A MICROCONTROLLER BASED SPEECH RECOGNITION SYSTEM USING MEL FREQUENCY CEPSTRAL COEFFICIENT ALGORITHM

ADEWUMI, NIYI BABATUNDE

CPE/13/1072

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 Speech Recognition System for Toilet Flushing using Mel Frequency Cepstral Coefficient (MFCC) Technique" was prepared and submitted by Adewumi, Niyi Babatunde to the Department of Computer Engineering, has been read and it serve as part of the partial requirement for the award of bachelor degree in Engineering in the Department of Computer Engineering, Federal University Oye – Ekiti, Ekiti State.

.... 08-04-2019

Engr. N.S Okomba **Date**

(Project Supervisor)

.... 10:04:19

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

(Head of Department)

DECLARATION

I Adewumi, Niyi Babatunde 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.

-- 08/04/2019

Adewumi, Niyi Babatunde **Date**

CPE/13/1072

DEDICATION

I dedicate this work to the glory of Almighty God the giver of life and sustenance of mankind, whose grace has kept me till this moment, to my family and friends whose motivation has thrived me to this feat in life. I also dedicate this to the entire human race, Computer Engineering department staff and the entire community of Federal University Oye-Ekiti, Ekiti –State, Nigeria.

ACKNOWLEDGEMENT

I cannot but 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. Kunle Adewumi, my siblings Ms. Adewumi Folashade, Ms. Adewumi Adeyinka and Master Adewumi Kunle (Jnr). 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.

My sincere gratitude goes to the following families for their love and compassion; Hon. Oseni Yusuf, Mr. and Mrs. A. O Ologunagba, Mrs Esan Oyemomi, Barrister & Mrs. P.P Monde, Chief Mrs. J. A Yusuf, Mr and Mrs Adetunji, Mr and Mrs Folayan and Ven. A. O Olusakin. A big thank you to my brother; Mr. Buraimoh Samuel and Mr. Adetunji Temitope Felix (CrownReawake).

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

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.

The development of a speech recognition system for toilet flushing using Mel Frequency Cepstral Coefficient (MFCC) technique, has become a necessity as the physically challenged ones (handicap), individuals suffering from osteonecrosis and osteoporosis, the aged ones in the society are not exempted in the agony and continuous pain experienced in flushing water closet toilet after usage. Mel Frequency Cepstral Coefficient (MFCC) technique will be used due to its wide acceptance in speech processing. It has high recognition accuracy in speech feature extraction and has high performance rate with low complexity.

This technology serve as means of controlling water closet toilet flushing operation without manual intervention. The project will make life more comfortable for the physically challenged ones (the handicap), people living with disease like osteonecrosis and osteoporosis and the aged ones. It will be of great help for them to use speech as a medium of controlling the flushing operation of their water closet toilet.

This project was developed with the use of MATLAB software, which was programmed to recognize spoken word for the flushing of water closet toilet, it replaces manual means through the use of speech for the user. The system is trained to be speaker independent. In the system training, MFCC feature extraction technique is employed. The components used are 5V DC submersible water pump, for controlling the water chamber of the water closet for flushing, 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 using a model water closet toilet system to show its reliability and efficiency.

TABLE OF CONTENTS

CHAPTER FOUR: IMPLEMENTATION AND RESULTS

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

- Automatic Speech Recognition (ASR)
- Artificial Intelligence (AI)
- Direct Current (DC)
- Dynamic Time Warping (DTW)
- Hidden Markov Model (HMM)
- Integrated Circuit (IC)
- Liquid Crystal Display (LCD)
- Linear Predictive Coding (LPC)
- Mel Frequency Ceptral Coefficient (MFCC)
- Printed Circuit Board (PCB)
- Probabilistic Linear Discriminate Analysis (PLDA)
- Vector Quantization (VQ)

CHAPTER ONE
INTRODUCTION

1.1 PREAMBLE

In this current era and today's world, speech recognition has become very popular and important in applying technology to solve human problem, speech plays a vital role in human to system and machine communication. Speech recognition system performs two fundamental operations: signal modeling and pattern matching (Picone, 1993). Signal modeling represents process of converting speech signal into a set of parameters. Pattern matching is the task of finding parameter set from memory which closely matches the parameter set obtained from the input speech signal.

The development of a speech recognition system for toilet flushing using Mel Frequency Cepstral Coefficient (MFCC) algorithm, has become a necessity as the physically challenged ones (handicap), individuals suffering from osteonecrosis and osteoporosis, the aged ones in the society are not exempted in the agony and continuous pain experienced in flushing water closet toilet after usage. Mel Frequency Cepstral Coefficient (MFCC) algorithm (Koustav, 2014) will be used due to its wide acceptance in speech processing. It has high recognition accuracy in speech feature extraction and has high performance rate with low complexity.

Water Closet (WC) toilet system is among the modern means of getting rid of human body waste products by mechanically realizing water from the stored chamber of the water closet.

In controlling the flushing of a water closet by speech will bring a major relief at home, offices, private places and public areas. This enable the water storage part to release water and also stop with the aid of speech control.

The project will be developed with the use of MATLAB software, which will be programmed to recognize spoken word for the flushing of water closet toilet, it replaces manual means through the use of speech for the user (Ashish *et al* 2007). The system is trained to be speaker independent. In the system training, MFCC feature extraction algorithm will be used. The components used are DC motor and 2N2222A Transistor for the control of flush pump for the flushing of the closet, 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 using a model water closet toilet system to show its reliability and efficiency.

1.2 STATEMENT OF PROBLEM

In the conventional way of flushing a toilet, the handle of the water chamber will be pressed in other for water to be released and flush the water closet toilet or by pouring water directly using a bucket. This make life more uncomfortable for the physically challenged ones (the handicap), the aged ones, people with diseases like osteonecrosis and osteoporosis. It will be of great help for the people with disability in their body to operate their water closet conveniently, using a speech recognition system.

This project will be developed with the use of MATLAB software, which will be programmed to recognize spoken word for the flushing of water closet toilet, it replaces manual means through the use of speech for the user. The system is trained to be speaker independent. In the system training, MFCC feature extraction algorithm is employed, Mel Frequency Cepstral Coefficient (MFCC) algorithm (Urmila and Vilas 2010) will be used due to its wide acceptance in speech processing. It has high recognition accuracy in speech feature extraction and has high performance rate with low complexity compared to other feature extraction algorithm.

1.3 AIM AND OBJECTIVES

The aim of this project is to make life more easier for the physically challenged, aged and those with disease that affect their bone to be able to flush a water closet toilet using speech recognition system.

The following are the objectives of this project:

a. To design a speech recognition system that flushes a water closet toilet.

b. To implement the design of the speech recognition system that flushes the water closet toilet using a hardware prototype.

c. To evaluate the effectiveness of the system developed.

1.4 SCOPE OF STUDY

The scope of the project is to design a speech recognition system that flushes a water closet toilet. To develop the whole project, there will be a code structure which will be done through MATLAB software that will be used in the training and testing of the speech recognition system.

Also there will be a small physical model of water closet toilet to demonstrate this project.

1.5 METHODS OF STUDY

There were several steps thât was followed in the designing of a speech recognition system that flushes the water closet toilet. This project was built using a microcontroller and MATLAB software. The stages in speech recognition are basically two stages which are the speech training stage, the speech verification and testing (recognition stage). In the training stage we have the following procedures; data acquisition stage (Speech input), the computer with the aid of a Matlab will perform the following operation for the first stage in processing the speech; analog to digital conversion, preprocessing and filtering, feature extraction using Mel Frequency Ceptral Coefficient (Hossan, 2010). The Next is the verification and testing (recognition stage) in which it undergoes the following steps with the aid of a Matlab on the computer system; acquiring the speech signal, performing the feature extraction using MFCC algorithm, reference model with speech identification to determine similarity with what is in the data base, decision making, verification of result for acceptance or rejection in order to drive the hardware operation. The final stage is the hardware development, making use of a microcontroller which is interfaced with the program developed in MATLAB which controls the operation and the LCD that display the state of operation.

1.6 SIGNIFICANCE OF THE PROJECT

The importance of the project cannot be overemphasized because it will bring a lot of relieve to water closet toilet user due to easy operation with the aid of speech recognition system.

The project will allow the physically challenge ones (handicap) in the society to be able to get rid of their body wastes by using water closet toilet with great ease.

The aged ones and those with diseases like osteonecrosis and osteoporosis will be able to flush their water closet toilet conveniently.

