ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

Wireless Monitoring For Industrial Automation Using Speech Recognition

Hussana Johar R.B, Asst Professor, Dept of TE, GSSSIETW, Mysore, Karnataka, India
Soujanya R, Student,Dept of TE, GSSSIETW, Mysore, Karnataka, India
Sushma.S, Student,Dept of TE, GSSSIETW, Mysore, Karnataka, India
Syeda Ansha Sami, Student,Dept of TE, GSSSIETW, Mysore, Karnataka, India
Supreetha M G, Student,Dept of TE, GSSSIETW, Mysore, Karnataka, India

Abstract: ***In this day and era of computers, industrial automation is becoming steadily more important in the industrialization process because computerized machines are able to handle recurring tasks faster and more effectively. This paper mainly focuses on reducing human efforts and risks in dangerous environments like in high voltage situations. The main aim is to design and develop a computer based interactive system with the help of speech recognition in order to monitor the power stations and also to control the same. Speech recognition is the translation of spoken words into text. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. Here the voltage is monitored and controlled from the remote locations using the wireless technology. When the user wants to know the voltage status, all that user needs to do is to speak as "status" through microphone which is stored in the database. After this, simulation is performed and will identify the speech and will read the status of the voltage from the microcontroller, then the data will be displayed on LCD and same will be fed to PC so that the person can get to know the status without touching the PC or by pressing any buttons. The voltage of device is controlled by saying as bulb on or off and fan on or off.***

Keywords: **MATLAB, PIC Microcontroller, RF Transmitter & RF Receiver Module, Speech Recognition.**

I. INTRODUCTION

Speech recognition is a technology that able a computer to capture the words spoken by a human with a help of microphone. These words are later on recognized by speech recognizer, and in the end, system outputs the recognized words. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a speech recognition engine. Speaker identity is correlated with the physiological and behavioral characteristics of the speaker. These characteristics exist both in the spectral envelop and in the supra- segmental features [1]. Speech is one of the natural forms of communication. In speaker identification, the task is to use a speech sample to select the identity of the person that produced the speech from among a population of speakers. In speaker verification, the task is to use a speech sample to test whether a person who claims to have produced the speech has in fact done so [1]. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers.

Speech has the potential to be a better interface than other computing devices used such as keyboard or mouse. Basically, the system is able to recognize the spoken utterances by translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique and increasing the efficiency using Hidden Markov Model (HMM).

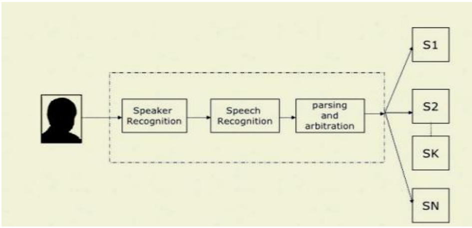

Fig 1: Process of Speech recognition

II. RELATED WORK

In earlier days analogue readers were being used in the industries to make a note of the information that involves status of the devices, these systems were time consuming and gave rise to manual errors. In speech recognition system LPC (Linear Predictive Coding) a speech recognition algorithm was limited to single user and had efficiency of only 57% and MFCC also had a limited efficiency of 85%. In order to overcome manual errors and also to provide security of data it was decided to work on PC. The main criteria of using PC are that the user can monitor and control the industrial device. The MFCC along with the advantages of HMM is used to increase the efficiency above 85%.Speaker recognition systems contain two main modules feature extraction and feature recognition.

A. Feature Extraction

The use of Mel Frequency Cepstral Coefficients can be considered as one of the standard method for feature extraction.Theoretically, it should be possible to recognize speech directly from the digitized waveform. However, because of the large variability of the speech signal, it is better to perform some feature extraction that would reduce that variability. Particularly, eliminating various source of information, such as whether the sound is voiced or unvoiced and, if voiced, it eliminates the effect of the periodicity or pitch, amplitude of excitation signal and fundamental frequency etc. The objective with feature extraction to attained are: 1.To untangle the speech signal into various acoustically identifiable components. 2.To obtain a set of features with low rates of change in order to keep computations feasible.

