Automation System using Speech Recognition for Speaker Dependency through MFCC

Prabhakar V. Mhadse
E & TC (Digital Electronics) Engg Department
SSBT's COET, Bambhori
Jalgaon, India.

Amol C. Wani
Electronics & Telecomm Engg Department
SSBT's COET, Bambhori
Jalgaon, India.

Abstract — This paper discusses the methodology for project named "Automation System Using Speech Recognition for Speaker Dependency through MFCC". Automation system is growing to improve the quality; the goal of design of speaker dependent automation is to provide an efficient integration and higher security among appliances. Such automation systems must comply with the household equipments. This paper discusses the overall design of Automation System Using Speech Recognition for Speaker Dependency through MFCC which is to be implemented. This system will have hardware for appliance interface using Zigbee wireless modules and speech processor. Software section includes development of Mel Frequency cepstral coefficient (MFCC) algorithm using MATLAB. The system focus on recognition of speech commands stored in data base of MATLAB and it is matched with incoming speech command of speaker. Using MFCC system will be speaker dependent, for speaker independent systems speech processor of will be used. It uses low-power RF ZigBee transceiver wireless communication modules which are relatively cheap. Such Automation system is developed to control lights and electrical appliances in a home or office with secure use. Wireless transmitter module is interfaced through serial communication with computer. Appliances to be controlled will be connected to zigbee receiver module.

Keywords — automation system, MFCC, speaker identification, Speech processing, ZigBee module.

I. Introduction

Speech Communication is one of the basic and most essential capabilities possessed by human beings. It is this amazing ability that sets him apart from other species living in this Earth. It is also the most natural form of communication between humans and thus as a subject, it has attracted attention over many years. The fundamental purpose of speech is communication, i.e., the transmission of messages. The speech wave itself conveys linguistic information, the speaker's vocal characteristics and emotion. Speech is one of the natural forms of communication. Advancements in scientific technology have made it possible to use this in security systems. Speaker recognition is a process that enables machines to understand and interpret the human speech by making use of certain algorithms and verifies the authenticity of a speaker with the help of a database. That is, speaker recognition or identification is essentially a method of automatically identifying a speaker from a recorded or a live speech signal by analyzing the speech signal parameters. First, the human speech is converted to machine readable format after which the machine processes the data. The data processing deals with feature extraction and feature matching. Then, based on the processed data, suitable action is taken by the machine. The action taken depends on the application.

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. 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. Some of this home automation systems target those seeking luxury and sophisticated home automation platforms; others target those with special needs like the elderly and the disabled. The aim of such Automation System is to provide those with special needs with a system that can respond to voice commands and control the on/off status of electrical devices, such as lamps, fans, television etc, in the home and office. The system should be reasonably cheap, easy to configure, and easy to run.

There are housing developments that have some degree of automation in domestic processes, focused on domestic water pumping and lights control. However, they are isolated elements with a certain degree of difficult use. There have been several commercial and research projects on smart homes and voice recognition systems which can wirelessly be connected to security alarms and other home appliances. The home automation through this system requires holding and interacting with a large panel which constraints the physical movements of the user.

Another popular commercially available system for home automation is from Home Automated Living (HAL) [3]. HAL software taps the power of an existing PC to control the home. It provides speech command interface. A big advantage of this system is it can send commands all over the house using the existing highway of electrical wires inside the home's walls. No new wires means HAL is easy and inexpensive to install. However, most of these products

sold in the market are heavily priced and often require significant home make over.

II. Objectives and Goals

The main objective of this research is to develop a Speaker dependent automation system capable of controlling many devices inside a home and office using speech commands with security of respective user only with speech commands transmitted and received wirelessly using Zigbee modules.

In order to achieve this objective, the following goals are defined and achieved:

• To create speech database of user to be recognized in order to characterize the speech signals to activate appliances.

• To design Wireless transmission media using ZIGBEE transceivers.

• To evaluate and select the most appropriate speech processing method for speaker identification like MFCC algorithm and selection of speech processor HM2007.

III. Review of the State of The Art

A speech activated automation system is integrated system to facilitate disabled, handicapped and elderly people at home and office with secure operation using speech identification and speaker dependent system. New technology discriminates elder and handicapped people because of the access to information barriers [8]. In this paper, the access barriers are grouped into three functional categories:

1) Barriers to provide input information to the computer,

2) Barriers because they cannot interpret the results displayed on screen,

3) Insecure activation of appliances by voice keyword by any user.

For solving all this problems it was proposed to use the speech as the best form of communication with the computer and bring together technologies used to work with voice to speech synthesis, speech recognition and speaker recognition. For people with mobility impairments are recommended to use speech technology as the optimal method of communication with their environment.

