DEVELOPMENT OF A MICROCONTROLLER BASED SPEECH RECOGNITION SYSTEM USING LINEAR PREDICTIVE CODING (LPC) ALGORITHM

ABATAN, OLUWAGBOTEMI DAMILOLA

CPE/13/1069

DEPARTMENT OF COMPUTER ENGINEERING

FACULTY OF ENGINEERING

FEDERAL UNIVERSITY OYE-EKITI,

EKITI STATE NIGERIA.

MARCH, 2019

CERTIFICATION

This is to certify that this project work titled "Development of a microcontroller based speech recognition system using Linear Predictive Coding (LPC) algorithm." was prepared and submitted by Abatan, Oluwagbotemi Ðamilola to the department of Computer Engineering, has been read and it is as meeting the partial requirement for the award of bachelor degree in Engineering in the department of Computer Engineering, Federal University Oye –Ekiti, Ekiti State.

.... 17 - 04 - 19

Engr. N. S. Okomba **Date**

(Project supervisor)

.... 16/4/19

Dr. (Engr.) O. M. Olaniyan **Date**

(Head of Department)

DECLARATION

I Abatan, Oluwagbotemi Damilola hereby declare that this project work carried out is the result of my personal effort under the supervision of Engr. N. S. 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

.... 16/04/2019

Abatan, Oluwagbotemi Damilola **Date**

(CPE/13/1069)

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.

ACKNOWLEDGEMENT

I appreciate the Almighty and awesome God who saw me through this program. I wish to express my profound gratitude to the contribution of various individuals who contributed to my acquisition of knowledge and successful completion of this project. My sincere appreciation goes to my parent Mr & Mrs. Adu Abatan, my Grandmother Mrs. Eunice Daramola, my Uncle Mr. Olumuyiwa Daramola, my siblings Ms. Abatan Ibironke and Master Komolafe Iyanuoluwa. Noteworthy thanks is extended to my project Supervisor Engr. N.S Okomba who has demonstrated great love and patience in supervising the project work. I really appreciate him for making himself available at all time despite his tight schedule for consultation. I really appreciate his helping hand both theoretically and practically.

I humbly appreciate the dean of the faculty of engineering Prof. Alabadan, the deputy dean Dr. (Engr.) I.A Adeyanju, the Head of Computer Engineering Department Dr. (Engr.) O.M Olaniyan, Engr. (Mrs.) Esan, Engr. (Mrs.) Omodunbi, Engr. Adegboye Mutiu, Engr. T. A Badmus, Engr. Adeleye, Engr. Adegboyega Otenaiki, Engr. Adebimpe, Engr. Candidus Okwor, Engr. Ms. Odiase and all non-academic staff of Computer Engineering Department.

I acknowledged my colleagues and friends namely; Abatan Oluwagbotemi, Bolaji Temitayo, Sunday John, Alebiosu Tobi, Alegbeye Bunmi, Adebayo Promise, Ayeni Samuel, Okunade Bolaji, Jayeoba Monisola, Fatai Isiaq, Olanitori Tobi, Olaniyi Michael, Toloruntomi Samuel, Nwafor Olive, Okunlola Saheed, Oyedokun Folakemi Ms. Oyebanji Adeola, Komolafe Funmileye, Nbakamma Osinachi, Ahmed Usman, Olorunfemi Daniel, Udor Oluwatimileyin and Late Ms. Ajetomobi Blessing.

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, friends, colleagues, and all not mentioned, I would like to say thank you and may God bless you all.

ABSTRACT

Speech Recognition is the process in which certain words of a particular speaker will be automatically recognized based on the information included in individual speech waves. In today's world, speech recognition has become very popular and important, speech plays a vital role in human to system and machine communication, speech signal is transformed from analog to digital wave form, which can be understood by the system or machine.

This project "Development of microcontroller based speech recognition system using Linear Predictive Coding (LPC) algorithm" is useful in instances when the user is unable to use keyboard for data entering because of old age, user's hands are occupied, disabled, or suffering from diseases such as quadriplegia, and also where speech command is more preferable to use of keyboard.

The system has two main parts: speech recognition and home appliances electronic control system. Speech recognition is implemented in MATLAB environment that contains two main modules: feature extraction and feature matching. Linear Predictive Coding (LPC) extraction techniques is used over others. e.g. PLP, MFCC due to its wide acceptance in speech processing for reducing the size of the transmitting signal and the encryption of data. Vector Quantization (VQ) approach using LBG algorithm (VQLBG) algorithm is applied for feature matching. Home appliances was modeled to control water pump and electric fan using 2N2222A Transistor for switch operation, PIC16F628A microcontroller and a Liquid Crystal Display (LCD, 16X2) to display the state of operation of the system.

The implementation of this project was done and tested to show its reliability and efficiency. The system was evaluated using Euclidean Distance. The registered speaker's Euclidean distance to spoken words such as "BREEZE" and "WATER" were 88.8% and 85.9% word accuracy.

TABLE OF CONTENTS

CHAPTER TWO: LITERATURE REVIEW

CHAPTER THREE: METHODOLOGY

CHAPTER FOUR: RESULTS AND DISCUSSIONS

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

- Automatic Speech Recognition (ASR)
- Word Error Rate (WER)
- Light Emitting Diode (LED)
- Complimentary Metal-Oxide Semiconductor (CMOS)
- Direct Current (DC)
- Operational Amplifier (OP-AMP)
- Linear Predictive Coding (LPC)

CHAPTER ONE

INTRODUCTION

1.1 Preamble

Biologically, speech is produced by the passage of air through various obstructions and routings of the human larynx, throat, mouth, tongue, lips, nose, and etcetera that will be decoded by brain. (Rabiner, Fundamental of Speech Recognition, 1993). Speech recognition is a field of computer studies aim to convert speech into a sequence of words by a computer program. The term "speech recognition began in the early 1960's with the exploration into voiceprint analysis which was somewhat similar to fingerprint concept.

Linear predictive coding (LPC) is chosen over others features extraction techniques because LPC is a powerful features extraction techniques which is used to reduce the bitrates of the speech i.e. reduces the size of the transmitting signal, LPC required less bandwidth and hence no. of users can be increased and LPC algorithm uses the encryption of data so the data is secured until the destination.

The desirable characteristics of speech recognition system such as scalability that is the efficiency of the communication, accuracy of recognition with ability to filter out noise, and adaptation to acoustic conditions, such as speaker's speech rate and accent that were embedded in Voice-activated routing systems at customer call centres, Voice dialling on mobile phones, and many other everyday applications. When a user speaks, a microphone convert the analog signal of his voice into digital chunks of data that can be analysed by computer (Fletcher, 1922.).

Application of speech recognition implemented using PIC16F628A microcontroller to home appliance such as fan and water pump as another means of control will somewhat different from ordinary fan and water pump systems. It will provide the facility of passing information and commands to home appliances system through microphone. Such facility and control supports special needs and services that support the old age, people with disability and parallel tasks (i.e. hands and eyes are busy elsewhere) while working. The user voice will be a more preferable input command to power ON/OFF the fan and water pump devices to the use of keyboard. Thus, it will potentially reduce power consumption by preventing occurrences of fan and water pump device being left on longer than necessary, saving money and the environment. (Lovine, 2013).

1.2 Problem Statement

Traditionally, fan, water pump system and others appliances used in many house/office and factory are controlled by using control switch and fan regulator mounted on the wall manually. This is a time and energy consuming method for people that are busy at work. It is a difficult task for elderly, handicap, disable or paralytic and asthmatic patient at dying state to be pressing/switching fan regulator and water pump control switch before getting his/her system power on and off (Ferrier, Shane, Ballard, Carpenter, & Benoit, 1995). The proposed speech recognition controlled fan and water pump system will allow the user to perform parallel tasks (i.e. hands and eyes are busy elsewhere) while working, their voices will be an alternative input command to ON/OFF the fan and water pump system. Also it will potentially saves money and the environment, and reduce power consumption by preventing fan and water pump system being left on longer than necessary (Flanagan, 1972).

1.3 Aim and Objectives

The aim of this project is to develop a microcontroller based fan and water pump system that is controlled using speech recognition.

To fully satisfy the aim of this project, as a minimum the following outlined objectives must be met:

1. To design a microcontroller based speech recognition system using LPC algorithm to control home fan and water pump.
2. To implement the design using a hardware prototype.
3. To evaluate the effectiveness of the system.

1.4 Scope of Study

This project concentrates on a development of a microcontroller based speech recognition system to control home fan and water pump. 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 will be handled using MATLAB software for coding of the speech recognition system, while the physical model which is a small model electric fan and water pump, in which the circuitry has to control. The speech recognition system will be speaker independent.

1.5 Significance of the Project

The project is intended to control home appliances such fan and water pump by the application of user's speech/commands rather than using wall mounted switch regulator and remote control. Also, this system will be designed for people with disability, old age, patients that are suffering from sickness and diseases, and users that undergo parallel tasking in his/her day to activities. E.g. banking officials, surgeon, industrial operators and others. This speech recognition controlling method is cost efficient and have a highly functional design which potentially reduce power consumption by preventing occurrences of fan and water pump device being left on longer than necessary. Thus, saves money and prevents the environment from electric hazards.

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 system controlling home appliances around the world.

1. The system will ensure easy operation of fan and water pump system.
2. The project will allow the physically challenge ones (handicap) to be able to operate their fan and water pump system conveniently.

1.6 Methods of Study

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:

i. Literature Survey
ii. Project Design (Hardware and Software)
iii. Circuit Testing
iv. Project Construction

v. Troubleshooting

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Early speech recognition systems tried to model the human articulatory channel. They didn't work. In 1970s, a transition was caused by the success of the HEARSAY and HARPY systems at Carnegie Mellon University (CMU) and a science fiction called "Star Trek to George Orwell's in 1984 after systems have been trained on example data rather than defined using rules (Thomas-Stonell *et al* 1998) before the best recognised speech recognition systems that were yielding a 15% error rate on a relatively small 20,000 dictation tasks were developed as early 1990s,. Human speech has change over many thousands years to become an efficient method developed for speech recognition, which could recognise the speech more efficiently with an error up to 2% for the benefit of sharing information and giving instructions to control appliances, tools and computers (Jain *et al* 2002).

The implementation of Speech recognition using different languages around the world has become practical concept in years. This chapter intends to discuss the relevant background to the field of speech recognition that has been used in real-world human language applications, such as information recovery. That is how the speech signal is produced and perceived by human beings. Speech recognition system can be characterised by environment, vocabulary, acoustic model, speaking style, speaking mode, language model, perplexity, Signal-to-Ratio (SNR) and transducer.

The Literature survey for research on the basics of speech recognition and different methods used in this field are done by referring to the journal paper, article, conference paper, books, internet and databases. This chapter describes a review of speech signal, speech production, properties of human voice, automatic speaker recognition system, speech recognition basics, classification of ASR system based on recognition style, ASR systems problems, speech analyser, speech classifier, speech recognition approaches, matching techniques, and speech controlling system.

2.2 The Speech Signal

Speech in human can be said as the most common means of the communication because the information maintains the basic role in conversation. The conversation that is captured by a microphone is converted from acoustic signal to a set of word that can either be the final result or it can then apply the synthesis to pronounce into sounds, meaning speech-to-speech. That is speech recognition is used as input to further linguistic processing in order to achieve speech understanding. The comprehensive diagram below demonstrates human communication process from speech production to speech perception between talker and listener. As stipulated in Figure 2.1.

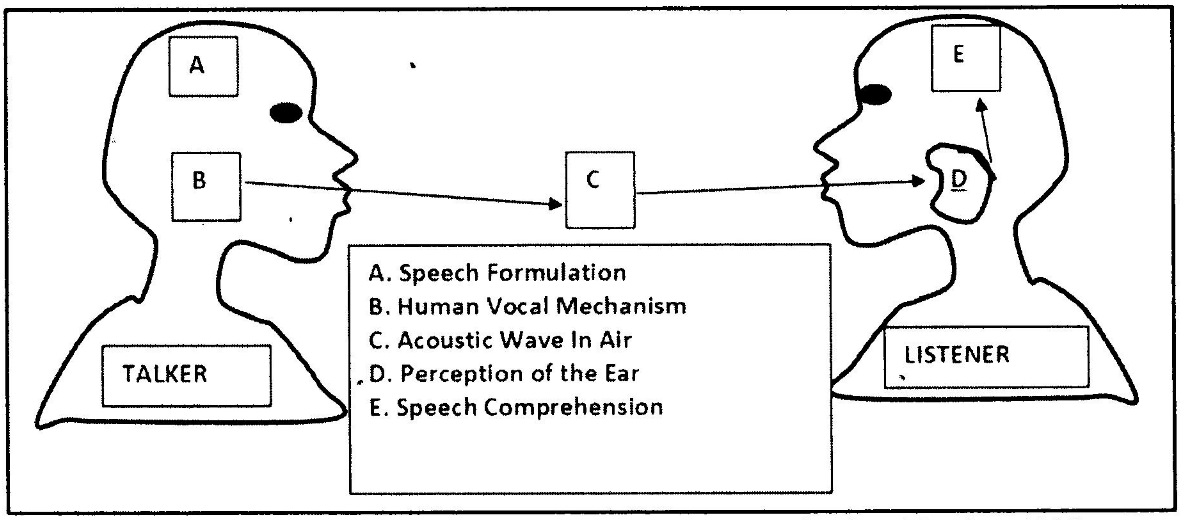


Figure 2.1: Five different elements of human communication (Fletcher, 1922).

The diagram describes five different elements of human communication. Namely,

A. Speech formulation: First element of communication that is associated with the formulation of the speech signaling the Talker's mind. This formulation is used by the second element (human vocal mechanism).

B. Human vocal mechanism: The function of this element is to produce the actual waveform transferred via air to the listener.

C. Acoustic air: The waveform transferred via air can be affected by external sources, such as noise, resulting in a more complex waveform.

D. Perception of the ear: The listener percepts the waveform and listener's mind when the wave reaches the Listener's hearing system.

E. Speech comprehension: This element processes the waveform into a comprehensive content for listener understanding of the talker's information. Speech recognition is needed to "simulate" how the listener process the speech produced by the talker. There are several actions taking place in the listeners head and hearing system during the process of speech signals. The perception process can be seen as the inverse of the speech production process. Phonemes that is defined as the basic theoretical unit for describing how to bring linguistic meaning to the formed speech, in the mind, can be grouped based on the properties of either the time waveform or frequency characteristics and classified in different sounds produced by the human vocal tract (Rabiner, 1993). Thus, speech has these followings characteristics:

i. Time-varying signal.
ii. Well-structured communication process.
iii. Depends on known physical movements.
iv. Composed of known, distinct units (phonemes).
v. Is different for every speaker.
vi. May be fast, slow, or varying in speed.
vii. May have high pitch, low pitch, or be whispered.
viii. Has widely-varying types of environmental noise.
ix. May not have distinct boundaries between units (phonemes).
x. Has an unlimited number of words.

2.3 Speech Production

Knowing how the human's vocal mechanism is being constructed will give the understanding on how the production of speech is performed. As stipulated in Figure 2.2.

Vocal tract together with nasal cavity are the key parts of the human vocal mechanism, which begins at the velum. The velum is a trapdoor-like mechanism that is used to formulate nasal sounds when needed. When the velum is lowered, the nasal cavity is coupled together with the vocal tract to formulate the desired speech signal. The cross sectional area of the vocal tract is limited by the tongue, lips, jaw and velum and varies from 0-20 cm2 (Nave .R, 2014).

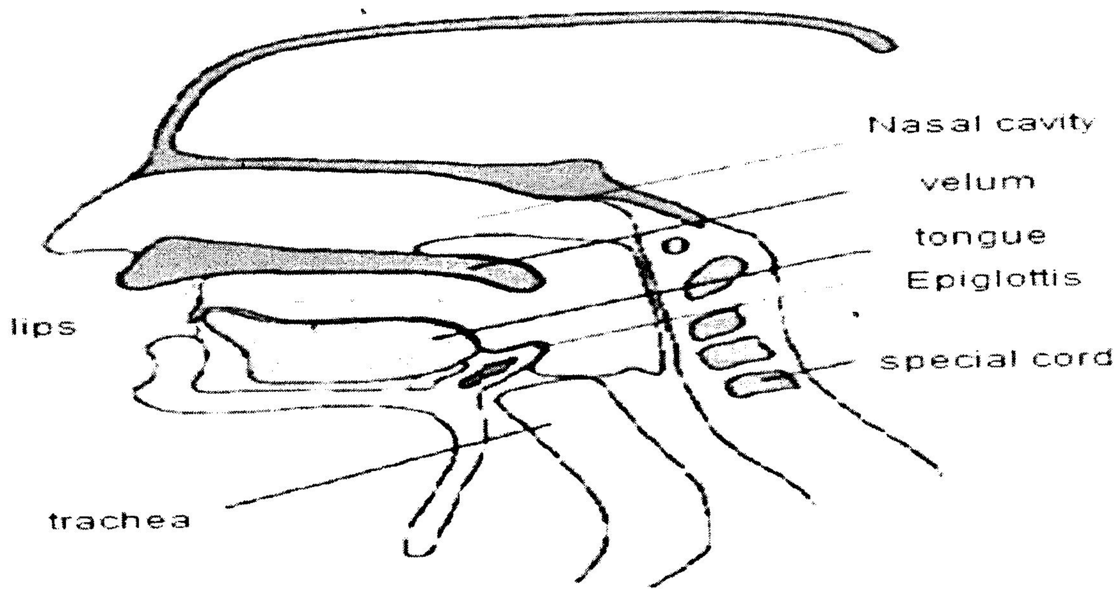


Figure 2.2 Human vocal mechanism (Rabiner, 1993)

2.4 Properties of Human Voice

One of the most important parameter of sound is its frequency. The sounds are discriminated from each other by the help of their frequencies. When the frequency of a sound increases, the sound gets high-pitched and irritating. When the frequency of a sound decreases, the sound gets deepen.

Sound waves are the resulted from vibration of the materials. Human can produces sounds of range 70Hz – 10 KHz. That is minimum and maximum values, and this frequency interval changes for every person. The magnitude of a sound is expressed in decibel (dB). A normal human speech has a frequency interval of 100Hz - 3200Hz and its magnitude is in the range of 30 dB - 90dB. A human ear can perceive sounds in the frequency range between 16 Hz and 20 kHz. And a frequency change of 0.5 % is the sensitivity of a human ear (Mann .W, 2005).

These following Speech characteristics varies from one speaker to another speaker:

i. Due to the differences in vocal tract length, male, female, and children's speech are different.

ii. Regional accents are the differences in resonant frequencies, durations, and pitch.

iii. Individuals have resonant frequency patterns and duration patterns that are unique (allowing us to identify speaker).

2.5 Automatic Speaker Recognition (ASR) System

Speech processing are done in a digital representation whereby speech being processed is seen as the interaction of digital signal processing and natural language processing. Natural language processing is a subfield of artificial intelligence and linguistics that studies the problems of automated generation and understanding of natural human languages. The role of natural language generation systems are to convert information from computer databases into normal-sounding human language, while conversion of samples of human language into more formal representations that are easier for computer programs to manipulate are done by natural language understanding systems. (Meysam, 2009).

2.5.1 Speech Coding

Speech coding is the compression of speech into a code for transmission with speech codecs that use audio signal processing and speech processing techniques.