B. Mel frequency Cepstral Coefficient (MFCC) Processor

It is used to extract the features from voice and Vector quantization technique to identify the speaker. Voice has an infinitive amount of information, user has to determine who is the person speaking based on the features of the person´s voice. An analysis for the voice in time domain will be very impartial. So an analysis in the frequency domain can be a more viable option. Extract the parametric representation of voice signals is a vital process for the recognition performance. MFCC algorithm is used to simulate feature extraction module. Using MFCC algorithm, the Cepstral coefficients are calculated on Mel frequency scale. VQ (vector quantization) method will be used for reduction of amount of data to decrease computation time. In the feature matching stage Euclidean distance is applied as similarity criterion. Because of high accuracy of used algorithms, the accuracy of this voice command system is high.

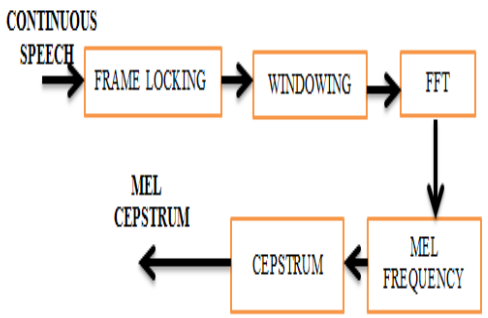

Fig 2: Block Diagram of MFCC

C. MFCC-HMM Algorithm

A Hidden Markov Model is a Finite State Machine having a fixed number of states. It is a statistical method of characterizing the spectral properties of the frames of a pattern. The underlying assumption of the HMM is that the speech can be well characterized as a parametric random process and that the parameters of the stochastic process can be estimated in a well defined manner. An approach to the recognition of speech signal using frequency spectral information with Mel frequency for the improvement of speech feature representation in a HMM based recognition approach. There are two strong reasons why Hidden Markov Model is used. First the models are very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. Secondly the models, when applied properly, work very well in practice for several important applications.

D. Mel-Frequency Wrapping

The speech signal consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured on the 'Mel' scale [8]. The Mel-frequency scale is linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. As a reference point, the pitch of a 1 kHz tone, 40dB above the perceptual hearing threshold, is defined as 1000 Mels. Therefore we can use the following formula to compute the Mels for a given frequency f in Hz:

$$\mathrm{Mel}(f) = 2595 * \log 10\,(1+f/700) \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

One approach to simulating the subjective spectrum is to use a filter bank, one filter for each desired Mel frequency component. The filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a Constant Mel-frequency interval [8].

E. Feature Recognition

The feature recognition process cuts the digitized audio signal, i.e. the sequence of sample values, into overlapping windows of equal length. The cut-out portions of the signal are called "frames", they are extracted out of the original signal every 10 or 20 ms. The length of a frame is about 30 ms. For speaker recognition tasks, sometimes longer frames are used in comparison to the feature extraction method used for speech recognition in order to increase spectral resolution. Each frame in the time domain is transformed to a MFCC vector. Therefore, the original speech signal is converted into a sequence of feature vectors, each vector representing cepstral properties of the signal within the corresponding window. The feature vector sequences of training and test utterances are the inputs of the classification step of a speaker recognition system.

F. Vector Quantization

Vector quantization (VQ) is a classical quantization technique from signal processing which allows the modeling of probability density functions by the distribution of prototype vectors. It was originally used for data compression. [3] It works by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms.

Figure 3 shows a conceptual diagram to illustrate this recognition process. In the figure, only two speakers and two dimensions of the acoustic space are shown. The circles refer to the acoustic vectors from the speaker 1 while the triangles are from the speaker 2.A speaker-specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors. The result code words (centroids) are shown in Figure 3 by black circles and black triangles for speaker 1 and 2,

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

respectively. The distance from a vector to the closest codeword of a codebook is called a VQ-distortion. In the recognition phase, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the total VQ distortion is computed. The speaker corresponding to the VQ codebook with smallest total distortion is identified as the speaker of the input utterance [3].

