# DEVELOPMENT OF A DOOR LOCK SECURITY SYSTEM BASED ON AUTOMATIC SPEECH RECOGNITION


**BY**

**NGENE, JOSHUA CHISOM**

**CPE/12/0888**


**AN UNDERGRADUATE PROJECT SUBMITTED**

**TO**


**THE DEPARTMENT OF COMPUTER ENGINEERING,**

**FACULTY OF ENGINEERING**

**FEDERAL UNIVERSITY, OYE-EKITI,**

**EKITI STATE, NIGERIA**


**IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF BACHELOR OF ENGINEERING (B.ENG) IN COMPUTER ENGINEERING**


**NOVEMBER, 2017**

# DECLARATION

I Ngene, Joshua Chisom hereby declare that this project work carried out is the result of my personal effort under the supervision of Engr. Nnamdi Okomba of the department of Computer Engineering, Federal university Oye-Ekiti, Ekiti State, as part of the requirement for the award of Bachelor Degree of Computer Engineering, and has not been submitted elsewhere for this purpose. All sources of information are explicitly acknowledged by means of reference.

**Copyright**



Student's full name: NGENE JOSHUA CHISOM ........ Sign. & Date: 06/12/2017

# CERTIFICATION

This project report with the title "Development of a door lock security system based on automatic speech recognition"

Submitted by

**NGENE, JOSHUA CHISOM**

**CPE/12/0888**

Has satisfied the requirements for the award of the degree of

Bachelor of Engineering (B.Eng.) in Computer Engineering

Federal University Oye-Ekiti, Ekiti State, Nigeria.

………………………………………… 15-12-2017
…………………………………………

ENGR. NNAMDI OKOMBA Date

Supervisor

………………………………………… 12-03-2018
…………………………………………

Dr. I ADEYANJU Date

Head of Department

DEPARTMENT OF COMPUTER ENGINEERING
HOD SIGN
DATE
FEDERAL UNIVERSITY OYE EKITI

# DEDICATION

I dedicate this work to the Almighty God, whose grace has kept me till this moment, to my family and friends whose motivation have moved me to this point. I also dedicate this to the entire Computer Engineering department staff and students of Federal University Oye-Ekiti, Ekiti.

# ACKNOWLEDGEMENTS

My first thanks goes to the Almighty God who saw me through this program.

I wish to express my genuine gratitude to the contribution of various individuals who contributed to my acquisition of knowledge and successful completion of this project.

Noteworthy thanks are extended to my Supervisor Engr. Nnamdi Okomba for his love, and patience in helping me practically, and gave his views as at when needed.

Also my appreciation goes to my parent, Mrs. Ngene, my brother and sisters, and my humble self, for our endless prayers.

I also thank The Federal Government of Nigeria for creating this platform for students to gain practical knowledge in their respective field of study.

To the rest of my wonderful family, friend, colleagues, and all not mentioned, I would like to say thank you and may God bless us all, Amen.

# ABSTRACT

The security of our homes and properties are of utmost importance as various individuals and engineers have worked on ways to improve the security of our homes. This project which is "Development of a door lock security system based on speech recognition" is aimed at introducing maximum security and assurance to individuals and their properties in their various homes.

This project was developed based on the use of the MATLAB software, which is programmed to recognize certain users and give a particular user (amongst the recognized users) control over the access to home for security purpose. In the training, MFCC feature extraction technique was used for extraction of appropriate features from the user's speech signal. Vector Quantization using LBG algorithm (VQLBG) was also employed in the recognition phase of the speaker recognition system.

This project was finally implemented/tested using a model home entrance door to show its reliability and efficiency. Other components such as DC motor, for movement of the door (open and close), H-bridge motor controller (L293D), for the directional control of the DC motor, PIC18F1X20 microcontroller and a display LCD, were used on the model house door to accomplish this project.

## TABLE OF CONTENT


Title Page……………………………………………………………………………………i

Declaration……………………………………………………………………………………ii

Certification…………………………………………………………………………………iii

Dedication……………………………………………………………………………………iv

Acknowledgment……………………………………………………………………………v

Abstract………………………………………………………………………………………vi

Table Of Content……………………………………………………………………………vii

List Of Figures………………………………………………………………………………xi

List Of Tables………………………………………………………………………………xiv

List of Abbreviations………………………………………………………………………xv

**CHAPTER ONE: INTRODUCTION**

1.1 Preamble……………………………………………………………………………1

1.2 Statement Of Problem………………………………………………………………1

1.3 Project Aim…………………………………………………………………………2

1.4 Objectives……………………………………………………………………………2

1.5 Scope of study………………………………………………………………………2

1.6 Importance and significance of study ……………………………………………3

1.7 Method of study……………………………………………………………………3

1.7.1 Project planning..........3

1.8 Report outline..........4

## CHAPTER TWO: LITERATURE REVIEW

2.0 Overview of Automatic Speech Recognition (ASR) System..........5

2.1 Human auditory system..........5

2.2 Elements of a Language..........6

2.3 Types of Speech Recognition..........6

2.4 Automatic Speech Recognition system classifications..........7

2.5 Steps in ASR..........8

2.5.1 Speech signal Acquisition..........9

2.5.2 Feature Extraction in speech recognition..........9

2.5.3 Recognition stage..........15

2.6 Approaches to ASR by Machine..........24

2.7 Matching Techniques..........26

2.8 Performance Evaluation of Speech Recognition Systems..........26

2.9 Security systems..........27

2.10 ASR projects..........31

2.11 Project work..........32

## CHAPTER THREE: METHODOLOGY

3.1 Introduction..........33

3.2 Components Theory..........33

3.2.1 The Microphone Circuit....................................................34
3.2.2 LCD Display....................................................35
3.2.3 Microcontroller (PIC18F1X20)....................................................36
3.2.3 H-bridge DC motor controller (L293D)....................................................37
3.2.4 Gear system and DC motor....................................................39
3.3 System analysis....................................................39
3.4 Design of power supply for circuit....................................................41
3.4.1 Power supply stage from ac power supply....................................................41
3.4.2 Power supply requirements....................................................42
3.4.3 Estimation of current requirement....................................................42
3.5 Product Operation flow....................................................44
3.5.1 Assembly process....................................................44
3.5.2 Design alternatives....................................................45
3.5.3 Operational Block Diagram....................................................45

**CHAPTER FOUR: SYSTEM EXECUTION, RESULTS, IMPLEMENTATION AND DISCUSSIONS**

4.1 Introduction....................................................48
4.2 Design of the speech system database ....................................................48
4.3 Functions used in the speech recognition system ....................................................52
4.4 Project execution and result evaluation of the speech system ....................................................54
4.4.1 Software execution and result discussion....................................................54
4.4.2 Hardware implementation and testing explanation....................................................63

**CHAPTER FIVE: CONCLUSION, PROBLEMS ENCOUNTERED AND RECOMMENDATIONS**

5.1 Conclusion........................................................................................69

5.2 Problems Encountered........................................................................70

5.3 Recommendations...............................................................................70

*References*.........................................................................................72

APPENDIX I: COST AND PROJECT COMPONENTS.........................................75

## LIST OF FIGURES

Figure 2.1: Speech processing classification........................................................8

Figure 2.2: Block Diagram of LPC..........................................................................11

Figure 2.3: Steps involved in MFCC Feature Extraction.......................................13

Figure 2.4: Block diagram of PLP speech analysis method....................................14

Figure 2.5: Block Diagram of RASTA......................................................................15

Figure 2.6: A simple Hidden Markov Model...........................................................16

Figure 2.7: Biological Neural Network.....................................................................19

Figure 2.8: Architecture of ANN..............................................................................19

Figure 2.9: VQ recognition process..........................................................................22

Figure 2.10: Speech Recognition Techniques...........................................................25

Figure 3.1: Microphone Circuit.................................................................................35

Figure 3.2: 16x2 LCD display...................................................................................36

Figure 3.3: PIC18F1X20 microcontroller.................................................................36

Figure 3.4: 16-pin L293 pin diagram........................................................................38

Figure 3.5: A Gearbox...............................................................................................39

Figure 3.6: Schematics of the hardware door control circuit ..................................40

Figure 3.7: Power Supply Circuit .............................................................................41

Figure 3.8: Power supply stage design......................................................................42

Figure 3.9: Rectified DC waveform....43

Figure 3.10: Data Acquisition and Storage flow diagram....45

Figure 3.11: Recognition and Access control flow diagram....46

Figure 3.12: Hardware Implementation flow diagram....47

Figure 4.1: Speaker recognition system menu....49

Figure 4.2: Empty database signal in system....51

Figure 4.3: Acquisition of Speech Data "Sound Id: 1"....55

Figure 4.4: Acquisition of Speech Data "Sound ID: 2"....56

Figure 4.5: Acquisition of Speech Data "Sound ID: 3"....56

Figure 4.6: Acquisition of speech data "sound ID: 4"....57

Figure 4.7: Acquisition of speech data "sound ID: 5"....58

Figure 4.8: Recognized speaker "sound ID: 4"....59

Figure 4.9: Recognized speaker "sound ID: 3"....60

Figure 4.10: Recognized speaker "sound ID: 5"....60

Figure 4.11: Recognized speaker "sound ID: 1", the authorized user....61

Figure 4.12: Recognized Speaker "Sound Id: 2"....62

Figure 4.13: PCB layout of door control circuit....65

Figure 4.14: Diagram of the Internal Arrangement of the Prototype/Model House Door Showing Its Components......66

Figure 4.15: Door Control Circuit......67

Figure 4.16: 3d Visualization of Door Control Circuit (Side View)......67

Figure 4.17: 3d Visualization of Door Control Circuit (Top View)......68

Figure 4.18: The Prototype/Model Door to Be Controlled By The Speech Recognition System......68

# LIST OF TABLES

TABLE 1. PIN FUNCTION OF L293D....................................................38

TABLE 2: PARAMETRIC REPRESENTATION OF RECOGNITION PROCESS OF USERS.........................................................................................63

# LIST OF ABBREVIATIONS

1. Automatic Speech Recognition (ASR)
2. Word Error Rate (WER)
3. Light Emitting Diode (LED)
4. Complimentary Metal-Oxide Semiconductor (CMOS)
5. Direct Current (DC)
6. Operational Amplifier (OP-AMP)

# CHAPTER ONE

# INTRODUCTION

## 1.1 PREAMBLE

Various control systems have been designed over the years to prevent access by unauthorized user. The main aim for providing locks for our home, school, office, and building is for security of our lives and property. It is therefore important to have convenient way of achieving this goal.

Door security system have become a standard feature on many different types of buildings and homes. And they are becoming popular every day to develop an effective electronic devices which provide security. Home security has been a major issue because of the increase in crime rate and everybody wants to take proper action to prevent unauthorized user.

## 1.2 STATEMENT OF PROBLEM

Generally, Security is a prime concern in our day-to-day life. Everyone wants to be as much secure as possible. An access control for doors forms a vital link in a security chain. The main purpose of these has in past few year been altered due to the breach of security of various home and offices. The increase in the rate of access to home or offices by unwanted or unauthorized person has drastically increased keeping properties and lives at risk, e.g. the 'Password security door lock system' has various times been by-passed due to the fact that it makes use of digits( ASCII characters) which can easily be manipulated, interpolated or even guessed.

Therefore it is necessary to improve the security systems so as to reduce or eradicate unwanted or unauthorized access to our concerned domain which is the home door security system. This will be done by introducing the 'voice/speech recognition based security system' which will greatly improve and make the security system more sophisticated and thereby making the door security system less vulnerable to unwanted and unauthorized users.

## 1.3. PROJECT AIMS

The main aim of this project is to improve the security of homes by the development of the 'security system based on speech recognition', through the use of readily available and low cost materials. This project is also aimed at neutralizing security to every class of individuals.

## 1.4. PROJECT OBJECTIVE

The main purpose of implementing this project is to have a working prototype of a speech recognition based security system providing a practical tool for access control of individual properties in our homes.

The main objectives are outlined as:

1. To design a security lock system using the speech/voice recognition as form of access control.
2. To implement the design using a hardware prototype.
3. To evaluate the effectiveness of the system.

## 1.5.SCOPE OF STUDY

Project concentrates on a development of a speech recognition based security system. To develop the whole project, it consists of two phases, which are the concept of the program/code structure and the construction of physical model.

The concept of code structure consist of software (MATLAB) used for coding of the speech recognition system, While the physical model which is a small model house entrance doors, in which the circuitry has to control.

## 1.6.IMPORTANCE AND SIGNIFICANT OF THE STUDY

The project is intended to make a door security system more secure to users which is cost efficient and have a highly functional design. It also reduces access as the mode of control is speech recognition based and it is known that no two persons in the world can have matching speech spectrum. This makes the access limited to the user only with 99% guarantee he/she is safe from security breach. The project is highly significant for security purposes as continuously, new technologies are been developed and introduced just to have more security at our homes.

The purpose of this project is to test and challenge the academic knowledge that has been gained so far throughout our entire course period of five years. This project requires us to utilize knowledge from all our subjects. All parts and parcels of this project are expected to work well in order to achieve perfect project function.

In upcoming years, an advanced version of this project will be built incorporated with the latest technology and more advanced features. This project is expected to be the model for future speech recognition door systems around the world.

## 1.7. METHODS OF STUDY

### 1.7.1 PROJECT PLANNING

Throughout this entire project, a proper planning has been carried out to identify the tasks and also as a guideline to complete this project. This project was organized based on ideas from literature survey and discussion through many source and information gather from internet web sites and books. The tasks are arranged as follow:

- Literature Survey
- Project Design (Hardware and Software)
- Circuit Testing
- Project Construction
- Troubleshooting

### 1.7.2 REPORT OUTLINE

Chapter one explains the preamble, statement of problem, the project aim, the project objective, the project scope, and project significance and method of study. Chapter two entails the literature review of speech recognition based security systems. The overview of speech recognition are the major element as a guide for the development of the speech recognition based security system.

Chapter three focuses on the methodologies for the development of the program/code structure and the implementations of driver circuitry. It gives a brief review on the components theory of the hardware for the speech recognition based security system implementation.

Chapter four discusses on the results obtained from the whole project. All discussions are concentrating on the result and performance of the speech recognition based security system. The discussion is valuable for future development of the proposed system.

Chapter five discusses about the conclusion on development of the speech recognition based security system. The recommendations and problems encountered with preferred solutions required on this project are stated in this chapter for further development.

## CHAPTER TWO

## LITERATURE REVIEW

### 2.0 An Overview of Automatic Speech Recognition (ASR) System.

Speech is the most natural form of human communication. Automatic Speech Recognition (ASR) is the process of converting a speech signal to a sequence of words, by means of an algorithm. ASR system involves two phases. Training phase and Recognition phase. In training phase, known speech is recorded and parametric representation of the speech is extracted and stored in the speech database. In the recognition phase, for the given input speech signal the features are extracted and the ASR system compares it with the reference templates to recognize the utterance. In a speech recognition system, many parameters affect the accuracy of recognition such as vocabulary size, speaker dependency, speaker independence, time for recognition, type of speech (continuous, isolated) and recognition environment condition. The extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any speech recognition system. Speech recognition algorithm consists of several stages in which feature extraction and classification are mainly important.

### 2.1 Human auditory system.

To model a human hearing system, it is important to understand the working of human auditory system. At the linguistic level of communication first the idea is formed in the mind of the speaker. The idea is then transformed to words, phrases and sentences according to the grammatical rules of the language. At the physiological level of communication the brain creates electric signals that move along the motor nerves. These electric signals activate muscles in the vocal tract and vocal cords. This vocal tract and vocal cord movements results in pressure changes within the vocal tract and in particular at the lips, initiates a sound wave that propagates in space. Finally at the linguistic level of the listener, the brain performs speech recognition and understanding. (Rabiner, 1993; Thomas, 2002; Lawrence, 2000).

## 2.2 Elements of a Language

A fundamental distinctive unit of a language is a phoneme. Different languages contain different phoneme sets. Syllables contain one or more phonemes, while words are formed with one or more syllables, concatenated to form phrases and sentences. One broad phoneme classification for English is in terms of vowels, consonants, diphthongs, affricates and semi vowels. (Thomas, 2002; Lawrence, 2000)

## 2.3 Types of Speech Recognition

Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize. These classes are classified as following.

1. **Isolated Words**

Isolated word recognizers usually require each utterance to have quiet (lack of an audio signal) on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses). Isolated Utterance might be a better name for this class.

2. **Connected Words**

Connected word systems (or more correctly 'connected utterances') are similar to isolated words, but allows separate utterances to be run-together with a minimal pause between them.

3. **Continuous Speech**

Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. (Basically, it's computer dictation). Recognizers with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine utterance boundaries.

4. **Spontaneous Speech**

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features (Mevsam, 2009; Simon, 2006).

## 2.4 Automatic Speech Recognition system classifications

The following tree structure emphasizes the speech processing applications. Depending on the chosen criterion, Automatic Speech Recognition systems can be classified as shown in Figure 2.1.

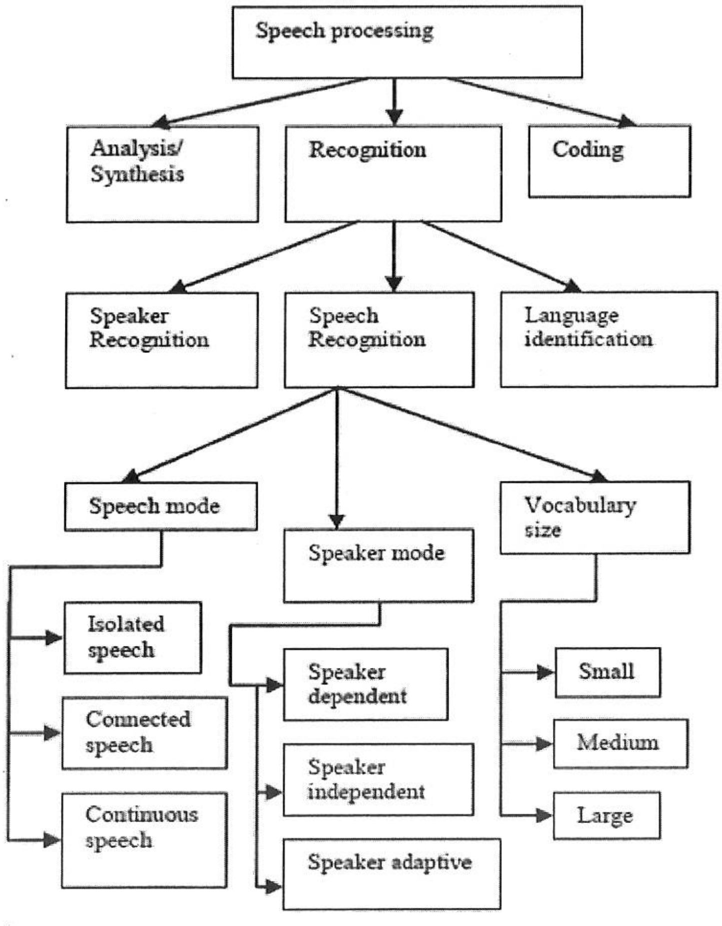


Figure 2.1: Speech processing classification

## 2.5 Steps in ASR

Modules that are identified to develop a speech recognition system are

1) Speech Signal acquisition

2) Feature Extraction

3) Recognition

### 2.5.1 Speech signal Acquisition

Much of the success of a speech recording depends on the recording environment and microphone placement. Ideally, speech recordings should take place in soundproof studios or labs. If those are not available, one should try to find a relatively quiet room with as little low-frequency noise as possible. Most typical sources of low-frequency noise include 60-Hz hum from electrical equipment, heating and air conditioning ducts, elevators, doors, water pipes, computer fans, and other mechanical systems in the building. If possible, those devices should be switched off during recording (Bartek, 2011), microphones, praat, audacity, sphinx, Julius are the various tools which are being used by researchers for recording speech database (Mathur, 2010).