The techniques used for speech coding are similar to that in audio data compression and audio coding where knowledge in psychoacoustics is used to transmit only relevant data to the human auditory system. For example, in narrow band speech coding, only information in the frequency band of 400 Hz to 3500 Hz is transmitted but the reconstructed signal is still adequate for intelligibility. However, the disparity between speech coding and audio coding is the availability of lot more statistical information that defines the properties of speech. The most important criterion in speech coding is the preservation of intelligibility and "pleasantness" of speech, with a constrained amount of transmitted data .The intelligibility of speech includes, the actual literal content, speaker identity, emotions, intonation, timbre etc. that are all important and needed to be emphasized for perfect intelligibility. The more abstract concept of pleasantness of degraded speech is a different property than intelligibility, since it is possible that degraded speech is completely intelligible, but subjectively annoying to the listener (Kharka et al, 2015).

2.5.2 Speech Synthesis

Speech synthesis is the artificial production of human speech. In the course of modelling human speech, there were approaches applied. Such as a Text-to-Speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Speech synthesis can also be implemented by concatenating pieces of recorded speech that are stored in a database. These systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. For example, since 1980s an intelligible text-to-speech program that allows people with visual impairments or reading disabilities to listen to written works on a home computer. Nowadays, many computer operating systems have included speech synthesizers (Shafiul, 2010).

2.5.3 Voice Analysis

As a result of voice problems originated from the vocal cords, voice analysis is required since it is the sound source and is thus most actively subject to tiring. However, the location of the vocal cords effectively prohibits direct measurement of movement. Imaging methods such as x-rays or ultrasounds do not work because the vocal cords are surrounded by cartilage which distorts image quality contribute to physical difficulties facing voice analysis.

Movements in the vocal cords are rapid, fundamental frequencies are usually between 80 and 300Hz, thus preventing usage of ordinary video. High-speed videos provide an option but in order to see the vocal cords the camera has to be positioned in the throat which makes speaking rather difficult.

Most important indirect methods are inverse filtering of sound recordings and Electro Glotto Graphs (EGG). In inverse filtering methods, the speech sound is recorded outside the mouth and then filtered by a mathematical method to remove the effects of the vocal tract. This method produces an estimate of the waveform of the pressure pulse which again inversely indicates the movements of the vocal cords. (Simon and Joe, 2006).

2.5.4 Speech Recognition

Speech recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means talking to your computer, and having it correctly recognize what you are saying. The basic principle to achieve speech recognition is to somehow extract certain key features from the uttered speech and then treat those features as the key to recognizing the word when it is uttered again. Speech recognition has become a controlling method for home appliances, toys, tools, computers and robotics, while increasing the effectiveness and efficiency of working with the device.

2.6 Speech Recognition Basics

2.6.1 Utterance

An utterance is the vocalization (speaking) of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences.

2.6.2 Speaker Dependency

Speech recognition is classified into two categories, speaker dependent and speaker independent.

Speaker dependent systems are designed around a specific speaker or group of speakers, i.e. speaker dependent system are trained by individual who will using the system. Generally, they are more accurate for the correct speaker, but much less accurate for other speakers. The system is designed with the assumption that the speaker will speak in a consistent voice and tempo. Speaker independent systems are designed for a variety of speakers with 95% accuracy for word recognition. Adaptive systems usually start as speaker independent systems and utilize training techniques to adapt to the speaker to increase their recognition accuracy. Thus, this approach is employed in software for personal computers.

Speaker independent system is designed or trained to respond to a word regardless of who speak to the system.

This is obviously much more complex than speaker dependent recognition. Because the system must respond to a large variety of speech patterns, inflations and enunciation's of the target word.

A problem of intermediate complexity would be to train with a group of speakers and recognize speech of a speaker within that group. Though the command word count is usually lower than the speaker dependent system, however high accuracy can still be maintained within processing limits (Mathur and Kansal, 2010). Speaker independent systems still impose many problems as follows:

i. There may be inconvenience to the user due to long training sessions.
ii. The processing time required is very large before the system can be used.
iii. Certain real-time applications cannot tolerate the delay of the training session.
iv. If each speaker's parameters are to be stored then a large amount of storage is needed separately for that.
v. There may be a change in the speaker's voice over time due to stress, exhaustion, illness, or variations in microphone positioning.

Speaker independent system are majorly implemented to meet industrial controlling requirements. We could call this speaker group dependent recognition.

2.6.3 Vocabularies

Vocabularies or dictionaries are lists of words or utterances that can be recognized by the Speech Recognition System. Generally, smaller vocabularies are easier for a computer to recognize, while larger vocabularies are more difficult. Unlike normal dictionaries, each entry doesn't have to be a single word. They can be as long as a sentence or two. Smaller vocabularies can have as few as 1 or 2 recognized utterances (e.g. "Pump water", "Run fast"), while very large vocabularies can have a hundred thousand or more!.

2.6.4 Accuracy

The ability of a speech recognition system can be examined by measuring its accuracy i.e. how well it recognizes utterances. This includes not only correctly identifying an utterance but also identifying if the spoken utterance is not in its vocabulary. Good ASR systems have an accuracy of 98% or more! The acceptable accuracy of a system really depends on the application.

2.6.5 Training

Some speech recognizers have been designed with the ability to adapt to a speaker. When the system has this ability, it may allow training to take place. In the course of training, a ASR system speaker or user repeat standard or common phrases and adjusting its comparison algorithms to match that particular speaker. Training a speech recognizer is usually done to improve its accuracy. And to be used by speakers that have difficulty speaking, or pronouncing certain words.

2.7 Classification of (ASR) System Based on Recognition Style.

One of the constraints of speech recognition systems lies in the style of speech they can recognise. A speech recognition system can operate in many different conditions and styles such as speaker dependent/independent, isolated/continuous speech recognition, for small/large vocabulary (Mann, 2005).

2.7.1 Isolated Speech Recognition System:

This is the most common speech recognizers, in which command word are spoken separately and usually require each utterance to have quiet (lack of an audio signal) on both sides of the sample window. It doesn't mean that it accepts single words, but does require a single utterance at a time. Often, these systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses).

2.7.2 Connected Speech Recognition System:

This is a half-way point between isolate and continuous speech recognition system. Because connected word systems are similar to isolated words, but allow separate utterances to be 'run-together' with a minimal pause between them. Also, it allows users to speak multiple words.

2.7.3 Continuous Speech Recognition System:

It is the natural conversational speech people are accustomed to in everyday life. Recognizers with continuous speech capabilities are some of the most difficult to create because they must utilize special methods to determine utterance boundaries. Also, it is extremely difficult for a recogniser to shift through the text as the words tend to merge together. Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Basically, it's computer dictation.

2.7.4 Spontaneous Speech:

There appears to be a variety of definitions for what spontaneous speech actually is. 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 such as words being run together, "ums" and "ahs", and even slight stutters.

2.8 Automatic Speech Recognition (ASR) Systems Problems

By now, almost problems in speech recognition had been discovered. However there are a number of problems that have been identified over the past few decades most of which still remain unsolved. Some of the main problems in ASR are:

i. **Determining word boundaries:** One of the common errors in continuous speech recognition is the missing out of a minuscule gap between words. This arises because speech is continuous in nature and word boundaries are clearly defined when the speaker is speaking at a high speed.

ii. **Varying Accents:** Different Pronunciation of People from different parts of the world for the same words leads to errors in ASR. However this is one problem that is not restricted to ASR but which plagues human listeners too.

iii. **Large vocabularies:** The higher the number of words in the speech recognition database for matching, the higher similarities sounding words generated and tend to cause a high amount of error i.e. there is a good probability that one word is recognized as the other.

iv. **Changing Room Acoustics:** Noise is a major factor in ASR. In fact it is in noisy conditions or in changing room acoustic that the limitations of present day ASR engines become prominent.

v. **Temporal Variance:** Different speakers speak at different speeds. Present day ASR engines just cannot adapt to that.

2.9 Speech Analyser

Speech analysis, also known as front-end analysis or feature extraction, is the first step in an automatic speech recognition system. This process is carried out to extract acoustic features from the speech waveform. The output of front-end analysis is a compact, efficient set of parameters that represent the acoustic properties observed from input speech signals, for subsequent utilization by acoustic modelling. There are three (3) major types of front-end processing techniques that contributes to the development of this prosed system, namely:

i. Linear Predictive Coding (LPC)
ii. Mel-frequency Cepstral Coefficients (MFCC)
iii. Perceptual Linear Prediction (PLP)

Where the MFCC and PLP are most commonly used in state-of-the-art ASR systems (Mann, 2005) However, others front-end processing techniques available includes:

iv. Power Spectral Analysis (FFT)
v. Mel Scale Cepstral Analysis (MEL)
vi. Relative Spectral Filtering of log domain coefficients (RASTA)
vii. First order derivative (DELTA)

2.9.1 Linear Predictive Coding (LPC)

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. LPC 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 minimise the sum of the squared differences between the original and the estimated speech signal over a finite duration (Jain and Saxena 2011).

LPC analyses the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal after the subtraction of the filtered modelled signal is called the residue.

LPC synthesizes the speech signal by reversing the process: use the buzz parameters and the residue to create a source signal, use the formants to create a filter (which represents the tube), and run the source through the filter, resulting in speech. Because speech signals vary with time, this process is done on short chunks of the speech signal, which are called frames; generally 30 to 50 frames per second give intelligible speech with good compression. LPC principle could be used to give a unique set of predictor coefficients that are estimated every frame, which is normally 20ms long.

$$H(z) = \frac{G}{1{-}\in akz - k} - - - - - (2.1)$$

This equation represent the transfer function of the time varying digital filter, where ak is the predictor coefficients, (G) is the gain, and k-1 to p will be 10 for LPC-10 algorithm and 18 for the improved algorithm that is utilised. Performance analysis of LPC recognition system are checked using these parameters namely:

i. Bit Rates.
ii. Overall Delay of the System
iii. Computational Complexity
iv. Objective Performance Evaluation.

2.9.1.1 Types of LPC

Linear Predictive Coding (LPC) extraction techniques can be one these types or hybridize two or more of these following:

i. Voice-excitation LPC
ii. Residual Excitation LPC
iii. Pitch Excitation LPC
iv. Multiple Excitation LPC(MPLPC)
v. Regular Pulse Excited LPC(RPLP)
vi. Coded Excited LPC(CELP

2.9.1.2 Advantages of LPC:

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

2.9.1.3 Disadvantages of LPC:

1. Due to reduce in the bitrates of the speech signal, the quality of voice signal is reduced.
2. This technique is lossy compression technique, hence data gets faded if transmitted to the long distance.

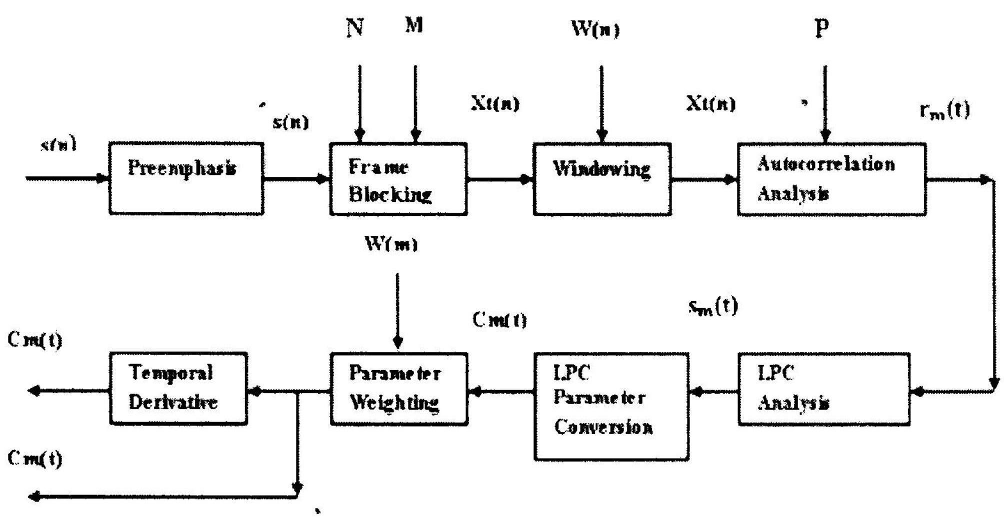


Figure 2.3: The LPC Processor (Mostafa et al, 2009)

The figure above can be broken into these following steps:

Step One

Signal Acquisition: Signal acquisition is the first step for the parametric analysis of speech. These are the steps for signal acquisition.

i. Record the word signal in the audio format by the recorder.
ii. Convert the audio format into .wav format by converter.
iii. Covert the sampling frequency to 8 kHz of word signal.
iv. Read the .wav file in Matlab.
v. This file is used for analysis through FFT and LPC.
vi. Spectra plus software is also used for analysis.

Step Two

Speech Processing: In speech processing the speech signal should be first transformed and compressed for further processing. There are many signal analysis techniques are available which are used to extract the important features and compress the signal without losing any important information. Among the most important technique are FFT and LPC. FFT is used to convert the word signal into spectrum but FFT require only complex value. LPC used to compress the signal and we get different spectrum from original signal spectrum. The analysis of word signal can be done by finding the different paràmeter of FFT and LPC spectrum.

Step Three.

Signal Analysis: Spectrum analysis is complex process of decomposing the speech signal into similar parts.

Step Four

Parameters of Speech Signal: FFT stands for finite Fourier transform it produce frequency spectrum which contains all the information about original signal, but in different form. It is frequency domain representation of signal. To find the FFT of signal, matlab algorithm can used.

After finding the FFT and LPC spectrum of signal next we find different parameters of signal using matlab and spectra plus software. For calculating the parameter of signal we take five sample of single signal word say u1, u2, u3, u4, u5.

After this step, we have use two signal analysis techniques FFT and LPC. Using these techniques the speech signal first analysed and different parameter of speech signal is obtained. If we find parameters for different words then, these parameters values can be used in many software for word recognition.

2.9.2 Mel- Frequency Cepstrum Coefficients (MFCC)

The use of MFCC can be considered as one of the standard method for feature extraction. These are derived from a type of cepstrum representation of the audio clip (a "spectrum-of-aspectrum"). The difference between the cepstrum and the Mel-frequency cepstrum is that in the MFCC, the frequency bands are positioned logarithmically (on the mel scale) which approximates the human auditory system's response more closely than the linearly-spaced frequency bands obtained directly from the FFT or DCT (Fang et al, 2001). This can allow for better processing of data, for example, in audio compression. However, unlike the sonogram, MFCCs lack an outer ear model and, hence, cannot represent perceived loudness accurately. The most notable downside of using MFCC is its sensitivity to noise due to its dependence on the spectral form. Methods that utilize information in the periodicity of speech signals could be used to overcome this problem, although speech also contains aperiodic content (Shaughnessy, 2001).

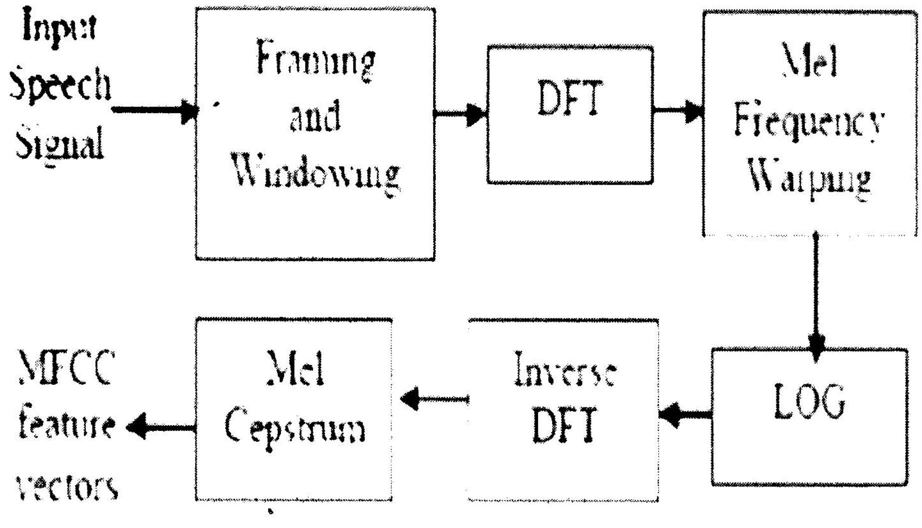


Figure 2.4: The MFCC Processor (Aseem, 2013)

2.9.2.1 Advantage of MFCC

As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system.

2.9.2.2 Disadvantage of MFCC

Application of MFCC techniques in speech recognition required normalization to lessen the influence of noise in MFCC values, since they are not very robust in the presence of additive noise.

2.9.2.3 Applications of MFCC

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

2.9.3 Perceptual Linear Prediction (PLP)

The Perceptual Linear Prediction PLP model developed by Hermansky 1990. The goal of the original PLP model is to describe the psychophysics of human hearing more accurately in the feature extraction process. PLP is similar to LPC analysis, is based on the short-term spectrum of speech.

In contrast to pure linear predictive analysis of speech, perceptual linear prediction (PLP) modifies the short-term spectrum of the speech by several psychophysically based transformations. This technique uses three concepts from the psychophysics of hearing to derive an estimate of the auditory spectrum namely:

i. The critical-band spectral resolution,
ii. The equal-loudness curve, and
iii. The intensity-loudness power law.

Figure 2.5 shows that, PLP involves two major steps namely: 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 (Hermansky, 1986; *et al*, 1974).

2.9.3.1 Advantages of PLP

i. 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.
ii. LPC, however, approximates the speech spectrum equally well at all frequencies, and this representation is contrary to known principles of human hearing.
iii. 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.
iv. 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

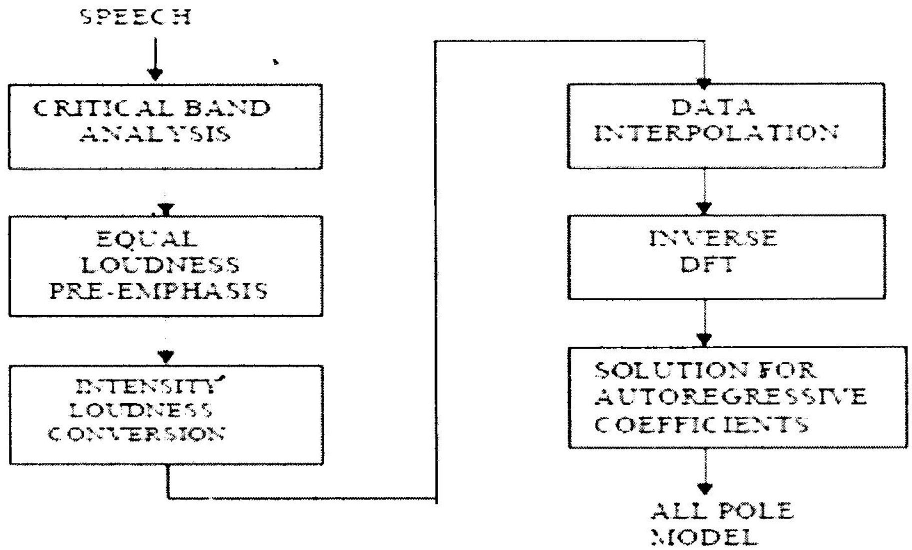


Figure 2.5: Block diagram of PLP speech analysis method (Hermansky, 1990)

2.10 Speech Classifier

The problem of Automatic Speech Recognition belongs to a much broader topic in scientific and engineering so called pattern recognition. The pattern recognition aims to classify objects of interest into one of a number of categories or classes which is also known as patterns or sequence of acoustic vector. These vectors are extracted from an input speech using the techniques described in the previous section. For the scope of this proposed work, classes here will be referred to individual speakers. Since the classification procedure in our case is applied on extracted features, it can be also referred to as feature matching (Prabhakar and Amol, 2013).