CHAPTER TWO

LITERATURE REVIEW

2.1 PREAMBLE

Speech recognition has in years become a practical concept, which is now being implemented in different languages around the world. Speech Recognition (is also known as Automatic Speech Recognition (ASR), is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program. It has been used in real-world human language applications (Wiqas and Navdeep 2012), such as information recovery and it has been applied in system as a means of controlling systems operation in place of switch or mechanical means of operation. Speech in human can be said as the most common means of communication because the information maintains the basic role in conversation (Anusuya and Katti 2009). Speech recognition system involves two phases; training phase and recognition phase. In training phase, known speech is recorded and parametric representation of the speech is extracted and stored in the speech database. In the recognition phase, for the given input speech signal the features are extracted and the ASR system compares it with the reference templates to recognize the utterance. In a speech recognition system, many parameters affect the accuracy of recognition such as vocabulary size, speaker dependency, speaker independence, time for recognition, type of speech (continuous, isolated) and recognition environment condition. The extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any speech recognition system (Koustav, 2014). Speech recognition algorithm consists of several stages in which feature extraction and classification are mainly important.

Speech recognition systems can be characterized by environment, vocabulary, acoustic model, speaking style, speaking mode, language model, perplexity, Signal to Noise Ratio (SNR) and transducer. The literature review for this project work was done by referring to the journal papers, conference papers, articles, books and internet. This chapter describes a review of speech recognition task, speech recognition approaches and current speech recognition system as well as different type of methods applied to speech recognition system. Also the review of the advantages and disadvantages of feature extraction techniques. This project work discusses the techniques and methods to develop a speech recognition system.

2.2 TYPES OF SPEECH RECOGNITION.

Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize (Shikha *et al* 2014). These classes are classified as the following:

2.2.1 Isolated Words:

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

2.2.2 Connected Words:

Connected word systems (or more correctly 'connected utterances') are similar to isolated words, but allows separate utterances to be 'run-together' with a minimal pause between them (Anusuya and Katti 2009).

2.2.3 Continuous Speech:

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

2.2.4 Spontaneous Speech:

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together

2.3 SPEECH RECOGNITION TECHNIQUES

The main objective of a speech recognition system is to have capacity to listen, understand and then after act on the spoken information. A speech recognition system includes four main stages which are;

- Analysis
- Feature Extraction
- Modelling
- Matching

The four main stages are further classified. The figure 2.1 below gives the description of the classification;

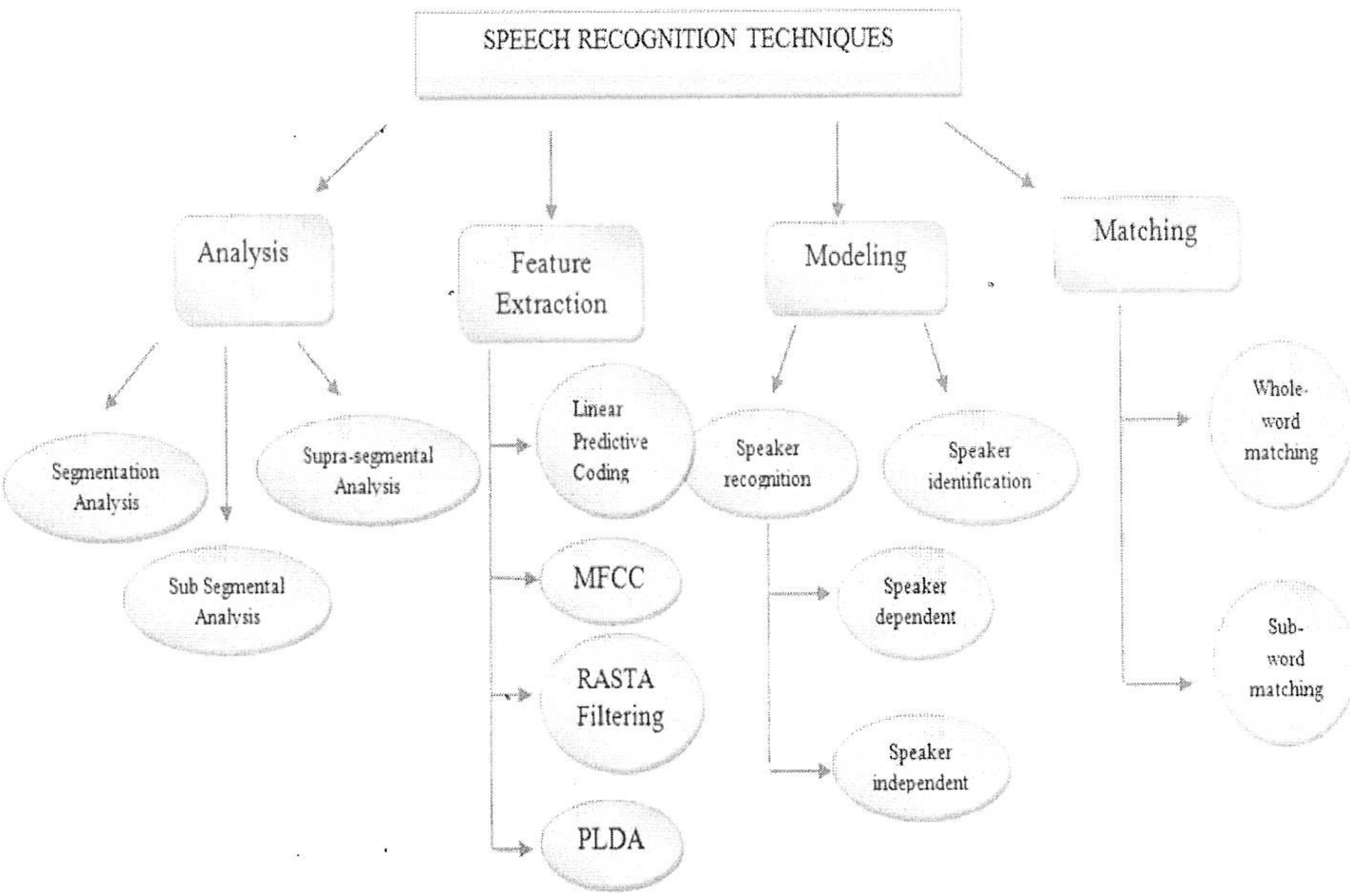


Figure 2.1 Speech recognition classification diagram (Santosh and Pravin 2010)

2.4 ANALYSIS

The first stage is analysis. When the speaker speaks, the speech includes different types of information that help to identify a speaker. The information is different because of the vocal tract, the source of excitation as well as the behaviour feature (Shreya *et al*, 2015). As shown in the figure above, the speech analysis stage can be further classified into three analyses:

a. Segmentation Analysis: In segmentation analysis, the testing to extract the information of speaker is done by utilizing the frame size as well as the shift which is in between 10 to 30 milliseconds (ms).

b. Sub-segmental Analysis: In this analysis technique, the testing to extract the information of speaker is done by utilizing the frame size as well as the shift which is in between 3 to 5 milliseconds (ms). The features of excitation state are analyzed and extracted by using this technique.

c. Supra-segmental Analysis: In Supra-segmental analysis, the analysis to extract the behaviour features of the speaker is done by utilizing the frame size as well as the shift size that ranges in between 50 to 200 milliseconds.

2.5 FEATURE EXTRACTION TECHNIQUES

Feature extraction is the main part of the speech recognition system. It is considered as the heart of the system. The work of this is to extract those features from the input speech (signal) that help the system in identifying the speaker. Feature extraction compresses the magnitude of the input signal (vector) without causing any harm to the power of speech signal. Some of the feature extraction techniques includes the following;

2.5.1 LINEAR PREDICTIVE CODING (LPC)

It is a tool which is used for speech processing. LPC is based on an assumption: In a series of speech samples, we can make a prediction of the nth sample which can be represented by summing up the target signal's previous samples (k). The production of an inverse filter should be done so that is corresponds to the formant regions of the speech samples. Thus the application of these filters into the samples is the LPC process. LP is a model based on human speech production. It utilizes a conventional source-filter model, in which the glottal, vocal tract, and lip radiation transfer functions are integrated into one all-pole filter that simulates acoustics of the vocal tract.

The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficients are estimated every frame, which is normally 20 ms long. The predictor coefficients are represented by a k. Another important parameter is the gain (G). The transfer function of the time varying digital filter is given by $H(z) = G/(1\text{-}\Sigma akz\text{-}k)$

Where k=1 to p, which will be 10 for the LPC-10 algorithm and 18 for the improved algorithm that is utilized. Levinson-Durbin recursion will be utilized to compute the required parameters for the auto-correlation method (Deller *et al.*, 2000).

2.5.1.1 Performance Analysis of LPC

Following parameters are involved in performance evaluation of LPC's

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

2.5.1.2 Types of LPC

Following are the types of LPC

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

2.5.1.3 Advantages of Linear Predictive Coding

a. It is a reliable, accurate and robust technique for providing parameters which describe the time-varying linear system which represent the vocal tract (Shanthi *e tal*, 2013).
b. Computation speed of LPC is good and provides with accurate parameters of speech.
c. Useful for encoding speech at low bit rate.

2.5.1.4 Disadvantages of Linear Predictive Coding

- It cannot distinguish the words with similar vowel sounds (Ghai & Singh 2012).
- Cannot represent speech because of the assumption that signals are stationary and hence is not able to analyze the local events accurately.
- LPC generates residual error as output that means some amount of important speech gets left in the residue resulting in poor speech quality.