In reference [11] is described the development of various efforts to control housing through voice activation systems focused on needs of people with special needs. And special emphasis on the need to develop interfaces that can respond to commands without the need to remember specific commands (natural language) and feedback in audio for people with some kind of visual weakness.

Moreover in [10] recognizes the increase has been the development of technology to improve life quality and independence for elder people, by using alarms for elderly people that live alone, these alarms are connected to hospitals and send a message to relatives cellular phones. It also refers

to use of biometric sensors for automatic detection of emergencies.

IV. System Overview

A speaker dependent automation system is integrated system including separate hardware and software sections and separate transmitter and receiver module connected wirelessly. Fig. 1 shows entire automation system.
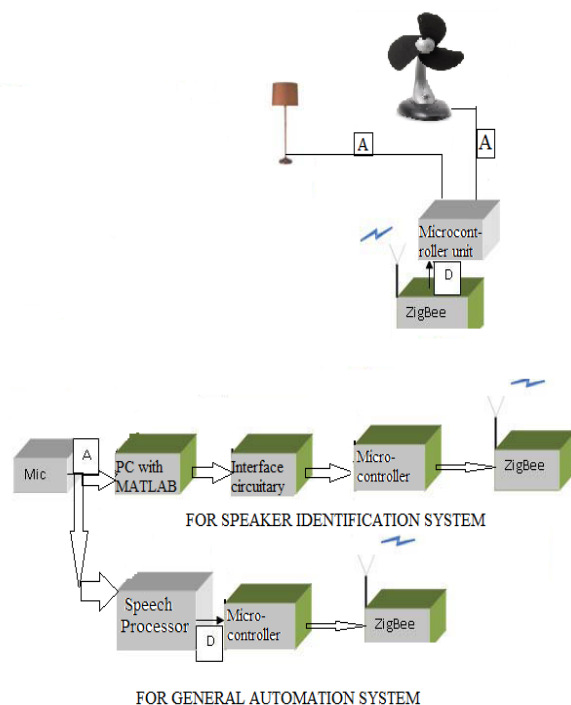

Figure 1. Functional Block Diagram of Automation System

This project aims to 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. Fig. 2 and Fig. 3 shows the architecture used for transmitter and receiver module, for speech interpretation design an environmental microphone and communication interface, which captures and transmit voice more clearly. Microcontroller unit is used to control every operation of speech processor. Voice command is further given to MATLAB software to improve authentication of user commands.

Intelligent automation system is an integrated system to facilitate user with an easy-to-use home automation system that can be fully operated based on speech commands. 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.
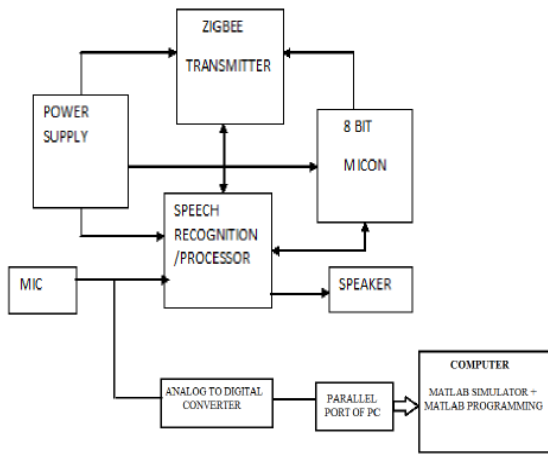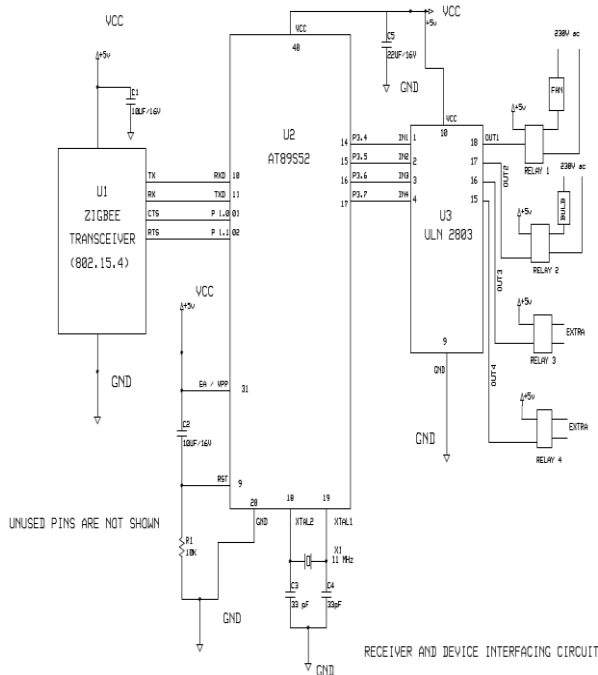
Figure 2. Transmitter module.