The state-of-the-art in feature matching techniques used in speaker recognition includes

i. Dynamic Time Warping (DTW).
ii. Hidden Markov Modelling (HMM).
iii. Vector Quantization (VQ) and
iv. Artificial Neutral Network.

2.10.1 Dynamic Time Warping (DTW)

Dynamic time warping is an algorithm for measuring similarity between two sequences which may vary in time or speed. For instance, similarities in walking patterns would be detected, even if in one video the person was walking slowly and if in another they were walking more quickly, or even if there were accelerations and decelerations during the course of one observation. DTW has been applied to video, audio, and graphics -indeed, any data which can be turned into a linear representation can be analysed with DTW.

A well-known application has been automatic speech recognition, to cope with different speaking speeds. In general, it is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions, i.e. the sequences are "warped" non-linearly to match each other. The goal of this matching techniques is complicated by a number of factors.

i. 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.
ii. Second, the rate of speech may not be constant throughout the word i.e. DTW is an efficient method for finding the optimal nonlinear alignment between a template and the speech sample.

In data mining and information retrieval fields, DTW has been successfully applied to automatically cope with time deformations and different speed associated with time-dependent data. This sequence alignment method is often used in the context of hidden Markov models.

2.10.2 Hidden Markov Model (HMM)

HMM is a statistical Markov model designed is assumed to be a Markov process with unobserved (hidden) states. A Hidden Markov Model is a collection of states connected by transitions The basic principle here is to characterize words into probabilistic models wherein the various phonemes which contribute to the word represent the states of the HMM while the transition 8 probabilities would be the probability of the next phoneme being uttered. Models for the words which are part of the vocabulary are created in the training phase. Now, in the recognition phase when the user utters a word it is split up into phonemes as done before and it's HMM is created.

2.10.2.1 Algorithms of HMM

There are three basic algorithms associated with Hidden Markov Models:

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

2.10.2.2 Limitations of HMM

i. Constant observation of frames
ii. The Markov assumption
iii. Lack of formal methods for choosing a model topology
iv. Large amounts of training data required
v. Weak duration modelling
vi. Restricted output PDFs
vii. The assumption of conditional independence

2.10.3 Vector Quantization (VQ)

VQ is a process of mapping feature vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its centre called a centroid. The collection of all code words is referred to a codebook. 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.

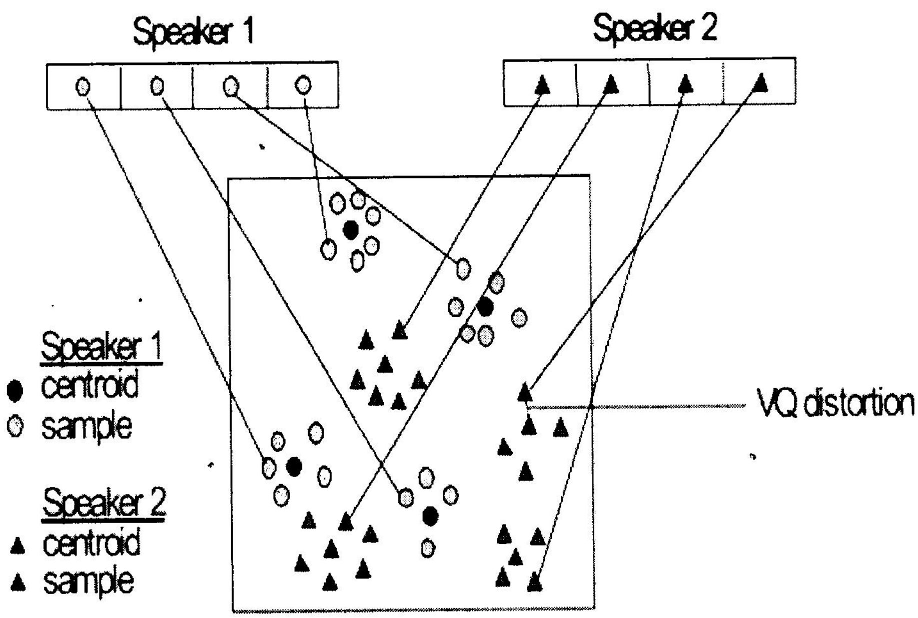


Figure 2.6: Vector quantization between two speakers (Linde, Buzo and Gray, 1980).

Figure 2.7, displays 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 speakęr 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 code word 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. Therefore is an economical compromise that we can live with. (Nave, 2014).

2.10.3.1 Optimization using LBG algorithm

The next important task after the enrolment session and extraction of feature vectors from the input speech provided by each speaker, is to build a speaker-specific VQ codebook for each speaker using the training vectors extracted (Nave, 2014). LBG is a well-known algorithm for clustering a set of L training vectors into a set of M codebook vectors (Linde, Buzo and Gray, 1980).

2.10.3.2 LBG Procedures

LBG algorithm is formally implemented by the following procedures:-

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

$$y_n^+ = y_n(1+\epsilon) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

$$y_n^- = y_n(1-\epsilon) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

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 code word 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 code word).

4. Centroid Update: update the code word 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 code word to initialize the search for a 2-vector codebook, and continues the splitting process until the desired M-vector codebook is obtained (Nave, 2014).

2.10.3.3 Advantages of VQ:

i. Reduced storage for spectral analysis information.
ii. 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.
iii. Discrete representation of speech sounds.

2.11 Speech Recognition Approaches

ASR techniques is a process used to transform a sequence of message from a speech signal which also referred to decoding. Decoded signal is converted into writing or commands to be processed (e.g. hands free dialling). One of the distinguishing characteristics of speech is that it is dynamic.

Distribution of the feature vectors can be modeled using these following approaches:

i. Acoustic Phonetic Approach
ii. Pattern Recognition Approach
iii. Artificial Intelligence Approach

2.11.1 Acoustic-Phonetic Approach

Among the three approaches, the acoustic-phonetic approach has been studied and researched more in the past 40 years. 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.

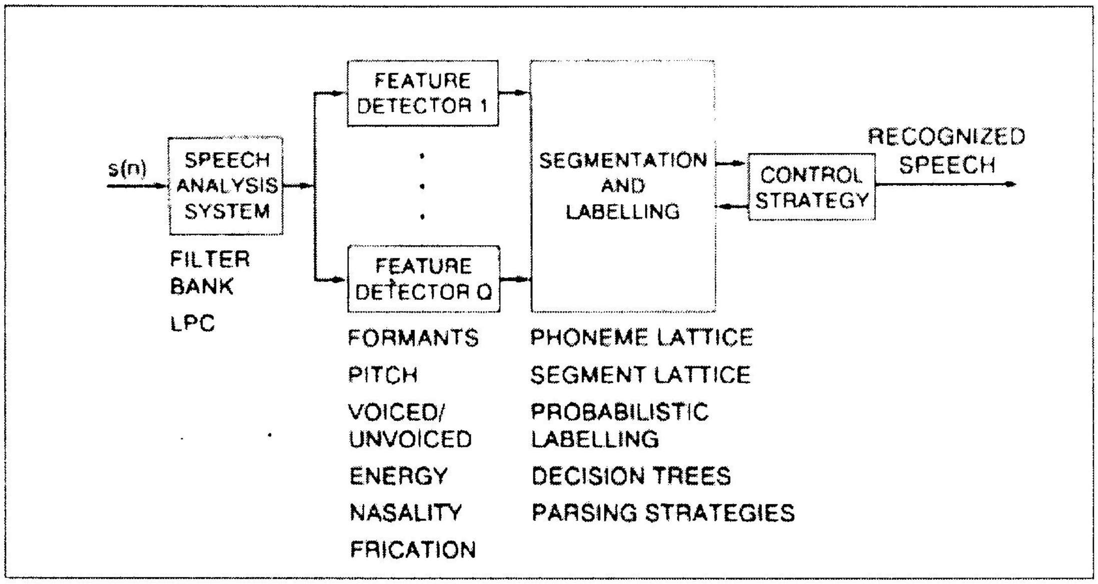


Figure 2.7: Acoustic-Speech recognition system

Figure 2.7 above shows the diagram of Acoustic-Phonetic Speech Recognition System. For a speech to be recognised, Acoustic-Phonetic speech system requires three (3) steps.

The first step is speech analysis or feature measurement method, which involves the filter bank processing. The most common spectral analysis methods are Discrete Fourier Transform (DFT), Linear Predictive Coding (LPC), or Mel Frequency Cepstral Coefficients (MFCC) methods.

Second step is feature-detection process. During this process, the spectral is converted to a set of features. Among the features are nasality (nasal resonance), frication (random excitation), format locations (frequencies of the first three resonances), voiced/unvoiced classification (Periodic or a periodic excitation), and energy ratios.

The third step is segmentation and labelling stage. In this stage, system find the feature stable regions and then label those regions accordingly in order to match each individual phonetic units.

Problems facing acoustic phonetic approach includes:

i. The need for extensive knowledge of acoustic properties of phonetic units.

ii. Features are often based on non-optimal ad hoc considerations rather than based on intuition.

iii. The choice of features is likely based on suboptimal and so optimal implementation of Classification And Regression Tree (CART) methods is rarely achieved.

iv. There is no well-defined, automatic procedure for tuning the labelled speech.

v. No standard way in labelling the training speech. Naturally, these problems need to be solved for it to be utilized practically.

2.11.2 Pattern Recognition Approach

Pattern recognition approach is the most common approach applied in most Automatic Speech Recognition (ASR) system. This approach involves four (4) essential steps namely;

i. Feature extraction steps is used to extract important features from the input signal and represent it a form of feature pattern. The feature extraction techniques include DFT, LPC, LPCC, MFCC etc.

ii. Pattern training, in which one or more test patterns corresponding to speech sounds of the same class are used to create a pattern representative of the features of that class.

iii. Pattern classification is used to compare the unknown test pattern with each class reference pattern, and a measure of similarity between the test pattern and each reference pattern is computed.

iv. Decision logic uses reference pattern similarity scores to decide which reference best matches the unknown test pattern.

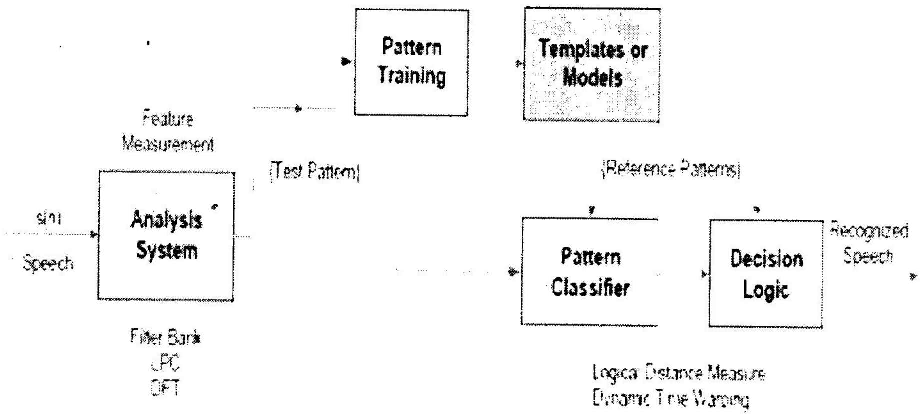


Figure 2.8: Pattern speech recognition system

The general strengths and weaknesses of the pattern recognition include the following:

i. The performance of the system is sensitive to the amount of training data available for creating Sound Class Reference Patterns; generally the more training, the higher the performance of the system for virtually any task.

ii. The reference patterns are sensitive to the speaking environment and transmission characteristics of the medium used to create the speech; this is because the speech spectral characteristics are affected by transmission and background noise.

iii. No speech-specific knowledge is used explicitly in the system; hence, the method is relatively insensitive to the choice of vocabulary words, task syntax, and task semantics.

2.11.3 Artificial Intelligence approach

This 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. The intention here is to mechanise the recognition procedure like the way a person applies his intelligence in analysing and making decision on the acoustic knowledge. 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 2.10.

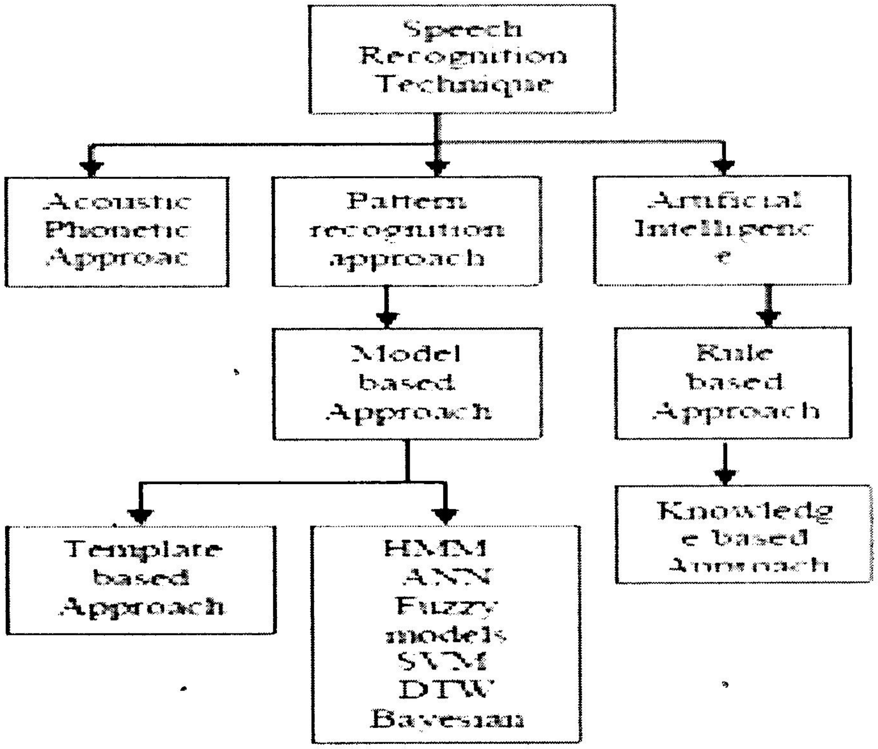


Figure 2.10: Speech Recognition Techniques

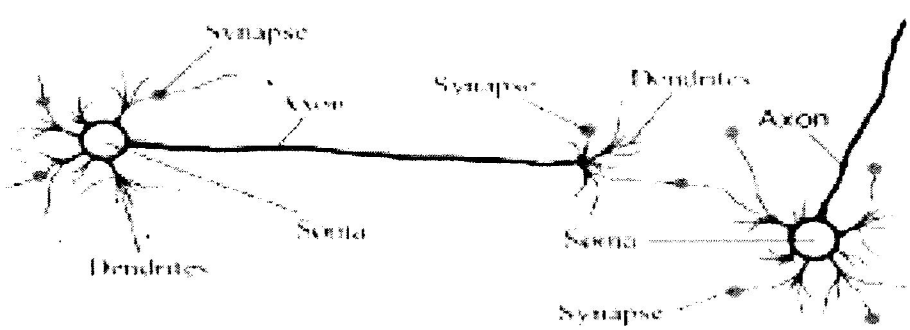


Figure 2.11: Biological Neural Network

2.12 Matching Techniques

Speech-recognition engines match a detected word to a known word using one of the following techniques (Prabhakar and Amol, 2013).

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.

2.13 Performance Evaluation of Speech Recognition Systems

The major performance measuring tools for speech recognition systems are 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).

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 is mathematically computed as:

$$WER = \frac{S + D + I}{N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

Where

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:

$$WRR = 1 - WER$$

$$= 1 - \frac{S + D + I}{N} \ldots\ldots\ldots\ldots (2.5)$$

$$= \frac{H - I}{N} \ldots\ldots\ldots\ldots\ldots\ldots (2.6)$$

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

2.14 Related Works

A There are some project that have been completed which regard to the application of the voice/speech in the controlling method.

The system presented in **(Yerrapragada & Fisher, 1993).** A voice controlled smart house. It works on the predefined set of voice commands for the defined areas in a house. The new commands can be added in as required. FIS, a pattern processing technology, is used that allows processing of complex patterns as experienced in speech and vision.

In **Gerhart, J. (1999),** an intelligent home navigation system (IHNS) is proposed to facilitate the elderly and the physically challenged persons. It proposes an idea of an automated voice based home navigation system. The system comprises over a wheelchair, navigation module and speech recognition module SR-07 and a line follower module for navigation. The system has predefined voice commands relating to different rooms, and predefined routes relating **to those** rooms for the navigation. There is also a collision avoidance system installed in the project.

An intelligent access control system is developed that is based on SPCE061A voice recognition chip by **(Mann, 2005)**. The supporting software comprises of the voice training module, the voice recognition module, the voice data processing module and the voice-playing module. The system completes the functions of collecting the voice data, distilling character, special voice recognition and voice playing in terms of initializing the system and the identification training.

Majgaonkar et al (2006), Worked on the project that utilizes the different electronic parts available in the market to build an integrated home appliances control system by using Bluetooth device and Microcontroller technology. This system gives service at low cost compared to the cost of the available controlling system.

A home appliance control system proposed by **(Nguyen et al, 2007)**. Used Infrared ray and power line communication to control the home appliances system. This system helps user to checks the status of appliances and controls them remotely from everywhere. And this is done through their cellular phone or Internet. The simple approach to control the home appliances is given in this paper.

(Khiyal et al., 2009) proposed a system for controlling home appliances remotely that is useful for the people who are not at home mostly. The main objective of the system is to provide security and control the home appliances such as AC, lights and alarms. The system is implemented by SMS technology that is used to transfer data from sender to receiver over GSM network.

(Nave, 2014) proposed a system comprises over a voice controlled wheel chair. The system is built using low cost speech recognition board and a microcontroller. The speech recognition board utilizing HM2007 speech recognition chip. It use san idea of paired-word for issuing a command to the system, so that if words similar to the voice commands are spoken within the vicinity of the voice recognition kit, they are not accidently detected. The system is speaker dependent and also an isolated speech system, to avoid further accidental voice detections.

CHAPTER THREE

DESIGN METHODOLOGY

3.1 Introduction

There are several steps and operations applied in designing microcontroller based speech recognition system for controlling home fan and water pump. The relevant is gathered through literature review from previous chapter. This project is going to be built using MATLAB software to produce the speech recognition system. A series of operations that resulted to recognition of the user of a home appliances is divided in three stages namely, acquisition stage, recognition stage and database. This operations involves capturing and representing the speaker speech in appropriate notations, creating feature vectors to represent time-slices of the converted input, clustering and purifying the input, matching the results against a library of known sound vectorized waveforms, choosing the most likely series of letter-sounds, and then selecting the most likely sequence of words.

LPC feature extraction techniques is applied to this work because 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 and to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. Also, VQ utilizing LBG algorithm is used for the recognition of speaker. Next is the hardware development, PIC16F628A microcontroller which receives signal from the speech recognition system developed in MATLAB and controls the on and off fan and water pump model operations is used in this work.