2.5.1.5 Characteristics of Linear Predictive Coding

- Provides auto-regression based speech features (Furui *et al*, 2005)
- It is a formant estimation technique
- The residual sound is very close to the vocal tract input signal.

2.5.2 Mel Frequency Cepstral Coefficients

The MFCC (Furui *et al*, 2005) (Gaikwad *et al*, 2010) is the most evident example of a feature set that is extensively used in speech recognition. As the frequency bands are positioned logarithmically in MFCC.

The figure 2.3 shows the steps involved in MFCC feature extraction.

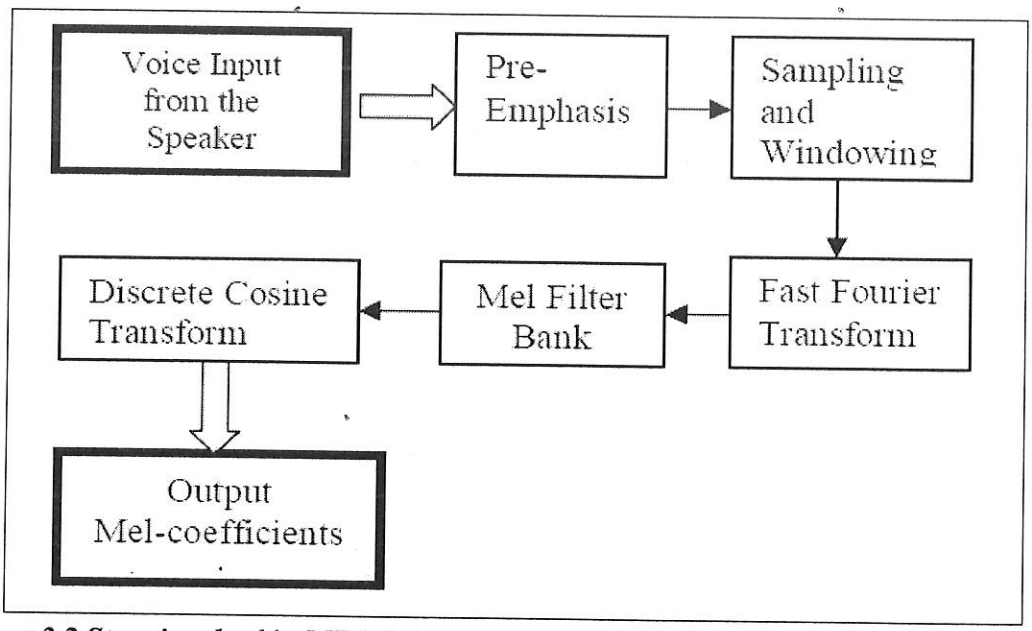


Figure 2.2 Steps involved in MFCC feature extraction (Koustav, 2014)

First Step

MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore is the best for Speech recognition. Taking voice input from the speaker.

Second Step

Pre Emphasis; The speech signal $x(n)$ is sent to a high-pass filter:

$y(n) = x - a * x(n-1)$...(2.1)

Where $y(n)$is the output signal and the value of a is usually between 0.9 and 1.0. The goal of pre-emphasis is to compensate the high-frequency part that was suppressed during the sound production mechanism of humans *(Lindasalwa et al, 2010).*

Third Step

Sampling and Windowing: The voice signal is divided into frames of N samples. Adjacent frames are being separated by$M(M < N)$. Typical values used are M = 100 and N= 256

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines. The Hamming window equation is given as:

If the window is defined as $W\ (n), 0 \leq n \leq N-1$ where

N = number of samples in each frame

$Y[n]$= Output signal

$X(n)$ = input signal

$W(n)$ = Hamming window

Fourth Step

Fast Fourier Transform: It convert each frame of N samples from time domain into frequency domain. The Fourier Transform is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response H[n] in the time domain. This statement supports the equation below:

$Y(w) = FFT[h\ (t) * X(t)] = H(w) * X(w)$..............(2.2)

If $X(w)$, $H(w)$ and $Y(w)$ are the Fourier Transform of $X(t)$, $H(t)$ and $Y(t)$ respectively.

Fifth Step

Mel filter Bank: The frequencies range in FFT spectrum is very wide and voice signal does not follow the linear scale. Then, each filter output is the sum of its filtered spectral components (Ginder and Ying, 2008). After that the following equation is used to compute the Mel for given frequency f in HZ:

$F(Mel) = [2595 * log10[1 + f]700]$..........................(2.3)

Sisth and Seventh Step

Discrete Cosine Transform This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector. The output of Mel Coefficient is obtained.

2.5.2.1 Advantages of Mel Frequency Cepstral Coefficient

- The recognition accuracy is high. That means the performance rate of MFCC is high.
- MFCC captures main characteristics of phones in speech.
- Low Complexity.

2.5.2.2 Dis Advantages of Mel Frequency Cepstral Coefficient

- In background noise MFCC does not give accurate results (Ghai & Singh, 2012.)
- The filter bandwidth is not an independent design parameter
- Performance might be affected by the number of filters (Juang, & Rabiner 2005)

2.5.2.3 Characteristics of MFCC

- Used for speech processing tasks.
- Mimics the human auditory system
- Mel frequency scale: linear frequency spacing below 1000Hz & a log spacing above 1000Hz.

2.5.3 Relative Spectral Transform (RASTA).

RASTA. It is a technique which is used to enhance the speech when recorded in a noisy environment. The time trajectories of the representations of the speech signals are band pass filtered in RASTA. Initially, it was just used to lessen the impact of noise in speech signal but know it is also used to directly enhance the signal (King *et al* 2005). The following figure shows the process of RASTA technique. The main thought here is to subdue the constant factors (Maheswari *et al* 2010).

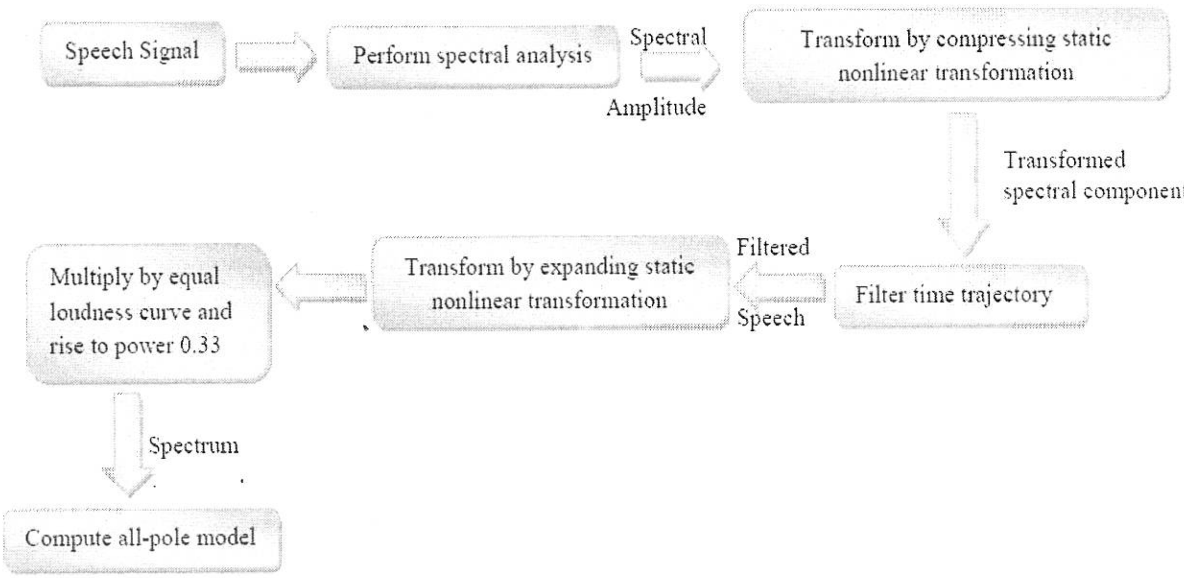


Figure 2.3: Block diagram of RASTA (Santosh and Pravin 2010)

To compensate for linear channel distortions the analysis library provides the power to perform rasta filtering (Shreya *et al*, 2015). The rasta filter is used either within the log spectral or cepstral domains. In result the rasta filter band passes every feature coefficient. Linear channel distortions seem as an additive constant in each the log spectral and therefore the cepstral domains. The high-pass portion of the equivalent band pass filter alleviates the result of convolution noise introduced in the channel. The low-pass filtering helps in smoothing frame to border spectral changes.

2.5.3.1 Advantages of RASTA

- Removes the slow varying environmental variations as well as the fast variations in artefacts. (Klevans & Rodman, 1997)

- This technique does not depend on the choice of microphone or the position of the microphone to the mouth, hence it is robust. (Maheswari *et al*, 2010)
- Captures frequencies with low modulations that correspond to speech (Moore, 1994)

2.5.3.2 Disadvantages of RASTA

- This technique causes a minor deprivation in performance for the clean information but it also slashes the error in half for the filtered case. (Maheswari *et al*, 2010) RASTA combined with PLP gives a better performance ratio (Moore, 1994).