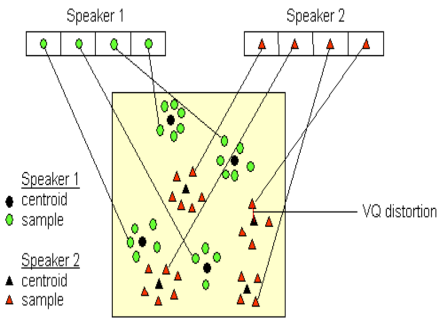

Fig 3: Conceptual diagram illustrating vector quantization codebook formation

III. DESIGN METHODOLOGY

The block diagram of "Wireless Monitoring for industrial automation is using Speech Recognition" as shown Figure 1, 2 and 3.

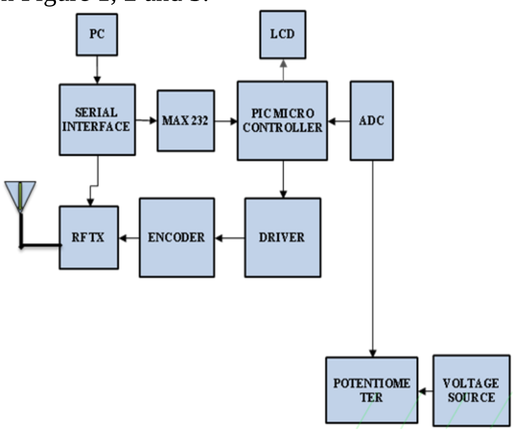

Fig 4: Block diagram of transmitter for Wireless Monitoring for industrial automation using speech recognition

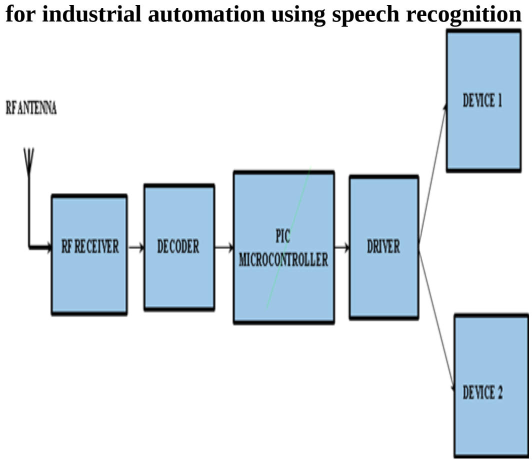

Fig 5: Block diagram of receiver for Wireless Monitoring for industrial automation using speech recognition

The block diagram of "Wireless Monitoring for industrial automation using Speech Recognition" consists of PIC Microcontroller, MAX 232, PC, RF Transmitter and Receiver, ADC, Potentiometer, encoder and decoder, Driver, Voltage source, LCD, MFCC, power supply and serial interface. Load the database through Matlab onto the PC. With MAX232 control voltage and reduce the loading effect to the Microcontroller. The MAX232 also acts as a buffer by converting the speech signal to digital value. Simultaneously apply the voltage from the source to potentiometer. Then apply this signal to the ADC in the microcontroller along with the signal from MAX 232. The single data value from the microcontroller is fed to the encoder where it will convert the data into multiple frequencies and is transmitted via RF antenna to the receiver antenna at a distance of 10-15meters. At the receiver, decoder converts multiple frequencies back to single value and finally passed onto a driver unit through which particular device will be selected. The voltage is monitored from the remote locations using the wireless technology. When the user wants to know the voltage status, that entire user needs to do is to speak as "status" in the data base. After this, the Matlab program will identify the speech and will read the status of the voltage from the microcontroller, then the data is then displayed on the LCD and the same is fed to the PC so that the person can get to know the status without touching the PC or by pressing any buttons.