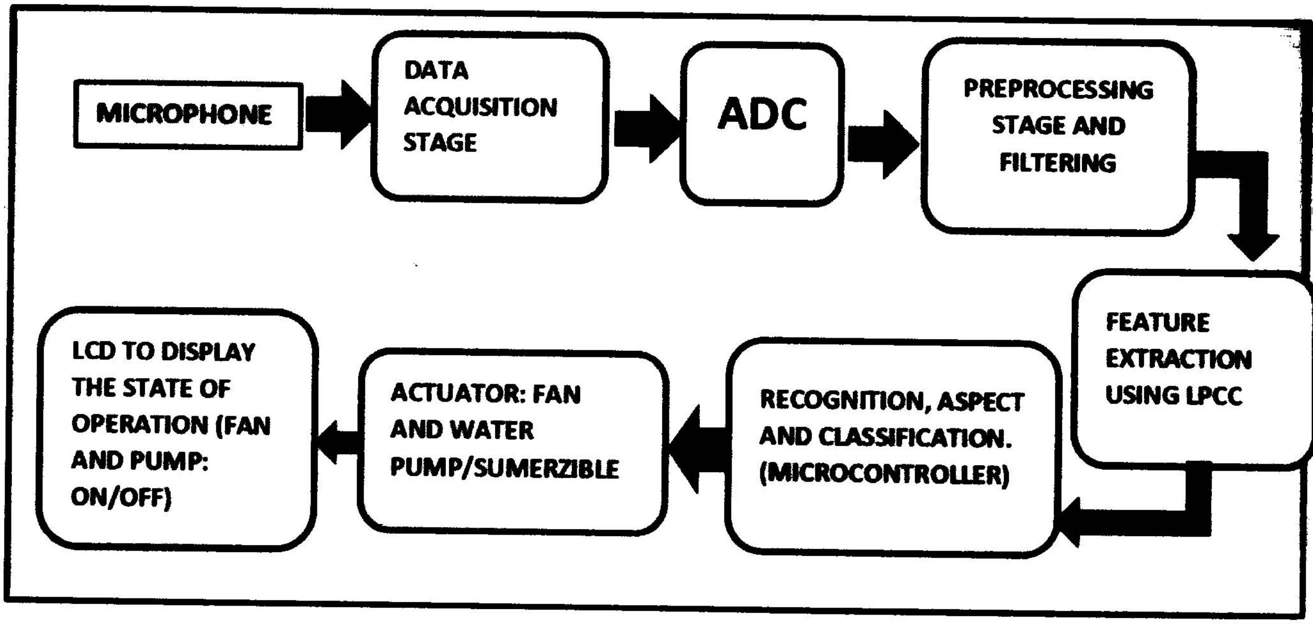


Figure 3.1: Speech recognition system block diagram

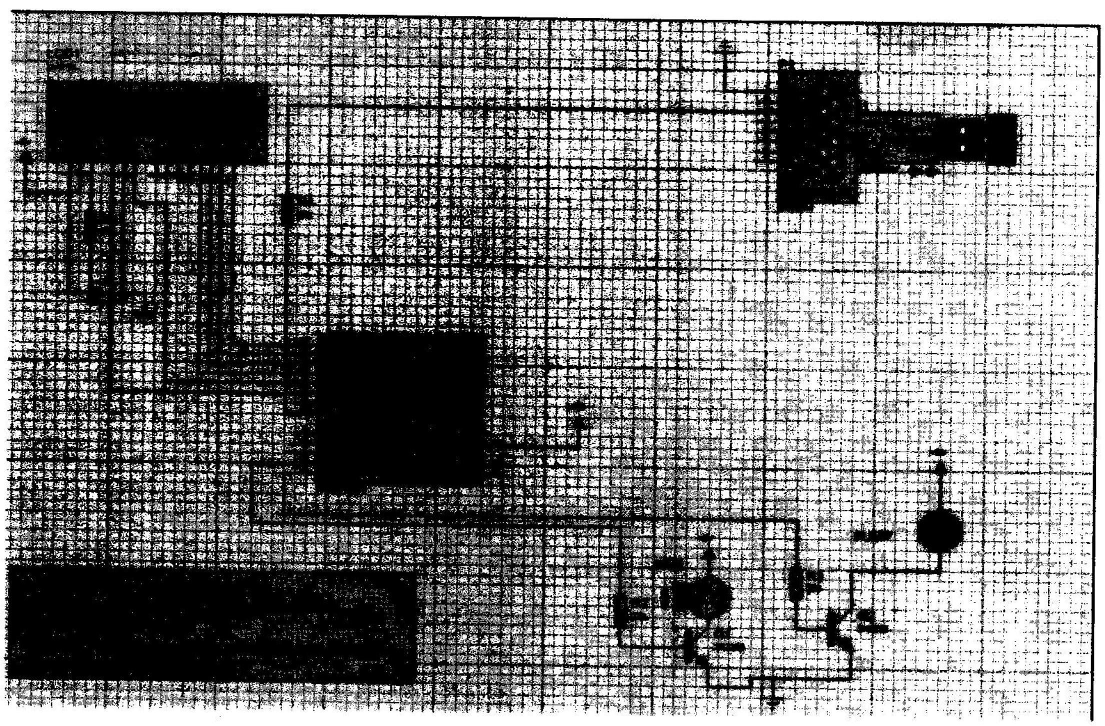


Figure 3.2: Schematic diagram of the Speech recognition system

3.3 COMPONENTS THEORY

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

1. MATLAB software for development of speech recognition system
2. The microcontroller (PIC16F628A)
3. The microphone circuit
4. 16x2 LCD display
5. 10K Resistor and two 4.7K resistors
6. 20Mhz Crystal
7. USB interface
8. DC motors for designing electric Fan
9. Water pump submersible

3.3.1 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 below.

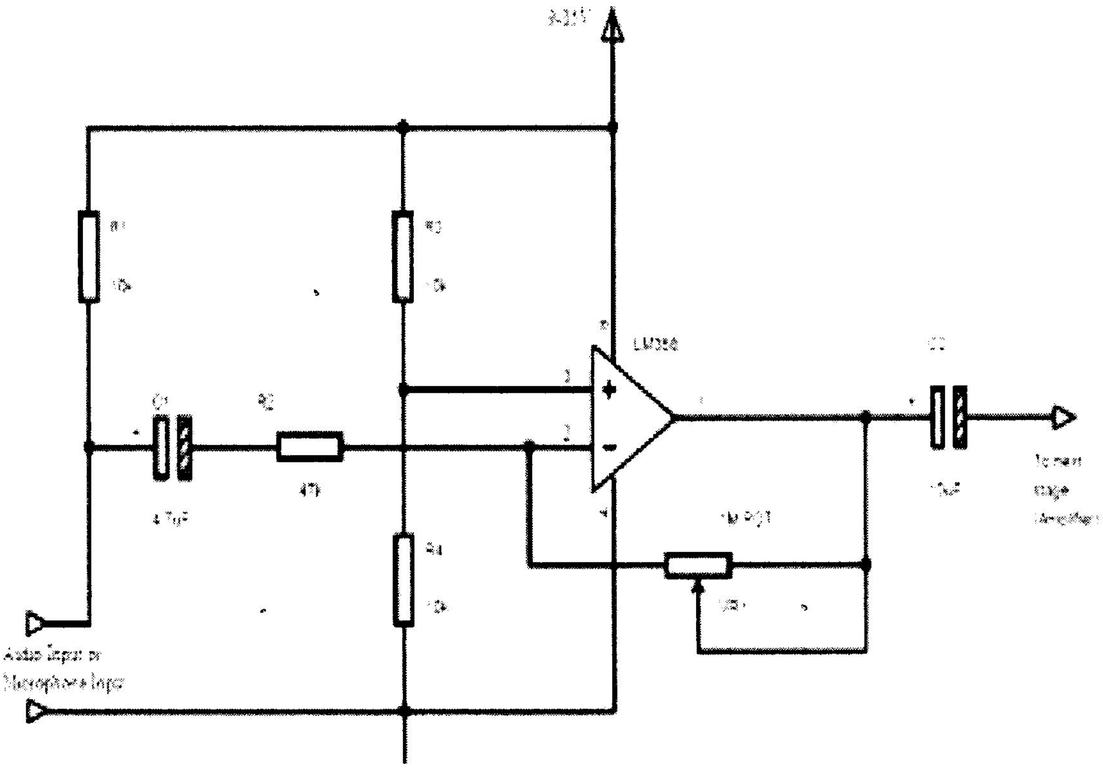


Figure 3.3: Microphone circuit

3.3.2 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 colour 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.

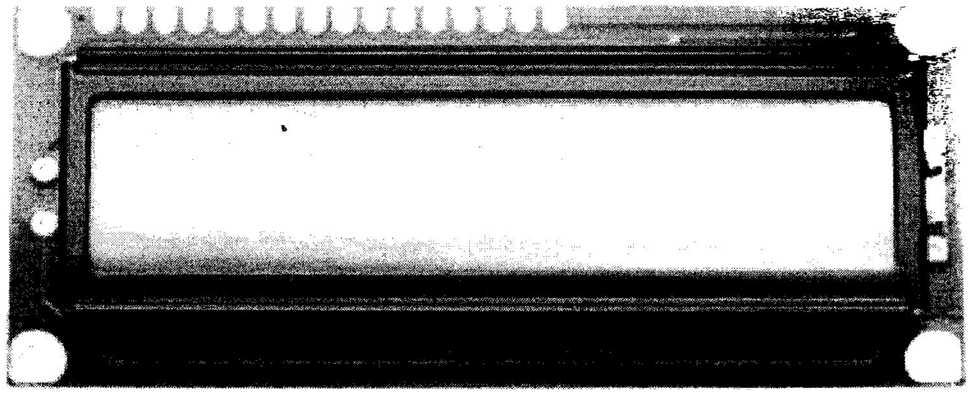


Figure 3.4: 16x2 LCD Display

3.2.3 Microcontroller (PIC16F628A)

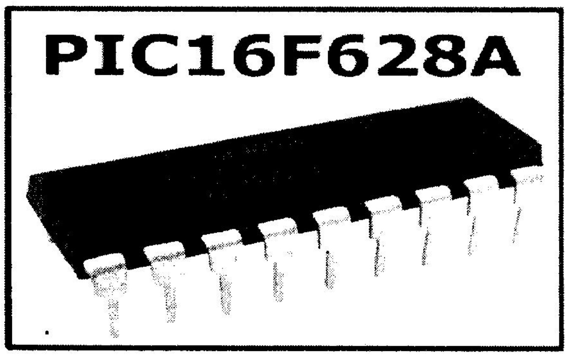


Figure 3.5: An image of PIC16F628A *(*Microchip Technology Inc., 2007*)*

The microcontroller (PIC16F28A) is 18-Pin FLASH based members of the versatile PIC16CXX family of low cost, high performance, CMOS, fully-static, 8-bit microcontrollers. The PIC16F628A has enhanced core features, eight-level deep stack, and multiple internal and external interrupt sources. The separate instruction and data buses of the Harvard architecture allow a 14-bit wide instruction word with the separate 8-bit wide data. The two-stage instruction pipeline allows all instructions to execute in a single cycle, except for program branches (which require two cycles). A total of 35 instructions (reduced instruction set) are available, complemented by a large register set. PIC16F628A microcontroller typically achieve a 2:1 code compression and a

4:1 speed improvement over other 8-bit microcontrollers in their class. PIC16F628A device has integrated features to reduce external components, thus reducing system cost, enhancing system reliability and reducing power consumption. A highly reliable Watchdog Timer with its own on-chip RC oscillator provides protection against software lockup.
It has the following additional features;

- Internal and external oscillator options
- Precision Internal 4 MHz oscillator factory calibrated to ±1%
- Low Power Internal 37 kHz oscillator
- External Oscillator support for crystals and resonators.
- Power saving SLEEP mode
- Programmable weak pull-ups on PORTB
- Multiplexed Master Clear/Input-pin

Table 3.1: PIC16F628A microcontroller features

NAME	VALUE
Program Memory Type	Flash
Flash Program Memory Size (words)	2048
Clock Maximum Frequency of Operation (MHz)	20
RAM Data Memory (bytes)	224
EEPROM Data Memory (bytes)	128
Serial Communications	USART
Capture/Compare/PWM modules	1
Timer Module(s)	TMR0, TMR1, TMR2
Number of Comparators	2
Internal Voltage Reference	Yes
Interrupt Sources	10
I/O Pins	16
Temperature Range (C)	- 40 to 85
Operating Voltage Range (V)	3.0 to 5.5

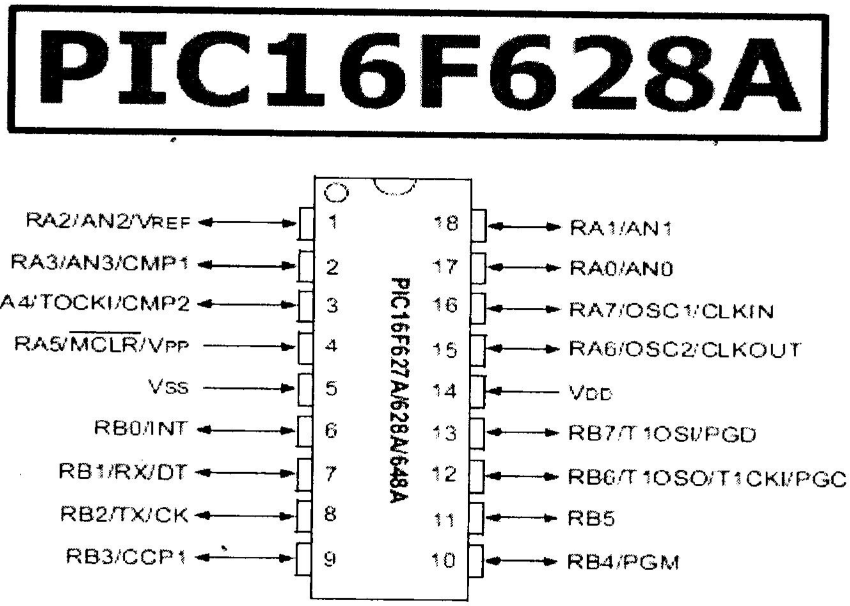


Figure 3.6: A diagram for Pin description of PIC16F628A (Microchip Technology Inc., 2007)

3.2.4 2N2222A NPN Transistor

The 2N2222A is a common NPN bipolar junction transistor (BJT) used for general purpose low – power amplifying or switching applications. It is designed for low to medium current, low power, medium voltage and can operate at moderately high speeds (10), the collector and emitter will be left open (Reverse biased) when the base pin is held at ground and will be closed (Forward biased) When a signal is provided to base pin. 2N2222A has a gain value of 110 to 800, this value determines the amplification capacity of the transistor. The maximum amount of current that could flow through the collector pin is 800mA.

Table 3.2: 2N2222A NPN Transistor Features:

Features	Value
DC Current Gain (h_{FE})	100
Continuous Collector current (I_C)	800mA
Emitter Base Voltage (V_{BE})	6V
Collector Emitter Voltage (V_{CE})	30V
Base Current (I_B)	5mA

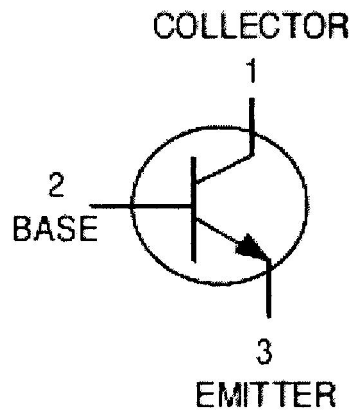


Figure 3.7: A description of 2N2222A NPN TRANSISTOR.

Table 3.3: Pin description of the 2N2222A Transistor.

Pin Number	**Pin Name**	**Description**
1	Collector	Current flows in through collector
2	Base	Controls the biasing of transistor
3	Emitter	Current drains out through emitter

3.2.5 Relationship between Resistor and Other Components (Microcontroller, Transistor and LCD)

- **Equation for relationship between resistor and LCD**

$$\mathrm{R} = \frac{V_{0H} - V_d}{I_d} \ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

R = Resistance

$V_{0H} = High\ output\ voltage$

$I_d = Resistor\ desire\ current$

$V_d = Desire\ voltage$

- **Equation for relationship between resistor and transistor**

$$\mathrm{R} = \frac{V_{DC} - V_{CE}}{R_C} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.2)$$

$\mathrm{R} = \mathrm{Resistance}$

$V_{DC} = Direct\ current\ voltage$

$V_{CE} = Collector\ Emitter\ voltage$

$R_C = Collector\ resistance$

3.2.6 DC Motor

The DC motor works in bidirectional. They work together with any controlling devices attached such as assisting in controlling the rotation speed and accuracy of the DC motor.

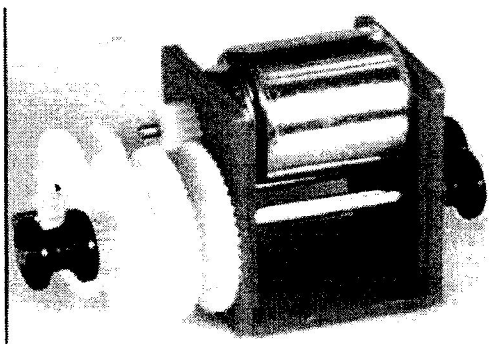


Figure 3.8: A DC Motor

3.4 SYSTEM ANALYSIS

The speech recognition system for controlling of home appliances such as fan and water pump is going to be built using MATLAB and implemented using a hardware model fan and water pump as discussed in the previous chapter.

From the schematic diagram above which shows the hardware with PIC16F628A 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 speaker by the speaker recognition system created with MATLAB, a controlling signal is sent through the

COM port to the PIC16F628A microcontroller which controls the powering on and of the attached devices through the transistor. The LCD displays the status of the system.

3.4.1 Design of Power Supply for Circuit

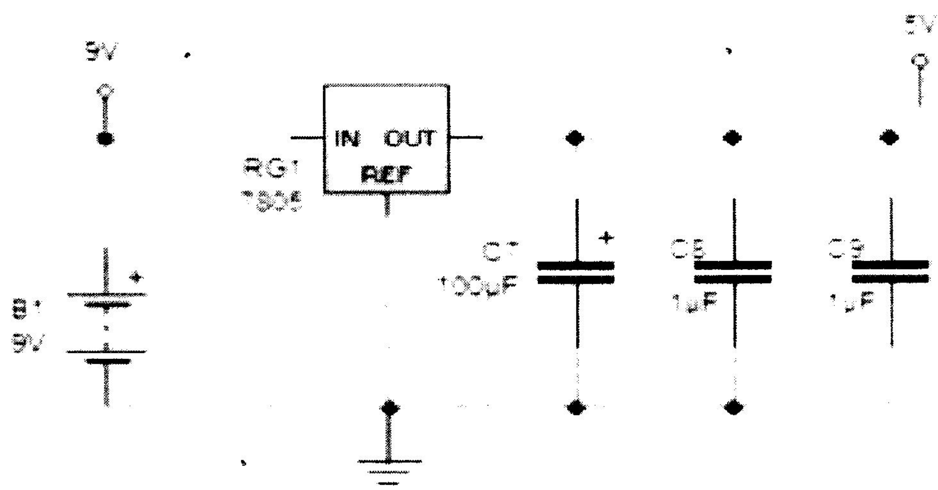


Figure 3.9: Power Supply Circuit

The 5V power supply uses a 78L05 regulator and filter caps to convert the 9V from the battery to 5V for the PIC16F628A circuits. Figure 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.

3.4.2 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.