2.5.3.3 Characteristics of RASTA

i. It is a band pass filtering technique.

ii. Designed to lessen impact of noise as well as enhance speech. That is, it is a technique which is widely used for the speech signals that have background noise or simply noisy speech.

2.5.4 Probabilistic Linear Discriminate Analysis (PLDA)

This technique is an extension for linear probabilistic analysis. Initially this technique was used for face recognition but now it is used for speech recognition (Morales et al, 2005).

2.5.4.1 Advantages of PLDA

1. It is a flexible acoustic model which makes use of variable number of interrelated input frames without any need of covariance modelling
2. High recognition accuracy (Morales et al, 2005).

2.5.4.2 Disadvantages of PLDA

1. The Gaussian assumption which are on the class conditional distributions. This is just an assumption and is not true actually.
2. The generative model is also a disadvantage. The objective was to fit the date which takes class discrimination into account (Reddy, 1966).

2.5.4.3 Characteristics of PLDA

1. Based on i-vector extraction. The i-vector is one which is full of information and is a low dimensional vector having fixed length.
2. This technique uses the state dependent variables of HMM. PLDA is formulated by a generative model.

2.5.5 Perceptually Based Linear Predictive Analysis (PLP)

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

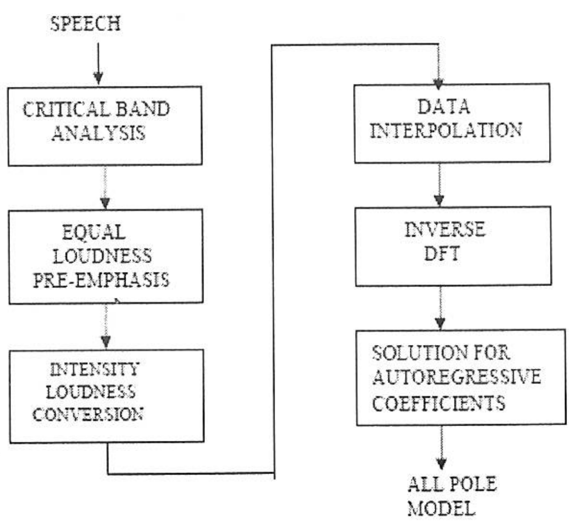


*Figure 2.4 Block diagram of PLP (*Hermansky *et al* 1985*)*

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

2.5.5.1 Advantages of PLP

i. 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.

ii. 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

2.6 APPROACHES TO SPEECH RECOGNITION:

- Basically there exist three approaches to speech recognition. These include;
- Acoustic Phonetic Approach
- Pattern Recognition Approach
- Artificial Intelligence Approach

2.6.1 *Acoustic Phonetic Recognition*

Acoustic phonetic recognition performs the function at phoneme level. It exist distinctive, finite phonemes which are characterized by a set of acoustic properties that occur in a speech signal (Juang & Rabiner 2005). English language includes forty different phonemes and doesn't depend on the vocabulary. It is the earliest approach of SRS to recognize speech by providing labels to the speech. This approach includes highly variable phonetic units, the variability in these unit are straight forward which are easily learned by machine (King *e tal*, 2007).

This approach is divided into three stages

- Feature Extraction
- Segmentation and Labeling
- Word-level recognition

Segmentation and labeling string of words are produced from phonetic label sequence.

In first step the spectral analysis of speech signal along with feature detection are performed which convert spectral measurements to set of features that provide vast acoustic properties to the signal. In second step attachment of phonetic label is done with segmentation region of speech signal. It gives phoneme lattice characterization of speech signal. In the last step by using

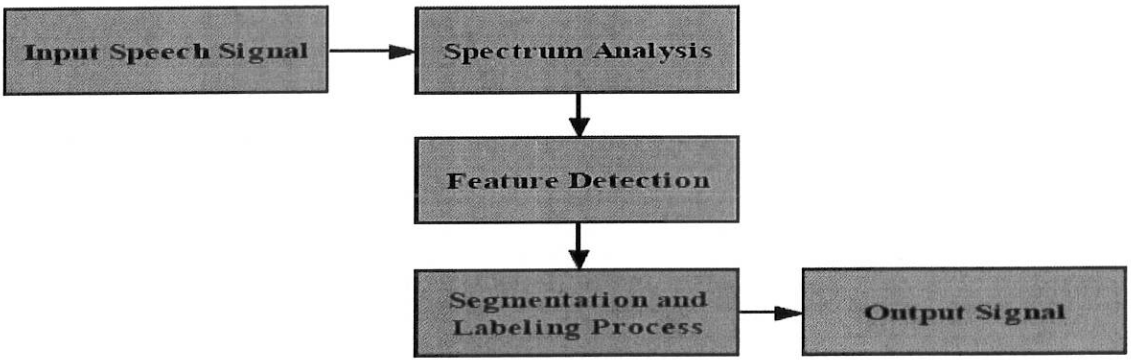


Figure 2.5: Acoustic Phonetic Recognition Block Diagram (Shaikh and Deshmukh 2016)

2.6.2 Artificial Intelligence approach (Knowledge Based approach)

The Artificial Intelligence approach (.Moore, 1994) is a hybrid of the acoustic phonetic approach and pattern recognition approach. In this, it exploits the ideas and concepts of Acoustic phonetic and pattern recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based approaches have been very effective in the design of a variety of speech recognition systems; they provided little insight about human speech processing, thereby making error analysis and knowledge-based system enhancement difficult. On the other hand, a large body of linguistic and phonetic literature provided insights and understanding to human speech processing. In its pure form, knowledge engineering design involves the direct and explicit incorporation of expert's speech knowledge into a recognition system. This knowledge is usually derived from careful study of spectrograms and is incorporated using rules or procedures. This form of knowledge application makes an important distinction between knowledge and algorithms. Algorithms enable us to solve problems. Knowledge enable the algorithms to work better.

2.6.3 Pattern Recognition Approach

By using mathematical framework this approach formulates speech pattern representation from formal training algorithms by set of labeled training samples via formal training algorithms (Klevans and Rodman 1997). This approach involves two major steps.

- Pattern training
- Pattern comparison

In pattern training process speech signals are shown in speech template or statistical modal. In pattern comparison process unknown speeches are compared with possible learned pattern which are formed from training stage.

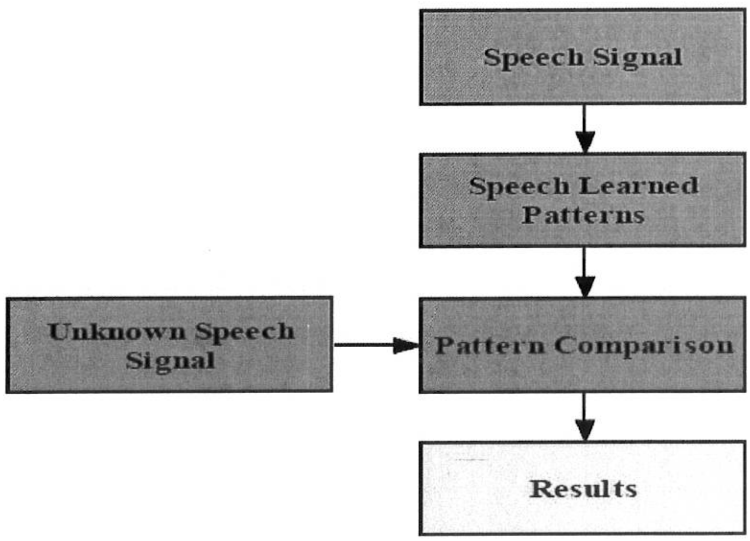
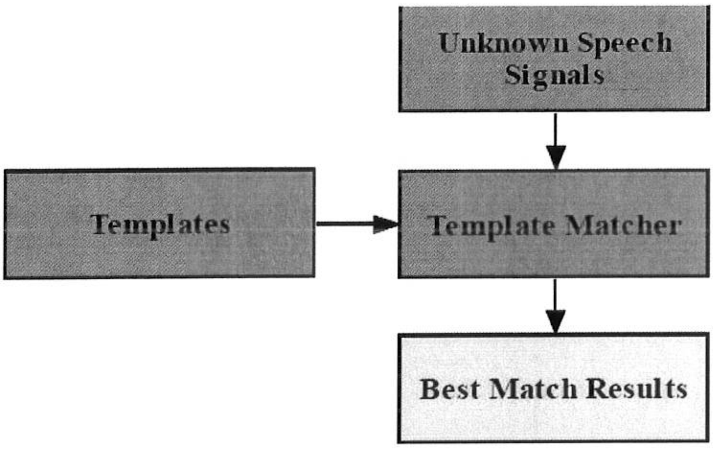


Fig 2.6: Pattern Recognition Approach Block Diagram (Shaikh and Deshmukh 2016)

There exists two methods in this approach namely;

2.6.3.1 Template approach and stochastic approach.

In template based approach a collection of prototypical speech patterns are stored as reference patterns representing the dictionary of candidate words (Annusuya & Katti 2009). Recognition is then carried out by matching an unknown spoken utterance with each of these reference templates and selecting the category of the best matching pattern. Usually templates for entire words are constructed. This has the advantage that, errors due to segmentation or classification of smaller acoustically more variable units such as phonemes can be avoided. In turn, each word must have its own full reference template; template preparation and matching become prohibitively expensive or impractical as vocabulary size increases beyond a few hundred words. One key idea in template method is to derive a typical sequences of speech frames for a pattern (a word) via some averaging procedure, and to rely on the use of local spectral distance measures to compare patterns.