Figure 3. Receiver module.

We have evaluated and selected the most appropriate speech recognition method using Processor HM2007 which will provide instructions for action the systems actuators. Here, speech signal is fed to speech processor for speech analysis and same signal is digitized to make it compatible for software processing by MATLAB using MFCC algorithm. In this case the actuators will be switches or motors which control lighting, heating and other features inside the home and office.

A MATLAB application program, running on the PC, uses MATLAB library for the speech commands recognition. Upon recognition of the commands, control characters are sent to the specified appliance address through microcontroller using wireless modules. Consequently,

appliances are turned ON or OFF depending on the control characters. Generally control characters sent will be same for appliance, but it will toggle the output of appliance through microcontroller.

After the above we have selected the signal bus and communication protocol between processor and peripherals for system. The most common are: dedicated wiring, twisted pair, ZigBee modules, serial cable, USB connector, power supply installed, etc.

V. Hardware Design

In this section we present the hardware description of the modules used for adaptation of speech processor HM 2007 [2] and wireless modules [7].

A. Speech Processor Module

This module is used when speaker dependent automation is not required. Speech processor is eligible to recognize the speech commands only it cannot recognize the speech commands of particular speaker.   This module basically includes microphone as input device used to record training words while record mode and input source for speech processor while operation. Speech processor HM 2007 is mother board for this module, which is responsible for speech recognition. HY-6264 8-bit static RAM device is used for data back-up stored in speech processor when power is cut.   Latch 74573 is used to separate data signals of processor to be fed to the microcontroller 89C51/52. Fig 4 shows complete speech processor module.

This unit consists of two vital subsystems, Speech Processor and Microcontroller. 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. Analog speech commands are digitized internally in processor. This learning process need to be done for a number of speech commands. 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 pertaining to a given equipment/appliance and speech command which will transmitted through RF channel using ZIGBEE transmitter.
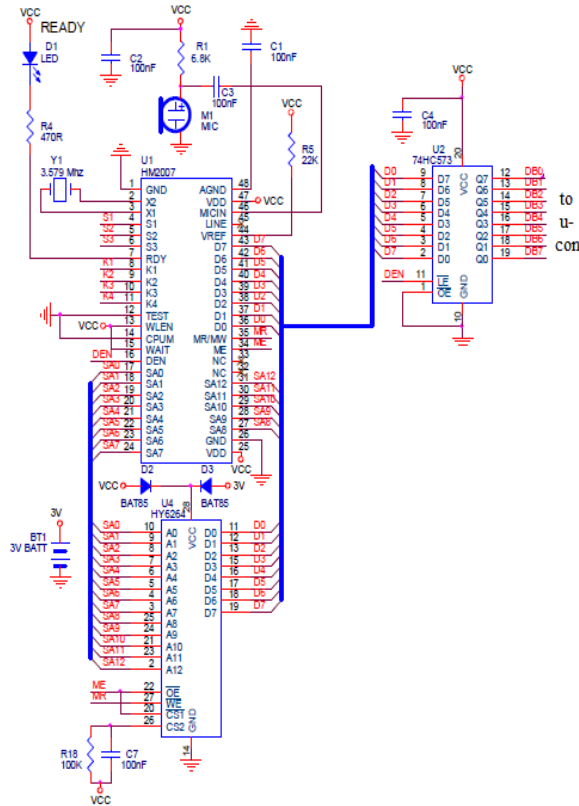
Figure 4. Speech Processor Module.

B. *Microcontroller Module*

Figure 5. Microcontroller Module

This module receives signal from speech processor and decision making regarding switching ON/OFF appliance is made here as shown in Fig. 6. Assembly program is used to do the same and burnt in microcontroller. Wireless module

is connected to one of the port of microcontroller. Generally port 3 is used for transmission of signal.

C. *Wireless (Zigbee) Module*

Zigbee protocol is the communication protocol that is used in this system for wireless communication. Zigbee offers 250 kbps as maximum baud rate, however, 115200 bps was used for sending and receiving as this was the highest speed that the UART of the microcontroller could be programmed to operate at.

