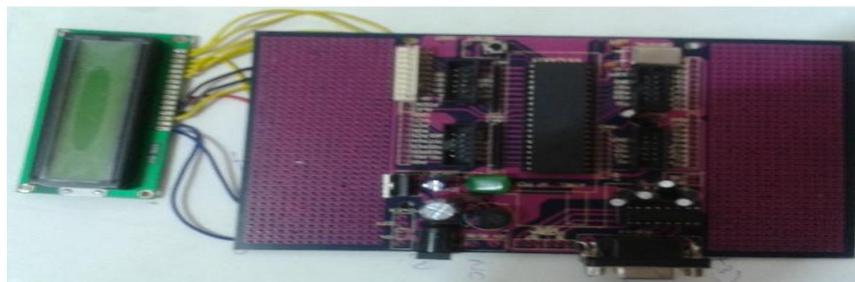


Fig. 8: Microcontroller-89c52 Board

This module receives signal from speech processor and decision making regarding switching ON/OFF appliance is made here as shown in figure 7. Either from speech processor or Matlab simulator output associated with appliance like Light or fan is given to microcontroller. Output of microcontroller will toggle for incoming signal from processors. Wireless module is connected to one of the input port of microcontroller and appliances are connected to output port.

4.3 Wireless (Zigbee) Module

Zigbee protocol is the communication protocol that is used in this system for wireless communication as shown in figure 9. Zigbee offers 250 kbps as maximum baud rate. Range of wireless unit is 100meters, but practically we get approximately about 30-40 meters.



Fig. 9: Zigbee Module.

V. Tests And Results

In this section, we need to train the system using speech commands, which are saved in database using .wav format. Following command is used to record speech command for every appliance.

```
y = wavrecord (30*8000, 8000);
```

after recording command it is checked whether it is in given sample format using,

```
Speechdetect (y);
```

If recording is ok then it is saved in database using command,

```
Wavwrite (y, 8000, 'speaker.wav');
```

For example, for light appliance it is saved as,

```
Wavwrite (y, 8000, 'light.wav');
```

After recording all commands for various appliances, single database called MODEL file is built, which will have all commands for specific speaker. This model file is built using following command in Matlab,

```
'transcript'.
```

Above steps are used to train and record the speech commands. Once commands are saved, our system is ready to work. At the serial port we use Zigbee transmitter module, which will send on/off commands to appliance, wirelessly. Using command 'digitrecgui' speech commands from users are received. If there is match found results will be shown on screen as shown in figure 10 and figure 11. Figure 10 shows matching for light command for particular speaker, where as figure 11 shows match for fan.

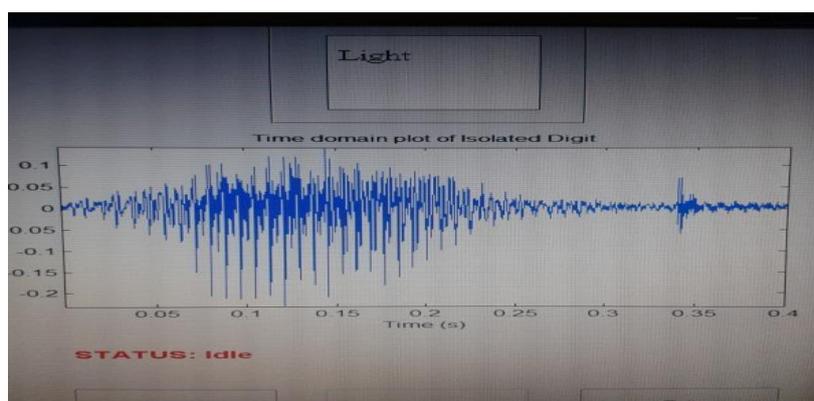


Fig. 10: command Light matched results.

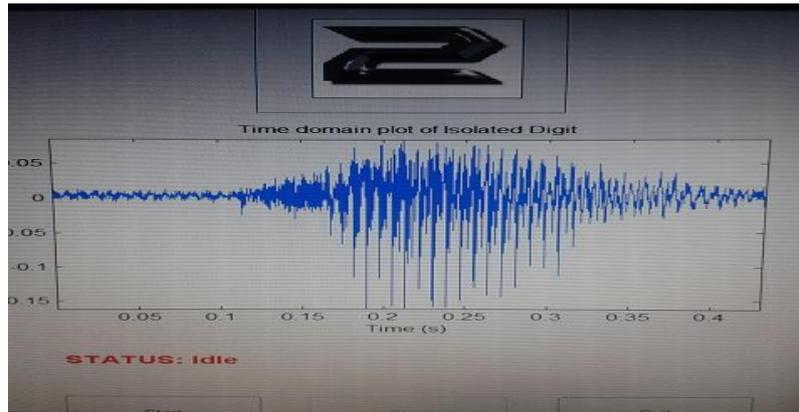


Fig. 11: command Fan matched results.

For 10 different speakers our system is tested, where we found following satisfactory results of speaker identification and appliances worked similarly, results are shown in Table 1.