*2.7: **Template Based Approach Block Diagram*** (Shaikh and Deshmukh 2016)

2.6.3.2 Stochastic Approach:

Stochastic modeling (Moore, 1994) entails the use of probabilistic models to deal with uncertain or incomplete information. In speech recognition, uncertainty and incompleteness arise from many sources; for example, confusable sounds, speaker variability s, contextual effects, and homophones words. Thus, stochastic models are particularly suitable approach to speech recognition. The most popular stochastic approach today is hidden Markov modeling. A hidden Markov model is characterized by a finite state markov model and a set of output distributions. The transition parameters in the Markov chain models, temporal variabilities, while the parameters in the output distribution model, spectral variabilities. These two types of variabilites are the essence of speech recognition. Compared to template based approach.

2.7 ALGORITHMS FOR SPEECH RECOGNITION

2.7.1 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, he or she were walking more quickly or even if there were accelerations and decelerations during the course of one observation (Shaikh and Deshmukh 2016). DTW has been applied to video, audio, and graphics indeed, any data which can be turned into a linear representation can be analyzed with DTW. A well-known application has been automatic speech recognition, to cope with different speaking speeds.In general, DTW is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension.

2.7.2 Vector Quantization (VQ):

Vector Quantization (VQ) (Moore, 1994) is often applied to ASR. It is useful for speech coders, i.e., efficient data reduction. Since transmission rate is not a major issue for ASR, the utility of VQ here lies in the efficiency of using compact codebooks for reference models and codebook searcher in place of more costly evaluation methods, for each vocabulary word gets its own VQ codebook, based on training sequence of several repetitions of the word. The test speech is evaluated by all codebooks and ASR chooses the word whose codebook yields the lowest distance measure. In basic VQ, codebooks have no explicit time information (e.g., the temporal order of phonetic segments in each word and their relative durations are ignored), since codebook entries are not ordered and can come from any part of the training words. However, some indirect durational cues are preserved because the codebook entries are chosen to minimize average distance across all training frames, and frames, corresponding to longer acoustic segments (e.g., vowels) are more frequent in the training data. Such segments are thus more likely to specify code words than less frequent consonant frames, especially with small codebooks. Code words nonetheless exist for constant frames because such frames would otherwise contribute large frame distances to the codebook.

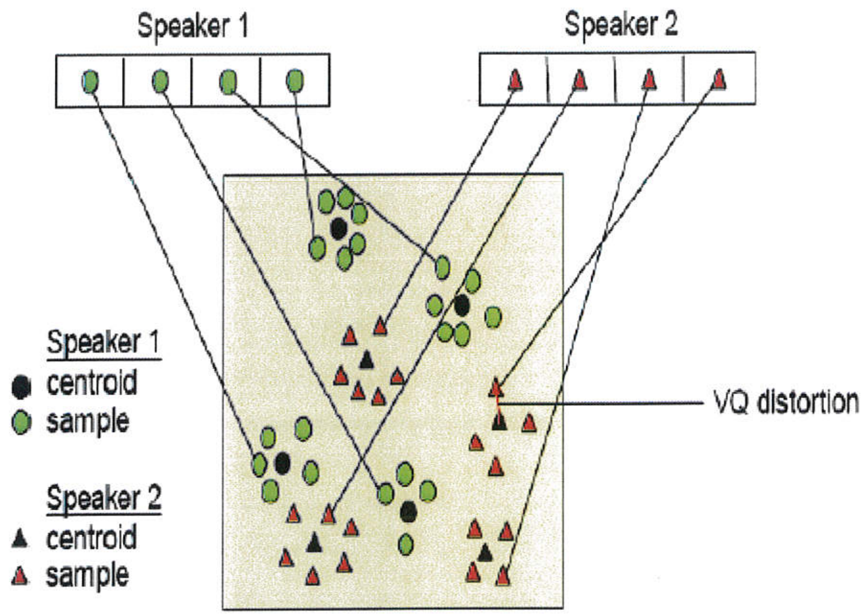


Figure 2.8: VQ recognition process (*Shikha *et al* 2014*).

Above figure shows a conceptual diagram to illustrate this recognition process. In the figure, only two speakers and two dimensions of the acoustic space are shown. The circles refer to the acoustic vectors from the speaker 1 while the triangles are from the speaker 2. In the training phase, a speaker-specific VQ codebook is generated for each known speaker by clustering his/her training acoustic vectors (Ashish, 2007). The result code words (centroids) are shown in Figure by black circles and black triangles for speaker 1 and 2, respectively. The distance from a vector to the closest codeword of a codebook is called a VQ-distortion. In the recognition phase, an input utterance of an unknown voice is "vector-quantized" using each trained codebook and the total VQ distortion is computed. The speaker corresponding to the VQ codebook with smallest total distortion is identified.

2.7.3 Artificial Neural Networks

The artificial intelligence approach (Lesser *et al.* 1975) attempts to mechanize the recognition procedure according to the way a person applies intelligence in visualizing, analyzing, and characterizing speech based on a set of measured acoustic features. Among the techniques used within this class of methods are use of an expert system (e.g., a neural network) that integrates phonemic, lexical, syntactic, semantic, and even pragmatic knowledge for segmentation and labeling, and uses tools such as artificial neural networks for learning the relationships among phonetic events (Weibel *et al,* 1989). The focus in this approach has been mostly in the representation of knowledge and integration of knowledge sources. A neural network can be defined as a model of reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells, or basic information-processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections, *synapses*, between them. By using multiple neurons simultaneously, the brain can perform its functions much faster than the fastest computers in existence today. Each neuron has a very simple structure, but an army of such elements constitutes a tremendous processing power. A neuron consists of a cell body, soma, a number of fibers called dendrites, and a single long fiber called the axon.

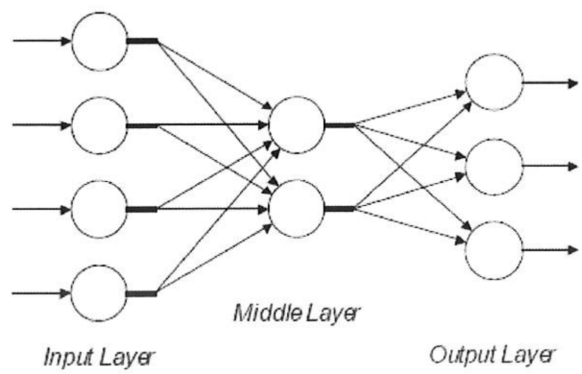


Figure 2.9: Artificial Neural Network

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

2.7.3.1 The three basic types of learning,

1. Supervised Learning
2. Un-Supervised Learning
3. Reinforced Learning

1. **Supervised Learning** - Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors (output vectors) are known as

supervised learning problems. Supervised learning is when the data you feed your algorithm is "tagged" to help your logic make decisions. Eg. Face recognition, perceptron

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

2.7.3.2 There are two types of ANN,

1. Feed-Forward NN
2. Recurrent NN

1. **Feed-Forward Neural Network (FFNN)** - A feed forward neural network is an artificial neural network where connections between the units do *not* form directed.
2. **Recurrent NN** - A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behaviour.

2.7.3.3 Advantages of ANN

i. ANNs are highly non-linear modelling

ii. ANN is nonlinear model that is easy to use and understand compared to statistical methods. ANN is non-parametric model while most of statistical methods are parametric model that need higher background of statistic.

iii. ANN with Back propagation (BP) learning algorithm is widely used in solving various classifications and forecasting problems. Even though BP convergence is slow but it is guaranteed.

2.7.3.4 Areas of Application of ANN

1. Character Recognition
2. Image Compression
3. Medicine and Security

2.7.4 Hidden Markov Model (HMM)

A Hidden Markov Model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (*hidden*) states. An HMM can be presented as the simplest dynamic Bayesian network. A Hidden Markov Model is a collection of states connected by transitions. It begins in a designated initial state. In each discrete time step, a transition is taken into a new state, and then one output symbol is generated in that state. The choice of transition and output symbol are both random, governed by probability distributions. The HMM can be thought of as a black box, where the sequence of output symbols generated over time is observable, but the sequence of states visited over time is hidden from view. This is why it's called a *Hidden* Markov Model. The figure 11 below is a general type of a HMM model, any state can be revisited and the revisits do not need to take place at specific time intervals.