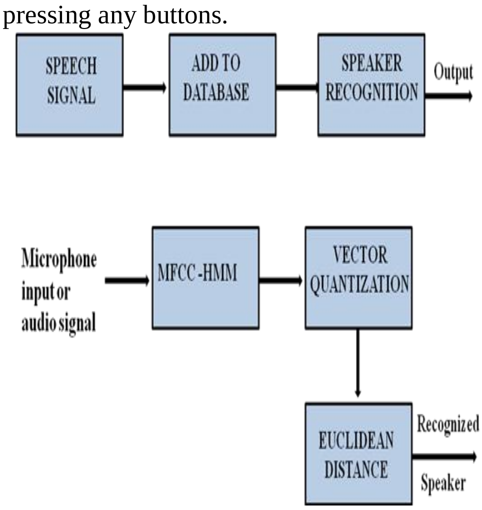

Fig 6: Block Diagram of Speech Recognition

The speech signal is added to the database. The new database added will be in the form of speech signal. The database contains recognition input for controlling and checking the status of the device Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker. Feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features from his/her voice input with the ones from a set of known speakers.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

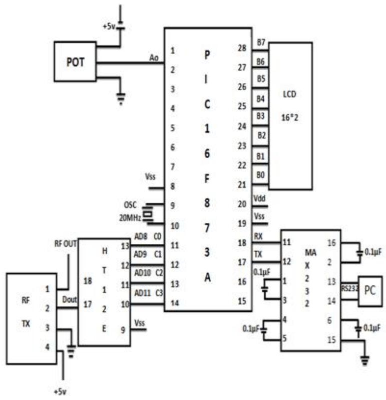

Fig 7: Circuit diagram of transmitter

Fig 7 describes the Speech signal is sent from the pc to the microcontroller. The Signal is in the form of digital data and the serial cable used is RS232.The Microcontroller and the PC is connected using the Max232 IC for the proper communication. The Microcontroller works with the logic of TTL logic, whereas the PC works with the RS232 logic hence to connect the PC and microcontroller a logic converter is needed here. Then the source voltage is displayed on the LCD and the microcontroller has in built ADC module.

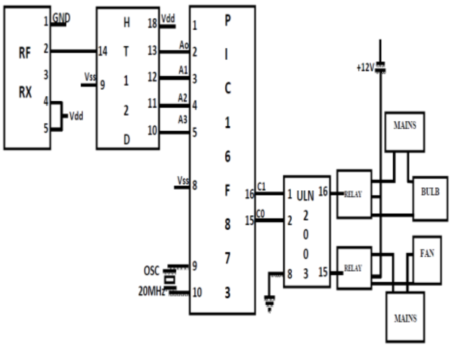

Fig 8: Circuit diagram of receiver

Receiver Circuit

Fig 8 describes the corresponding signal is received by the receiver antenna of the RF receiver module and is given to the decoder (HT12D).From the decoder it is given to the microcontroller and the signal is in the form of digital data. This data is given to the driver which is ULN2003 and the driver is nothing but an amplifier and from driver it is given to relay which is then given to corresponding devices**.**

A. PIC16F873A MICROCONTROLLER

PIC is a family of modified Harvard architecture microcontrollers made by Microchip Technology. The term PIC referred as Peripheral Interface Controller. PIC16F873A is a 28-pin 8-bit CMOS flash microcontroller which offers high performance for very low power consumption and price. It includes a very small instruction set of 35 single word instructions and a two stage pipeline concept fetch and execution of instructions. The core architecture is high performance Reduced Instruction Set Computer (RISC) CPU. PIC16F873A operates in operating speed of DC-20Mhz clock input at 5V, each instruction cycle takes 200ns. Program memory and data memory are two types of internal memories. Program memory is provided by 4K words of flash memory and data memory has two sources. One type of data memory is a 192-byte Random Access Memory (RAM) and other is 128-byte Electrically Erasable Programmable ROM (EEPROM) [6]. It includes 3 I/O ports such as port A, port B, port C. Port A is a bidirectional I/O port and is reserved for Analog to Digital conversion. Port B can be software programmed for internal weak pull-ups on all inputs and is responsible for LCD. Port C is a bidirectional I/O port and is responsible for USART (serial communication), timers and transmission/reception. It has five A/D input channels and has fourteen interrupt [6].