Sr. No.	SPEAKER	LIGHT (%)	FAN (%)
1	Actual Speaker	80	85
2	Speaker 1	10	10
3	Speaker 2	0	0
4	Speaker 3	0	20
5	Speaker 4	10	0
6	Speaker 5	0	10
7	Speaker 6	20	10
8	Speaker 7	0	10
9	Speaker 8	10	0
10	Speaker 9	0	0

Table 1: Results of Speech recognized Automation system using speaker identification.

For actual speaker and other speakers, statistics is calculated for speech command matching.

VI. Conclusions And Future Work

A speech recognized automation system using speaker identification through wireless communication is built and implemented. The prototype developed can control electrical devices in a home or office wirelessly. The system implements Automatic Speech Recognition using speech processor and Speaker Identification through MATLAB coding using MFCC algorithm. The system implements the wireless network using ZigBee RF modules for their efficiency and low power consumption.

After analyzing all the information from our project, we found that our algorithm improves the speaker recognition task. Because the MFCC uses the Mel scale, the approximation and maximum likelihood nature of speech to the human voice behavior is good, it represent in a better way the voice. We improved the precision in the recognition task. We found that as long as the number of speakers increases, the number of errors increases; this is because the distance to each centroid of each speaker can be similar to another and maximum likelihood of speech is checked by algorithm.

Future work will entail:

- Adding confirmation commands to the voice recognition system from appliance.
- Integrating appliance control functions to improve the system versatility such as providing control commands other than ON/OFF commands such as slow down, dim, etc.

Acknowledgements

I, Prabhakar V. Mhadse acknowledge with thanks the guidance received from Prof. A. C. Wani and prof. S. R. Suralkar, Head of the Department in preparation of this Project.

I would also give my sincere thanks to other staff of college for their warm support. The matter taken from the various sources in preparing this report has been duly cited in reference.

References

- [1] Abhishek Thakur, Neeru Singla, "Design of Matlab-Based Automatic Speaker Recognition and Control System", *International Journal Of Advanced Engineering Sciences And Technologies*, Vol no. 8, issue no. 1.
- [2] Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm, Ch.Srinivasa Kumar et al. / *International Journal on Computer Science and Engineering (IJCSE)*.
- [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*, Volume 2, Issue 3, March 2010, Issn 2151-9617.
- [4] Rojas-Rodríguez Rafael, Aceves-Pérez Rita1, Cortés-Aburto Obed, García-Meneses Carlos1, Vela-Valdés Luis, "Development and integration of automated systems aimed to the comfort of seniors and people with disabilities" *International Conference on Intelligent Environments*, pp339-342,2012.
- [5] Rabiner L. R. and Schafer R. W., *Digital Processing of Speech Signals*, New Jersey, US: Prentice Hall Inc, 1978.
- [6] Augusto J. C. and Nugent C. D., "Smart homes can be smarter", in *Lecture Notes in Artificial Intelligence*. Germany vol. 4008, pp 1-15, May 2006.
- [7] "XBee-2.5-Manual," ZigBee RF communication protocol. (2008).
- [8] Algimantas R., Ratkevicius K. and Rudzionis V., "Voice interactive systems", in *The Engineering Handbook of Smart Technology for Aging, Disability and Independence USA*, pp 281-296, 2008.
- [9] Vovos A. Kladis B. and Fakotakis N.D., "Speech operated smart-home control system for users with special needs", in *Proc. INTERSPEECH*, pp.193-196, 2005.
- [10] Sixmith A., "Modeling the well being of older people" in *The Engineering Handbook of Smart Technology for Aging, Disability and Independence USA*, pp 569-584, 2008.
- [11] Muñoz C., Arellan D, Perales F. J. and Fontaner G. "Perceptual and intelligent domotics system for disabled perople" in *proceedings of Sixth IASTED International Conference, USA pp 70-75 September 2011*.
- [12] Jorge MARTINEZ*, Hector PEREZ, Enrique ESCAMILLA; Masahisa Mabo SUZUKI; "Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques" *IEEE 2012*.
- [13] Shannon B.J., Paliwal K.K., "A comparative study of filter bank spacing for speech recognition", *Proc. of Microelectronic engineering research conference, Brisbane, Australia, Nov. 2003*.
- [14] T. Hill and P. Lewicki, *Statistics Methods and Applications*. Tulsa, OK: StatSoft Inc., 2006.
- [15] S. Furui, *Digital Speech Processing, Synthesis, and Recognition*. New York: Marcel Dekker, 2001.
- [16] Fezari, M.; Khati, A.-E, "New speech processor and ultrasonic sensors based embedded system to improve the control of a motorized wheelchair," *Design and Test Workshop, 2008. IDT 2008*.
- [17] A. Spanias, T. Painter and V. Atti *Audio Signal Processing and Coding*. John Wiley & Sons, 2007.
- [18] Mohd. Rihan, M. Salim Beg, Narayan Gehlot; "Developments in Home Automation and Networking", *Proc. National Conference on "Computing for Nation Development"*, Bharati Vidyapeeth Institute of Computer and Management, New Delhi, 23-24 Feb. , 2007, pp. 61-64.