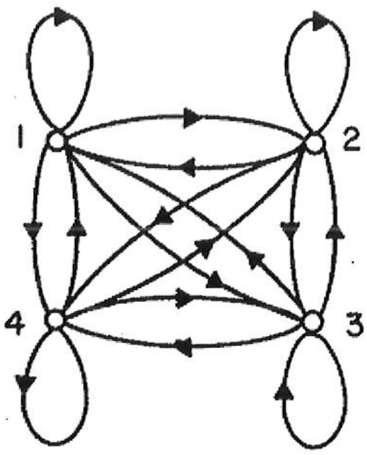


*Fig 2.10: A General Hidden Markov Model (*Rabiner and Junang 1986)

2.7.4.1Algorithms of HMM

There are three basic algorithms associated with Hidden Markov Models:

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

2.7.4.2 Limitations of HMM

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

2.7.5 Matching Techniques: The word that has been detected is used by the engine of speech recognizer to a word that is already known by making use of one of the following techniques:

i. **Sub word matching**: Phonemes are looked up by the search engine on which the system later performs pattern recognition. These phonemes are the sub words thus the name sub word matching. The storage that is required by this technique is in the range 5 to 20 bytes per word which is much less in comparison to whole word matching but it takes a large amount of processing.

ii. **Whole word matching**: In this matching technique there exists a pre-recorded template of a particular word according to which the search engine matches the input signal. The processing that this technique takes is less in comparison to sub word matching. A disadvantage that this technique has is that we need to record each and every word that is to be recognized beforehand in order for the system to recognize it and thus it can only be used when we know the vocabulary of recognition beforehand. Also these templates need storage that ranges from 50 bytes to 512 bytes per word which very large as compared to sub word matching technique (Abdulla, et al, 2003).

2.7.6 PERFORMANCE ASSESSMENT OF SPEECH RECOGNITION

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

Word Error Rate (WER): Word error rate is a common metric of the performance of a speech recognition or machine translation system (Anusuya and Katti 2009). The general difficulty of measuring performance lies in the fact that the recognized word sequence can have a different length from the reference word sequence. The WER is derived from the Levenshtein distance, working at the word level instead of the phoneme level. This problem is solved by first aligning the recognized word sequence with the reference (spoken) word sequence using dynamic string alignment.

Word error rate can then be computed as

$$WER = (S+D+I)/N$$

S is the number of substitutions,

D is the number of the deletions,

I is the number of the insertions,

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

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

$$WRR = 1 - WER = 1 - (S+D+I) / N \ldots\ldots\ldots\ldots \quad (2.4)$$
$$= (H - I) / N$$
Where $H = (N-S-D)$ is the correctly recognized words.

2.7.7 APPLICATION AREA OF SPEECH RECOGNITION

Table 2.1: Application area of Speech Recognition

AREA OF APPLICATION	USES
Speech/Telephone/ Communication Sector/Recognition	Telephone directory enquiry without operator assistance
Education area	Teaching students of foreign languages to pronounce Vocabulary correctly. Teaching overseas students to Pronounce English correctly.
Artificial Intelligence aspect	Robotics
Physically Handicapped	Useful to the people with limited mobility in their arms and hands or for those without sight
Health	Health care, Medical Transcriptions (digital speech to text).

2.8 REVIEWED WORK

In 2016 there was a work on Voice Controlled Home Automation by Akshay *et al* (2016)
With the rapid increase in the number of elderly ones in the society and the handicapped, there is great need for speech control systems. This project is about controlling home appliances with the aid of speech. The automation recognizes voice commands given by the user and transfers it to a microcontroller which detects the voice command and proceeds with the switching accordingly. Raspberry Pi microcontroller module & Bluetooth module HC05 was used to implement the vision. The home automation system is intended to control all lights and other electrical appliances in a home or office using voice commands.

A voice control wheel chair project was done by Bhargavi *et al* (2015), to help the physically challenged and old people who face many problems in life regarding to mobility and enable them not to be dependent relatives or individuals in the society to be moving them from one place to another. The invention of the wheel chair is of a great help to them but it still restricts their movement. Voice controlled wheel chair was made using HM2007 voice recognition kit.

In 2015 also, a project work was implemented using Voice Controlled Wheel Chair Using Arduino by Karka *et al* (2015). The project is to help the physically challenged people and the elderly people who need the assistance of others to help them in moving from one spot to another. The project makes use of Arduino Uno board and other components. Voice Recognition Kit (HM2007 Module) is being used to recognize the voice command. The voice command given is converted to binary numbers by Voice Recognition Kit and those binary data is given to the arduino board for the control of the wheel chair. For example when the user says "forward" then chair will move in Forward direction and when he says "Backward" than the chair will move in backward direction and similarly for left, right and stop. We have used LCD display unit to display the direction in which direction of the wheel chair.

In 2008 there was a work on Robot Control with Voice Command by Muhammed, (2008).
The project was on robot control with voice command. The system developed in this project is a robot controlled with the speech commands. Speech commands are taken by a microphone. The features of the commands are extracted with MFCC algorithm. The commands are recognized using Neural Network. The recognized command converted to the form in which the robot can recognize. The final form of the commands is sent to the robot and the robot move accordingly. The set of command the robot was trained to recognize are; Move Forward, Move Backward, Turn Right, Turn Left, Stop and Continue.

In 2007 there was a work on a project that is SOPC-based Voiceprint Identification System which was done by Huan *et al* (2007). It is an efficient means of recognition compare to other forms of biometric means of identification. Voiceprint, as a basic human physiological characteristics, possess a unique role which is difficult to manipulate, imitate and replace. Voice authentication refers to the process of accepting or rejecting the identity claim of a speaker on the basis of individual information present in the speech waveform. It has received increasing attention over the past two decades, as a convenient, user-friendly way of replacing (or supplementing) standard password-type matching. The authentication procedure requests from the user to pronounce a random sequence of digits. After capturing speech and extracting voice features, individual voice characteristics are generated by registration algorithm. The central process unit decides whether the received features match the stored voiceprint of the customer who claims to be, and accordingly grants authentication.

In 2004 there was a work on a Voice Activated Door Control System which allow the opening of home doors and office doors by the use of voice control which was implemented by Emerson *et al* (2004). Words are programmed and stored on the microprocessor of the voice recognition circuit and while the circuit is powered on it constantly listens for external commands, when a command is recognized it sends an appropriate signal to its output port to activate attached interface cards and/or devices. The components used include the following; Voice Recognition Circuit, Relay Interface Circuit (SPDT Switches), H-Bridge DC Motor, Control Circuit, 6 Speed Gear System & DC Motor, SN74LS86 Quad 2-Input Exclusive OR Gate and LM555 Precision Timer - 8 Pin, DIP (Used in conjunction with the H Bridge, for motor timing configuration

CHAPTER THREE

DESIGN METHODOLOGY

3.1 PREAMBLE

The project is on the designing and development of water closet toilet that will be flushed through speech recognition system. In the development and implementation of the project several steps will be taking into consideration. This project will be implemented with the aid of a MATLAB software to produce the speech recognition system which comprises of acquisition stage, recognition stage with the database. Mel Frequency Cepstral Coefficient (MFCC) is used in which FFT Fast Fourier Transform is used in the speech signal analysis, for the speech recognition. Next is the hardware development, making use of PIC16F628A microcontroller which receives signal from the speech recognition system developed in MATLAB and controls the transistor for the flushing of the water closet.

3.2 Analyzing the System

The flushing of water closet toilet system based on speech recognition is built using MATLAB and implemented using a hardware model.

The schematic diagram for the water closet toilet flushed through speech recognition.