Fig 9: PIC16F873A Microcontroller

B. MAX232

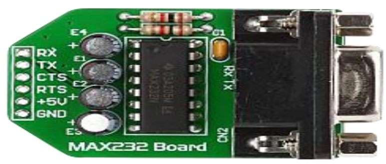

Fig 10: MAX232

The MAX 232 is an IC, first created in 1987 by maximum integrated products that convert signals from an RS232 serial port to signal suitable for use in TTL compatible digital logic circuit. The MAX232 is a dual driver/receiver and typically converts the RX, TX, CTS and RTS signal. The drivers provide RS232 voltage level output from a signal +5v supply via on chip charge pumps and external capacitors This makes it useful for implementing RS232 in devices that otherwise do not need any voltage outside the 0v to +5v range as power supply design doesn't need to be made mode complicated just for driving the RS232.The receiver reduce RS232 input to standard 5v TTL levels, these receivers have a typical threshold of 1.3v and a typical hysteresis of 0.5v.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

C. LCD Display

Fig 11 : 16X2 LCD display

Liquid crystals were first discovered in 1888. LCD (Liquid Crystal Display) screen is an electronic display module and find a wide range of applications. A 16x2 LCD display is very basic module and is very commonly used in various devices and circuits. A **16x2 LCD** means it can display 16 characters per line and there are 2 such lines. In this LCD each character is displayed in 5x7 pixel matrix. This LCD has two registers, namely, Command and Data.

The command register stores the command instructions given to the LCD. A command is an instruction given to LCD to do a predefined task like initializing it, clearing its screen, setting the cursor position, controlling display etc. The data register stores the data to be displayed on the LCD. The data is the ASCII value of the character to be displayed on the LCD. The LCD screen is more energy efficient and can be disposed of more safely than a CRT. Its low electrical power consumption enables it to be used in battery-powered electronic equipment.

D. Relay Driver (ULN2003) and Relay

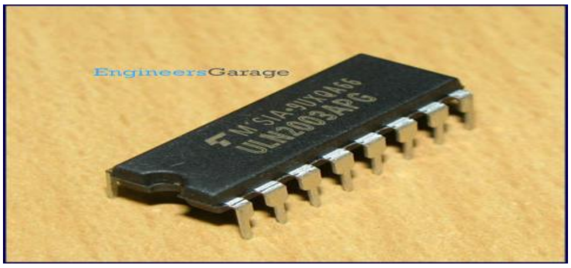

Fig 12: Relay driver uln2003

ULN2003 is a high voltage and high current Darlington array IC. It contains seven open collector Darlington pairs with common emitters. A darlington pair is an arrangement of two bipolar transistors. ULN2003 belongs to the family of ULN200X series of ICs. Different versions of this family interface to different logic families. ULN2003 is for 5V TTL, CMOS logic devices. These ICs are used when driving a wide range of loads and are used as relay drivers, display drivers, line drivers etc. Relay is one of the most important electromechanical devices highly used in industrial applications specifically in automation.. A relay is used for electronic to electrical interfacing i.e. it is used to switch on or off electrical circuits operating at high AC voltage using a low DC control voltage A relay generally has two parts, a coil which operates at the rated DC voltage and a mechanically movable switch. an be divided into two parts: input and output. The input section has a coil which generates magnetic field when a small voltage from an electronic circuit is applied to it. This voltage is called the operating voltage. Operating voltage used in this relay is 12V. There are 5 Pins in a relay: 2 pins for Control and 3 Pins for the Switching. 3 pins are NC (normal closed), NO (normal open) and C(com/pole) and 2 pins are +VCC and –VCC. NC and NO always connect to com pin; it is the moving part of the switch. Normally Open (NO), or "make" contact is open when the coil is de-energized and closes when the coil is energized. Normally Closed (NC), or "break" contact is closed in the de-energized position and opens when the coil is energized [7].