### 2.5.2 Feature Extraction in speech recognition.

In speech recognition, feature extraction requires much attention because recognition performance depends heavily on this phase. The main goal of the feature extraction step is to compute a parsimonious sequence of feature vectors providing a compact representation of the given input signal. The feature extraction is usually performed in three stages. The first stage is called the speech analysis or the acoustic front end temporal analysis of the signal and generates raw features describing the envelope of the power spectrum of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features. Finally, the last stage transforms these extended feature vectors into more compact and robust vectors that are then supplied to the recognizer. (Jain, 2011).

### A. Mel Cepstrum Analysis

This analysis technique uses cepstrum with a nonlinear frequency axis following *mel*scale (Shaughnessy, 2001). For obtaining *mel*cepstrum the speech waveform *s(n)* is first windowed with analysis window *w(n)* and then its DFT *S(k)* is computed. The magnitude of *S(k)* is then weighted by a series of *mel*filter frequency responses whose center frequencies and bandwidth roughly match those of auditory critical band filters.

### B. Linear Prediction Coefficient

LPC is one of the most powerful speech analysis techniques and is a useful method for encoding quality speech at a low bit rate. The basic idea behind linear predictive analysis is that a specific speech sample at the current time can be approximated as a linear combination of past speech samples. LP is a model based on human speech production. It utilizes a conventional source-filter model, in which the glottal, vocal tract, and lip radiation transfer functions are integrated into one all-pole filter that simulates acoustics of the vocal tract.

The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficients are estimated every frame, which is normally 20 ms long. The predictor coefficients are represented by ak. Another important parameter is the gain (G). The transfer function of the time varying digital filter is given by

$$H(z) = G/(1-\Sigma akz\text{-}k) \;\ldots\ldots\ldots\ldots (2.1)$$

Where k=1 to p, which will be 10 for the LPC-10 algorithm and 18 for the improved algorithm that is utilized. Levinsion-Durbin recursion will be utilized to compute the required parameters for the auto-correlation method.

The LPC analysis of each frame also involves the decision-making process of voiced or unvoiced. A pitch-detecting algorithm is employed to determine to correct pitch period / frequency. It is important to re-emphasis that the pitch, gain and coefficient parameters will be varying with time from one frame to another.

In reality the actual predictor coefficients are never used in recognition, since they typical show high variance. The predictor coefficient is transformed to a more robust set of parameters known as cepstral coefficients.

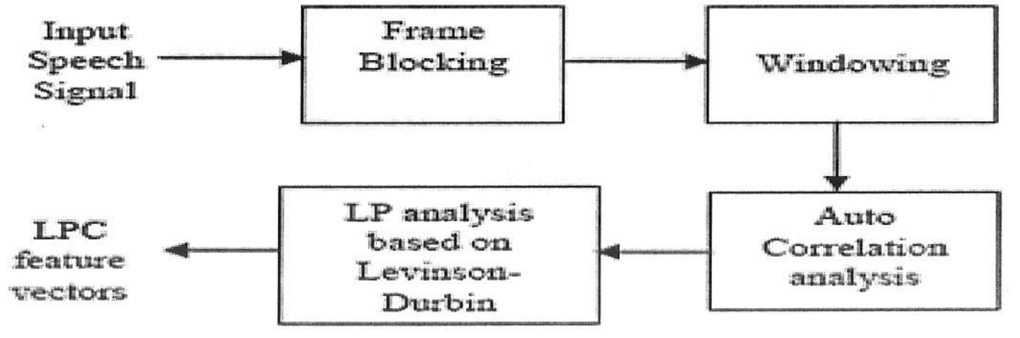


Figure 2.2: Block Diagram of LPC (Atal, 1974).

The parameters involved in performance evaluation of LPC's are Bit rates, Overall delay of the system, Computational complexity and Objective Performance evaluation.

Following are the types of LPC

- Voice-excitation LPC
- Residual Excitation LPC
- Pitch Excitation LPC
- Multiple Excitation LPC(MPLPC)
- Regular Pulse Excited LPC(RPLP)
- Coded Excited LPC(CELP)

Below are some advantages of LPC

The main advantage of linear predictive coding is to reduce the bitrates of the speech i.e. reduces the size of the transmitting signal. The signal transmitted through LPC required less bandwidth and hence no. of users can be increased. Finally, this method of coding uses the encryption of data so the data is secured until the destination.

Some disadvantages are also seen which are; Due to reduce in the bitrates of the speech signal, the quality of voice signal is reduced. This technique is lossy compression technique, hence data gets faded if transmitted to the long distance.

### C. Mel-Frequency Cepstrum Coefficients (MFCC)

A compact representation would be provided by a set of Mel-Frequency Cepstrum Coefficients (MFCC), which is the result of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale. (Pols, L.C.W, 1966). The performance of the Mel-Frequency Cepstrum Coefficients (MFCC) may be affected by the number of filters, the shape of filters, the way that filters are spaced and the way that the power spectrum is warped. The traditional MFCC calculation excludes the 0th coefficient. Fang Zheng, Guoliang Zhang and Zhanjiang Song have proposed that it can be regarded as the generalized Frequency Band Energy (FBE) and is hence useful, which results in the FBE-MFCC. (Fang, 2001; Hossan, 2010).

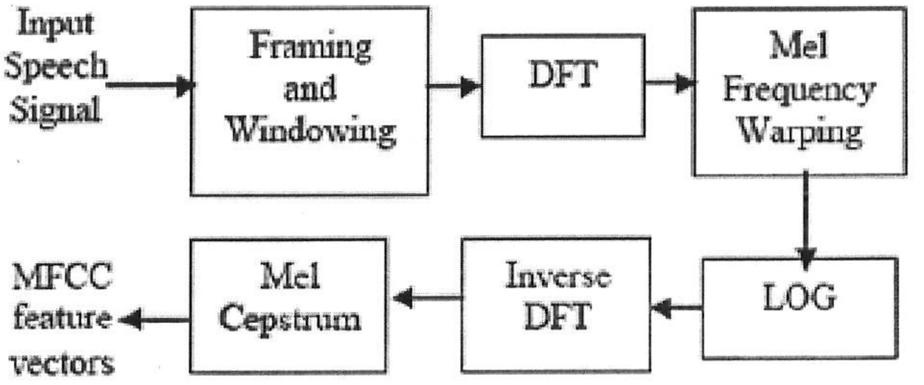


Figure 2.3: Steps involved in MFCC Feature Extraction (Pols, L.C.W, 1966).

As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system and this is an advantage of MFCC.

On disadvantage, MFCC values are not very robust in the presence of additive noise, and so it is common to normalize their values in speech recognition systems to lessen the influence of noise.

In application, MFCCs are commonly used as features in speech recognition systems, such as the systems which can automatically recognize numbers spoken into a telephone. They are also common in speaker recognition, which is the task of recognizing people from their voices.

MFCCs are also increasingly finding uses in music information retrieval applications such as genre classification, audio similarity measures, etc.

**D. Perceptually Based Linear Predictive Analysis (PLP)**

H.Hermansky, B. A. Hanson, H. Wakita proposed a new PLP analysis (Hermansky et al, 1985), which models perceptually motivated auditory spectrum by a low order all pole function, using the autocorrelation LP technique. This technique was mainly focused in cross-speaker isolated word recognition. PLP analysis results also demonstrated that speech representation is more consistent than the standard LP method. Basic concept of PLP method is shown in block diagram of Fig. 2.4.

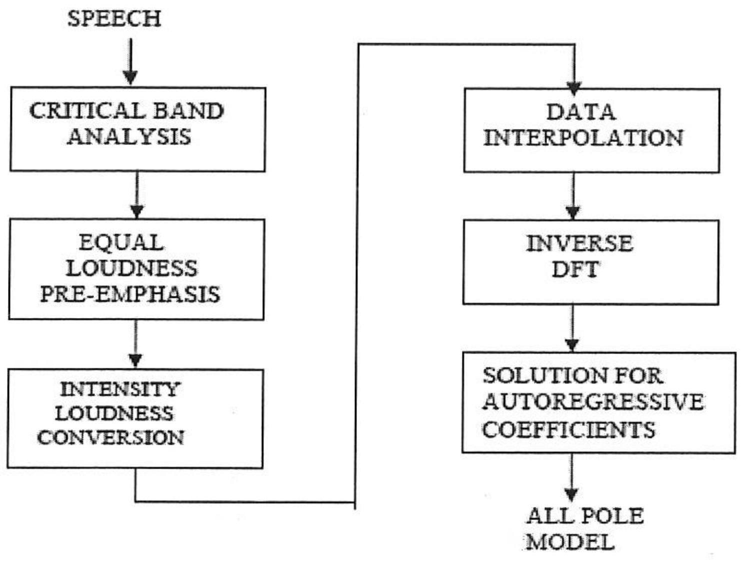


Fig 2.4: Block diagram of PLP speech analysis method (Hermansky et al, 1985).

It involves two major steps: Obtaining auditory spectrum and approximating the auditory spectrum by an all pole model. Auditory spectrum is derived from the speech waveform by critical-band filtering, equal loudness curve pre-emphasis, and intensity loudness root compression. The PLP analysis provides similar results as with LPC analysis but the order of PLP model is half of LP model. This allows computational and storage saving for ASR. (Hermansky, 1986; Atal, 1974).

Advantages

- PLP coefficients are often used because they approximate well the high-energy regions of the speech spectrum while simultaneously smoothing out the fine harmonic structure, which is often characteristic of the individual but not of the underlying linguistic unit.
- LPC, however, approximates the speech spectrum equally well at all frequencies, and this representation is contrary to known principles of human hearing.

- The spectral resolution of human hearing is roughly linear up to 800 or 1000 Hz, but it decreases with increasing frequency above this linear range.
- PLP incorporates critical-band spectral-resolution into its spectrum estimate by remapping the frequency axis to the Bark scale and integrating the energy in the critical bands to produce a critical-band spectrum approximation

**E. Relative Spectra Filtering of Log Domain Coefficients (RASTA)**

To compensate for linear channel distortions the analysis library provides the power to perform rasta filtering. The rasta filter is used either within the log spectral or cepstral domains. In result the rasta filter band passes every feature coefficient. Linear channel distortions seem as an additive constant in each the log spectral and therefore the cepstral domains. The high-pass portion of the equivalent band pass filter alleviates the result of convolution noise introduced in the channel. The low-pass filtering helps in smoothing frame to border spectral changes.

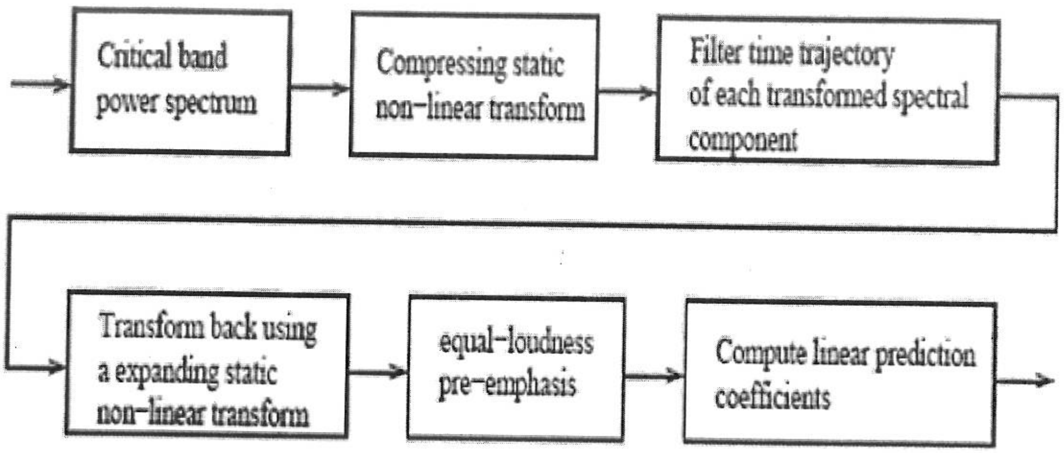


Figure 2.5: Block Diagram of RASTA (Jane and Saxena, 2011).

### 2.5.3 RECOGNITION STAGE

There are mainly 3 algorithms that are used for SR. Those are given below:

a) Hidden Markov Model(HMM)

b) Dynamic Time Warping(DTW)

c) Artificial Neural Networks(ANN)

d) Vector Quantization(VQ)

Above algorithms are explained in detail in further sections.

### a) HIDDEN MARKOV MODEL (HMM)

A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (*hidden*) states. An HMM can be presented as the simplest dynamic Bayesian network. A Hidden Markov Model is a collection of states connected by transitions, as illustrated in Figure 2.6. It begins in a designated initial state. In each discrete time step, a transition is taken into a new state, and then one output symbol is generated in that state. The choice of transition and output symbol are both random, governed by probability distributions. The HMM can be thought of as a black box, where the sequence of output symbols generated over time is observable, but the sequence of states visited over time is hidden from view. This is why it's called a *Hidden* Markov Model.

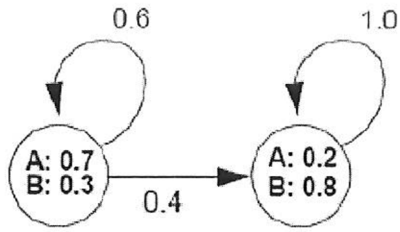


Figure 2.6: A simple Hidden Markov Model, with two states and two output symbols, A and B. (Meysam and Fardad, 2009)

HMMs have a variety of applications. When an HMM is applied to speech recognition, the states are interpreted as acoustic models, indicating what sounds are likely to be heard during their corresponding segments of speech; while the transitions provide temporal constraints, indicating how the states may follow each other in sequence. Because speech always goes forward in time, transitions in a speech application always go forward.

Algorithms of HMM;

There are three basic algorithms associated with Hidden Markov Models:

- Forward algorithm, useful for isolated word recognition;
- Viterbi algorithm, useful for continuous speech recognition; and
- Forward-backward algorithm, useful for training an HMM.

Limitations of HMM;

- Constant observation of frames
- The Markov assumption
- Lack of formal methods for choosing a model topology
- Large amounts of training data required
- Weak duration modelling
- Restricted output PDFs
- The assumption of conditional independence

### b) DYNAMIC TIME WARPING (DTW)

The simplest way to recognize an *isolated* word sample is to compare it against a number of stored word templates and determine which the "best match" is. This goal is complicated by a number of factors. First, different samples of a given word will have somewhat different durations. This problem can be eliminated by simply normalizing the templates and the unknown speech so that they all have an equal duration. However, another problem is that the rate of speech may not be constant throughout the word; in other words, the optimal alignment between a template and the speech sample may be nonlinear. Dynamic Time Warping (DTW) is an efficient method for finding this optimal nonlinear alignment.

Dynamic time warping (DTW) is a well-known technique to find an optimal alignment between two given (time-dependent) sequences under certain restrictions intuitively; the sequences are warped in a nonlinear fashion to match each other. Originally, DTW has been used to compare different Speech patterns in automatic speech recognition. In fields such as data mining and information retrieval, DTW has been successfully applied to automatically cope with time deformations and different speeds associated with time-dependent data.

In time series analysis, dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed. For instance, similarities in walking patterns could be detected using DTW, even if one person was walking faster than the other can also detected by DTW.

Problem of finding an average sequence for a set of sequences. The average sequence is the sequence that minimizes the sum of the squares to the set of objects.

### c) ARTIFICIAL NEURAL NETWORKS(ANN)

A neural network can be defined as a model of reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells, or basic information-processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections, *synapses*, between them. By using multiple neurons simultaneously, the brain can perform its functions much faster than the fastest computers in existence today.

Each neuron has a very simple structure, but an army of such elements constitutes a tremendous processing power. A neuron consists of a cell body, soma, a number of fibers called dendrites, and a single long fiber called the axon.

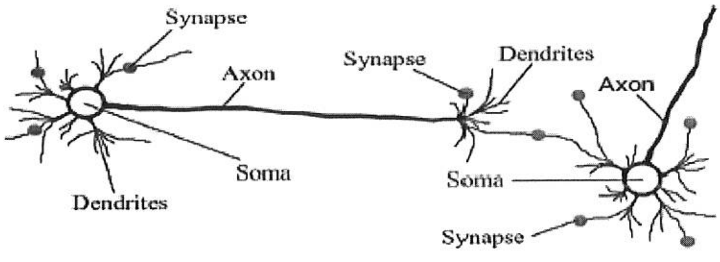


Figure 2.7: Biological Neural Network (Meysam and Fardad, 2009)

An artificial neural network consists of a number of very simple processors, also called neurons, which are analogous to the biological neurons in the brain. The neurons are connected by weighted links passing signals from one neuron to another. The output signal is transmitted through the neuron's outgoing connection. The outgoing connection splits into a number of branches that transmit the same signal. The outgoing branches terminate at the incoming connections of other neurons in the network.

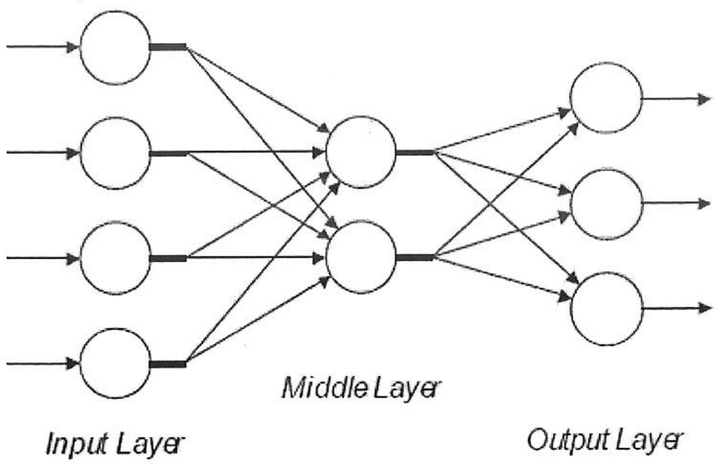


Figure 2.8: Architecture of ANN (Meysam and Fardad, 2009)

Learning in ANN;

There are basically three types of learning,

1. Supervised Learning

2. Un-Supervised Learning

3. Reinforced Learning

- Supervised Learning - Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors (output vectors) are known as supervised learning problems. Supervised learning is when the data you feed your algorithm is "tagged" to help your logic make decisions. Eg. Face recognition, perceptron
- Un- Supervised Learning - In other pattern recognition problems, the training data consists of a set of input vectors x without any corresponding target values. The goal in such unsupervised learning problems may be to discover groups of similar examples within the data, where it is called clustering. Clustering is unsupervised learning: you let the algorithm decide how to group samples into classes that share common properties. Eg. Hopfield Network
- Reinforced Learning - In reinforcement learning, data are usually not given, but generated by an agent's interactions with the environment. At each point in time, the agent performs an action and the environment generates an observation and an instantaneous cost, according to some (usually unknown) dynamics. The aim is to discover a *policy* for selecting actions that minimizes some measure of a long-term cost; i.e., the expected cumulative cost. The environment's dynamics and the long-term cost for each policy are usually unknown, but can be estimated.

Types of ANN;

There are two types of ANN,

1. Feed-Forward NN
2. Recurrent NN

- Feed-Forward Neural Network (FFNN) - A feed forward neural network is an artificial neural network where connections between the units do *not* form directed.

- Recurrent NN - A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior.

Advantages of ANN;

- ANNs are highly non-linear modelling
- ANN is nonlinear model that is easy to use and understand compared to statistical methods.
- ANN is non-parametric model while most of statistical methods are parametric model that need higher background of statistic.
- ANN with Back propagation (BP) learning algorithm is widely used in solving various classifications and forecasting problems. Even though BP convergence is slow but it is guaranteed.
- Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms.