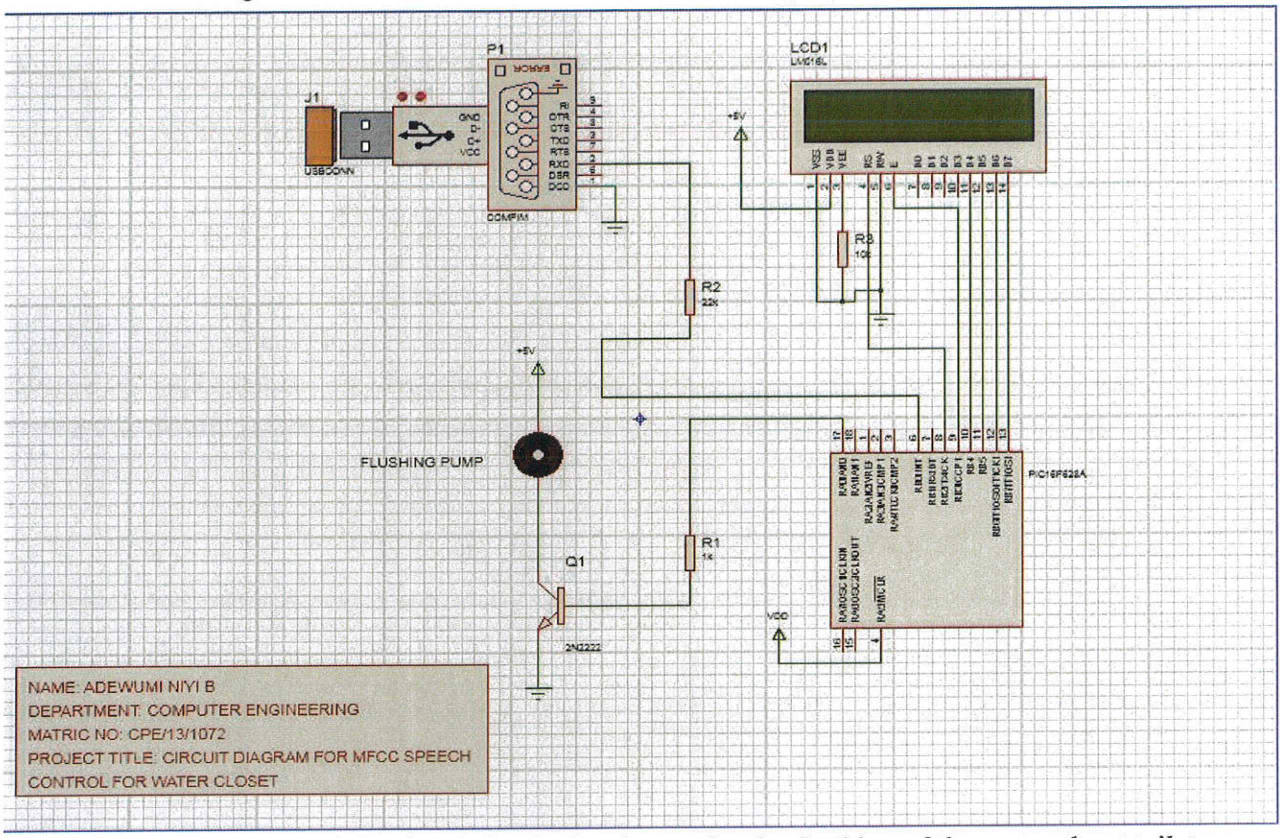


Figure 3.1: Schematics of the circuit for the hardware, for the flushing of the water closet toilet.

From the schematic diagram above which shows the hardware with PIC16F628A microcontroller which receives an access signal from the MATLAB after the recognition stage of the MATLAB through a COM port (serial program cable). Here, after recognition of the speech by the speech recognition system created with MATLAB, an access signal is sent through the COM port to the PIC16F628A microcontroller which controls the opening and closing of the model water closet toilet through the Transistor (2N2222A) in which the DC water pump mechanism works with the DC Motor to control the water chamber for flushing. The LCD displays the status of the system operation.

3.3.1 THE OPERATION AND THE SYSTEM BLOCK DIAGRAM OF THE PROJECT

The development and operation of the system is described below;

Stages in speech recognition are basically two stages which are the speech training stage and the speech verification and testing (recognition stage). In the training stage we have the following procedures; data acquisition stage (Speech input), the computer with the aid of a Matlab will perform the following operation for the first stage in processing the speech; analog to digital conversion, preprocessing and filtering (Ashish and Amit 2007), feature extraction using Mel Frequency Ceptral Coefficient in speech processing, the speech signal will 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.

FFT (Fast Fourier Transform) will be used in this project to convert the word signal into spectrum (*Hossan, 2010*). The Next is the verification and testing (recognition stage) in which it undergoes the following steps with the aid of a Matlab on the computer system; acquiring the speech signal, performing the feature extraction using MFCC algorithm, reference model with speech identification to determine similarity with what is in the database, decision making, verification of result for acceptance or rejection in order to drive the hardware operation. The final stage is the hardware development, making use of a microcontroller which is interfaced with the program developed in MATLAB which controls the operation and the Liquid Crystal Display (LCD) that shows the state of operation.

3.3.2 System Block Diagram

The stages in speech recognition are described below in the figure 3.7, the computer with the aid of a Matlab will perform the following operation for the first stage in processing the speech; analog to digital conversion, preprocessing and filtering. The Next is the verification by performing the feature extraction using MFCC algorithm, reference model with speech identification to determine similarity with what is in the data base, decision making and verification of result for acceptance or rejection in order to drive the hardware operation. The final stage is the hardware development, making use of a microcontroller which is interfaced with the program developed in MATLAB which controls the operation and the LCD that display the state of operation.

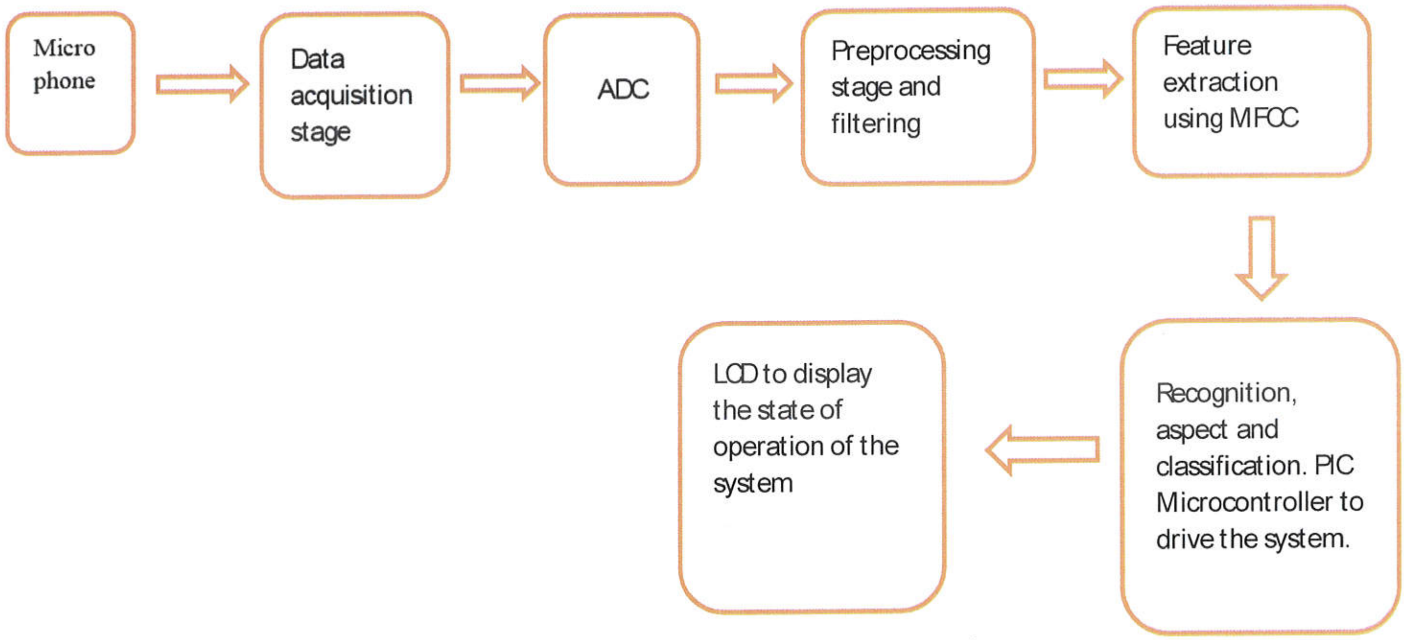


Figure 3.2: A block diagram of the system operation.