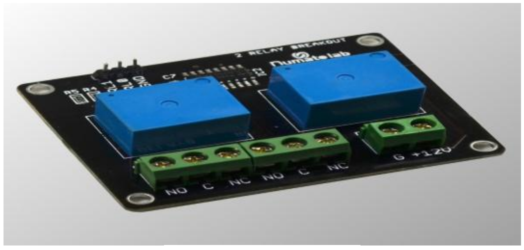

Fig 13: 5 pin relays

IV. SOFTWARE IMPLEMENTATION

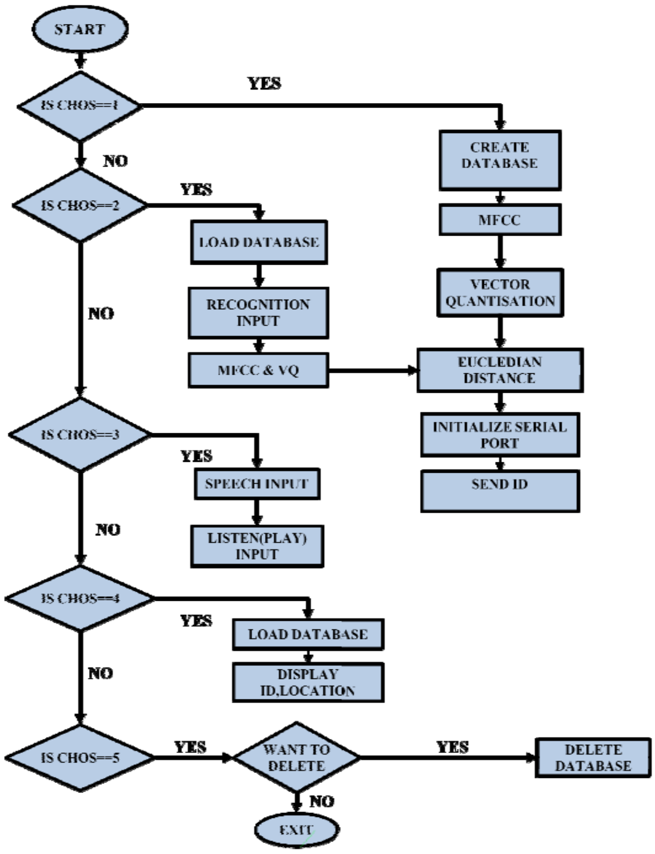

Fig 15: Flow chart of speech recognition

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

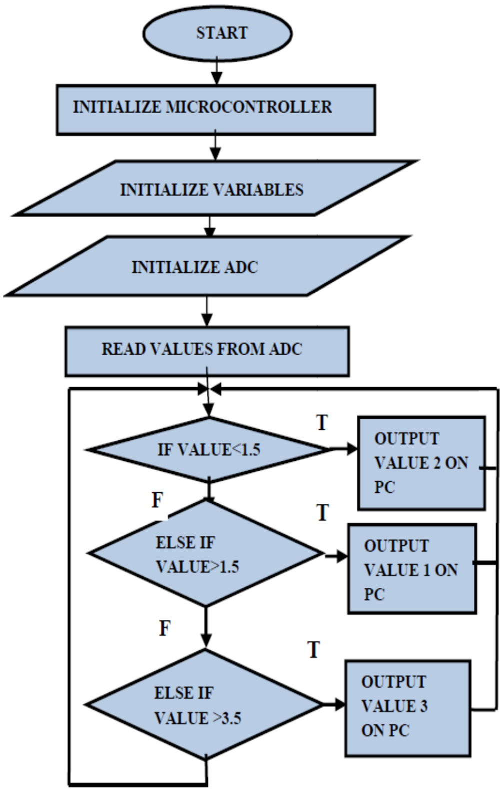

Fig 16: Flow chart of transmitter

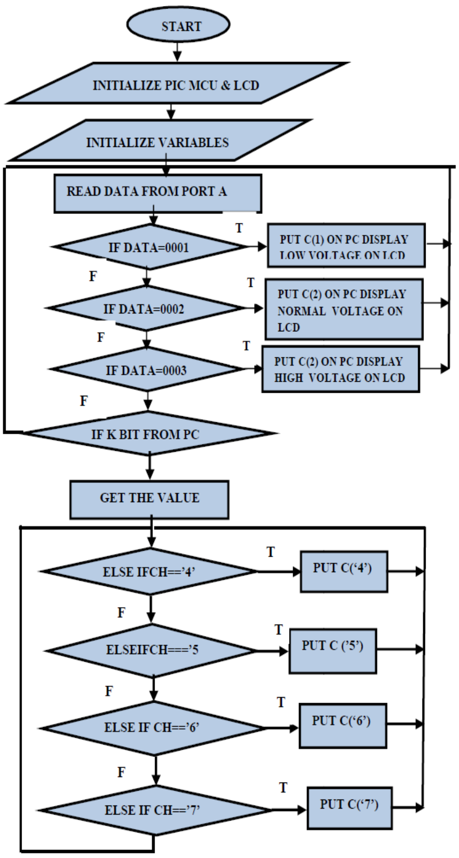

Fig 17: Flow chart of receiver

V. RESULTS

This Paper discusses about the result of the project. It consists of results and necessary outcomes of the project implementation. Figure 18 shows the voltage status of devices when spoken as "status" and corresponding ID 9 is generated.