3.4.3 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)

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 below shows the power supply stage design

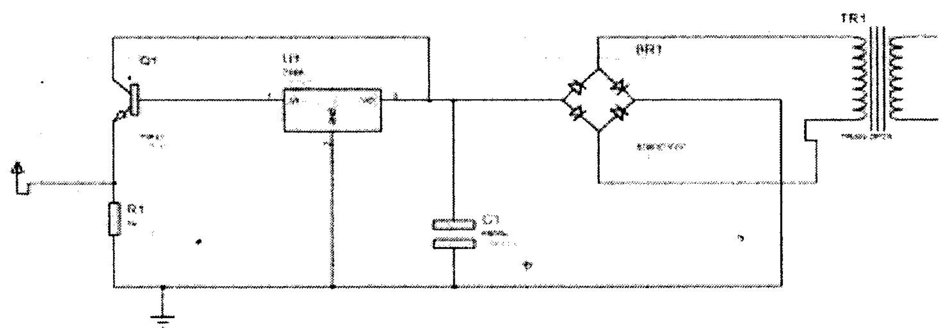


Figure 3.10: Power supply stage design

3.5 Product Operation Flow

Speech recognition system using LPC features extraction techniques required both hardware and software model testing in which some vital parameters and recommended values needs to be supplied

3.5.1 Assembly Process

- Assemble and test the speech recognition system
- Setup and test the Transistor control circuit
- Assemble the model fan and the water pump frame
- Interface all the various components and test the overall system

3.5.2 Operational Block Diagram

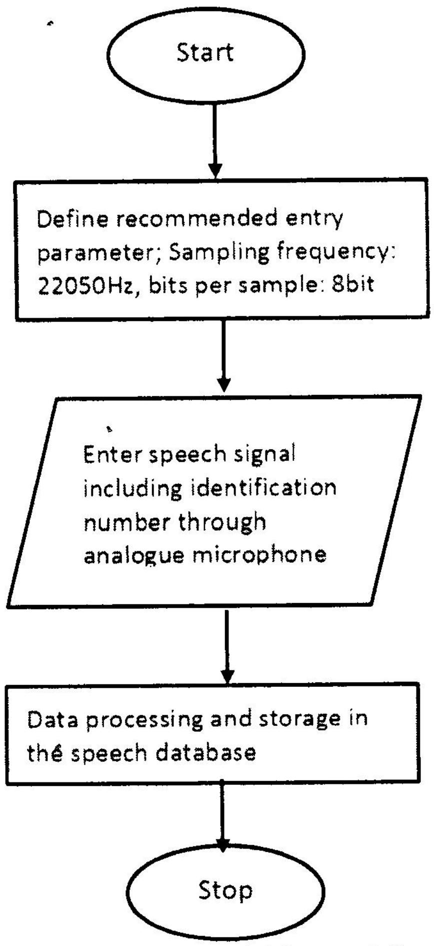


Figure 3.11: Data Acquisition and Storage flow diagram

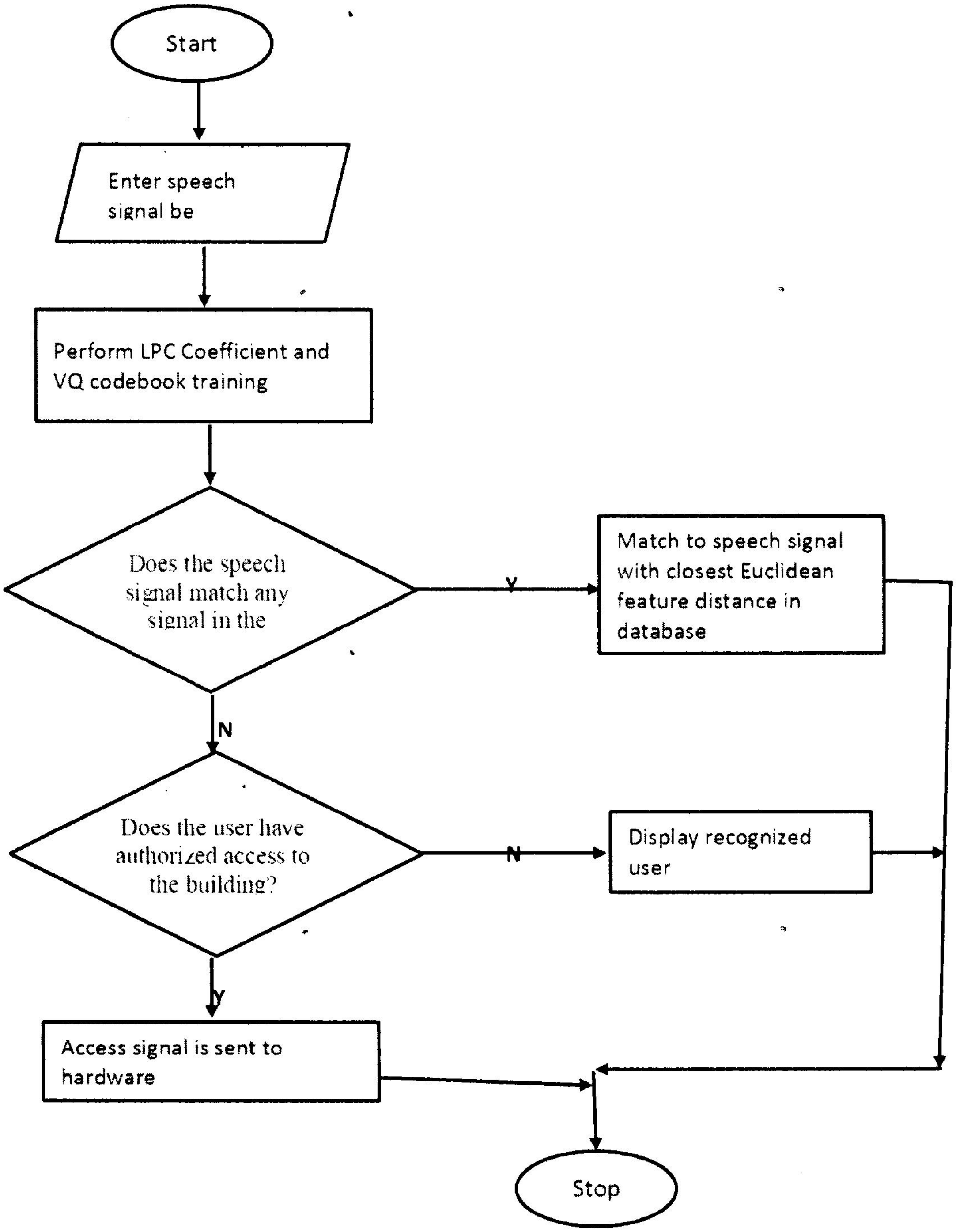


Figure 3.12: Recognition and Access control flow diagram

CHAPTER FOUR

IMPLEMENTATION AND RESULTS

4.1 Introduction

This chapter will briefly discuss on the execution, results and discussion of the software and hardware implementation speech recognition system developed. It discuss the main speech recognition system, database design of the speech recognition system and the different steps taken in the execution and results phase using MATLAB 2015Ra software. The hardware implementation is done using a model electric fan and water pump, designed to be controlled by the speech recognition system built in MATLAB.

4.2 Design of the Speech System Database

The speech system database was programmed using structural programming method whereby each section (function) leads to another in the whole speech program in MATLAB 2015 software due to its high level dynamic nature. The Matlab code controlling the speech recognition system, allows user interaction with through GUI rather than thousands of code. This GUI has some press-buttons menu that represent some key steps taken in achieving high efficiency and optimization of the system as listed below with their individual code sections too.

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

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

```
while chos~=possibility,
chos=menu('Speaker Recognition System','Add a new speech from microphone',...
        'Speaker recognition from microphone','LPC Representation',
        'DatabaseInfo','Delete database','Exit');
```

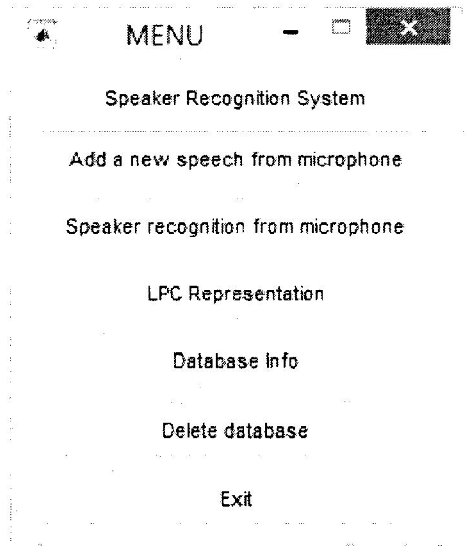


Figure 4.1: Pop-up Menu of the Speech Recognition

1. **Add a new speech from microphone;**

This is the first press-button on the pop-up menu of the speech recognition system. It is pressed to add individual user speech signal to the system database. This section represent the training phase of the speech recognition system where the system is trained with different user speech signals and then stored in the speech sample database that will be used for reference model in the recognition phase of the system. The acquisition stage of the speech signal, requires some parameters to be input for pre-processing the user speech. These are the parameters together with their recommended values in bracket; a class number or sound ID (1, 2, ... etc) for recognizing individual speech in the database; the sampling frequency (22050Hz); Sampling bit (8); duration of recording (3 seconds). After recording, the user(s) speech are added to the system database.

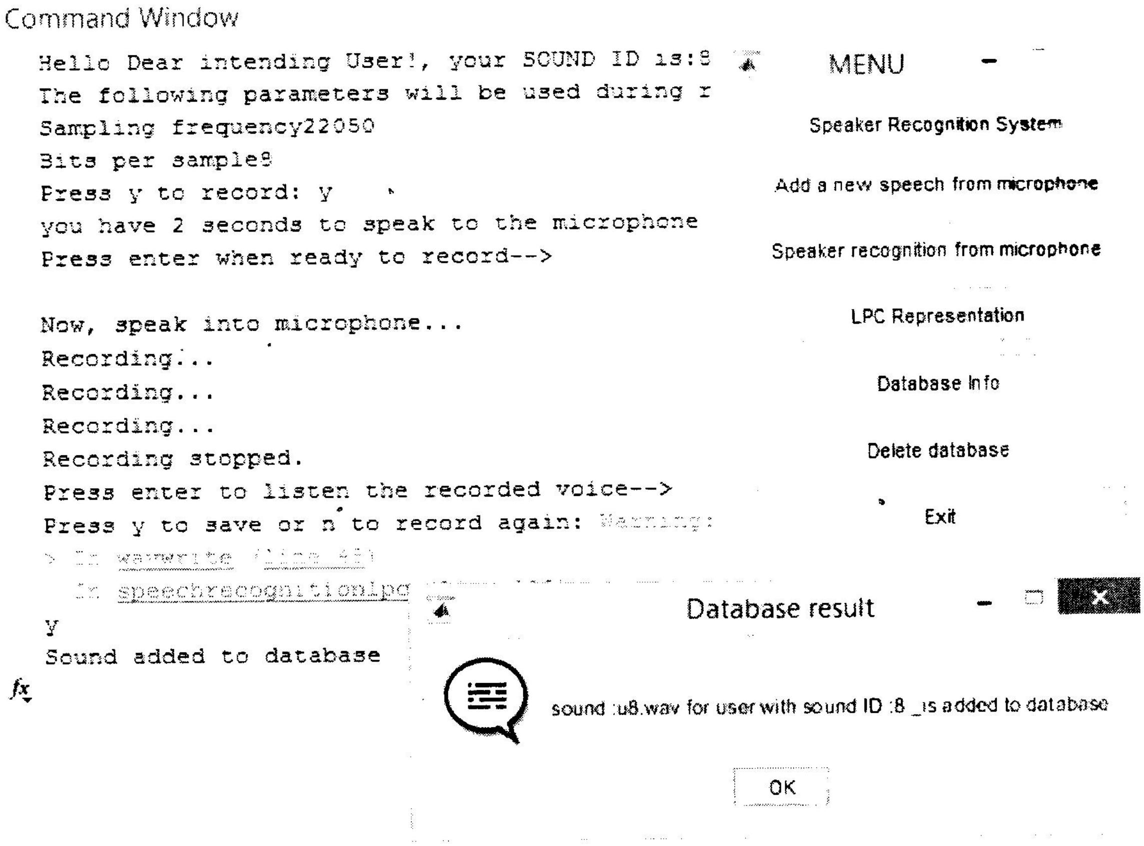


Figure 4.2: Command Prompt and Pop-up Message Box display for the Speech Acquisition

2. **Speech recognition from microphone;**

This second press-button of the pop-up menu is the essential section of the whole speech recognition system sections. This section controls the operation of the hardware such as powering on and off of the model electric fan and water pump (hardware) after the user's Sound ID is recognized. During the recognition phase of the speech system, speech signals recorded into the system are processed and compared to see if there is a match for the speech signal in the system database. These are the mathematical tools and algorithm used while processing the user(s) speech signal namely; Linear Predictive Coding (LPC) algorithm for feature extraction and VQLBG (vector quantization using LBG algorithm) for easy recognition of user 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.

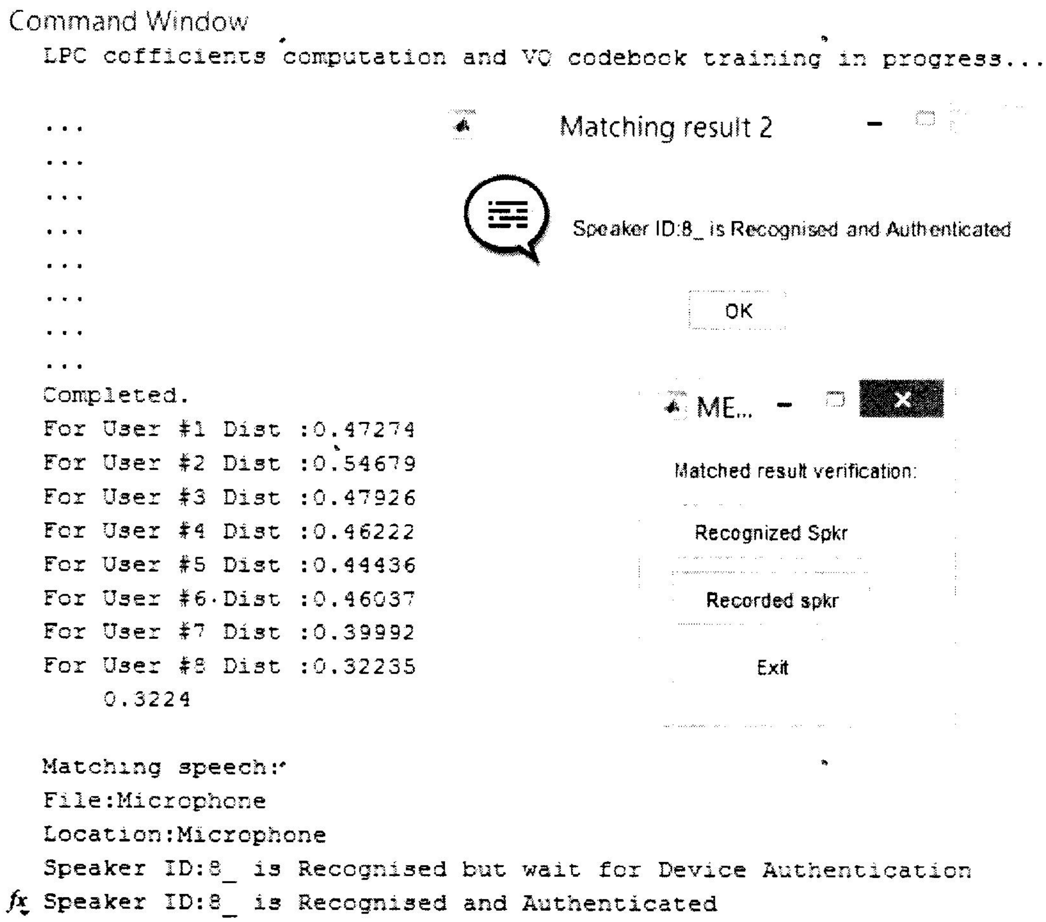


Figure 4.3: Command Prompt and Pop-up Message Boxes display for the Speaker ID: 8 recognition using LPC computation and VQLBG method

3. **LPC representation**;

This is the third section of the pop-up GUI menu of the speech recognition system that is pressed to display the effect of the LPC (Linear Predictive Coding) algorithm on the user's input speech sample during recognition phase. After series mathematical computation on the user speech signal, this press-button display waveform corresponding to the effect of zero-crossing (silence removal), filtering, LPC cepstral co-efficients (speech features) and its spectrum effect.

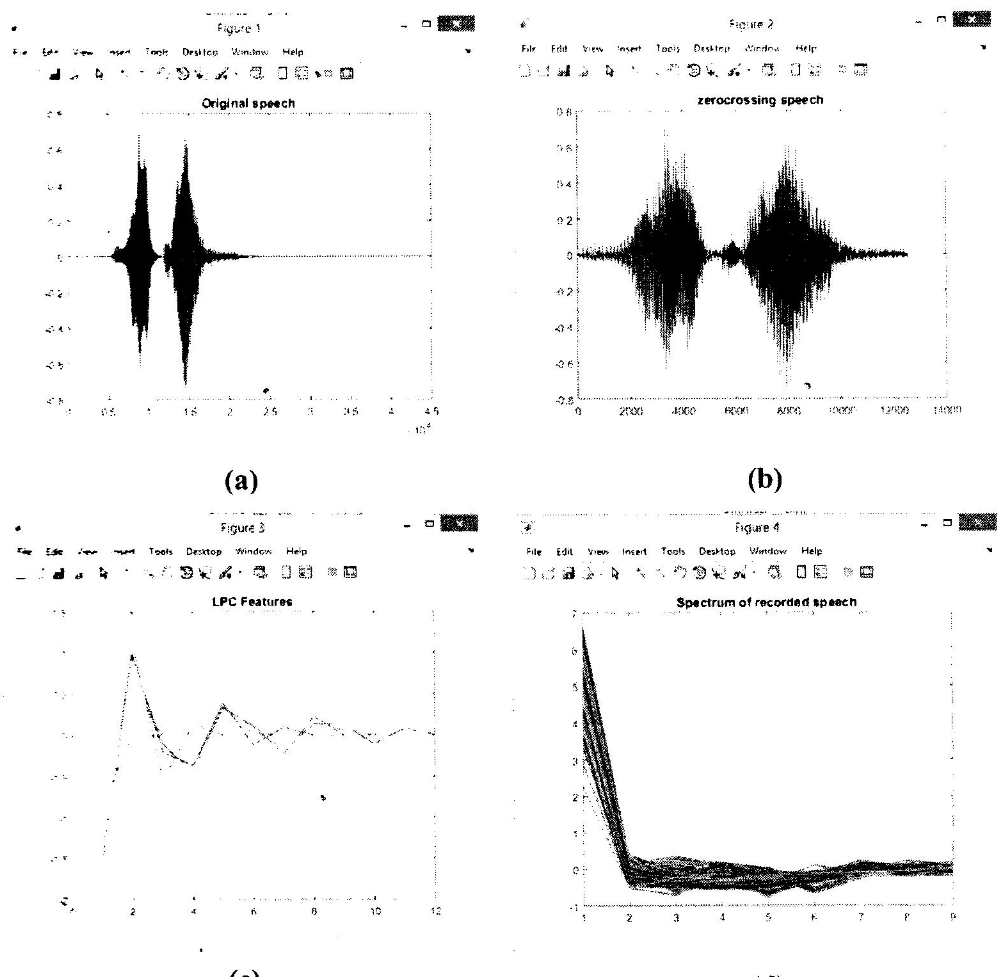


Figure 4.4: Waveform Result of LPC Representation for the Speaker ID: 8 Speech Sample (a): Original Speech. (b): Silence Removal sample. (c): Speech Sample cepstral features. (d): Spectrogram of the Cepstral features

4. **Database information;**

This fourth press-button of the pop-up GUI menu of the speech recognition system which is used to display the details of the user(s) speech stored in the database of the system during the acquisition phase. Such that when any Speaker ID is supplied, this section play the sound wav file pertaining the ID number. And if there is no speech signal in the database, it prompts up a message saying "Database is empty". Else it displays the content of the Sound ID in the database.