Applications of ANN;

Found in character Recognition, Image Compression, Stock Market Prediction, Travelling Salesman Problem, Medicine and Security

### d) Vector Quantization

Vector quantization (VQ in short) involves the process of taking a large set of feature vectors of a particular user and producing a smaller set of feature vectors that represent the centroids of the distribution, i.e. points spaced so as to minimize the average distance to every other point. Vector quantization is used since it would be highly impractical to represent every single feature vector in feature space that we generate from the training utterance of the corresponding speaker. While the VQ algorithm does take a while to generate the centroids, it saves a lot of time during the testing phase as we are only considering few feature vectors instead of overloaded feature space of a particular user. Therefore is an economical compromise that we can live with. A vector quantizer maps k-dimensional vectors in the vector space *Rk* into a finite set of vectors Y = {yi: *i* = 1, 2... *N*}. Here k-dimension refers to the no of feature coefficients in each feature vector. Each vector y*i* is called a code vector or a codeword and the set of all the code words is called a codebook (Ashish, 2007).

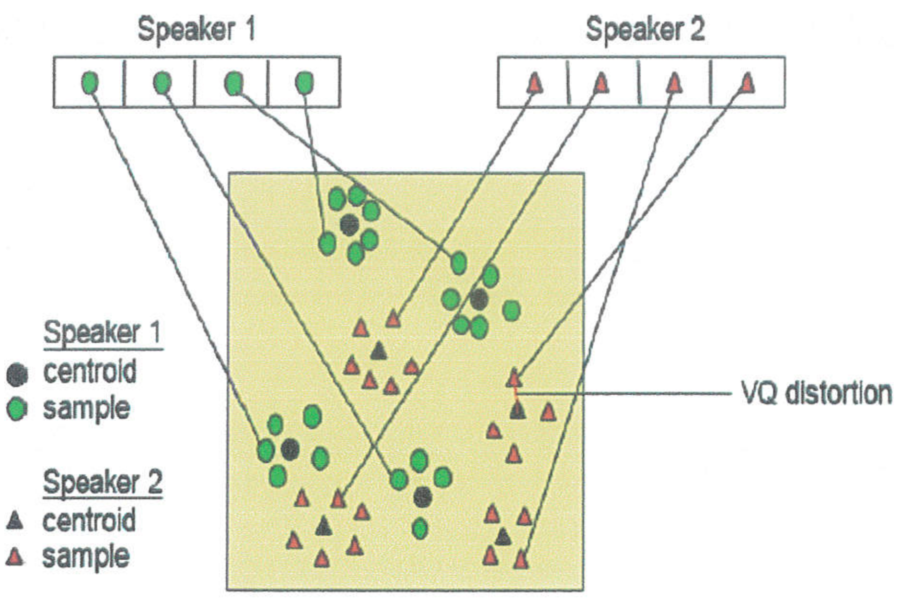


Figure 2.9: VQ recognition process (Ashish, 2007).

Above figure 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. In the training phase, 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 by black circles and black triangles for speaker 1 and 2, 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.

Optimization using LBG algorithm: After the enrolment session, the feature vectors extracted from input speech of each speaker provide a set of training vectors for that speaker. The next important task is to build a speaker-specific VQ codebook for each speaker using the training vectors extracted (Ashish, 2007). There is a well-known algorithm, namely LBG algorithm [Linde, Buzo and Gray, 1980], for clustering a set of L training vectors into a set of M codebook vectors. The algorithm is formally implemented by the following procedure:-

1. Design a 1-vector codebook; this is the centroid of the entire set of training vectors (hence, no iteration is required here).

2. Increase the size of the codebook twice by splitting each current codebook "yn" according to the rule

$$y_n^{+} = y_n(1 + \varepsilon)$$

$$y_n^{-} = y_n(1 - \varepsilon)$$

where n varies from 1 to the current size of the codebook, and is a splitting parameter (we choose =0.01).

3. Nearest-Neighbour Search: for each training vector, find the codeword in the current codebook that is the closest

(in terms of similarity measurement), and assign that vector to the corresponding cell (associated with the closest codeword).

4. Centroid Update: update the codeword in each cell using the centroid of the training vectors assigned to that cell.

5. Iteration 1: repeat steps 3 and 4 until the average distance falls below a preset threshold

6. Iteration 2: repeat steps 2, 3 and 4 until a codebook size of *M* is designed. Intuitively, the LBG algorithm generates an *M*-vector codebook iteratively. It starts first by producing a 1-vector codebook, then uses a splitting technique on the codeword to initialize the search for a

2-vector codebook, and continues the splitting process until the desired *M*-vector codebook is obtained (Ashish, 2007).

The advantages of VQ are:

- Reduced storage for spectral analysis information.
- Reduced computation for determining similarity of spectral analysis vectors. In speech recognition, a major component of the computation is the determination of spectral similarity between a pair of vectors. Based on the VQ representation this is often reduced to a table lookup of similarities between pairs of codebook vectors.
- Discrete representation of speech sounds (Ashish, 2007).

**2.6 Approaches to ASR by Machine**

One of the distinguishing characteristics of speech is that it is dynamic. Even within a small segment such as a phone, the speech sound changes gradually (Dongsuk, 2003). The beginning of a phone is affected by the previous phones, the middle portion of the phone is generally stable, and the end is affected by the following phones. The temporal information of speech feature vectors plays an important role in recognition process. After feature extraction, to model the distribution of the feature vectors x (1:N) any of the following modelling technique can be used .Basically there exist three approaches of speech recognition. They are

- Acoustic Phonetic Approach
- Pattern Recognition Approach
- Artificial Intelligence Approach

In the Acoustic Phonetic approach the speech recognition were based on finding speech sounds and providing appropriate labels to these sounds. This is the basis of the acoustic phonetic approach which postulates that there exist finite, distinctive phonetic units (phonemes) in spoken language and that these units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time.

The pattern-matching approach involves two essential steps namely, pattern training and pattern comparison. The essential feature of this approach is that it uses a well formulated mathematical framework and establishes consistent speech pattern representations, for reliable pattern comparison, from a set of labelled training samples via a formal training algorithm.

The Artificial Intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach. In this, it exploits the ideas and concepts of Acoustic Phonetic and Pattern Recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Existing modelling approaches for speech recognition have been represented diagrammatically in the following Figure 2.9.

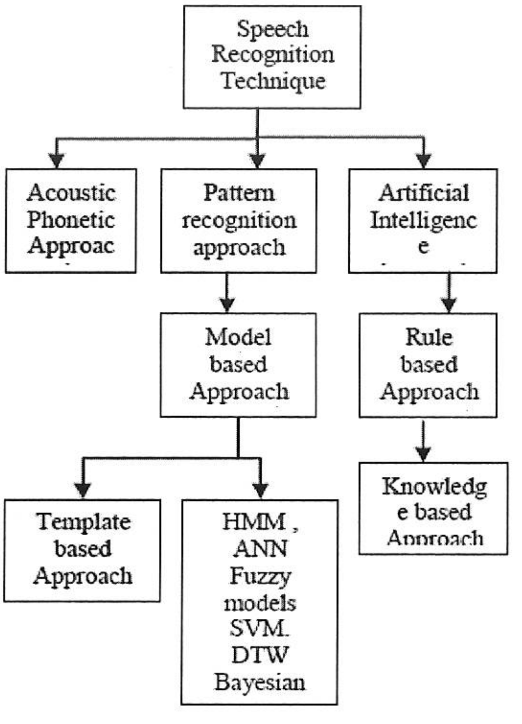


Figure 2.10: Speech Recognition Techniques (Dongsuk, 2003).

## 2.7 Matching Techniques

Speech-recognition engines match a detected word to a known word using one of the following techniques (Svendsen et al., 1989).

I. **Whole-word matching**. The engine compares the incoming digital-audio signal against a pre-recorded template of the word. This technique takes much less processing than sub-word matching, but it requires that the user (or someone) prerecord every word that will be recognized - sometimes several hundred thousand words. Whole-word templates also require large amounts of storage (between 50 and 512 bytes per word) and are practical only if the recognition vocabulary is known when the application is developed.

II. **Sub-word matching**. The engine looks for sub-words – usually phonemes and then performs further pattern recognition on those. This technique takes more processing than whole-word matching, but it requires much less storage (between 5 and 20 bytes per word). In addition, the pronunciation of the word can be guessed from English text without requiring the user to speak the word beforehand. Rabiner (1993), discuss that research in the area of automatic speech recognition had been pursued for the last three decades.

## 2.8 Performance Evaluation of Speech Recognition Systems

The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy which is usually rated with Word Error Rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR).

Word Error Rate (WER): Word error rate is a common metric of the performance of a speech recognition or machine translation system. The general difficulty of measuring performance lies in the fact that the recognized word sequence can have a different length from the reference word sequence. The WER is derived from the Levenshtein distance, working at the word level instead of the phoneme level. This problem is solved by first aligning the recognized word sequence with the reference (spoken) word sequence using dynamic string alignment.

Word error rate can then be computed as

$$WER = (S+D+I)/N \ldots (2.2)$$

*S* is the number of substitutions,

*D* is the number of the deletions,

*I* is the number of the insertions,

*N is* the number of words in the reference.

When reporting the performance of a speech recognition system, sometimes Word Recognition Rate (WRR) is used instead:

$$WRR = 1 - WER = 1 - (S+D+I) / N \ldots (2.3)$$

$$= (H - I) / N \ldots (2.4)$$

Where H = (N-S-D) is the correctly recognized words.

## 2.9 Security System

***Lia et al (2003),*** worked on a project that gives overall idea of how to control home security for smart homes especially for door key locks. We use android door lock system for indoor and outdoor key lock system. It also provides a security for Android phone users. This project based on Android platform which is Free Open Source. So the implementation rate is inexpensive and it is reasonable by a common person. With the wireless Bluetooth connection in microcontroller permits the system installation in more easy way. The system has been designed successfully. And aimed to control the door condition using an Android phone which is Bluetooth-enabled via Bluetooth HC-05.We have discussed a simple prototype in this paper but in future it can be extended to many other regions.

***Shilpi (2009),*** worked on a project that gives information about system has been given in which we can unlock the door by using pre-decided password. It increases the security level to prevent an unauthorized unlocking done by attacker. In case the user forgets the both passwords, this system gives the flexibility to the user to change or reset the password. This automatic password based lock system will give user more secure way of locking-unlocking system. First the user combination will be compared with prerecorded password which are stored in the system memory. User can go for certain number of wrong combinations before the system will be temporarily disabled.

The door will be unlocked if user combination matches with the password. The same password can be used to lock the door as well. This system will give the user an opportunity to reset his own password if he wants.

***Arpita et al (2009),*** worked on a project that has considered about day to day life security of any object or place password based system and created a secure access for a door which needs a password to unlock the door. Using keypad it enters a password to the system and if entered password is correct then door is open by motor which is used to rotate the handle of the door lock. When it is entered incorrectly at the first time it will give three attempts to enter the password. Some extra features like adding new users and changing old password are configured by the keypad as usual. To display messages to the user LCD module is used. Now adays most of the systems are automated in order to face new challenges to achieve good results. These systems have less manual operations, so the flexibility, reliabilities are high and accurate are there characteristics. Hence every field prefers automated control systems especially in the field of electronics.

***Majgaonkar et al (2006),*** Worked on this project that utilizes the different electronic parts available in the market to build an integrated home security system by using Bluetooth device and Microcontroller technology. This system gives service at low cost compared to the cost of the available security system. A system was made that will give 24 into 7 service by using registered password in this system we can unlock the door by which it increases the security level to prevent an unauthorized unlocking.

If the user forgets the combination of password this system gives the flexibility to the user to change or reset the password.

Security measure is very high as provided in two ways. First we have to enter password for blue-tooth connection and second is for unlocking the door in application. Both passwords can be changed as and when required. This automatic password based lock system will give user more secure and low cost way of locking-unlocking system.

***Assaf et al (2001),*** worked on this project aiming at the conventional design of home security systems which typically monitors only the property and lacks physical control aspects of the house itself. Also, the term security is not well defined because there is a time delay between the alarm system going on and actual arrival of the security personnel. This project discusses the development of a home security and monitoring system that works where the traditional security systems that are mainly concerned about curbing burglary and gathering evidence against trespassing fail. The project presents the design and implementation details of this new home control and security system based on field programmable gate array (FPGA). The user here can interact directly with the system through a web-based interface over the Internet, while home appliances like air conditioners, lights, door locks and gates are remotely controlled through a user-friendly web page. An additional feature that enhances the security aspect of the system is its capability of monitoring entry points such as doors and windows so that in the event any breach, an alerting email message is sent to the home owner instantly.

***Qasim et al (2003),*** worked on this project that implements the face recognition techniques in an embedded system which is a very important aspect of various security applications, such as authorization identification in cash machines and door access for employee attendance. The purpose of this project is to survey the existing hardware implementations of a face recognition system, particularly those who used for developing an embedded door access control system. This includes a brief discussion about face recognition algorithms and the hardware platforms outlining the importance of using a Field Programmable Gate Array (FPGA) device, which can be used to design an embedded door access control system. It is found that the success of any door access control system depends on the type of face recognition algorithm in the feature extraction phase of the face recognition system and the selected hardware device.
Based on a comparison between different feature extraction algorithms, the use of a hybrid feature extraction technique can improve the accuracy of face recognition system. An efficient door access control system can be obtained using a FPGA device, which is preferable because of its technical characteristics of parallelism, re-programmability and very high speed, in the implementation of a face recognition system.

***Shafiul (2010),*** Worked on this project with the aim of using the PIC18F452 microcontroller for hardware and software implementation of our home security system design. Based on

PIC18F452, systems can monitor doors and windows of a house and can set alarm and warming signal to a nearest police station if anybody tries to break in. This security system also provides the functionality to identify the residents ID card to get access to the house without turning on the warning signal and alarm. Also, the security system provides a status that is not monitoring the door and windows since there is some possibility that the host do not want the system always checks the status of their house.

## 2.10 ASR projects

**Kharka *et al* (2015),** worked on a project which represent the "voice control wheel chair" for the physically handicapped person were the voice command controls the movements of the wheel chair. Voice Recognition Kit (HM2007 Module) is being used to recognize the voice command. The voice command given is converted to binary numbers by Voice Recognition Kit and those binary data is given to the Arduino board for the control of the wheel chair. For example when the user says „forward" than chair will move in Forward direction and when he says „Backward" than the chair will move in backward direction and similarly for left, right and stop. We have used LCD display unit to display the direction in which direction of the wheel chair.

**Bhargavi *et al* (2015),** worked on a project to help the physically challenged and old people who face many problems in life and enable them not to be dependent on a third person to move from one place to another. Many scientists have been working for this solution for a long time. The invention of wheel chair is a great boon to them but it still restricts their motion. In order to make their life a bit easier, many developments in wheel chairs came into existing such as electric-powered, gesture based etc.,. Voice controlled wheel chair was made using HM2007 voice recognition kit.

## 2.11 project work

This system which is the development of speech recognition based security door lock system, was developed considering some factors such as economy, availability of components, efficiency, compatibility, portability and also durability.

It was built using readily available MATLAB software, within which the program for the speaker dependent speech recognition was written to aim at the project being dynamic (i.e. the data in the database of the system can be erased and reprogrammed multiple number of times). In the speaker recognition system, acquisition of speech data was done using an analogue microphone and stored in database, MFCC feature extraction technique was used during the feature extraction stage in training and recognition phase, lastly vector quantization utilizing LBG algorithm was employed in the recognition phase for matching/recognition of the supposed speaker with data in the database of the speaker recognition system. The project was implemented using MATLAB and tested using a hardware model door with PIC18F1X20 microcontroller and L293D H-bridge driver which is interfaced to control the opening and closing of the model house door.

# CHAPTER THREE

# DESIGN METHODOLOGY

## 3.1 Introduction

There are several steps applied in designing an automatic speech recognition based door lock system. The relevant is gathered through literature review from previous chapter. This project was built using MATLAB software to produce the speech recognition system which have acquisition stage, recognition stage and database. It was programmed to be dynamic such that the data in database can be erased and reprogrammed multiple number of times while still retaining the systems functional efficiency, making use of MFCC for feature extraction and VQ utilizing LBG algorithm for the recognition of speaker. Next is the hardware development, making use of PIC18F1X20 microcontroller which receives signal from the speech recognition system developed in MATLAB and controls the L293D H-bridge driver in opening and closing of the model house door.

## 3.2 Components Theory

The materials used in the development and implementation of the speech recognition door lock system are listed below;

- MATLAB software for development of speech recognition system
- The microcontroller (PIC18F1X20)
- The microphone circuit
- LCD display
- Gear system and DC motor
- H-bridge DC motor driver (L293D)

### 3.2.1 THE MICROPHONE CIRCUIT (PRE-AMPLIFIER)

The main function of a Preamplifier is to amplify small and weak signals for further amplification. Generally, weak signals from microphones, audio sources and other sound detectors must be extracted with compromising the intrinsic signal to noise ratio (SNR). Hence, the best position of a preamplifier is close to the sensor or detector. The output of the preamplifier is further amplified by Power Amplifiers. Preamplifier amplifies the signal with very high gain but doesn't have the drive current or current gain to drive the output. Hence, the boosted signal from preamplifier is given to a power amplifier where the current is amplified. If the input signal is subject to filtering, the filtering circuit may add noise to the signal. When a preamplifier is used, the noise can be considerably reduced. Preamplifier also helps in minimizing the noise in the lines when the sensor and power amplifier are placed at a distance.

In the preamplifier circuit, LM358 Op-Amp is used. It is connected in a negative feedback fashion with a 1MΩ POT connected in the feedback path. The input from the microphone is given to the inverting terminal of the op amp while the non-inverting terminal is given with a constant input from the voltage divider formed by R3 and R4 as shown in figure 3.1 below

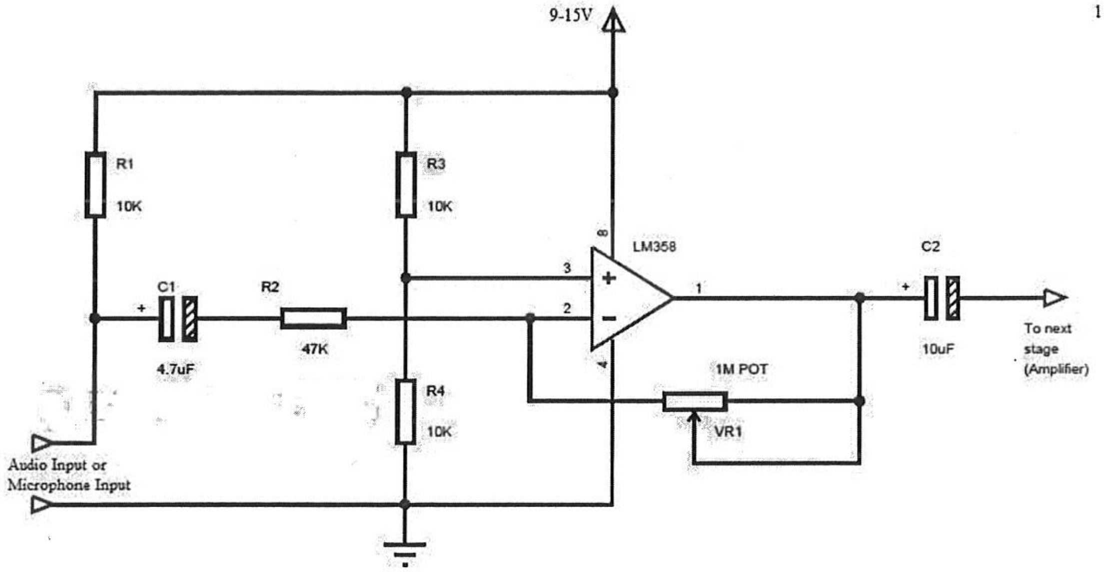


Figure 3.1: Microphone circuit (Wikipedia.com)

### 3.2.2 LCD DISPLAY

A liquid-crystal display (LCD) is a flat-panel display or other electronically modulated optical device that uses the light-modulating properties of liquid crystals. Liquid crystals do not emit light directly, instead using a backlight or reflector to produce images in color or monochrome. LCDs are available to display arbitrary images (as in a general-purpose computer display) or fixed images with low information content, which can be displayed or hidden, such as preset words, digits, and 7-segment displays, as in a digital clock. They use the same basic technology, except that arbitrary images are made up of a large number of small pixels, while other displays have larger elements.