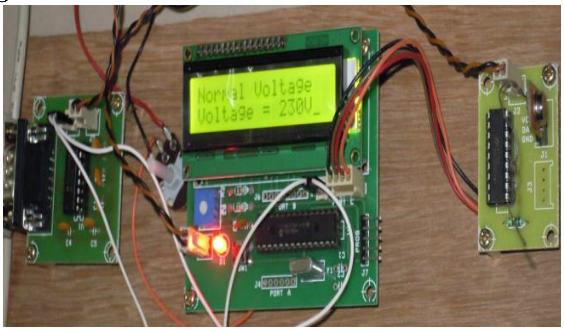

Fig 18: Transmitter module

Figure 19 shows energized devices when spoken as "bulb on" and "fan on" through microphone and corresponding ID 4 and 6 is generated that is stored in database.

Fig 19: Receiver module

VI. CONCLUSION

The main aim of this project is that the user can only use speech to monitor and also to know the status at any given point of time without much intervention with the PC. This project in designed a program to read the commands from the user, and LCD to display voltage status such as: Low Voltage varying from 0-1.5 Normal Voltage varying from 1.5-3.5 High Voltage varying from 3.5-6.5With the design of the interface to the PC using the micro controller and the associated circuits and developed with wireless communication technology for the same in the real time environment. The efficiency of the system can be increased by improving speaker identification and by developing the Mel scale.

REFERENCES

[1] "Fundamental of Speech Recognition", "Lawrence Rabiner and Biing-Hwang Juang, Prentice-Hall, Englewood Cliffs".

[2] "Vector Quantization"," IEEE ASSP Magazine", pp. 4--29, April 1984.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 11, May 2014

[3] "Dipmoy Gupta, Radha Mounima C,Navya Manjunath, Manoj PB", "Isolated Word Speech Recognition Using Vector Quantization (VQ)", International Journal of Advanced Research in Computer Science and Software Engineering, vol 2, pp. 166-167, May 2012.

[4] "F. Soong, E. Rosenberg, B. Juang, and L. Rabiner", "A Vector Quantization Approach to Speaker Recognition", "AT&T Technical Journal", vol. 66, pp. 14-26,March/April 1987.

[5] "Ben Gold and Nelson Morgan, John Willy & Sons", "Speech and Audio Signal Processing", Part- IV, Chap.14, pp. 189-203, 2002.

[6] http://akizukidenshi.com//PIC16F877A

[7] http://www.pc-control.co.uk/relays

[8] "Vibha Tiwari", "MFCC and its applications in speaker recognition", pp. 20-21, Feb-2010.

AUTHOR'S PROFILE

Mrs. Hussana Johar R.B joined the faculty of Telecommunication Department at GSSSIETW, Mysore in 2012. She has secured 3rd rank in her M.Tech in 2011 from Department of Electronics and communication at MCE Hassan, she also received UG in E&C from NIE, Mysore. Her recent work besides multiple sign language into voice message has focused on the theory of body area network, wireless sensor networks to provide coverage and connectivity in sensor network devices. She also published a paper on Modified dynamic evaluation model for threshold cryptography in MANET's and received best paper award.

Ms. Soujanya R pursuing her final year engineering in telecommunication stream at GSSSIETW Mysore, Karnataka, India.

Ms. Sushma S pursuing her final year engineering in telecommunication stream at GSSSIETW Mysore, Karnataka, India.

Ms. Syeda Ansha Sami pursuing her final year engineering in telecommunication stream at GSSSIETW Mysore, Karnataka, India.

Ms. Supreetha M G pursuing her final year engineering in telecommunication stream at GSSSIETW Mysore, Karnataka, India.