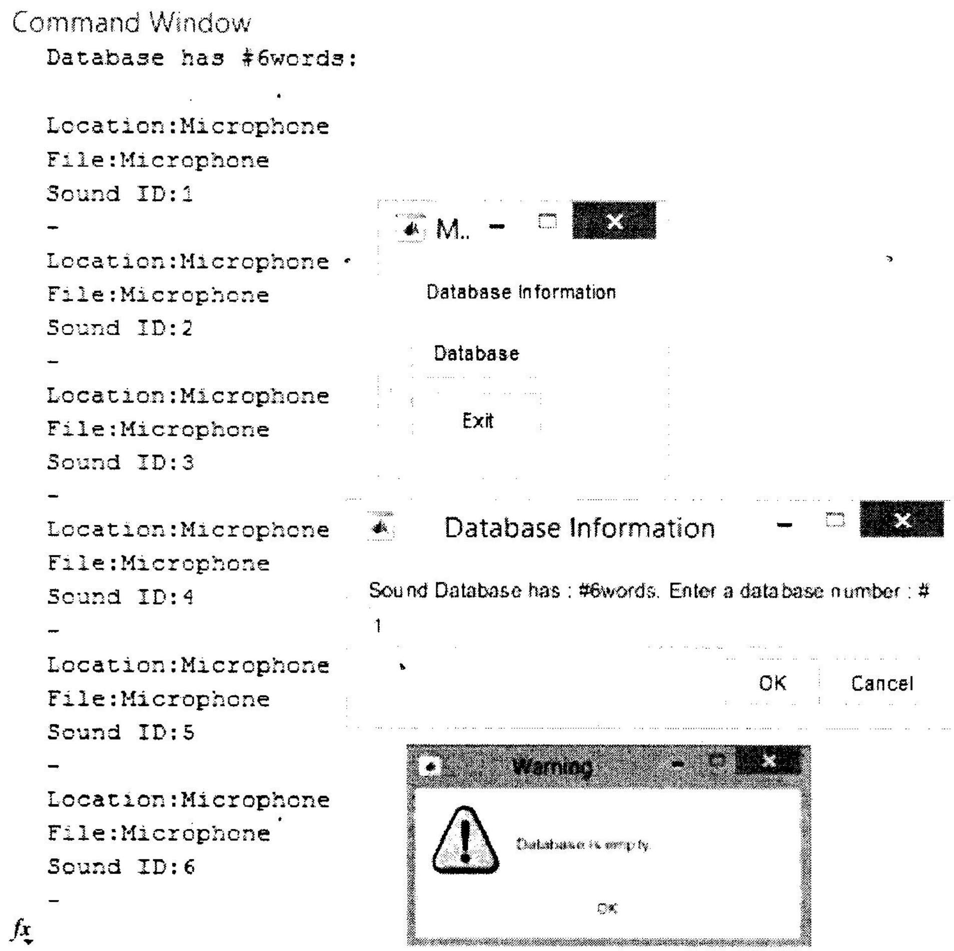


Figure 4.5: Display of the "Database Info" operation

5. **Delete database;**

This fifth press-button on the popup menu of the speech recognition system is used to erase delete the entire content of the database. This is included to address the need of capturing different set and countless number of speech signals also to remove sets of speech sample 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".

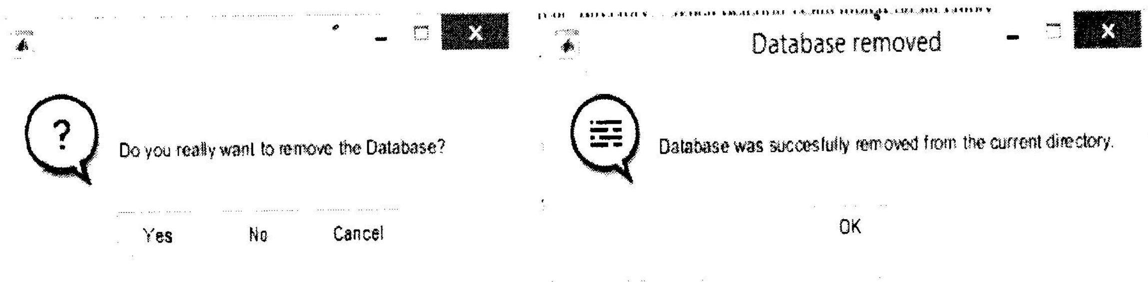


Figure 4.6: Status Display of the database deletion process.

6. **Exit;**

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

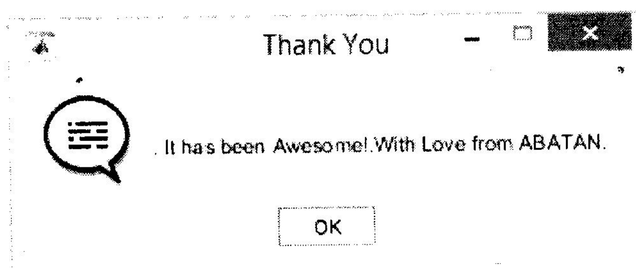


Figure 4.7: Exit Operation Display

4.3 Matlab Functions used in the Speech Recognition System

Matlab function is a sets of code that perform one or more task. It is declared using keyword "function" to define the function name, input parameters, and output variable.

Below are the metrics used in achieving different phase of this project and the matlab codes are in appendix.

- **Speech Recording/Capturing function**

This code capture the user input speech samples. The function requests the user to input some recording parameters and their recommended value in case the user failed to response.

- **zerocrossing [endpoint detection] function;**

This function remove silence from the signal "y" speech sample, amplify the frame and waveform and filter out the unwanted high frequency component to be processed into frames, having a sampling rate "fs" of the signal.

- **LPC function;**

This function is used for accurate extraction of features of a given speech signal "samples" having a sampling frequency "fs" and "cesptra" which contains the transformed signal and the function

- **Vector Quantization (VQ) codebook matlab function**

This code implement the process of mapping feature vectors from a large vector space to a finite number of regions in that space using the Linde-Buzo-Gray algorithm. Each region is called a cluster and can be represented by its centre called a centroid. The collection of all code words is referred to a codebook.

- **Euclidean distance function**

This function computes the Euclidean distance (sum((x-y).^2).^0.5) between column of two matrices X and Y which each column is a vector data. "d" contains training data vector (one per column), "k" is number of centroids required, "c" contains the result VQ codebook (k columns, one for each centroids).

4.4 Project Execution and Result Evaluation of the Speech Recognition System.

4.4.1 Software Execution and Result Discussion

The software execution and result discussion of the speech recognition system basically cover these phases namely: Data acquisition and pre-processing stage, Feature extraction and data storage and Recognition using VQLBG algorithm

1. **Data acquisition and pre-processing**

This is the first section of the MATLAB program process, in which user(s) speech 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) in a conducive environment such as a silent room to prevent acquisition of unwanted signals called noise that may have effect on the speech system. Data acquisition is done. After data acquisition, the speech 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, the user speech has been transformed from analog form to digital form, then the vocal characteristics/features of the speech are extracted from the speech signal and stored in the system database speech in form of SOUND ID.

LPC 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.

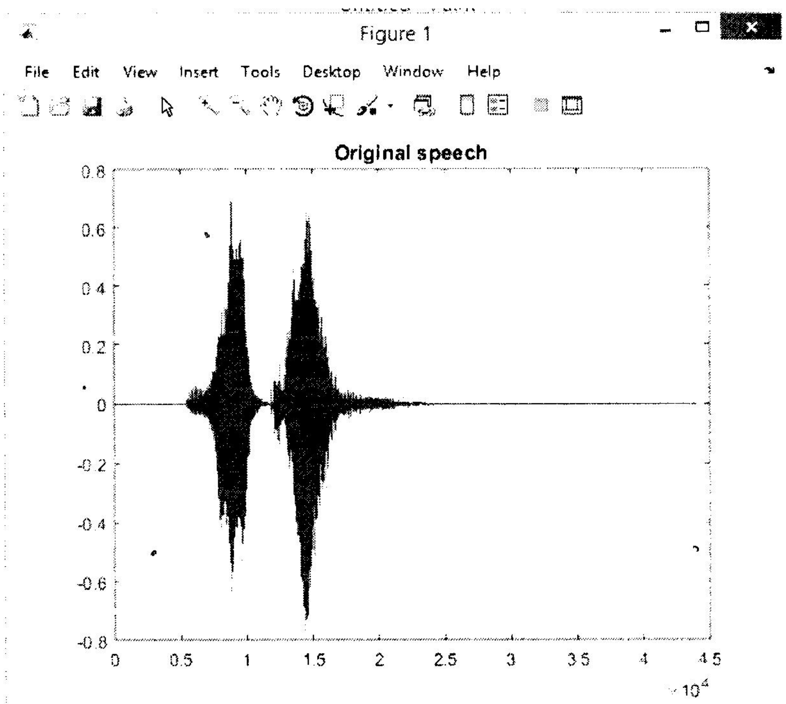


Figure 4.8: Original Speech Sample

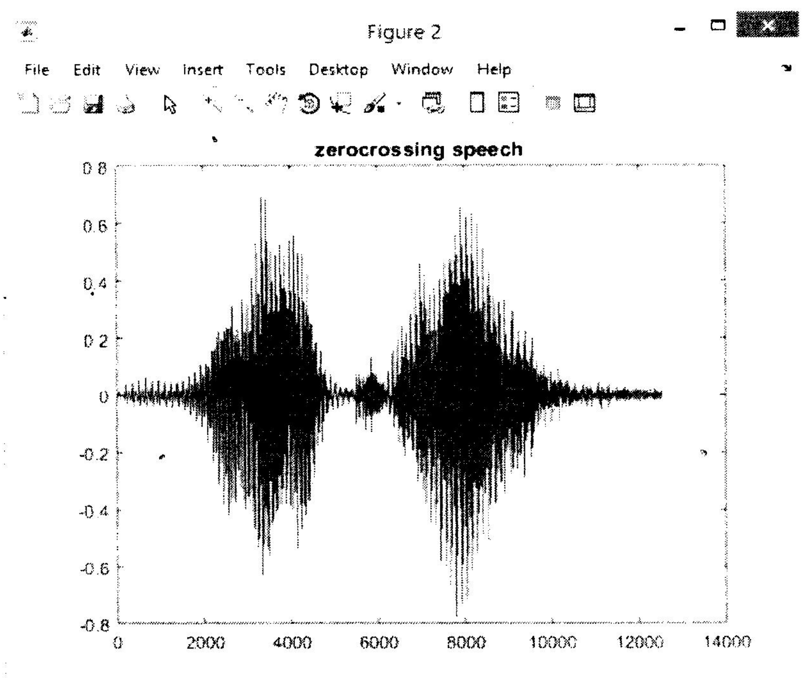


Figure 4.9: Filtered and Zero-crossed Speech Sample

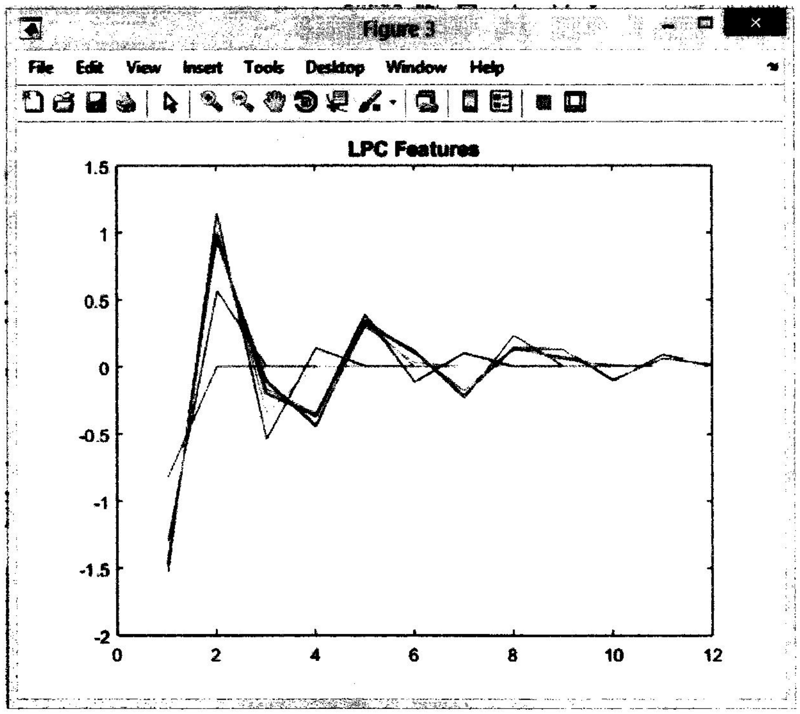


Figure 4.10: Extracted LPC Coefficients/Features of the User Speech Sample

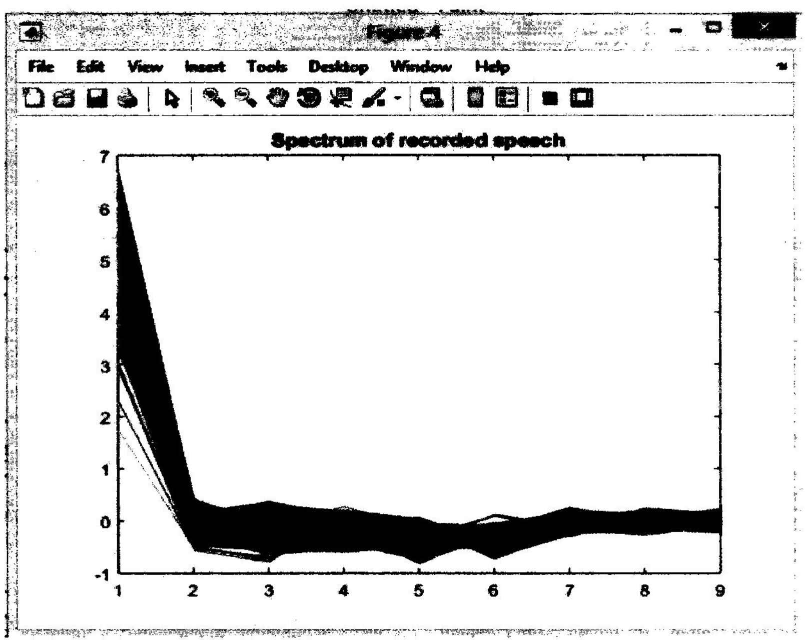


Figure 4.11: Spectrum Representation of the LPC Features

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 with those in the database to know which speaker it is and also, activate the powering of the fan and water pump for the authorized user. VQLBG algorithm simply matches the inputted speech signal to a speech signal in the database based on the closeness of the parametric distance (Euclidean distance).

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 LPC 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 to the operation control of the hardware prototype requested (either Fan or Water Pump). On recognition of the authorized user, a signal is sent from the MATLAB through the RS232 program cable to the hardware prototype.

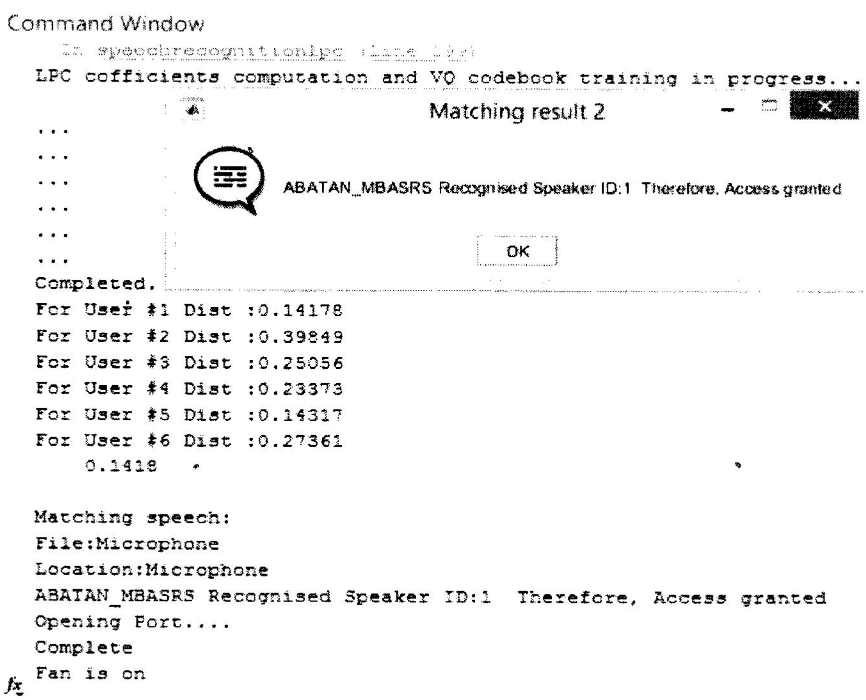


Figure 4.12: Recognized speaker "sound ID: 1", the authorized user

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

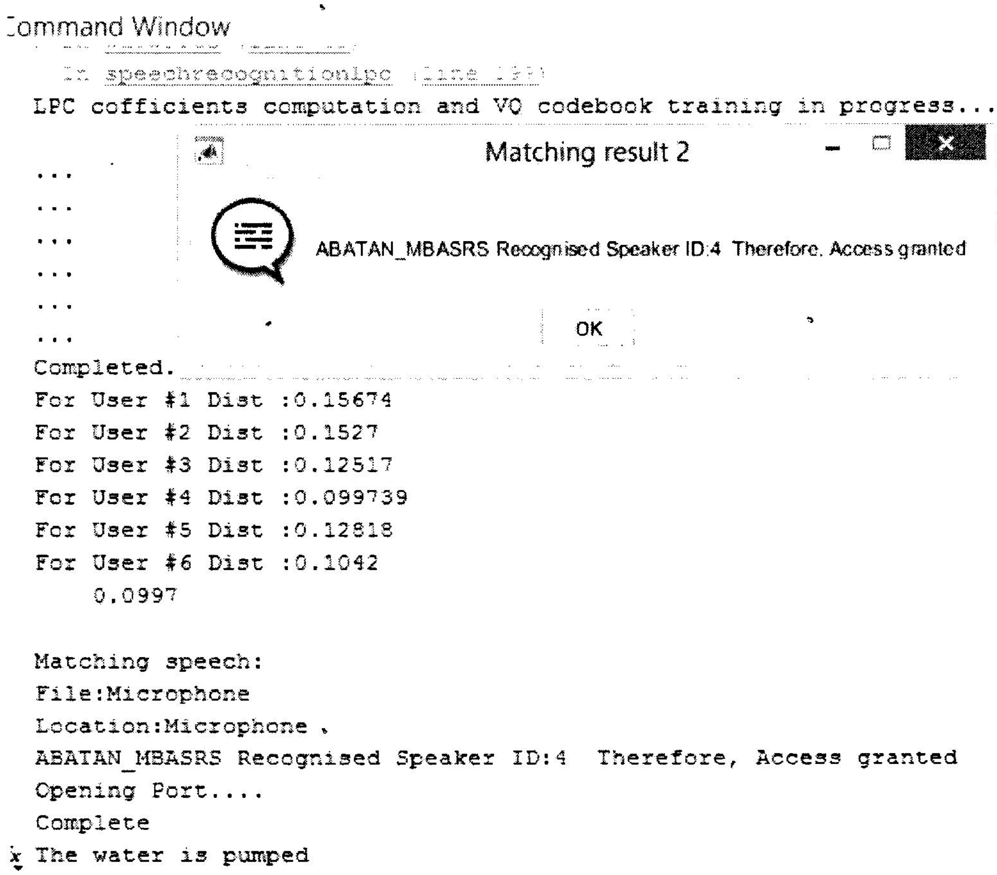


Figure 4.13: Recognized speaker "sound ID: 4"

Here, it matches the speech signal to the user ID: 4 of with closest Euclidean distance: 0.099739. Therefore give water pump access control to the Speaker that matches user ID: 4. User 1 distance is 0.15674, user 2 distance is 0.15270, user 3 distance is 0.12517.