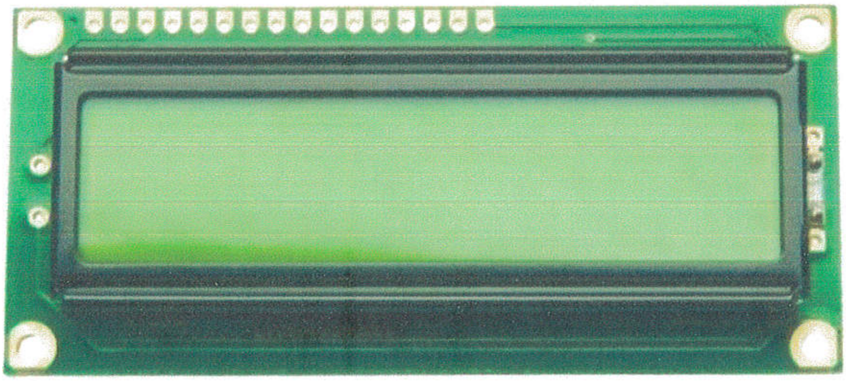

Figure 3.2: 16x2 LCD display (Wikipedia.com)

### 3.2.3 Microcontroller (PIC18F1X20)

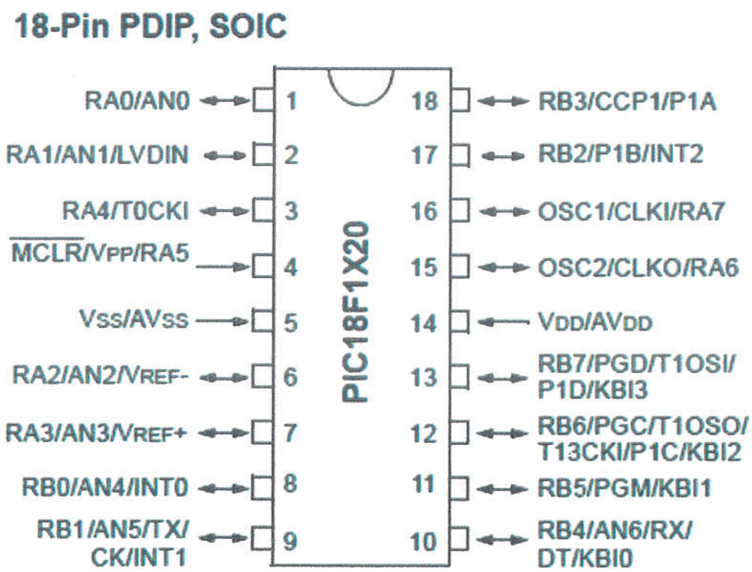


Figure 3.3: PIC18F1X20 microcontroller (Wikipedia.com)

Specification;

C compiler optimized architecture with optional extended instruction set. USB V2.0 compliant (i.e. it has on-chip USB transceiver with on-chip voltage regulator).Operating speed: for External oscillator, has two External RC modes up to 4MHz, and two External Clock modes, up to 40MHz. Also for Internal oscillator, allow up to 8 user-selectable frequencies: 31 kHz, 125 kHz, 250 kHz, 500 kHz, 1 MHz, 2 MHz, 4 MHz, 8 MHz. It has

In-Circuit Serial Programming™ (ICSP™), It also has Programmable code protection. it is an 18 pin version which Contains 16 I/O pins with individual direction control, 1-3 timers, Dual analog comparators with input multiplexing, has Power-Saving sleep mode and Wake-up from sleep on pin charge, 4 Kbytes Flash program memory, 2048 bytes SRAM data memory and 256 bytes EEPROM data memory, Operating voltage: 2V to 5.5V.

It can be programmed with Velleman VM134 (=K8076) and our USB "in-circuit" programmer PICKIT2

### 3.2.4 H-bridge DC motor Driver (L293D)

The L293 and L293D are quadruple high-current half-H drivers. The L293 is designed to provide bidirectional drive currents of up to 1 A at voltages from 4.5 V to 36 V. The L293D is designed to provide bidirectional drive currents of up to 600-mA at voltages from 4.5 V to 36 V. Both devices are designed to drive inductive loads such as relays, solenoids, dc and bipolar stepping motors, as well as other high-current/high-voltage loads in positive-supply applications.

All inputs are TTL compatible. Each output is a complete totem-pole drive circuit, with a Darlington transistor sink and a pseudo-Darlington source. Drivers are enabled in pairs, with drivers 1 and 2 enabled by 1,2EN and drivers 3 and 4 enabled by 3,4EN. When an enable input is high, the associated drivers are enabled, and their outputs are active and in phase with their inputs. When the enable input is low, those drivers are disabled, and their outputs are off and in the high-impedance state. With the proper data inputs, each pair of drivers forms a full-H (or bridge) reversible drive suitable for solenoid or motor applications. On the L293, external high-speed output clamp diodes should be used for inductive transient suppression.

A VCC1 terminal, separate from VCC2, is provided for the logic inputs to minimize device power dissipation. The L293and L293D are characterized for operation from 0°C to 70°C.

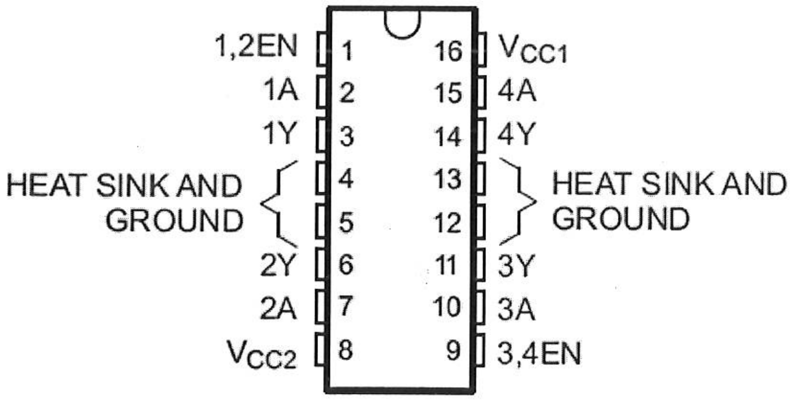


Fig 3.4: 16-pin L293 pin diagram (Wikipedia.com)

Table 1: PIN FUNCTION OF L293

| PIN | | | |
|---|---|---|---|
| NAME | NO. | TYPE | DESCRIPTION |
| 1,2EN | 1 | INPUT | Enable driver channels 1 and 2 (active high input) |
| <1:4>A | 2,7,10,15 | INPUT | Driver inputs, noninverting |
| <1:4>Y | 3,6,11,14 | OUTPUT | Driver outputs |
| 3,4EN | 9 | INPUT | Enable driver channels 3 and 4 (active high input) |
| GROUND | 4,5,12,13 | NILL | |

| | | | |
|---|---|---|---|
| | | | Device ground and heat sink pin. Connect to printed-circuit-board ground plane with multiple solid vias |
| VCC1 | 16 | NILL | 5-V supply for internal logic translation |
| VCC2 | 8 | NILL | Power VCC for drivers 4.5 V to 36 V |

### 3.2.5 Gear System and DC Motor

The gear system and DC motor works in bidirectional. They work hand-in-hand as the gear system assists in controlling the rotation speed and accuracy of the DC motor so as to easily open and close the model door to be used.

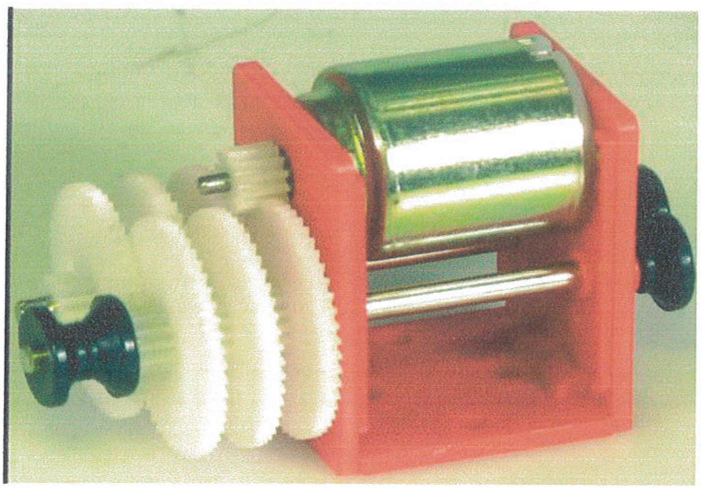

Fig 3.5: A Gearbox (Wikipedia.com)

## 3.3 SYSTEM ANALYSIS

The security system based on speech recognition is going to be built using MATLAB and implemented using a hardware model door as discussed in the previous chapter is shown in figure 3.6 below;

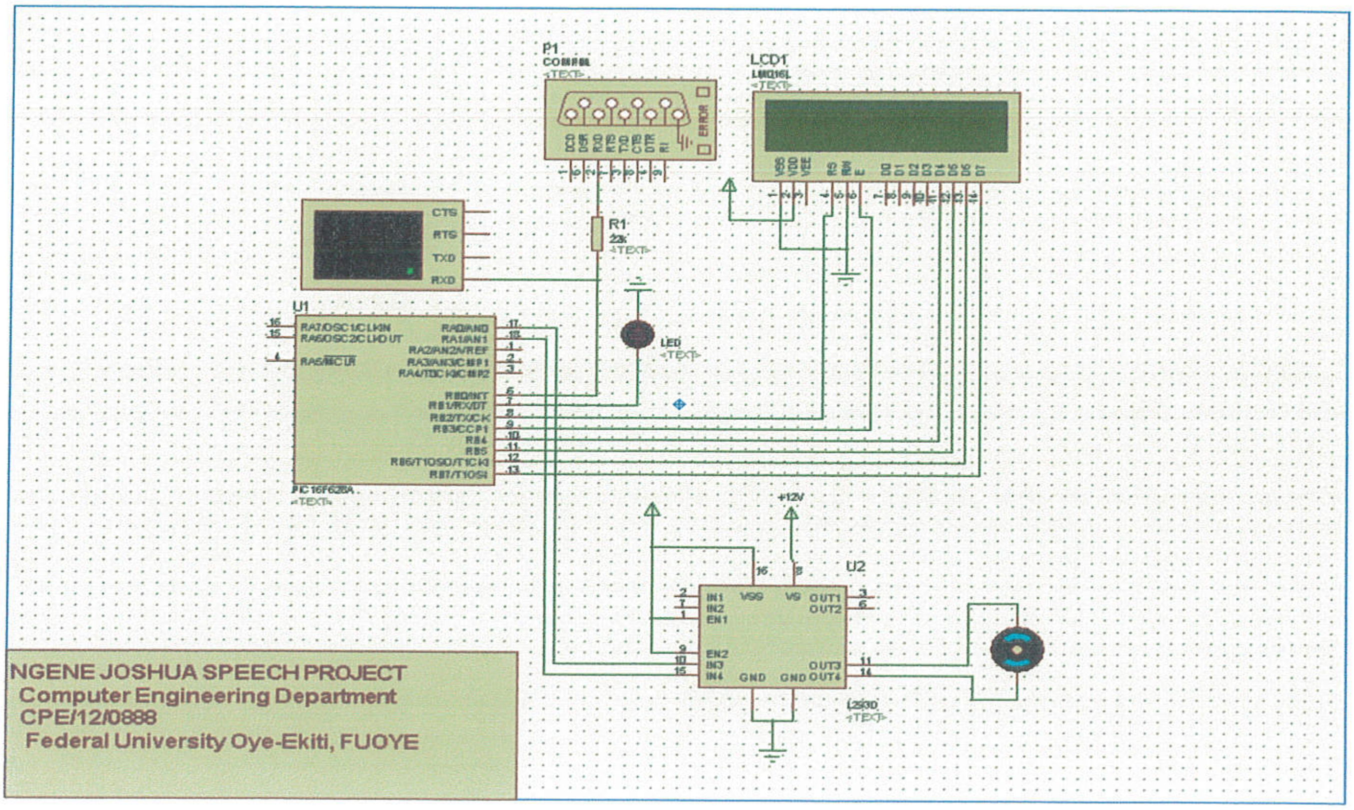


Figure 3.6: Schematics of the hardware door control circuit (proteus software)

From the schematic diagram above which shows the hardware with PIC18F1X20 microcontroller which receives access signal from the MATLAB after the recognition stage of the MATLAB through a COM port (serial program cable). Here, after recognition of the authorized access speaker by the speaker recognition system created with MATLAB, an access signal is sent through the COM port to the PIC18F1X20 microcontroller which controls the opening and closing of the model house door through the H-bridge door driver mechanism which works in the forward and backward direction to physically open and close the door. The LCD displays the status of the system.

## 3.4 DESIGN OF POWER SUPPLY FOR CIRCUIT

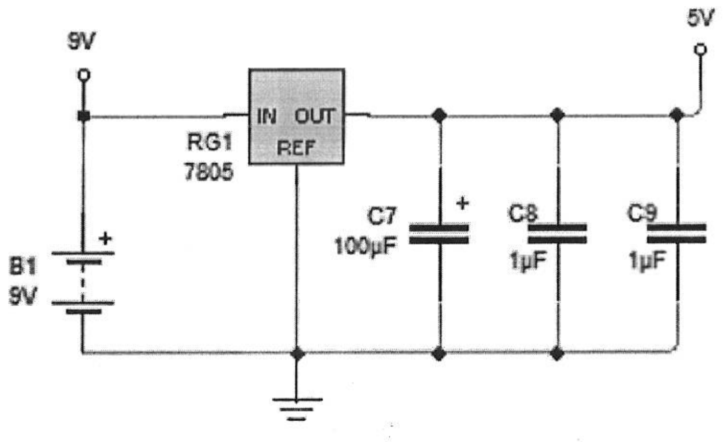


Figure 3.7: Power Supply Circuit

The 5V power supply uses a 78L05 regulator and filter caps to convert the 9V from the battery to 5V for the PIC16F1X20 and the L293D circuits.

Figure 3.7 above shows the LM7805 which is connected to a 9V source and provides a constant 5V output. The output is used to supply power to the MCU. The output is also used to power up the voice playback device on the PCB board.

### 3.4.1 POWER SUPPLY STAGE FROM AC POWER SUPPLY

The talking voltmeter design uses 5V dc power supply rail. The need for the power supply stage is to provide the voltage and current requirements for the circuit since all electronic components work with D.C voltages. The required dc voltage and current of the power supply for the project is dependent on the component specifications and the nature of the circuit to be powered.

For this project the following power requirements were estimated for the circuit components and stepper motor driven requirements.

### 3.4.2 POWER SUPPLY REQUIREMENTS.

Supply voltage: DC 5V

Maximum current: 2A

The maximum current is estimated based on the PIC microcontroller used which can operate on a supply voltage from 3-6V DC, specifically 5V (see datasheet in appendices)

### 3.4.3 ESTIMATION OF CURRENT REQUIREMENT

PIC microcontroller and LCD display = 400mA

Estimated current requirement = 1.1 A

To create some margin for error and tolerance, 1.5A would be used. Hence our power supply must be capable of sourcing 1.5A at 5V.

Figure 3.8 below shows the power supply stage design

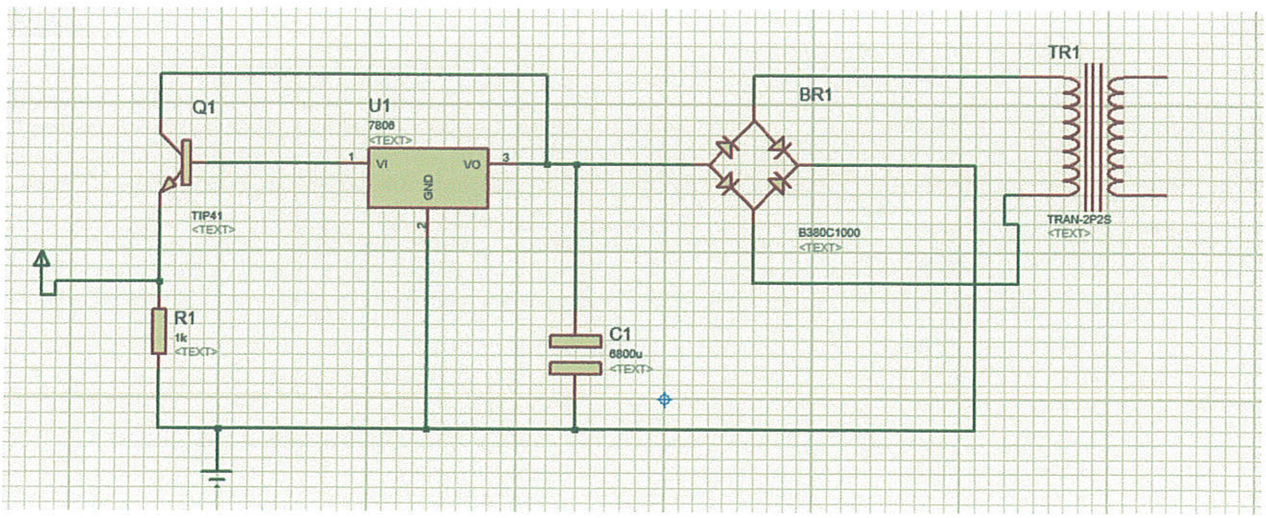


Figure 3.8: Power supply stage design

Figure 3.9 shows the rectified DC waveform (before filtering with a capacitor C1)

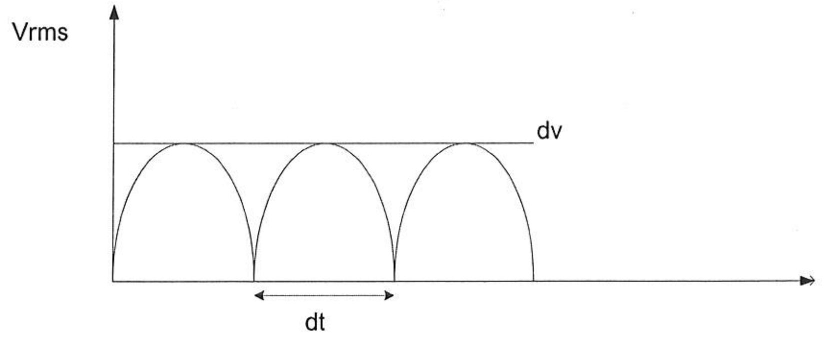


Figure 3.9: Rectified DC waveform

The electric charge, Q = It

Since Q = It

CV= It (since Q = CV) ---------- Equation (3.1)

Where C is the capacitance

V is the voltage

I = current and t = period of one cycle of the AC waveform

From (1) $C = I\frac{t}{V}$

=> $C = I\frac{dt}{dV}$ ------------Equation (3.2)

Since C is proportional to the current and also inversely proportional to the ripple gradient of the voltage with time.