D. *Appliance Control Module*

Once the speech commands are recognized, control charterers are sent to the specified appliance address through ZigBee communication protocol. Each appliance that has to be controlled has a relay controlling circuit shown in Fig. 8.
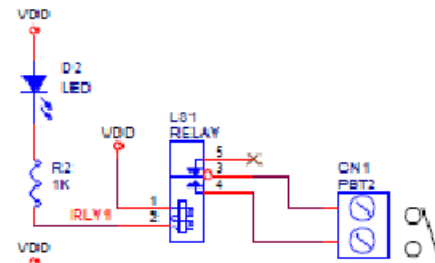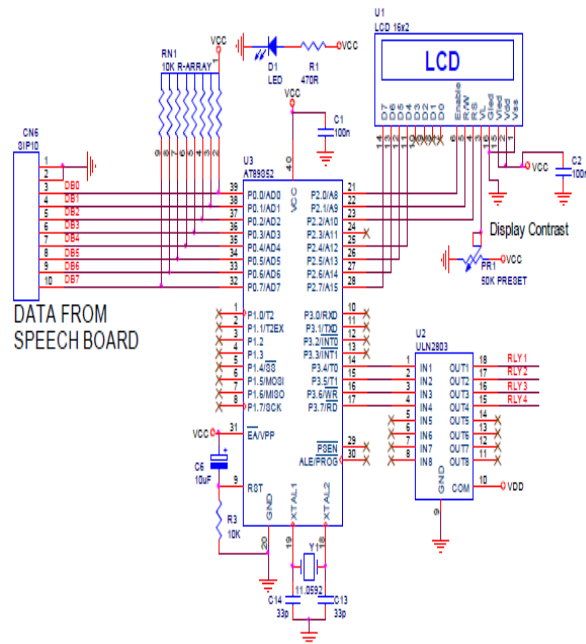

Figure 6. Appliance Control Module.

VI. Software Design

Software design for this project involves training for speech processor as HM 2007 is ASIC processor and assembly programming for microcontroller to work on incoming signals from speech processor. It receives control commands from processor and switches ON/OFF relay connected to the particular appliance. Our project simultaneously works with MATLAB code, where speech recognition will be carried out using MFCC algorithm. If voice command is matched with the corpus stored in system then and then only code will generate matched waveform and appliance will be activated.

A. *MFCC 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. 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. The mel cepstrum is a common transform used to gain information from a person's speech signal. It can be

78

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.

B. *MFCC steps*

1) Automatically detects an isolated word from your speech (input utterance). It does this by calculating energy and the number of zero crossings on a frame-by-frame basis, and compares these values to a threshold. It uses number of iterations for calculation of zero crossings so that maximum likelihood speech is matched.

2) For each frame in a word, it applies a window function followed by a pre-emphasis filter. It then calculates Mel Frequency Cepstral Coefficients (MFCC) and their delta and delta-delta coefficients for each frame, and uses these as feature vectors.

3) For each word's training data, we estimate the parameters of a Gaussian Mixture Model to fit the distribution of training vectors. We therefore train a model to represent each word. This model contains ten different appliances words which are recorded in particular users voice commands.

VII. **Preliminary Results**

We are currently working on the algorithm for speaker dependency using MFCC. At the same time wireless modules using Zigbee transceiver are built.

Speech processor module is assembled and working on speech commands. MFCC algorithm is simulated on MATLAB, it partially identifies speaker. Training and database maintenance is in process.

Acknowledgment

References

[1] Humaid AlShueili, Gourab Sen Gupta, Subhas Mukhopadhyay, "Voice Recognition Based Wireless Home Automation System", 4th International Conference on Mechatronics (ICOM), 17-19 May 2011, Kuala Lumpur, Malaysia, pp 1-6 ,2011.

[2] Datasheet HM 2007, speech processor.

[3] Home Automated Living website. [Cited 2010 14th Oct], http://www.homeautomatedliving.com/default.htm.

[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 engi- neering 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, &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. 3rd International 20-22 Dec. 2008 Page(s):345- 349.

[17] A. Spanias, T. Painter and V. Atti Audio Signal Processing and Coding. John Wiley & Sons, pp. 91-105, 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.

[19] M. Salim Beg, Mohd. Fahad Imteyaz : "Development of Wireless Sensor Networks using Bluetooth", National Conference on Advancement of Technologies – Global Scenario (ADTECH – GLOS), Mathura, 25-26 Feb 2007, pp. 14-20.

[20] Yeon-Joo Oh, Eui-Hyun Paik, and Kwang-Roh Park, " Design of a SIP-based Real-time Visitor Communication and Door Control Architecture using a Home Gateway ", IEEE Transactions on Consumer Electronics, pp.1256-1260, Vol.52, No.4, Nov. 2006.

[21] D. Persson, T. Eriksson, and P. Hedelin, "Packet Video Error Concealment with Gaussian Mixture Models," IEEE Trans. on Image Processing, Vol. 17, No. 2, pp. 145-154, Feb. 2008.

[22] R. Gadalla, "Voice Recognition System for Massey University Smart house," M. Eng thesis, Massey University, Auckland, New Zealand, 2006.