4. Parametric Value of Users Speech Signal.

Table 4.1: The LPC coefficients of Speech Samples [BREEZE and WATER] for two Speakers in a Noisy Environment

	No of Trials	Speech Sample: [BREEZE]	Speech Sample: [Water]
SPEAKER 1	1	7.2033 -0.0266 -0.6530	7.0499 0.0470 -0.4099
	2	6.8861 0.0459 -0.2512	7.2485 0.0654 -0.4107
	3	7.1046 0.0356 -0.4330	7.8700 0.2526 -0.4656
	4	6.6582 -0.0457 -0.5206	6.5171 0.1187 -0.2738
	5	7.3113 0.0450 -0.3245	6.9053 -0.0652 -0.6109
SPEAKER 2	1	6.6802 0.0440 -0.4454	6.6943 -0.0174 -0.3318
	2	6.7932 -0.0813 -0.6058	6.7742 -0.0252 -0.3276
	3	6.7874 0.0581 -0.2692	6.6332 0.0482 -0.3426
	4	6.6746 0.1455 -0.2800	6.8602 0.0096 -0.4060
	5	6.8457 0.0318 -0.4609	6.8580 0.0365 -0.4152

Table 4.2: The LPC coefficients of Speech Samples [BREEZE and WATER] for two Speakers in a Noiseless Environment

	No of Trials	Speech Sample: [BREEZE]	Speech Sample: [Water]
SPEAKER 1	1	6.5595 -0.0034 -0.3078	6.4691 -0.0287 -0.2600
	2	6.4001 -0.0362 -0.2793	6.4844 -0.0402 -0.2725
	3	6.4697 -0.0417 -0.2658	6.3976 -0.0510 -0.2773
	4	6.4968 -0.0367 -0.2742	6.2892 -0.0421 -0.2713
	5	6.4653 -0.0430 -0.2776	6.4164 -0.0712 -0.2645
SPEAKER 2	1	6.3937 -0.0450 -0.2732	6.5051 -0.0450 -0.2734
	2	6.4194 -0.0511 -0.2750	5.4467 -0.3094 -0.1874
	3	5.9018 -0.1217 -0.1858	6.5516 -0.0418 -0.2635
	4	6.3776 -0.0511 -0.2713	6.5255 -0.0499 -0.2533
	5	6.5348 -0.0255 -0.2965	6.7918 -0.0013 -0.4070

As stipulated in Table 4.1 and Table 4.2, the LPC vector features/coefficients in the Table 4.1 by the two speakers in the noisy area are relatively vary due to distortion arises from different incoming speech signal. In comparison with Table 4.2 the difference in the LPC coefficients is sufficient to analyze the speech samples (BREEZE and WATER).

Table 4.3: Comparison of the Three Speech Features Extraction Techniques

Feature Extraction Techniques	Speech Sample: [BREEZE]	Speech Sample: [Water]
LPC	5.3684 -0.2960 -0.0867	2.8106 -0.5935 -0.2286
MFCC	-10.5096 -5.3939 -3.8267	-5.2831 2.6378 0.3616
PLP	0.0853 -0.2950 -0.2995	-0.5958 -0.3858 -0.2580

The table above shows the 3 features vector coefficients of three features extraction techniques (Linear Predictive Coding, Mel-Frequency Cepstra Coefficients and Perceptual Prediction) for speech sample BREEZE and WATER.

Table 4.4: Euclidean Distance of Speakers to Speech Samples [BREEZE and WATER] For Noisy Environment.

User 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			
	Speech Sample: [BREEZE]			
	Distance to User-1	Distance to User-2	Distance to User-3	Distance to User-4
Speaker-1	0.27574	0.37012	0.72682	0.29976
Speaker -2	0.22788	0.38932	0.71056	0.31292
Speaker -3	0.21898	0.36737	0.60802	0.28124
Speech Sample: [WATER]				
Speaker -1	0.33991	0.34818	0.61768	0.27299
Speaker -2	0.28496	0.29728	0.53835	0.22606
Speaker -3	0.33461	0.30088	0.60194	0.23555

Table 4.5: Euclidean Distance of Speakers to Speech Samples [BREEZE and WATER] For Noiseless Environment

User 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			
	Speech Sample: [BREEZE]			
	Distance to User-1	Distance to User-2	Distance to User-3	Distance to User-4
Speaker-1	0.15822	0.35403	0.63086	0.26652
Speaker-2	0.11250	0.27497	0.51767	0.19527
Speaker-3	0.12753	0.29227	0.5681	0.21607
Speech Sample: [WATER]				
Speaker-1	0.18092	0.16473	0.30805	0.15117
Speaker-2	0.17536	0.16876	0.33452	0.14101
Speaker-3	0.19704	0.18837	0.37269	0.16415

The tables (Table 4.3 and Table 4.4) above summarize the result and evaluate the system based on the Euclidean distance of the individual speech signal to be recognized both in noisy and quiet area. Here, the parametric representation of the distances of the speech signal is shown as calculated when the LPC and VQ codebook training function are applied. The speaker 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.

4.4.2 System Evaluation

As stipulated in figure 4.5, the minimum value of Euclidean distance of User-1 and User-4 was used to evaluate the accuracy of the speech recognition to speech samples BREEZE and WATER respectively.

$$\text{Word Error Rate } (WER) = Edit\ Distance * 100 \ldots\ldots\ldots\ldots (4.1)$$

$$\text{Word Accuracy } = 100 - \text{WER} \ldots\ldots\ldots\ldots (4.2)$$

Sample: BREEZE

Speaker-2:

Euclidean distance: 0.11250

WER=0.11250 * 100

= 11.25%

Word Accuracy = (100-11.25) %

=88.75%

Sample: WATER

Speaker-2

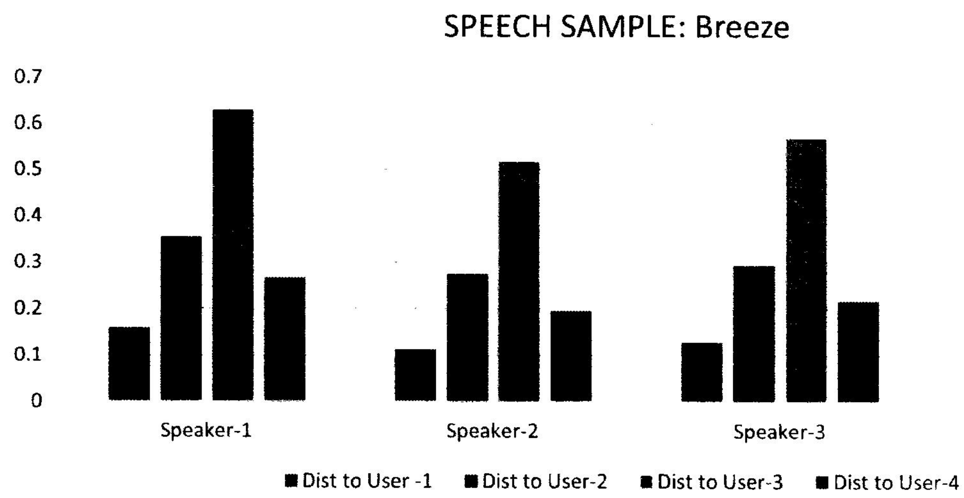
Euclidean distance: 0.14101

WER = 0.14101 * 100

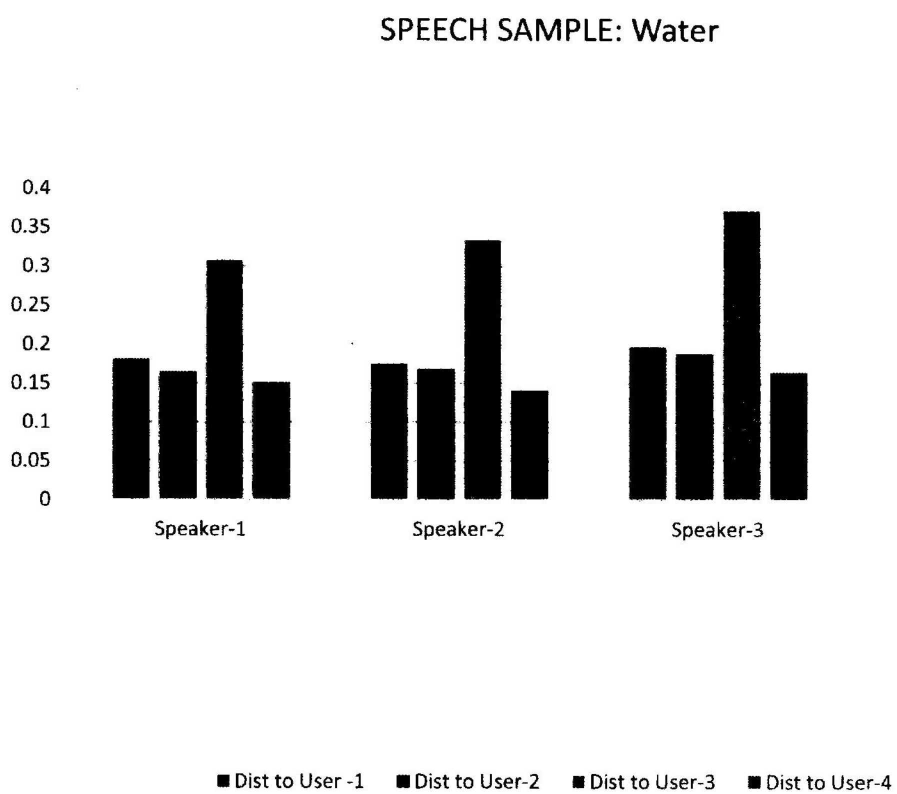
= 14.101%

Word Accuracy = (100-14.101)%

= 85.899%



(a)



(b)

Figure 4.14: Euclidean distance for BREEZE and WATER Speech Samples

4.4.3 Hardware Implementation and Testing Explanation

The microcontroller PIC16F628A receives the signal from the MATLAB software through the RS232 cable and sends a signal to the submersible water pump and fan motor for the their operation.

A 4MHz crystal oscillator provides speed for the microcontroller and is connected to two 22uF capacitor connected in parallel.

A 10k resistor is used to provide the contrast for the LCD

A 22K 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.

Two 4.7k resistors are connected to control the contrast regulation of the LCD. These resistors connected at the base of the transistor that serves as a buffer, the pullup 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.

The PCB layout diagram of the fan and water pump control circuit is shown above. A 5V DC is supplied into the circuit by a Computer System 1through a Universal Serial Bus cable interfaced with a serial port connector. The voltage in the circuit passed through 2200uF capacitor (a condenser microphone), which work by removing the lumps/ripples present in the voltage. The connection of pins between the different materials are listed below;

Pin 17 and Pin 18 of the PIC16F628A is connected to the 2N2222A transistors which controls the DC motor to power on and off the submersible water pump for water flow and fan rotation respectively. Pin 10-13 of PIC16F628A is connected to pin 6-3 of the LCD connector which is equivalent to pin 11-14 on the LCD itself, it is used to send command or data to the LCD. Pin 15 and 16 of PIC16F628A is connected to 4MHz crystal oscillator, which is also connected to two 22uF capacitors in parallel. Pin 4 of PIC16F628A to 4.7kohms pull-up resistor. Pin 6 of the lower side of the LCD connector which is equivalent to pin 3 on the LCD itself is connected to 10kohm resistor for contrast of the LCD. Pin 6-7 is used as RX pin which serve as USART receive pin or synchronous data I/O which allow serial communication of the microcontroller in which an RS232 is interfaced with a Universal Serial Bus module through which the control of the hardware operation was implemented.

Pin 5 of the PIC16F628A is connected to ground while Pin 4 of the PIC16F628A is connected to 5V in which Pin 14 can also be connected to 5V as power source. Pin 8 of PIC16F628A is connected to Pin 5 of the lower part of the connector for LCD which is equivalent to Pin 4 of the LCD itself which is the Register Select control pin which toggles the LCD between commands or data register. Pin 9 of PIC16F628A is connected to Pin 3 of the lower part of the connector for LCD which is equivalent to Pin 6 of the LCD itself which is the enable control pin, it needs to be held high to perform read or write operation by the LCD.

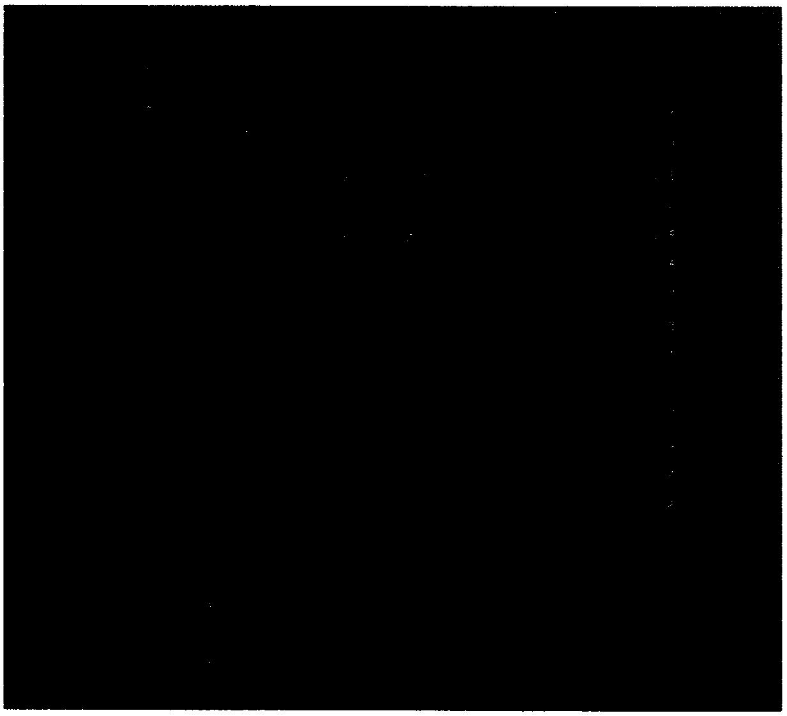


Figure 4.15: PCB layout of door control circuit

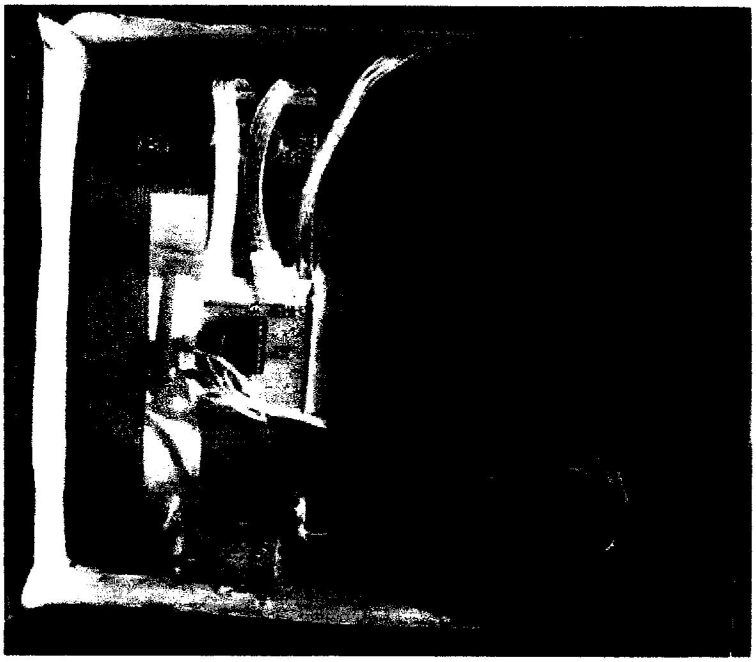


Figure 4.16: Diagram of the Internal Arrangement of the Prototype/Model electric water pump and fan

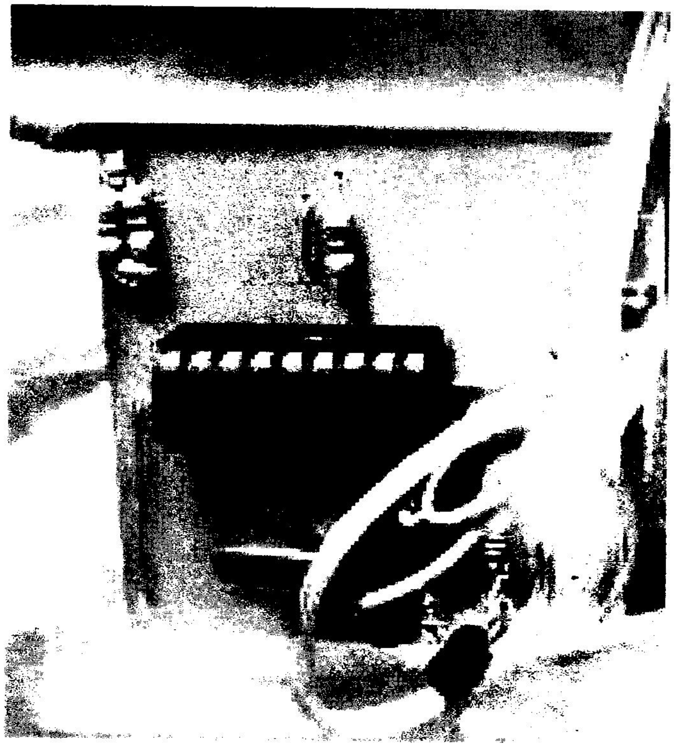
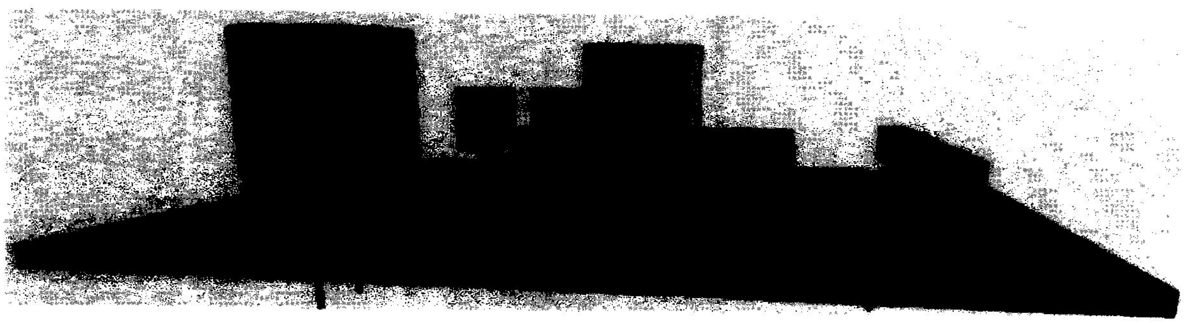
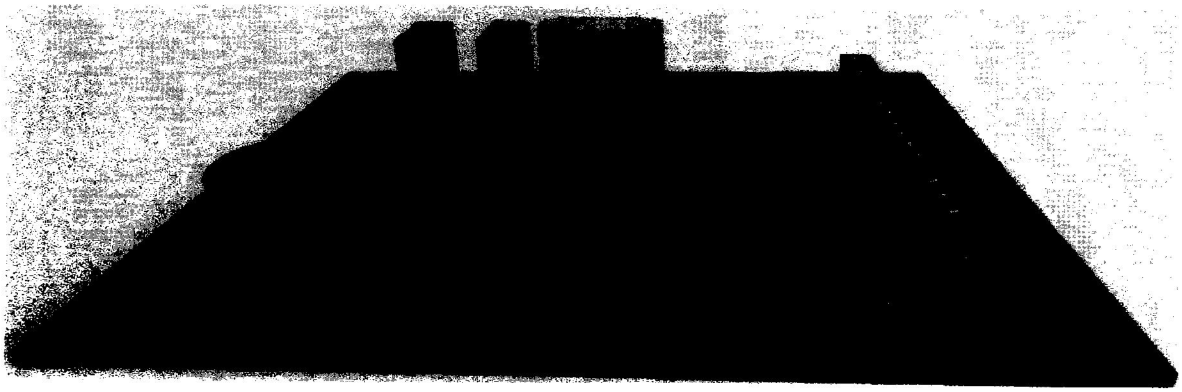