The peak unregulated voltage is given by

$V_{PEAK} = Vrms\ X\ \sqrt{2}$ ---------------- Equation (3.3)

Where $V_{RMS}$ is the AC voltage stepped down on the transformer

Hence considering a peak voltage of 12V dc, from Equation (3.3)

$V_{RMS} = \frac{12V}{\sqrt{2}} = 8.48 \approx 9V$ AC

This implies a step down transformer of 9V at 1.5A

From (2) $C = I\frac{dt}{dV}$

I = 1.5A as required from PSU requirement for design

dt = 0.01s (this is the time duration of the duty cycle of half the waveform)

dv = ripple factor

which is approximately 20% of the peak voltage

from (2) $C = 1.5A \text{ X } \frac{0.01}{5V}$

= 3000 UF

= 2200 Uf preferred value

Hence the value of C1 in the Figure 3.1 is 2200uf. (Robert, 1999)

## 3.5 PRODUCT OPERATION FLOW

### 3.5.1 ASSEMBLY PROCESS

- Assemble and test the speech recognition system
- Setup and test the H-Bridge motor control circuit
- Assemble the model door and the door frame
- Assemble the gear system and the motor
- Interface all the various components and test the overall system

### 3.5.2 DESIGN ALTERNATIVES FOR HARDWARE

- Stepper motor and stepper motor driver circuit instead of H-bridge and a DC motor.
- Pulley system instead of a gear system to open and close the door.

### 3.5.3 OPERATIONAL BLOCK DIAGRAM

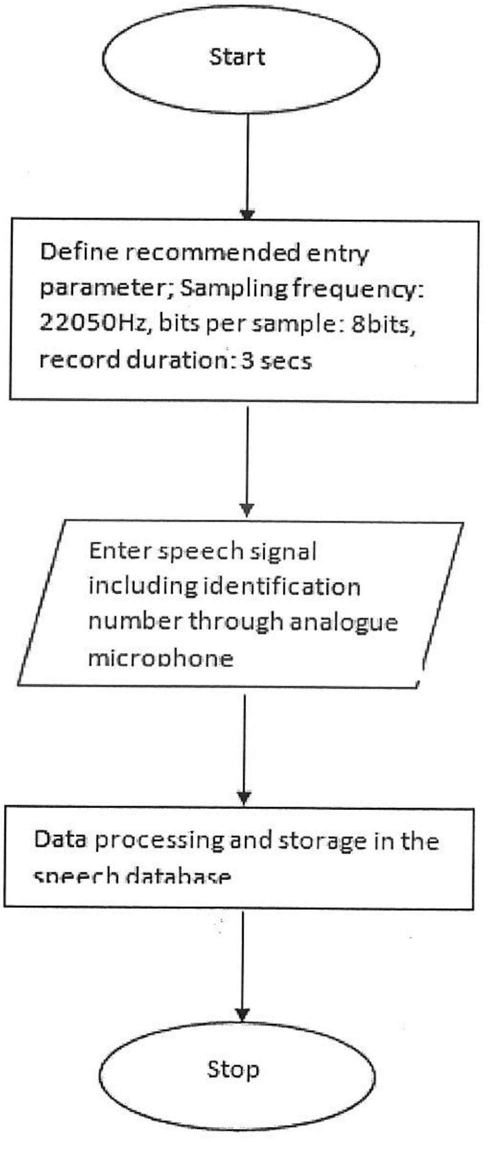


Figure 3.10: Data Acquisition and Storage flow diagram

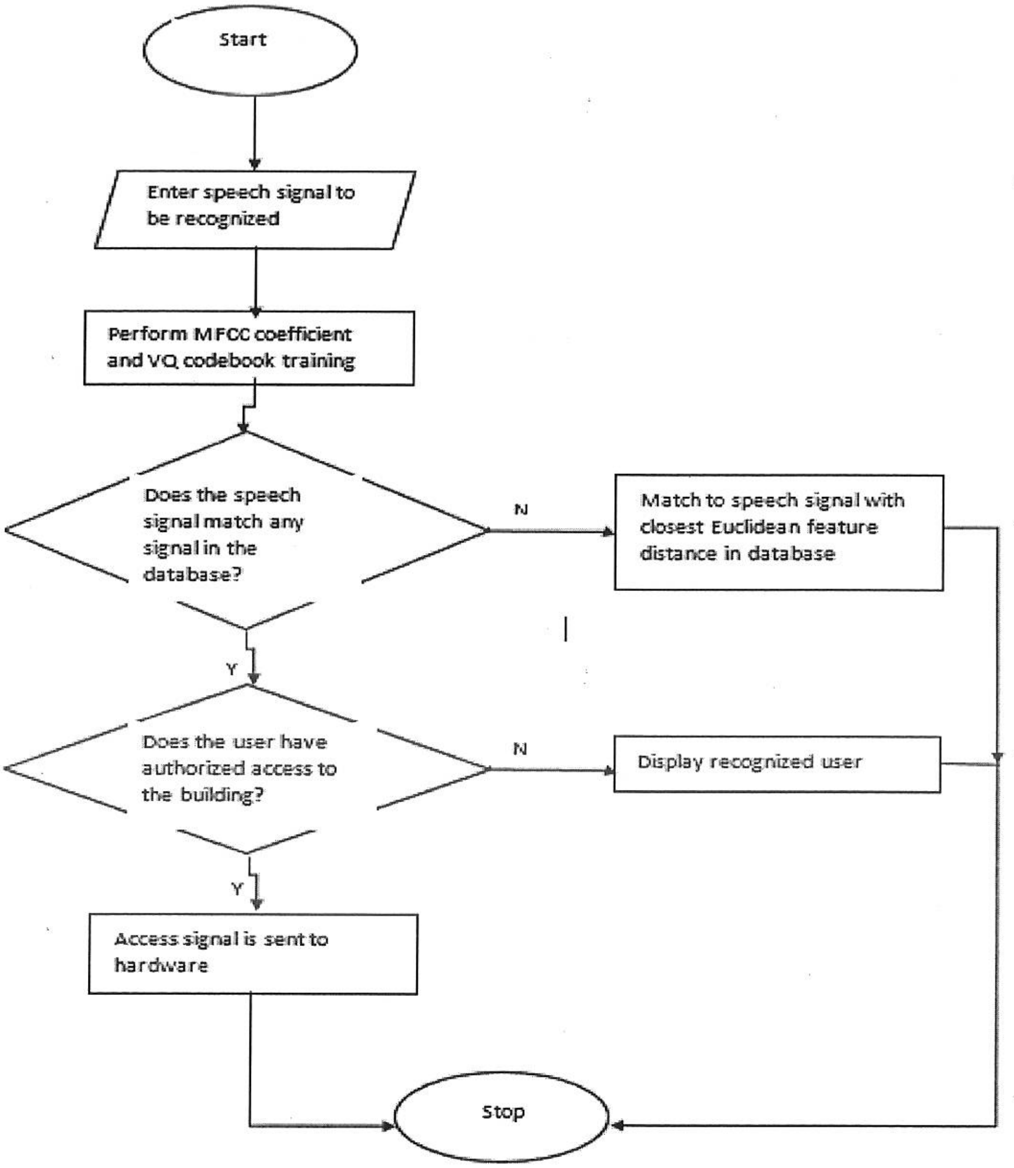


Figure 3.11: Recognition and Access control flow diagram

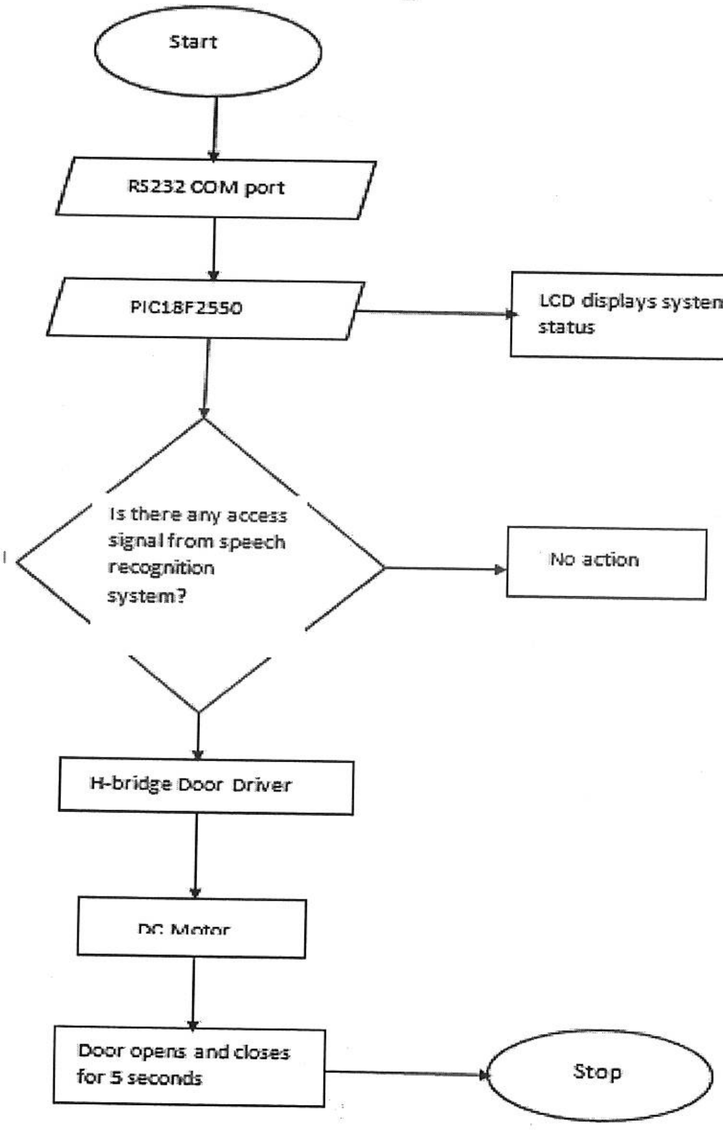


Figure 3.12: Hardware Implementation flow diagram

# CHAPTER FOUR

# SYSTEM EXECUTION, RESULTS, IMPLEMENTATION AND DISCUSSIONS

## 4.1 INTRODUCTION

This chapter will briefly discuss on the execution, results and discussion of the system developed. This consists of software and hardware implementation, where the main speech recognition system is done using MATLAB software and also, explains the database design of the speech recognition system and the different steps in execution and results gotten from the MATLAB. The hardware implementation is done using a model house door, built to be used in implementation if the speech recognition system built in MATLAB.

## 4.2 DESIGN OF THE SPEECH SYSTEM DATABASE

The speech system database was designed using a sequentially structural method whereby each section leads to another in the whole speech program. The database was done using the MATLAB software (MATLAB 2015) due to its high level dynamic nature. The speech system database code written in MATLAB, has some sections which a listed below and these sections are backed up with their individual code sections too.

- Add a new speech from microphone
- Speaker recognition from microphone
- Database information
- Delete database
- Exit

Below is shown, a picture of the menu/section code and the execution result

```
chos=menu('Speaker Recognition System','Add a new speech from microphone',...
    'Speaker recognition from microphone',...
    'Database Info','Delete database','Exit');
```

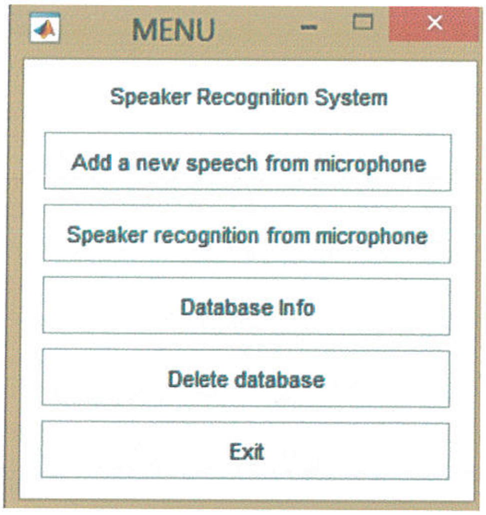


Figure 4.1: speaker recognition system menu

1. **ADD A NEW SPEECH FROM MICROPHONE;**

This is the first chosen section of the speech recognition system where speech signals or data in the form of speech is added to the database of the system. This can also be said to be the training phase of the whole speech system where the system is trained with different individual's speech signals which are then stored in the speech database, and used as a reference in the recognition phase of the system. In this stage, during the acquisition of the speech signal, some parameter are to be inputted for good acquisition of speech signal, these parameter are; a class number (sound ID) that will be used to differentiate different speech in database and also will be used for recognition; the sampling frequency; bits per sample; duration of recording which is in seconds. There are some recommended values for these parameters listed above, sampling frequency (22050 recommended), sampling bits (8 recommended), and duration of recording in seconds (3 seconds recommended). After recording, the sounds are added to the database of the speech system.

2. **SPEECH RECOGNITION FROM MICROPHONE;**

This is the second chosen section in the speech system database which the core section of the whole project. This section controls the opening and closing of the model house door (hardware) as it assigns priority to only one sound ID to have access to a containment by opening its entrance door. During the recognition phase of the speech system, speech signals entered into the system are processed and compared to see if there is a match for the signal in the database of the system. The processing of these signals include the use of MFCC for feature extraction and VQLBG (vector quantization using LBG algorithm) for easy recognition of speech signals. In recognition, VQLBG was used instead of HMM, ANN, Gaussian method as it has faster processing and dynamic to changes in the use of different speech signals. VQLBG makes erasing and re-addition of speech signals very easy and faster unlike HMM or ANN which needs must training and is not easy to erase data, they are not as dynamic in change of data during erasing process as the system must be configured to be able to erase speech signals and also reprogram new speech signals countless number of times.

3. **DATABASE INFORMATION;**

This is the third chosen section of the speech system database which is used to show the speech signals stored in the database of the system during the acquisition phase. If there are no speech signal in the database, it prompts up a message saying "*Database is empty*".

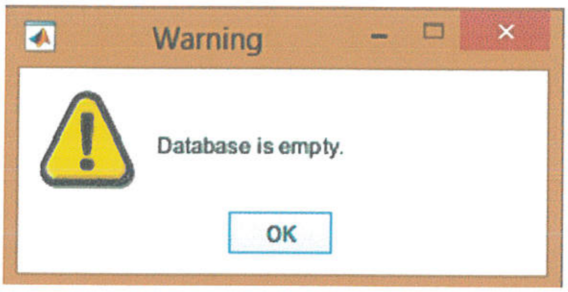


Figure 4.2: Empty database signal in system.

Else it displays the content of database based on the sound ID present.

4. **DELETE DATABASE;**

This is the fourth chosen section of the speech system database which is used to erase/delete the entire content of the database should there be any need for a change in sound ID. This section is very necessary as it makes the system very dynamic to taking in different and countless number of speech signals since they can always be removed on the users will. On deletion of speech signals from the database, it prompts up a message saying "*Database was successfully removed from the current directory.*" When deletion is attempted on an empty database, it also prompts a warning message "*Database is empty*".

5. **EXIT;**

This is the final section of the speech system database which is used to close all operation of the database.

## 4.3 FUNCTIONS USED IN THE SPEECH RECOGNITION SYSTEM

There are various metrics used in the achievement of the project in different phase. They are blockFrame function; MFCC function for feature extraction; and the VQLBG Vector quantization using the Linde-Buzo-Gray algorithm for recognition phase.

- BLOCKFRAME FUNCTION;

  This function puts signal "s" to be processed into frames, and within a frame, individual signals are broken into a number of sample "n" having a sampling rate "fs" of the signal. The distance between the beginnings of two frames is denoted by "m". A matrix "M3" contains all frames and the function of MATLAB code is briefed below;

```
function M3 = blockFrames(s, fs, m, n)
`l = length(s);
nbFrame = floor((l - n) / m) + 1;
for i = 1:n
   for j = 1:nbFrame
      M(i, j) = s(((j - 1) * m) + i); %#ok<AGROW>
   end
end
h = hamming(n);
M2  = diag(h) * M;
for i = 1:nbFrame
   M3(:, i) = fft(M2(:, i)); %#ok<AGROW>
end
end
```

- MFCC FUNCTION;

  This function is used for accurate extraction of features of a given speech signal "s" having a sampling frequency "fs" and "r" which contains the transformed signal and the function MATLAB code is briefed below;

```
function r = mfcc(s, fs)
m = 100;
n = 256;
frame=blockFrames(s, fs, m, n);
m = melfb(20, n, fs);
```

```
n2 = 1 + floor(n / 2);
z = m * abs(frame(1:n2, :)).^2;
r = dct(log(z));
end
```

- VQLBG VECTOR QUANTIZATION USING THE LINDE-BUZO-GRAY ALGORITHM;

This function is used for recognition of features of a speech signal which relates to a stored feature of speech signal in the speech database. it basically checks the distance between parameters of the verifying speech signal features and matches it to a speech signal stored in the database based on the Euclidean distance closeness. Below is the brief MATLAB code as "d" contains training data vector (one per column), "k" is number of centroids required, "r" contains the result VQ codebook (k columns, one for each centroids);

```
function r = vqlbg(d,k)

e = .01;
r = mean(d, 2);
dpr = 10000;
for i = 1:log2(k)
   r = [r*(1+e), r*(1-e)];
   while (1 == 1)
      z = disteu(d, r);
      [m,ind] = min(z, [], 2);
      t = 0;
      for j = 1:2^i
         r(:, j) = mean(d(:, find(ind == j)), 2); %#ok<FNDSB>
         x = disteu(d(:, find(ind == j)), r(:, j)); %#ok<FNDSB>
         for q = 1:length(x)
            t = t + x(q);
         end
      end
      if (((dpr - t)/t) < e)
         break;
      else
         dpr = t;
      end
   end
end
end
```

## 4.4 PROJECT EXECUTION AND RESULT EVALUATION OF THE SPEECH SYSTEM.

### 4.4.1 SOFTWARE EXECUTION AND RESULT DISCUSSION

1. Data acquisition and pre-processing stage
2. Feature extraction and data storage
3. Recognition using VQLBG algorithm

1. Data acquisition and pre-processing

   This is the first section of the MATLAB program process, in which data is acquired by the use of analogue microphone using some recommended properties such as; Sound ID (to represent the speech signal), Duration of recording, Sampling frequency (22050Hz recommended) and Number of bits per sample (8bits recommended). Data acquisition is done under a conducive environment such as a silent room to prevent acquisition of unwanted signals called noise that may have effect on the speech system. After data acquisition, the data is pre-processed by filters and converters which are embedded in the system used. Noise filter which removes unwanted signals from actual signals needed and ADC that converts the signal from analogue form received from microphone to digital process-able signal by computer.

2. Feature extraction and Data storage

   After the pre-processing and signal is now in computer process-able form (digital form), some features of the vocal characteristics of the speech are extracted from the speech signal and the speech in form of SOUND ID is stored in the speech system database. The MFCC feature extraction technique is used for the extraction of features as it has been seen to have high accuracy compared to other techniques of feature extraction.