Figure 4.17: Fan and Water Pump Control Circuit



a. (Side View)



b. (Top View)

Figure 4.17: 3D Visualization of Fan and Water Pump Control Circuit

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

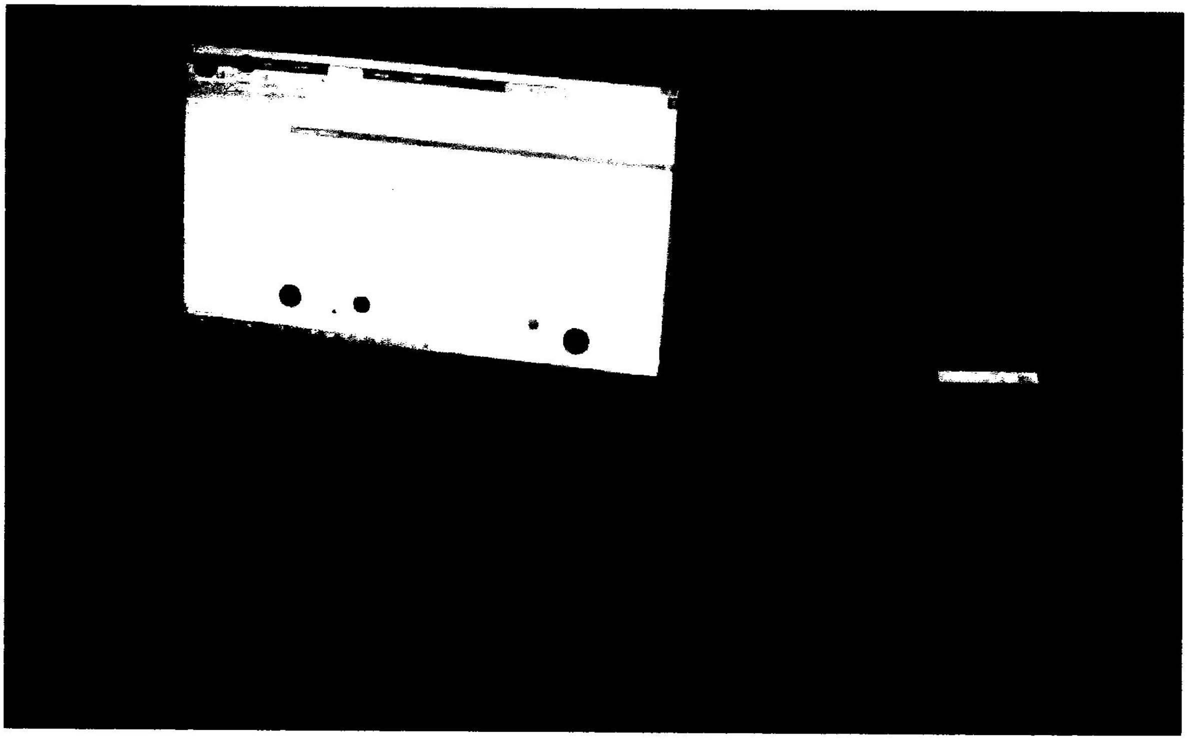


Figure 4.18: The Prototype/Model Fan and Water Pump to be controlled by the Speech Recognition System.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The system developed which is "Microcontroller Based Speech Recognition" is based on recognizing an unknown speaker from given a set of registered speakers in the system database.

The unknown speaker is assumed 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 Linear Predictive Coding. These features act as a basis for further development of the speaker recognition process. Next on recognition phase, we went for feature mapping using the vector Quantization using LBG algorithm. The results obtained using LPC and VQ are appreciable. The LPC coefficients and vectors for each speaker were computed and quantized respectively for efficient representation. The code books were generated using LBG algorithm which optimizes the quantization process. VQ distortion between the resultant codebook and LPC coefficients 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 89% was obtained using VQLBG algorithm and Euclidean distance.

Finally, the functionality of the project was implemented and tested using a model electric water pump and fan system built, controlled by PIC16F628A microcontroller which receives access signal from the speaker recognition system design on MATLAB and drives the Bipolar Transistor(NPN) which trigger the electric fan and water pump.

5.2 Problems Encountered

Problems encountered during this project 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. PIC16F628A microcontroller and other 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 can be 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

Anamul Haque, S. M., Kamruzzaman, S. M., & Ashraful, I. (n.d.). A System for Smart Home Control of Appliances Based on Timer and Speech Interaction. *4th International Conference on Electrical Engineering & 2nd annual Paper Meet 2.*

Aseem, S., Amit, K. S., Shashank, C., & Surabhi, C. (2013). Cloud Computing for Better IT Development. *International Journal of Advances In Computer Science and Cloud Computing. Vol- 1,Issue- 2,ISSN: 2321-4058.*

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,* , 1304-1313.

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *Journal of Acoustical Society of America., Vol 24, No. 6,*, 627-642.

Fang, Z., Guoliang, Z., & Zhanjiang, S. (2001). Comparison of Different Implementations of MFCC. . *The journal of Computer Science & Technology* , 582-589.

Ferrier, L. J., Shane, H. C., Ballard, H. F., Carpenter, T., & Benoit, A. (1995). Dysarthric Speakers' Intelligibilityand Speech Characteristics in Relation to Computer Speech Recognition. *Journal of Augmentative and Alternative Communication, Vol 11*, 165-174.

Gerhart, J. (n.d.). *(1999). Home automation and wiring.* Mcgraw-hill.

H., F. (1922.). The Nature of Speech and its Interpretations, . *Bell System Technology. Journal., Vol 1*, 129 144.

Harvey, F. (1922). *Nature of Speech and Its Interpretation.*

Hermansky, .., Hanson, B., & Wakita, .. (1985). Perceptually based Linear Predictive Analysis of Speech. *IEEE Int. Conf. On Acoustic, speech, and Signal processing.* 509-512.

Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis for speech. *Journal of Acoustic Society of America, Vol. 87*, 1738-1752.

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

Jawarkar, N. P., Vasif, A., & Thakare, R. (2007). Remote Control using Mobile through Spoken Commands. *IEEE - International Consortium of Stem Cell Networks (ICSCN) 2007.*, 622-625.

Kharka, B. R., Jeetendra, T., & Nirmal, R. (2015). "Voice controlled wheel chair using Arduino.

L., F. J. (1972.). *Speech Analysis, Synthesis and Perception* (Second ed.). Springer-Verlag.

Lovine, J. (2013). *http://www.talkingelectronics.com/John/SpeechRecognition.html.*

APPENDIX

A: Matlab code for LPC Speech Recognition System

```
function []=recognitionlpc()
% For clear screen
clc;
ronaldo=10;
char st;
disp('Project: Microcontroller Based Speech Recognition Identification
system');
disp('Electronic Fan and Water Pump Control System By ABATAN OLUWAGBOTEMI D.
');
disp('LOADING ');
 disp('> ABATAN LPC SPEECH RECOGNITION SYSTEM')
        pause(2);
        chos=0;
        possibility=7;
        while chos~=possibility,
chos=menu('Speaker Recognition System','Add a new speech from microphone',...
                'Speaker recognition from microphone',...
                'Speaker recognition ftom microphone2','LPC
Representation','Database Info','Delete database','Exit');
%% 10.1 Add a new sound from microphone
            if chos==1

                if (exist('sound_database.dat','file')==2)
                    load('sound_database.dat','-mat');
classe = input('Insert a class number (sound ID) that will be used for
recognition:');
                    if isempty(classe)
                        classe = 1;
                        disp( num2str(classe) );
                    end
disp('you have 1.5seconds to speak to the microphone');
samplingfrequency = input('Insert the sampling frequency (22050
recommended):');
                    if isempty(samplingfrequency )
                        samplingfrequency = 22050;
                        disp( num2str(samplingfrequency) );
                    end
samplingbits = input('Insert the number of bits per sample (8
recommended):');
                    if isempty(samplingbits )
                        samplingbits = 8;
                        disp( num2str(samplingbits) );
                    end
                    micrecorder =
audiorecorder(samplingfrequency,samplingbits,1);
                    disp('Now, speak into microphone...');
                    pause(0.5);
                    record(micrecorder,durata);

                    while (isrecording(micrecorder)==1)
                        disp('Recording...');
```

```
                        pause(0.5);
                    end
                    disp('Recording stopped.');
                    y1 = getaudiodata(micrecorder);
                    y = getaudiodata(micrecorder, 'uint8');

                    if size(y,2)==2
                        y=y(:,1);
                    end
                    y = double(y);
                    sound_number = 1;
                    data{sound_number,1} = y;
                    data{sound_number,2} = classe;
                    data{sound_number,3} = 'Microphone';
                    data{sound_number,4} = 'Microphone';
                    st=strcat('u',num2str(sound_number));
                    wavwrite(y1,samplingfrequency,samplingbits,st)

save('sound_database.dat','data','sound_number','samplingfrequency','sampling
bits');
                   des=strcat('sound : ',st,'.wav for user with sound ID :
',num2str(classe),' is added to database');
                    msgbox(des,'Database result','help');
                    disp('Sound added to database');
                end

            end
%%  Speech Recognition from microphone
            if chos==2
                if (exist('sound_database.dat','file')==2)
                    load('sound_database.dat','-mat');
                    Fs = samplingfrequency;
                      durata=1.5;
                    micrecorder =
audiorecorder(samplingfrequency,samplingbits,1);
                    disp('Now, speak into microphone...');
                    pause(0.2);
                    record(micrecorder,durata);
                    while (isrecording(micrecorder)==1)
                        disp('Recording...');
                        pause(0.5);
                    end
                    disp('Recording stopped.');
                    y1 = getaudiodata(micrecorder);
                    wavwrite(y1,samplingfrequency,samplingbits,'v');
                    y = getaudiodata(micrecorder, 'uint8');
                    if size(y,2)==2
                        y=y(:,1);
                    end
                    y = double(y);
                    %----- code for speaker recognition -------
disp('LPC cofficients computation and VQ codebook training in progress...');
                    disp(' ');
                    % Number of centroids required
                    k =16;
                    for ii=1:sound_number
```

```
% Compute LPC cofficients for each sound present in database
                    v = autocortest(data{ii,1},12);
                    % Train VQ codebook
                    code{ii} = vqlbg(v, k);
                    disp('...');
                end
                disp('Completed.');
                % Compute LPC coefficients for input sound
                v = autocortest(y,12);
                % Current distance and sound ID initialization
                distmin = Inf;
                k1 = 0;
                for ii=1:sound_number
                    d = disteu(v, code{ii});%code{ii} is output of VGLB
code book and v is mfcc coefficient
                    dist = sum(min(d,[],2)) / size(d,1);
                    message=strcat('For User #',num2str(ii),' Dist :
',num2str(dist));
                    disp(message);
                    if dist < distmin
                        distmin = dist;
                        k1 = ii;
                    end
                end
                if distmin < ronaldo
                    min_index = k1;
                    speech_id = data{min_index,2};
                    disp('Matching speech:');
                    message=strcat('File:',data{min_index,3});
                    disp(message);
                    message=strcat('Location:',data{min_index,4});
                    disp(message);
 message = strcat('Speaker ID: ',num2str(speech_id),'_ is Recognised but wait
for Device Authentication');
                    disp(message);
                    msgbox(message,'Matching result 1','help');
                    if speech_id==3
   message = strcat('Speaker ID: ',num2str(speech_id),'_ is Recognised and
Authenticated');
                    disp(message);
                    msgbox(message,'Matching result 2','help');
% COM1 CONTROL
clc; disp('BEGIN')
SerPIC = serial('COM1');
set(SerPIC,'BaudRate',2400);
set(SerPIC,'DataBits',8);
set(SerPIC,'Parity','none');
set(SerPIC,'StopBits',1);
set(SerPIC,'FlowControl','none');
fopen(SerPIC);
%*-*-*-*-*-*-
fprintf(SerPIC,'%s','F');pause(0.2)
fclose(SerPIC);
delete(SerPIC)
clear SerPIC
disp('STOP')
```

```
                    end
                   ch3=0;
                   while ch3~=3
                   ch3=menu('Matched result verification:','Recognized
spkr','Recorded spkr','Exit');
                        if ch3==1
                            st=strcat('u',num2str(sound_number));
                            [s fs nb]=wavread(st);
                            p=audioplayer(s,fs,nb);
                            play(p);
                        end
                        if ch3==2
                            [s fs nb]=wavread('v');
                            p=audioplayer(s,fs,nb);
                            play(p);
                        end
                   end
                else
                 warndlg('Wrong User . No matching Result.',' Warning ')
                end
            else
      warndlg('Database is empty. No matching is possible.',' Warning ')
          end
        end

%% 10.3 LPC Representation
        if chos==3
             if (exist('sound_database.dat','file')==2)
                 load('sound_database.dat','-mat');
            st=strcat('u',num2str(sound_number));
                               [regsig fs nb]=wavread(st);
                                regobj=autocortest(regsig,12);
                                figure(1);
                                plot(regobj);
                               title('Autocor analysis of recorded speech');
                                figure(2);
                                 plot(regsig);
                                 title('recorded speech');
                               [versig fs nb]=wavread('v');
                               verobj=autocortest(versig,12);
                               figure(3);
                               plot(verobj);
                              title('Autocor analysis of verified speech');
                                figure(4)
                                plot(versig);
                                title('verified speech');
             end
        end
%% 10.4 Database Info
         if chos==4
             if (exist('sound_database.dat','file')==2)
                 load('sound_database.dat','-mat');
                 message=strcat('Database has
#',num2str(sound_number),'words:');
                 disp(message);
                 disp(' ');
```

```
                for ii=1:sound_number
                    message=strcat('Location:',data{ii,3});
                    disp(message);
                    message=strcat('File:',data{ii,4});
                    disp(message);
                    message=strcat('Sound ID:',num2str(data{ii,2}));
                    disp(message);
                    disp('-');
                end
                ch32=0;
                while ch32 ~=2
                    ch32=menu('Database Information','Database','Exit');
                    if ch32==1
    st=strcat('Sound Database has : #',num2str(sound_number),'words. Enter a
database number : #');
                        prompt = {st};
                        dlg_title = 'Database Information';
                        num_lines = 1;
                        def = {'1'};
                        options.Resize='on';
                        options.WindowStyle='normal';
                        options.Interpreter='tex';
                        an = inputdlg(prompt,dlg_title,num_lines,def);
                        an=cell2mat(an);
                        a=str2double(an);
                        if (isempty(an))
                        else
                            if (a <= sound_number)
                                st=strcat('u',num2str(an));
                                [s fs nb]=wavread(st);
                                p=audioplayer(s,fs,nb);
                                play(p);
                            else
                                warndlg('Invalid Word ','Warning');
                            end
                        end
                    end
                end

            else
                warndlg('Database is empty.',' Warning ')
            end
        end
%% 10.4 Delete database
    if chos==5
            close all;
            if (exist('sound_database.dat','file')==2)
       button = questdlg('Do you really want to remove the Database?');
                if strcmp(button,'Yes')
                    load('sound_database.dat','-mat');
                    for ii=1:sound_number
                        st=strcat('u',num2str(ii),'.wav');
                        delete(st);
                    end
                    if (exist('v.wav','file')==2)
                        delete('v.wav');
```

```
                    end
                    delete('sound_database.dat');
msgbox('Database was succesfully removed from the current
directory.','Database removed','help');
                end
            else
                warndlg('Database is empty.',' Warning ')
            end
        end
    end
msgbox('. It has been Awesome!.With Love from ABATAN.','Thank You','help');
```

B: PROGRAM CODE FOR (PIC18F1X20 MICROCONTROLLER)

```
include "modedefs.bas"
DEFINE OSC 4
valor var byte
trisb =%00000001
trisa =%00000000
DEFINE LCD_DREG PORTB
DEFINE LCD_BITS 4
DEFINE LCD_DBIT 4
DEFINE LCD_RSREG PORTB
DEFINE LCD_RSBIT 2
DEFINE LCD_EREG PORTB
DEFINE LCD_EBIT 3
DEFINE LCD_LINES 2
CMCON = 7
PORTA = 0
LCDOUT $FE,$80,"    SPEECH    "
LCDOUT $FE,$C0,"  RECOGNITION  "
pause 4000
LCDOUT $FE,$80,"  PUMP AND FAN  "
LCDOUT $FE,$C0,"    CONTROL     "
pause 4000
D1:
LCDOUT $FE,$80," PUMP AND FAN  "
LCDOUT $FE,$C0,"  READY TO GO  "
pause 4000

inicio:
serin portb.0,T2400,valor
if valor == "F" then GOTO FA
if valor == "P" then GOTO pmp  'A->Apagado
goto inicio

FA:
high porta.0
LCDOUT $FE,$80,"     FAN      "
LCDOUT $FE,$C0,"    OPERATED    "
HIGH PORTA.0
pause 10000
```

```
LOW porta.0
LCDOUT $FE,$80,"   FAN TURNED      "
LCDOUT $FE,$C0,"    OFF           "
LOW PORTA.0
PAUSE 500
GOTO D1

PMP:
high porta.1
LCDOUT $FE,$80,"   WATER PUMP    "
LCDOUT $FE,$C0,"    OPERATED    "
HIGH PORTA.1
pause  10000
low porta.1
LCDOUT $FE,$80,"   PUMP TURNED     "
LCDOUT $FE,$C0,"    OFF           "
LOW PORTA.1
PAUSE 500
GOTO D1
END
```

of Engineering for Measurement and Evaluation (BEME)

[illegible]	COMPONENT DESCRIPTION	QUANTITY	UNIT COST(₦)	TOTAL COST(₦)
1	IC, Transistor(2N2222A), 5V	1	200	200
2	IC, PIC16F268A	1	1,200	1,200
3	Submersible water pump DC	1	1,000	1,000
4	Capacitor, 22uF	2	50	100
5	Capacitor, 2200uF	1	150	150
6	Resistor, 4.7K Ω	4	50	200
7	LCD(16X2)	1	900	900
8	4MHz Crystal Oscillator	1	400	400
9	RS232 Com Cable	1	1,500	1,500
10	Casing	1	3,000	3,000
11	Water and Fan Model	1	2,000	2,000
12	Fiber Clad Board	1	500	500
13	Soldering Lead	4	100	400
14	Connector/Jumper wire	8	50	400
15	Male and Female USB Cord	1	1500	1500
16	USB Module	1	450	450
17	DC Motor	1	350	350
	TOTAL			**₦14,250**