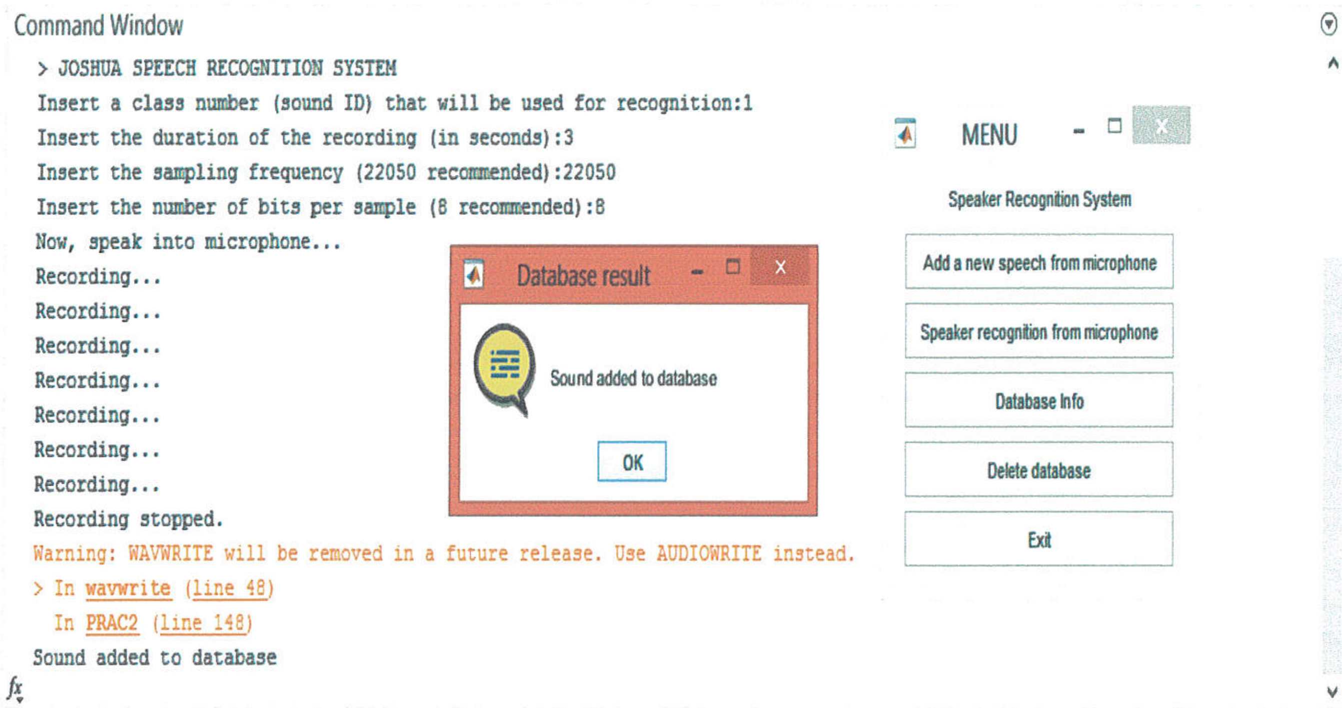


Figure 4.3: Acquisition of Speech Data "Sound Id: 1"

Figure above shows the addition of speech signal into the database, speech signal with representation "SOUND ID: 1", sampling frequency of 22050Hz, 8bits per sample, and 3 second duration of recording. These value are taken constant for subsequent speech signals to be added to the database.

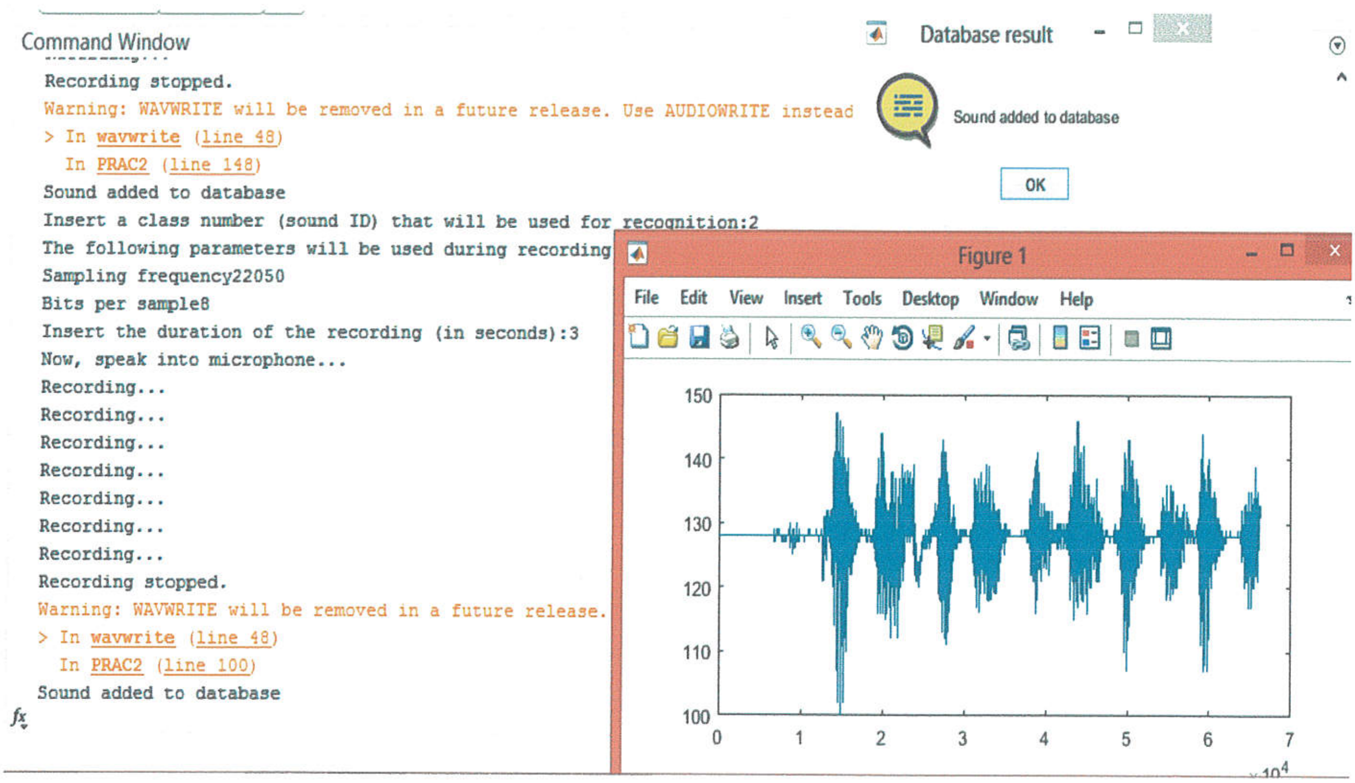


Figure 4.4: Acquisition of Speech Data "Sound ID: 2"

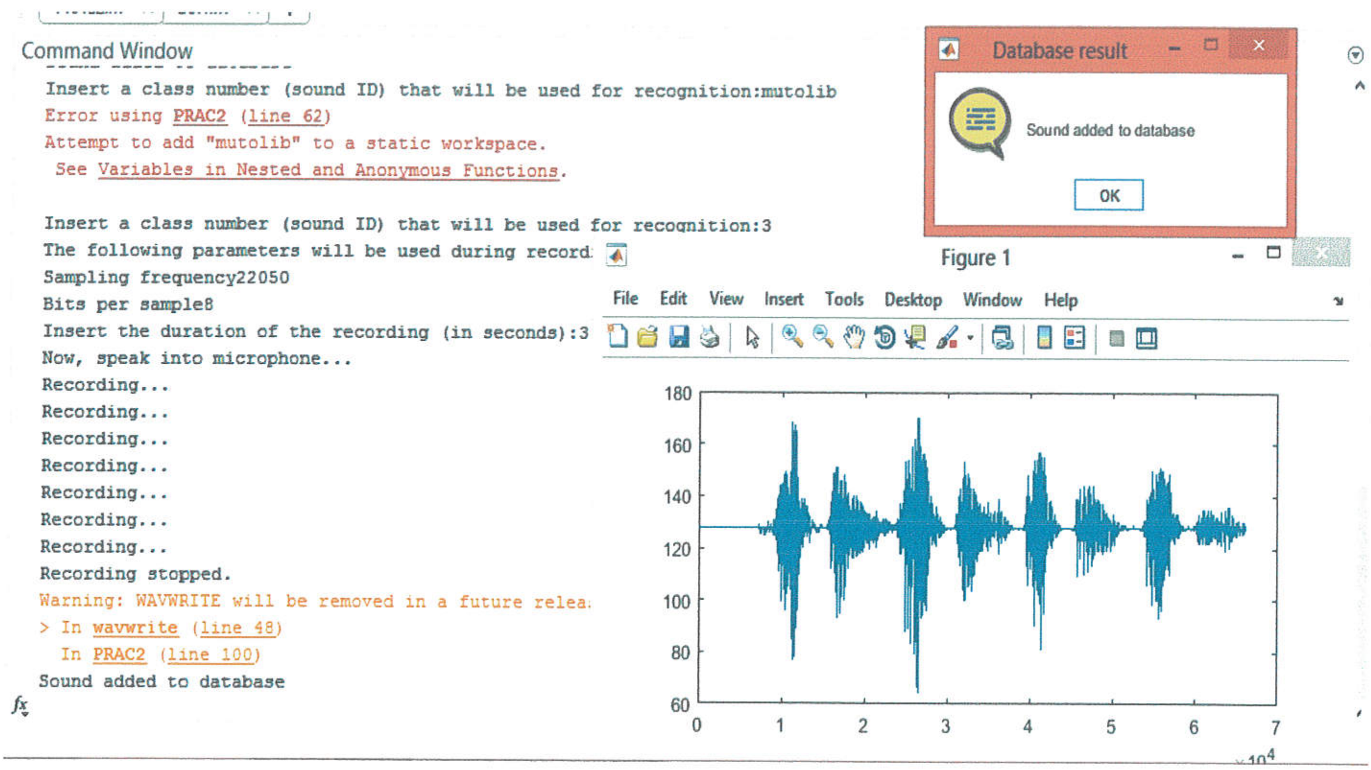


Figure 4.5: Acquisition of speech data "sound ID: 3"

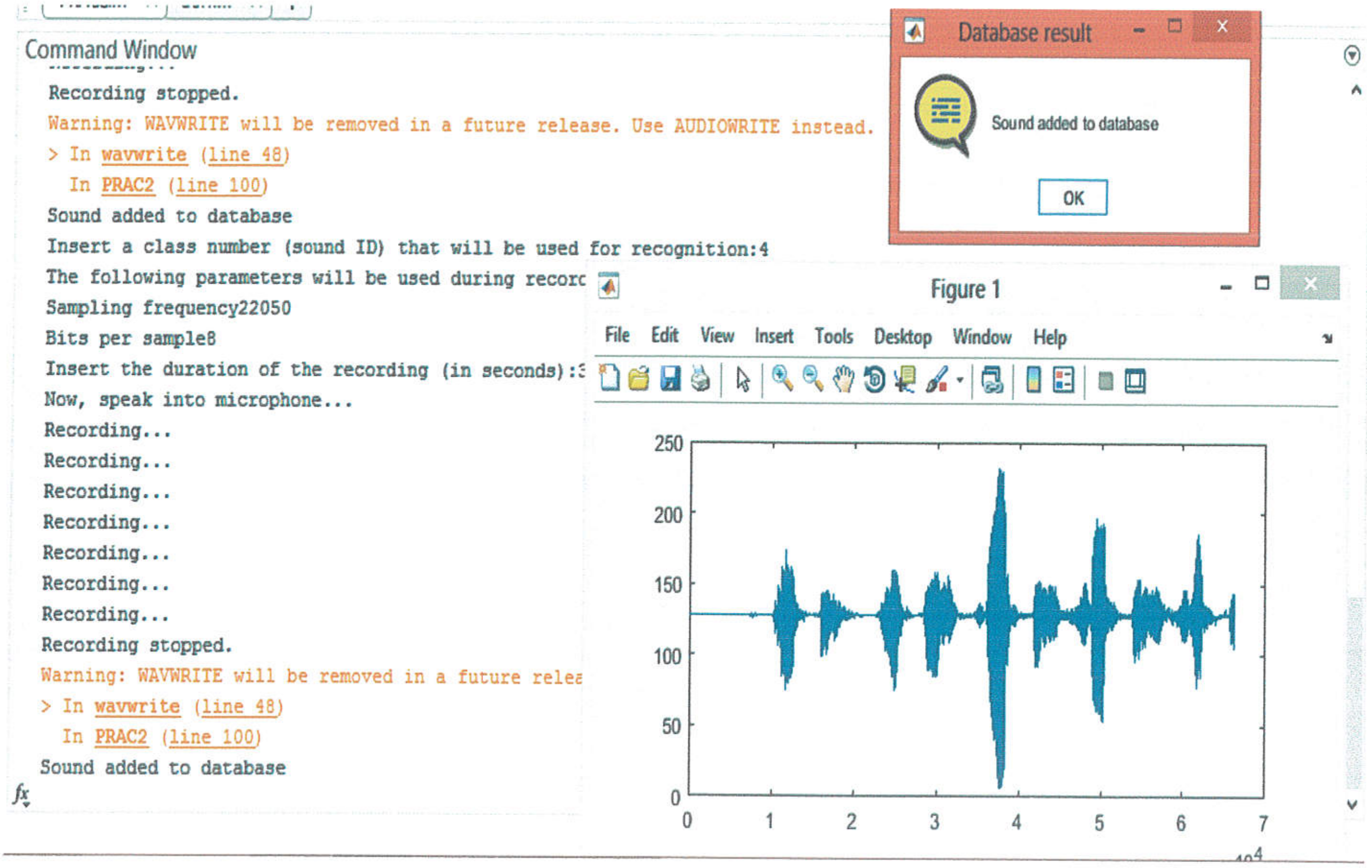


Figure 4.6: Acquisition of speech data “sound ID: 4”


Figure 4.7: Acquisition of speech data "sound ID: 5"

The above diagrams represent speech signals from five (5) different individuals tagged as sound ID 1 to sound ID 5, collected during the acquisition stage and stored in the speech system database.

3. Speaker recognition using VQLBG algorithm and access control stage

This is the final stage where the recognition tasks place. The aim is to identify the speaker's claimed identity by comparing the features of the input speech to those in the database to know which speaker it is and also, activate the opening of the door for the authorized user. It uses VQLBG algorithm which simply matches the inputted speech signal to a speech signal in the database based on the closeness of the parametric distance (Euclidean distance).

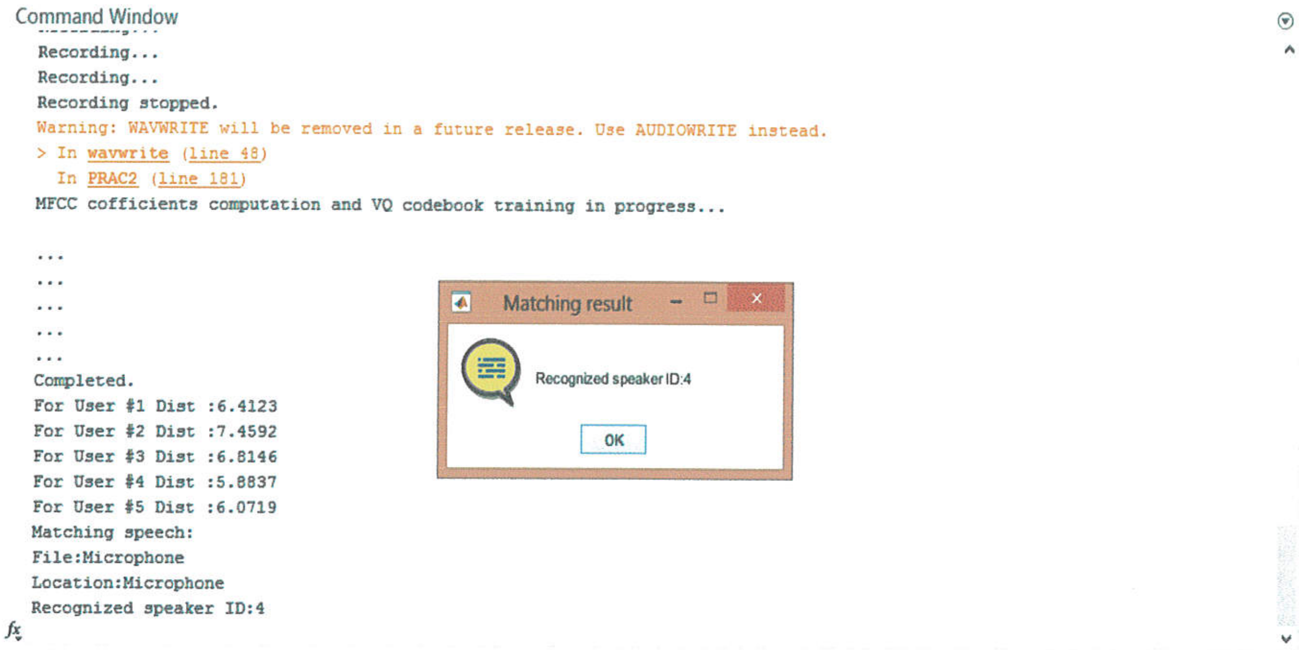


Figure 4.8: Recognized speaker "sound ID: 4"

On insertion of a speech signal to be recognized, as seen in the figure above, speaker ID: 4 is the recognized speaker as the system performs MFCC coefficient and VQ codebook training on the speech signal to be recognized and after the processing, recognition is done as it matches the speech signal to the user ID of which it is closest to in Euclidean distance which is user 4 (5.8837). User 1 distance is 6.4123, user 2 distance is 7.4592, user 3 distance is 6.8146, and user 5 distance is 6.0719.

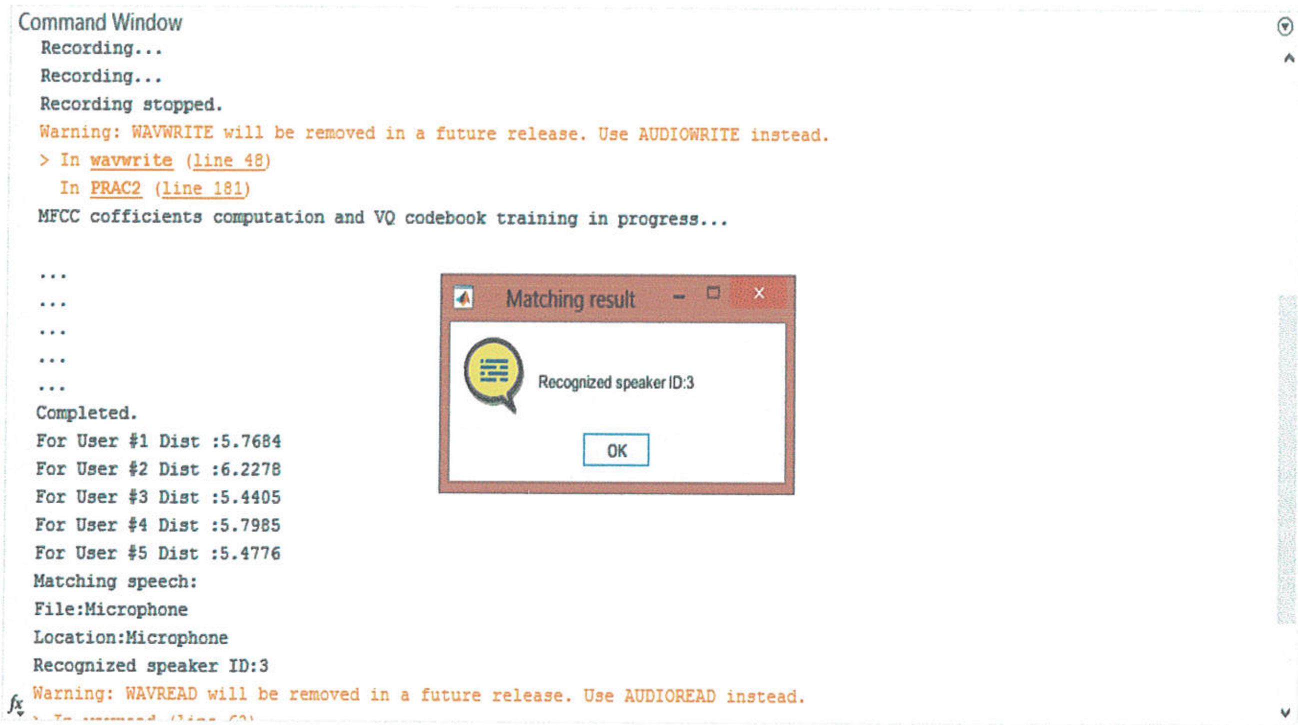


Figure 4.9: Recognized speaker "sound ID: 3"

Here, it matches the speech signal to the user ID of which it is closest to in Euclidean distance which is user 3 (5.4405). User 1 distance is 5.7684, user 2 distance is 6.2278, user 4 distance is 5.7985, and user 5 distance is 5.4776.

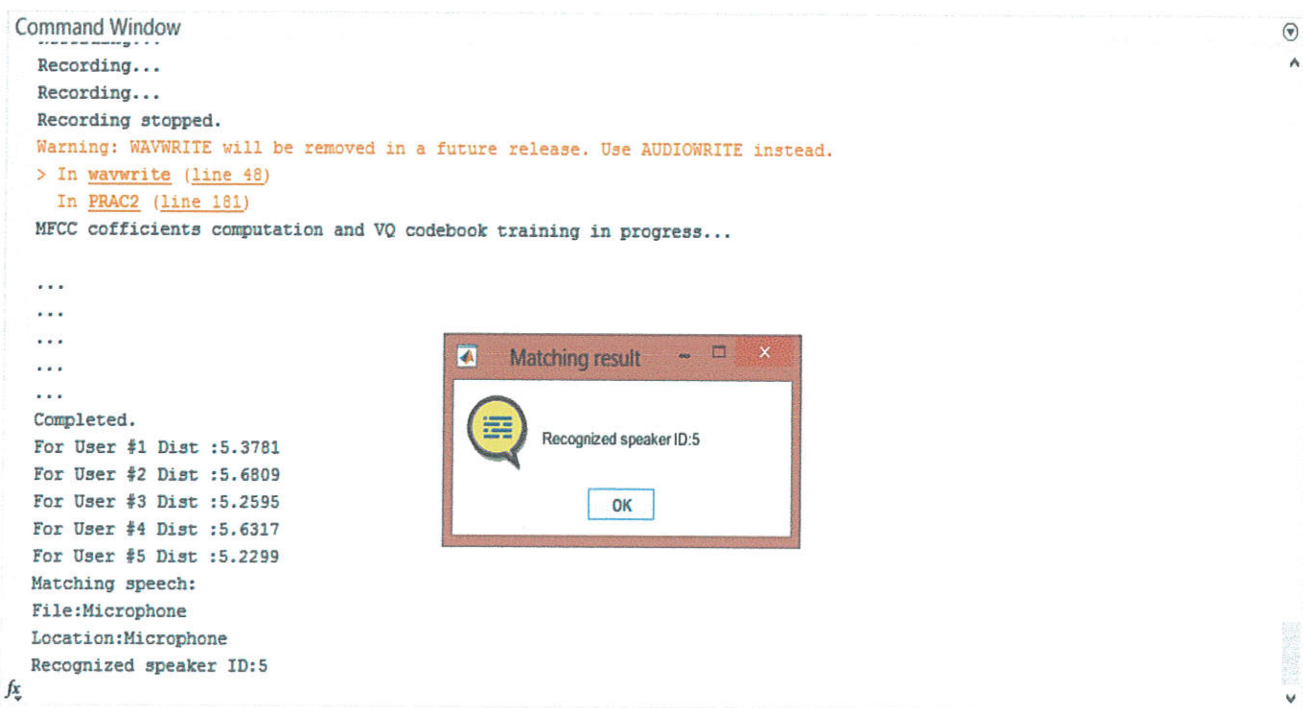

Figure 4.10: Recognized speaker "sound ID: 5"

Here, it matches the speech signal to the user ID of which it is closest to in Euclidean distance which is user 5 (5.2299). User 1 distance is 5.3781, user 2 distance is 5.6809, user 3 distance is 5.2595, and user 4 distance is 5.6317.

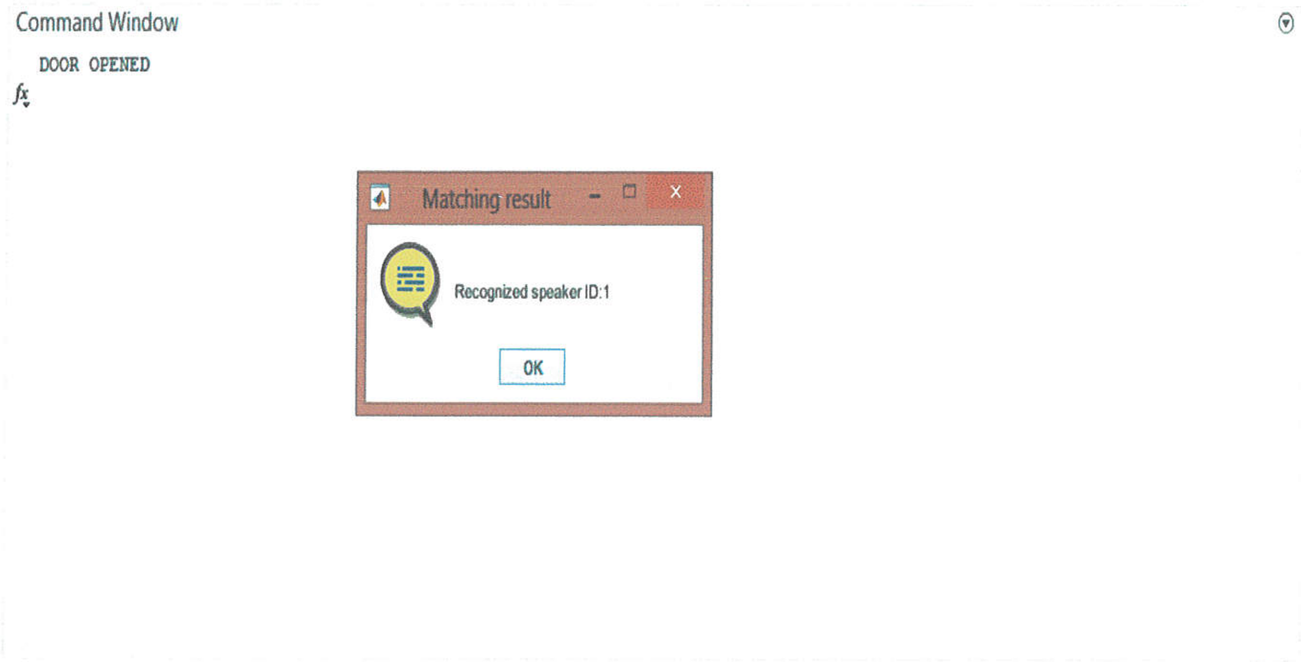


Figure 4.11: Recognized speaker "sound ID: 1", the authorized user

Here, the system recognizes the speaker as user 1, which is configured as the authorized user to gain access control to open the door.

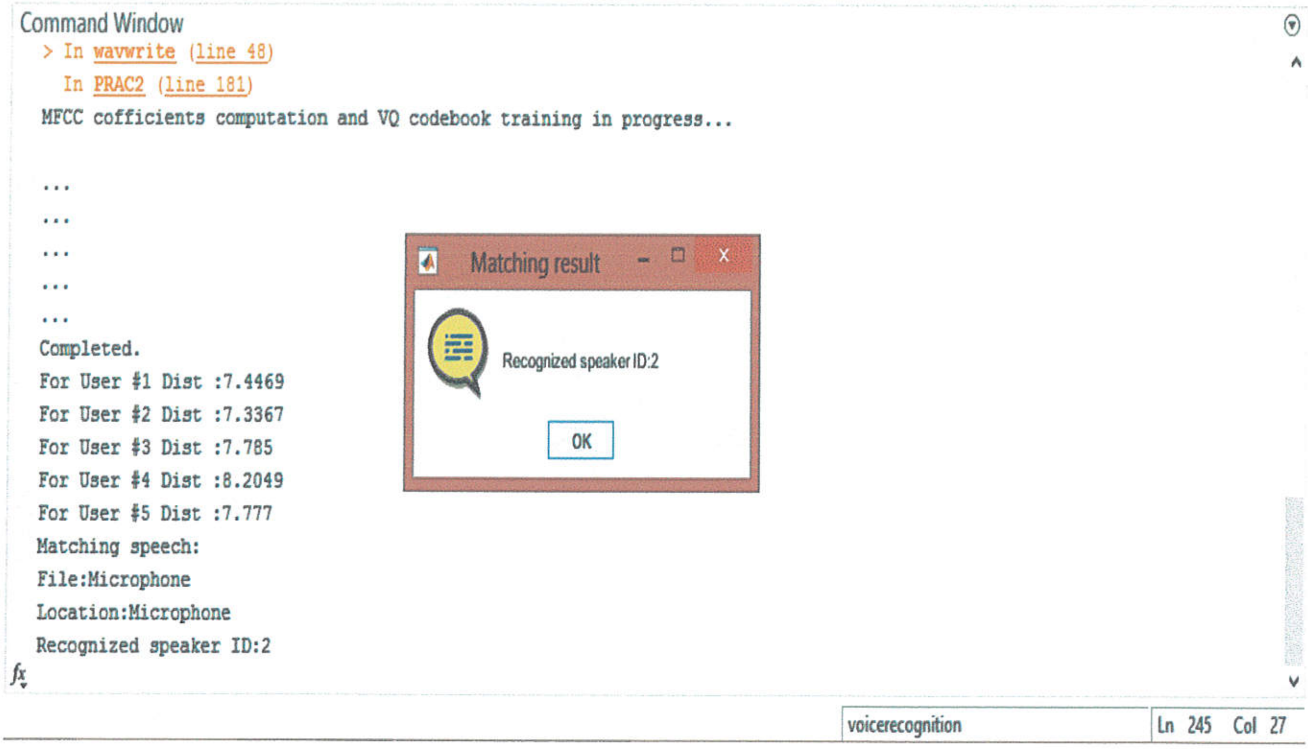


Figure 4.12: Recognized Speaker "Sound Id: 2"

Here, it matches the speech signal to the user ID of which it is closest to in Euclidean distance which is user 2 (7.3367). User 1 distance is 7.4469, user 3 distance is 7.785, user 4 distance is 8.2049, and user 5 distance is 7.777.

In this stage as shown in the diagrams above, the five different speakers inserts their speech data to the system through an analogue microphone and these speech signals are then processed as the system performs MFCC coefficients computation and VQ codebook training to get the features from the speech signals and match the speech signal to the appropriate user sound ID stored in the database of the system. The access control allows only one authorize user to have access. Looking at the diagrams above, speaker with sound ID 1 is the authorized user with the access to the building through the door. When speaker with sound ID 1 is recognized, the door automatically opens for 5 seconds and close after that period of time. On recognition of the authorized user, a signal is sent from the MATLAB through the RS232 program cable to the hardware prototype.

Table 2: Parametric Representation Of Recognition Process Of Users.

| **Users in Database** | **Speakers to be recognized using the individual Euclidean distance of features extracted to the features of data stored in the speech system database.** | | | | |
|---|---|---|---|---|---|
| | **Distance to User-1** | **Distance to User-2** | **Distance to User-3** | **Distance to User-4** | **Distance to User-5** |
| **User-1** | 5.1527 | 7.4469 | 5.7684 | 6.4123 | 5.3781 |
| **User-2** | 5.8374 | 7.3367 | 6.2278 | 7.4592 | 5.6809 |
| **User-3** | 6.1213 | 7.7850 | 5.4405 | 6.8146 | 5.2595 |
| **User-4** | 5.3778 | 8.2049 | 5.7985 | 5.8837 | 5.6317 |
| **User-5** | 6.1183 | 7.7770 | 5.4776 | 6.0719 | 5.2299 |

The table above summarizes the result and evaluates the system based on the Euclidean distance of the individual speech signal to be recognized. Here, the parametric representation of the distances of the speech signal is shown as calculated when the MFCC and VQ codebook training function are applied. The user having the shortest feature distance is matched to a user stored in the database of the speech system, and thus, is the recognized user. In the table above, the red sections simply indicate the recognized users.

Word Recognition Rate (WRR) = S/H X 100%; where S is number of success in H number of recognition trials.

WRR = 15/20 X 100%; 0.75 X 100%, 75% accuracy.

### 4.4.2 HARDWARE IMPLEMENTATION AND TESTING EXPLANATION

- A slider mechanism with slide gear fixed with a motor which is gotten from a CD player, is used to slide the door opened and close in response to signal from the microcontroller.

- Lastly, the transformer, an AC transformer, can take in up to 220V and give out 9V.

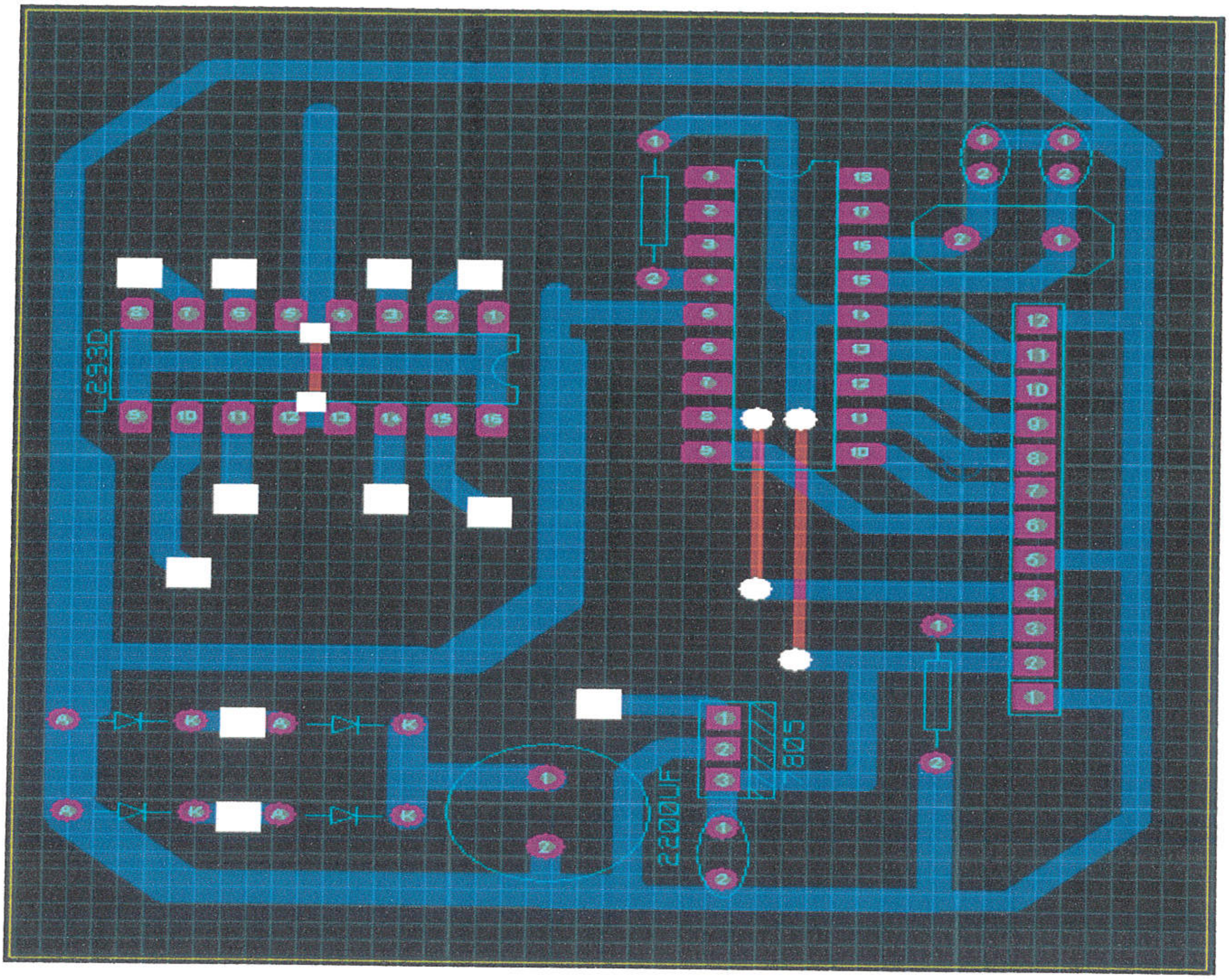


Figure 4.13: PCB layout of door control circuit

The PCB layout diagram of the door control circuit is shown above. Firstly, the adapter section where the AC voltage is received by the four (4) 1N4001 diodes for full wave rectification. This voltage after rectification is passed to the recommended 2200uF capacitor (a condenser microphone), which work by removing the lumps/ripples present in the voltage from the rectification. A 5V regulator is used to convert the 9V DC to 5V required voltage by the different materials used in the circuit. The connection of pins between the different materials are listed below;

For L293D Pin connections, Pin 4, 5, 11, 14 are connected to GND, Pin 8 is connected to VCC (can take 4.5V to 36V). Pin 3, 6, 11, 14 (Driver outputs) are connected to the DC motor to control its forward and reverse movement. Pin 17 and 18 from PIC18F1X20 is connected to pin 1 and 2 (Driver inputs) of L293D H-bridge driver.

Pin 9-14 of PIC18F1X20 is connected to pin 6-11 of the LCD. Pin 4 of PIC18F1X20 to 10kohms pull-up resistor. Pin 5 to VCC and 14 to GND for PIC18F1X20. Pin 15 and 16 of PIC18F1X20 is connected to 4MHz crystal oscillator, which is also connected to two 30uF capacitors in parallel. Pin 3 of LCD is connected to 10kohms resistor for contrast of the LCD.

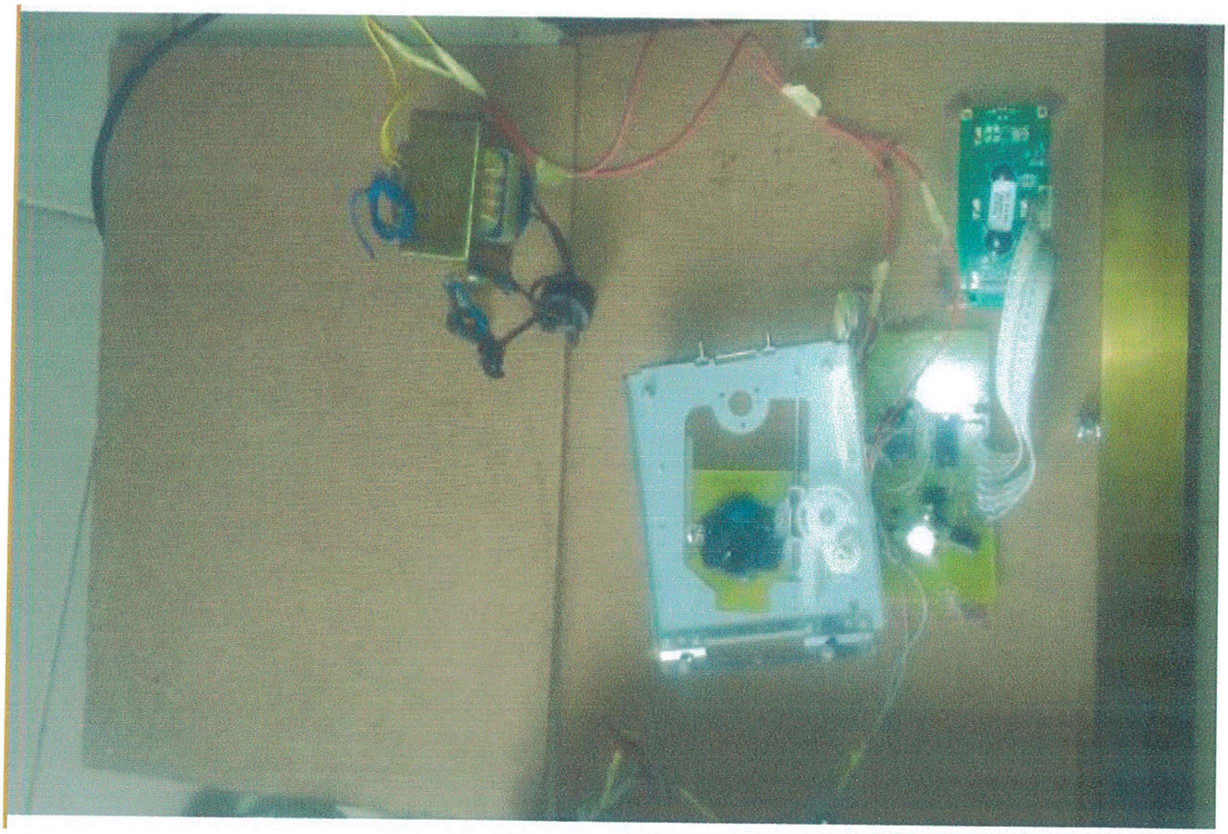

Figure 4.14: Diagram of The Internal Arrangement Of The Prototype/Model House Door Showing Its Components

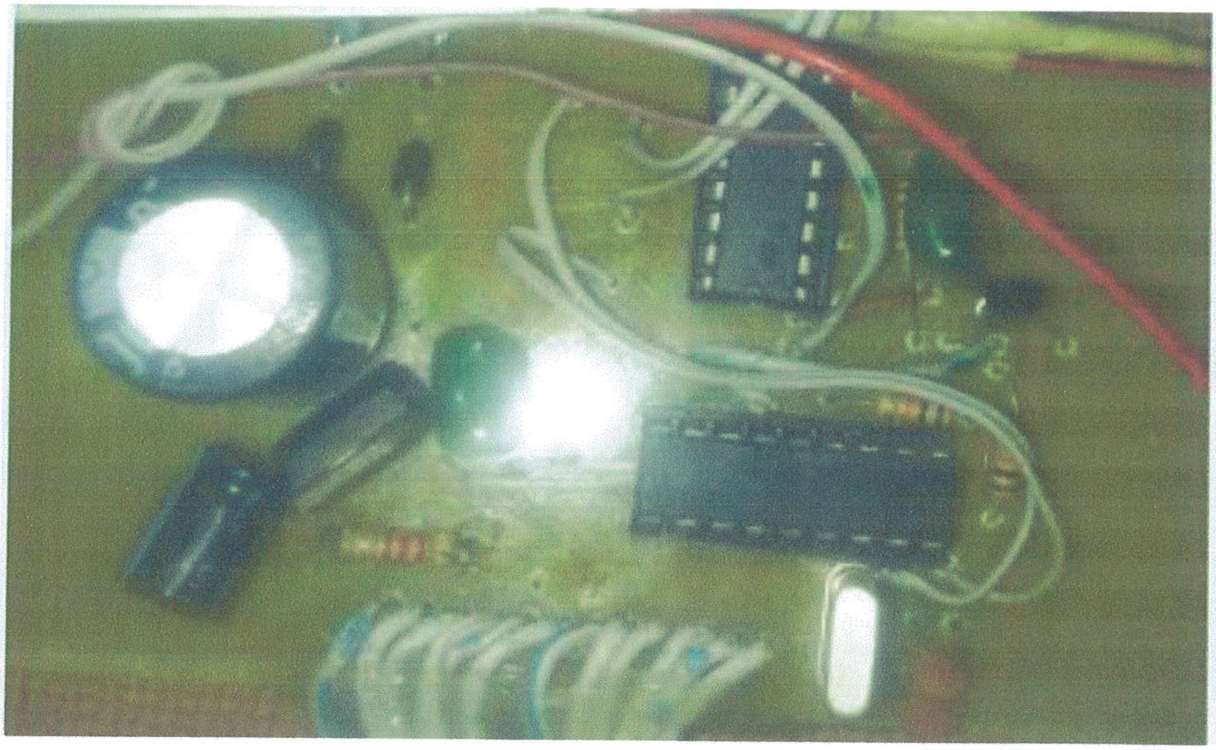

Figure 4.15: Door Control Circuit

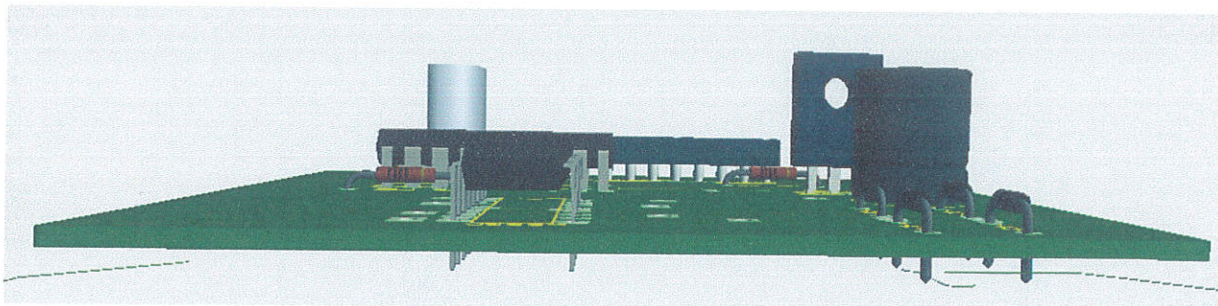

Figure 4.16: 3d Visualization Of Door Control Circuit (Side View)

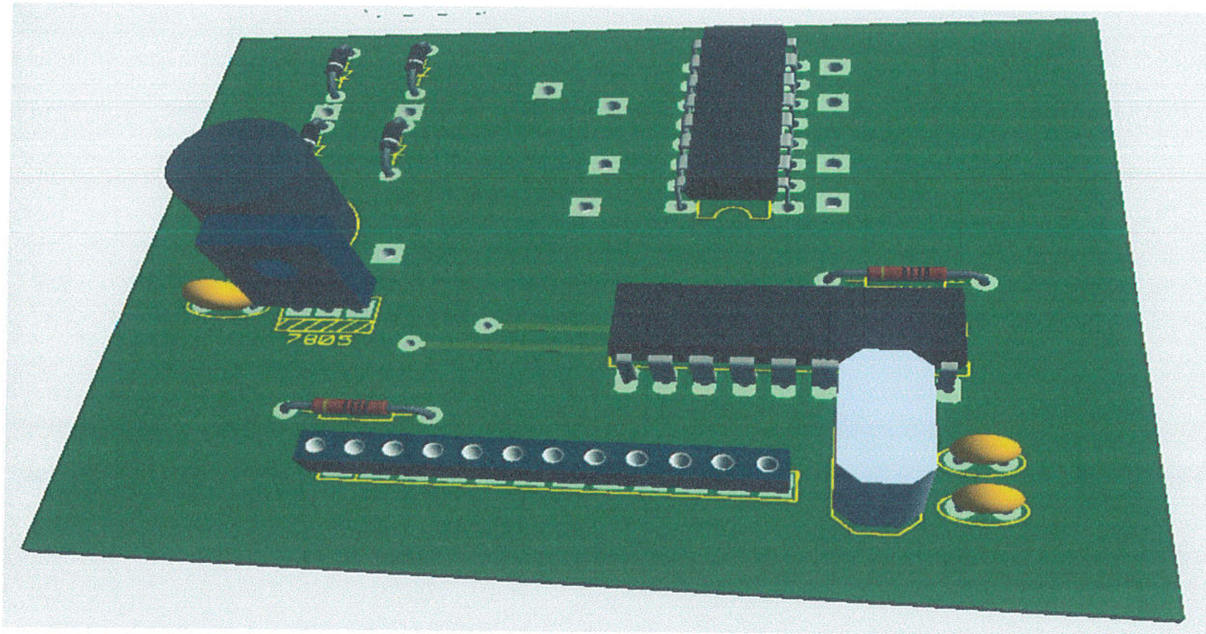


Figure 4.17: 3d Visualization of Door Control Circuit (Top View)

The 3D visualization of the door control circuit was done using the ARES professional in the proteus software.

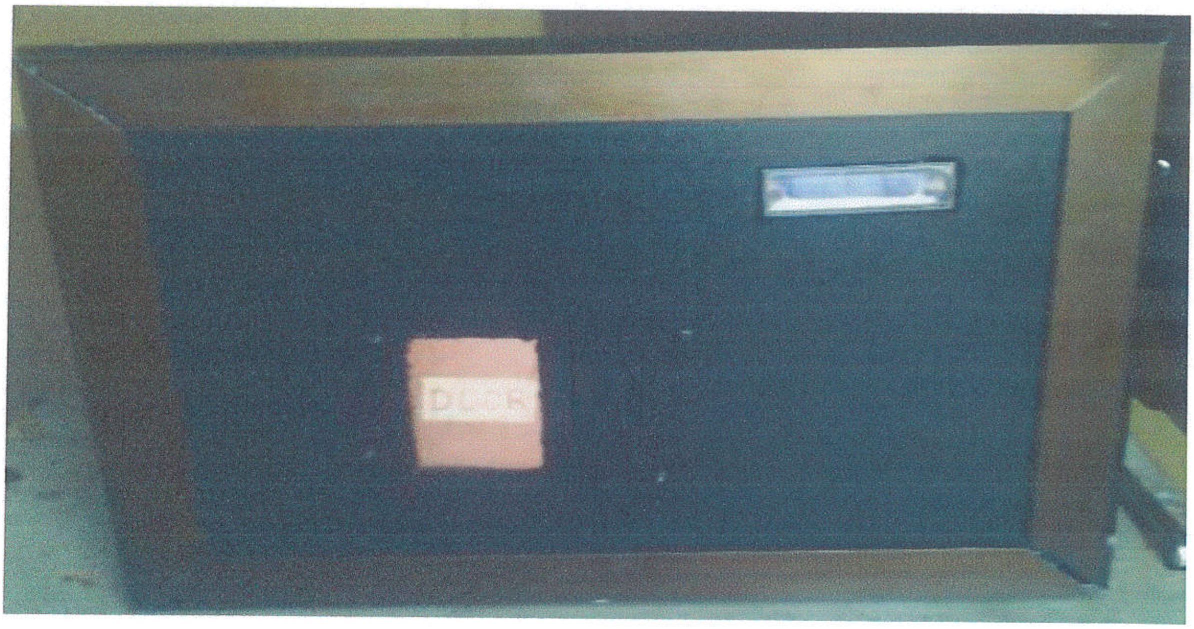

Figure 4.18: The Prototype/Model Door To Be Controlled By The Speech Recognition System

- The microcontroller PIC18F1X20 receives the signal from the MATLAB software through the RS232 cable and sends a signal to the motor for the door to be opened.
- The H-bridge driver controls the forward and backward movement of the door, i.e. controls the opening and closing of the model house door.
- A 4MHz crystal oscillator provides speed for the microcontroller and is connected to two 30uF capacitor connected in parallel.
- Rectifying diodes are provided for full wave rectification (1N4001 rectifying diode (4)). The rectified power of the AC will still have some ripples, after passing through the rectifier, becomes a pulsating DC, meaning it has ripples. The voltage is then sent to the condenser microphone, a 2200uF capacitor.
- The condenser microphone takes away the ripples making it smooth to get the pure DC voltage at 9V. It is the 9V that is fed into the 3pin regulator (a 5V regulator with path number LM7805). To recover the voltage after it has passed through the regulator, a 100uF capacitor is used. The 100uF capacitor close to the regulator acts as a storage that always makes 5V available to the microcontroller at any time. A 0.1uF capacitor is recommended as a noise filter for every IC (one for the microcontroller and the other for the H-bridge).
- A 10k resistor is used to provide the contrast for the LCD
- A 10K pull-up resistor is used for the microcontroller as it has a port with open circuit, they are not connected internally, and they are used to provide the missing filter when the port is in a high state.
- There are resistors connected at the base of the transistor that serves as a buffer, the pull-up and series resistors from the RS232 port form a buffering unit for the RS232 signal that is coming into the PIC, it provides the voltage level that the PIC needs to communicate with the PC.

# CHAPTER FIVE

# CONCLUSION, PROBLEMS ENCOUNTERED AND RECOMMENDATIONS

## 5.1 Conclusion

The system developed which is "speaker recognition security door lock system" is based on recognizing an unknown speaker from given a set of registered speakers in the system database. Here we have assumed the unknown speaker to be one of the known speakers and tried to develop a model to which it can best fit into. This project was accomplished using MATLAB software in creating the speaker recognition system. In the first step of generating the speaker recognition model, speech data acquisition and pre-processing was done, we then went for feature extraction using Mel Frequency Cepstral Coefficients. These features act as a basis for further development of the speaker identification process. Next on recognition phase, we went for feature mapping using the vector Quantization using LBG algorithm. The results obtained using MFCC and VQ are appreciable. MFCCs for each speaker were computed and vector quantized for efficient representation. The code books were generated using LBG algorithm which optimizes the quantization process. VQ distortion between the resultant codebook and MFCCs of an unknown speaker was taken as the basis for determining the speaker's authenticity. Euclidean distance was calculated for each signal to be recognized and matched to the closest speech signal in the speech database. Accuracy of 75% was obtained using VQLBG algorithm.

Finally, the functionality of the project was implemented and tested using a model house door system built, which has PIC18F1X20 microcontroller that receives access signal from the speaker recognition system design on MATLAB and controls the H-bridge driver which opens and closes the door.

## 5.2 Problems Encountered

Several problems were encountered during the project. The problems ranges from code problems to implementation problems and also construction problems. The major problems are as follows:

(1) The problem of representing the different speech samples in the speech system database with characters like words such as exact names of the users for better identification. This problem was solved by using numbers to represent each speech sample in the speech system database, such as sound ID 1, sound ID 2, etc.

(2) The problem of how large the database can take in users i.e. if there was to be a required number of user in the speech system database. This problem was solved as the system built was limited to maximum of 8 users for better efficiency of the system.

(3) For the hardware, the ICs i.e. PIC18F1X20 microcontroller and L293D H-bridge driver circuit generated noise which affected the logic count sequence. This was solved by proper filtering of the outputs using capacitors. Other problems include soldering and measurement errors but these problems were solved by proper troubleshooting serious care in the construction of the hardware.

## 5.3 Recommendations

For future works or improvements related to this project, the following suggestions can be applied;

(1) A system can be developed which the speech samples to be stored in the database will be represented with the user's exact name for easy identification.

(2) Also, the system developed to be able to contain wide range of users and still retain it efficiency and accuracy in recognition and feature mapping.

(3) The system can also include a voice output which states the operation of the system alongside the LCD which makes the system better usable for visually disabled individuals.

(4) The whole project can be made standalone by implementing it on a very high speed microcontroller interfaced with some necessary component which would cost more to achieve.

(5) Along with this system, the additional use of Face-detection can be implemented to enhance the security of the system.

## *REFERENCES*


Atal B.S. (1974). *"Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification,"* Journal of the Acoustical Society of America, *vol. 55, no. 6, pp. 1304–1312.*

Arpita Mishra, Siddharth Sharma, Sachin Dubey & Dubey S.K (2009). Paper named as *"Password based security lock system"*.

Ashish Kumar Panda, Amit Kumar Sahoo *"Study of speaker recognition systems"* Department Of Electronics And Communication National Institute Of Technology, Rourkela 2007.

BartekPlichta (2011). *"Best Practices in the Acquisition, Processing, and Analysis of Acoustic Speech Signals"*, Michigan State University. Historicalvoices,org/flint/extras/Audio-technology.pdf.

Bhargavi .Y, Hariprasad Reddy, Ravi TejaCh.V (2015). Paper named as *"Voice based wheel chair for physically challenged"*

DongsukYook (2003). *"Introduction to Automatic Speech Recognition"* Department of computer science, Korea University.

Deller J.R., Hansen J.H.L. & Proakis J.G. (2000): *Discrete-Time Processing of Speech Signals.* IEEE Press.

Fang Zheng, Guoliang Zhang, and Zhanjiang Song (2001). *"Comparison of Different Implementations of MFCC"*. The journal of Computer Science & Technology, *pp. 582-589.*

Hossan M.A. (2010). *"A Novel Approach for MFCC Feature Extraction"*, 4th International Conference on Signal Processing and Communication Systems (ICSPCS), pp. 1-5.

Hermansky .H, Hanson B.A, Wakita .H (1985). "*Perceptually based Linear Predictive Analysis of Speech,*" Proc. IEEE Int. Conf. on Acoustic, speech, and Signal processing, *pp. 509-512.*

Hermansky .H, Hanson B.A, & Wakita .H (1986). "*Perceptually based Processing in Automatic Speech Recognition,*" Proc. IEEE Int. Conf. on Acoustic, speech, and Signal processing, *pp. 1971-1974.*

Jain R. & Saxena S. K. (2011). *"Advanced Feature Extraction & Its Implementation In Speech Recognition System"*, IJSTM, Vol. 2 Issue 3.

Kharka Bahadur Rai, Jeetendra Thakur, Nirmal Rai (2015). Paper named as *"Voice controlled wheel chair using Arduino".*

Lawrence R. & Rabiner L (2000). *Speech Recognition by Machine.* CRC Press LLC.

Lia Kamelia, Alfin Noorhassan S.R, MadaSanjaya W.S. & Edi Mulyana. (2003). Paper named as *"Door-automation system using Bluetooth based android for mobile phone".*

Mansour Assaf, Ronald Mootoo, Sunil R.D, Emil M. Petriu, Voicu Groza, & Satyendra Biswas (2001). Paper named as "*Sensor based home automation and security system.*"

Mathur R., Babita, & Kansal A. (2010). *"Domain specific speaker independent continuous speech recognizer using Julius*", Proceedings of ASCNT – CDAC, Noida, India, *pp. 55 – 60.*

Md. Shafiul Islam (2010). Paper named as "*Home security system based on pic18f452 microcontroller*".

Meysam Mohamad pour & FardadFarokhi. (2009). "*An Advanced Method for Speech Recognition*", World Academy of Science, Engineering and Technology.

Neelam Majgaonkar, Ruhina Hodekar, Priyanka Bandagale. (2006). Paper named as *"Automatic door locking system".*

Pols, L.C.W. (1966). *"Spectral analysis and identification of Dutch vowels in monosyllabic words,"* Doctoral Dissertation, Free University, Amsterdam, The Netherlands.

Shaughnessy D.O. (2001). *Speech communication: Human and machine. Second Edition India:* University Press (India) Private Limited.

Shilpi Banerjee (2009). Paper named as *"Automatic password based door lock system"*.

Qasim Al-Shebani, Prashan Premaratne & Peter Vial (2003). Paper named as *"Embedded door access control systems based on face recognition"*.

Rabiner & Juang B-G. (1993). *"Fundamentals of Speech approach"*, Prentice Hall PTR.

Robert L. Boylestad, Louis Nashelsky. (1999). Electronic Devices and Circuit Theory (Eight Edition).

Simon Kinga & Joe Frankel (2006). *"Speech production knowledge in automatic speech recognition"*, Journal of Acoustic Society of America.

Thomas Quatieri (2002). *Discrete-Time Speech Signal Processing*: Principles and Practice, Pearson Education.

**APPENDIX I: COST AND PROJECT COMPONENTS.**

| Qty. | Description | Cost ( Naira ) | Total ( Naira ) |
|---|---|---|---|
| 1 | IC, LM7805, 5V | 200 | 200 |
| 1 | IC, PIC18F1X20 | 1,200 | 1,200 |
| 1 | IC, L293D, H-BRIDGE DRIVER | 750 | 750 |
| 4 | DIODE,1N4001 | 50 | 200 |
| 1 | DC MOTOR | 350 | 350 |
| 1 | SLIDER MECHANISM | 800 | 800 |
| 1 | GEAR SYSTEM | 1,050 | 1,050 |
| 1 | MICROPHONE (ANALOGUE) | 800 | 800 |
| 2 | CAPACITOR, 0.1uF | 50 | 100 |
| 2 | CAPACITOR, 30uF | 50 | 100 |
| 1 | CAPACITOR, 100uF | 50 | 50 |
| 1 | CAPACITOR, 2200uF | 150 | 150 |
| 4 | RESISTOR, 10K OHM | 50 | 200 |
| 1 | LCD( 16X2 ) | 850 | 850 |
| 1 | TRANSFORMER 220V/9V, 2A | 1,100 | 1,100 |
| 1 | 4MHZ CRYSTAL OSCILLATOR | 400 | 400 |
| 1 | RS232 COM CABLE | 1,500 | 1,500 |
| 1 | VERO BOARD | 200 | 200 |
| 1 | CASING | 3,500 | 3,500 |
| 4 | SOLDERING LEAD | 100 | 400 |
| 50 | CONNECTOR/JUMPER WIRE | 10 | 500 |
| | **TOTAL ( Naira)** | | 14,400 |