3.3.3 Flow Chart Diagram for the Speech Recognition

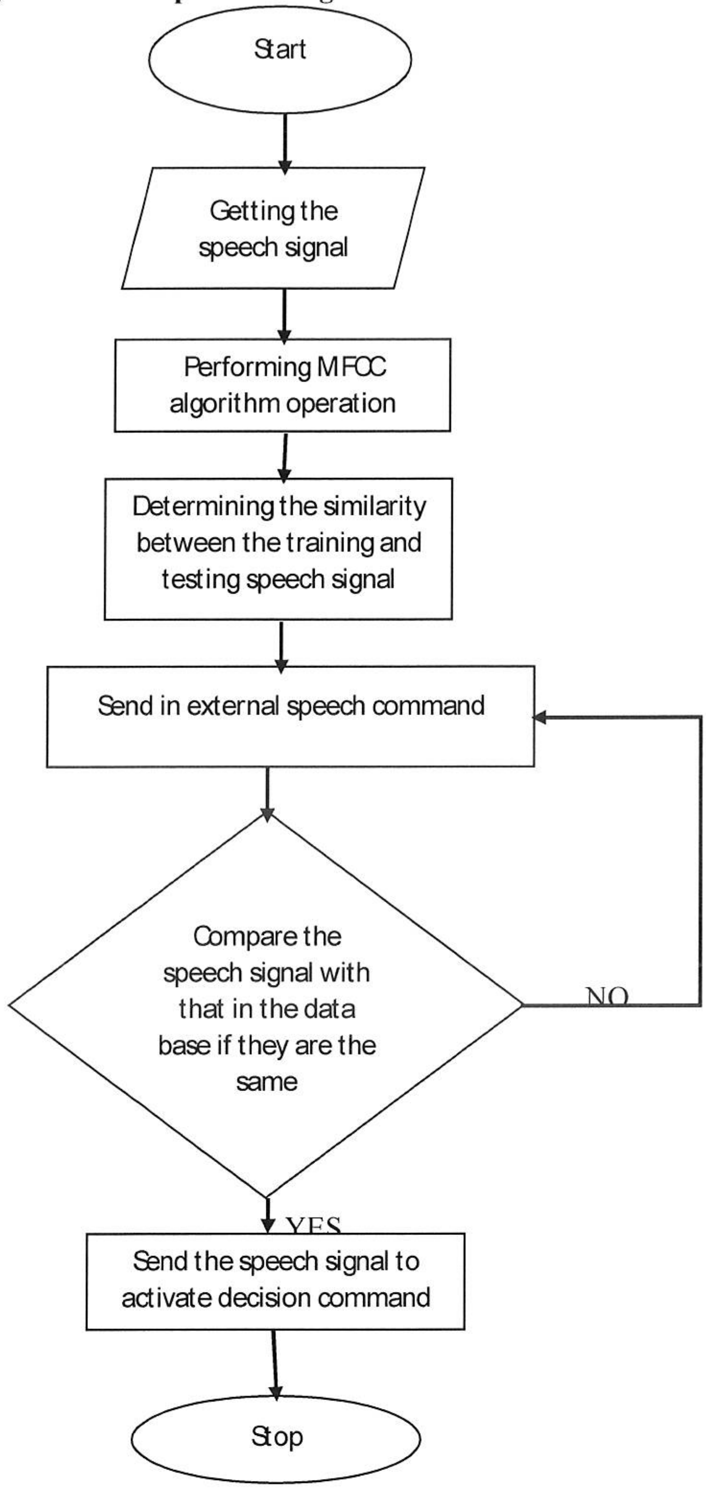


Figure 3.3: Flow chart diagram for the speech recognition

3.4 Steps for signal acquisition and processing.

- Recording the word signal in the audio format by the recorder.
- The audio format is being converted into .wav format by converter.
- The sampling frequency was converted to 8 kHz of word signal.
- The .wav file will be read in Matlab.
- The file is used for analysis through Fast Fourier Transform (FFT).

3.5 The basic operation that follows in the system implementation are;

- Testing the speech recognition system
- Setting up and testing the transistor, the DC water pump which works with the DC motor.
- Assembling the model for the water closet toilet.
- Interfacing all the various components and testing the overall system.

3.6 COMPONENTS THEORY

The materials used in the development and implementation of a water closet system are listed below;

- MATLAB software for development of speech recognition system
- Microcontroller is (PIC16F628A)
- Microphone
- LCD display
- 5v DC Submersible water pump for flushing
- 2N2222A Transistor
- Resistor (1k, 10k, 22k, 4.7k)
- USB Module Interfaced with RS232
- Crystal Oscillator (4 MHz)
- Capacitors (22pf)

3.6.1 The Microphone.

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 need to 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 current gain to drive the output. Furthermore, the boosted signal from preamplifier is given to a power amplifier where the current is amplified.

3.6.2 5V DC Submersible Water Pump

A submersible pump is a device which has a hermetically sealed motor close-coupled to the pump body. The whole assembly is submerged in the fluid to be pumped. The main advantage of this type of pump is that it prevents pump cavitation, a problem associated with a high elevation difference between pump and the fluid surface. Submersible pumps push fluid to the surface as opposed to jet pumps having to pull fluids. Submersible are more efficient than jet pumps. It uses a direct current of 5V in its operation. Its application in the system is to supply water for flushing of the water closet.

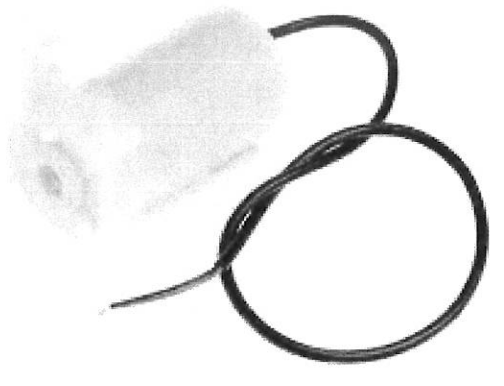


Figure 3.4: 5V DC Submersible Water Pump.

3.6.3 Liquid Crystal Display (LCD)

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

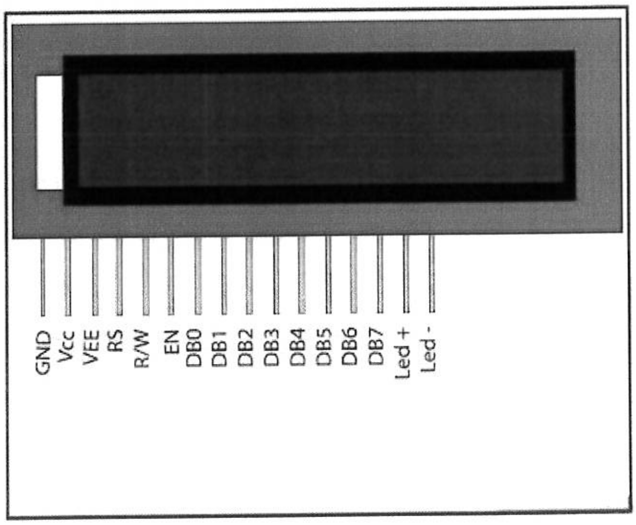


Figure 3.5: An LCD diagram (16X2)

3.6.4 Microcontroller (PIC16F628A)

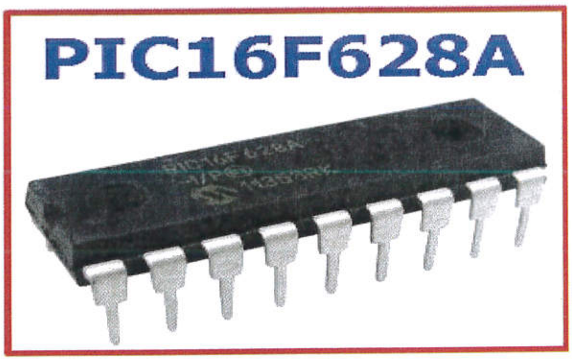


Figure 3.6: An image of PIC16F628A (Microchip Technology Inc., 2007)

The microcontroller (PIC16F628A) 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: Parametric description of the microcontroller.

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

18-Pin PDIP, SOIC

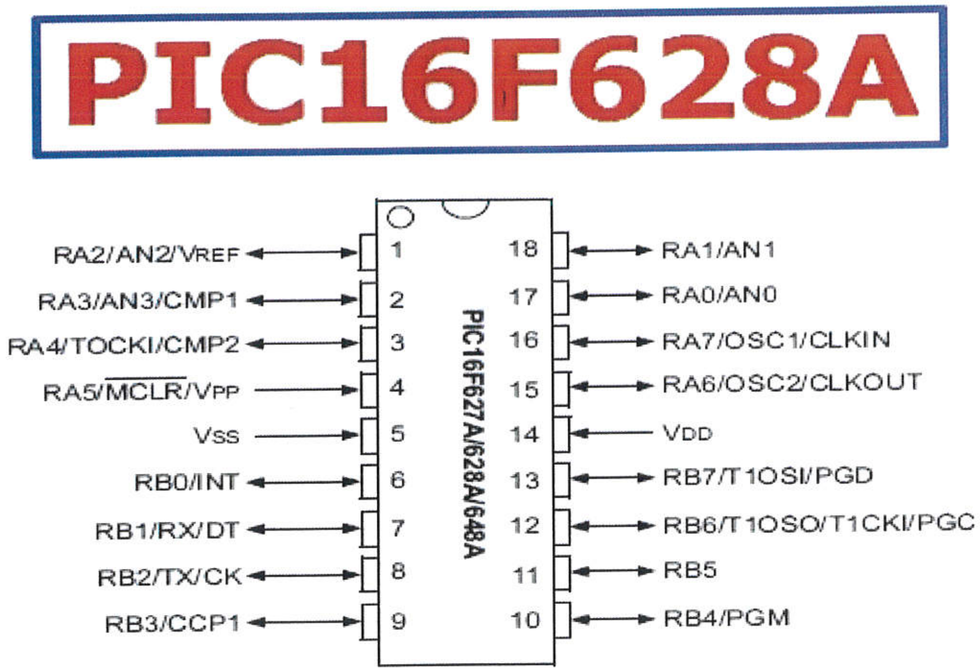


Figure 3.7: A diagram for Pin description of PIC16F628A (Microchip Technology Inc., 2007*)*

3.6.5 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.

Features:

Bi-polar high current NPN Transistor

DC Current Gain (h_{FE}) is 100

Continuous Collector current (I_C) is 800mA.

Emitter Base Voltage (V_{BE}) is 6V

Collector Emitter Voltage (V_{CE}) is 30V

Base Current (I_B) is 5mA maximum current

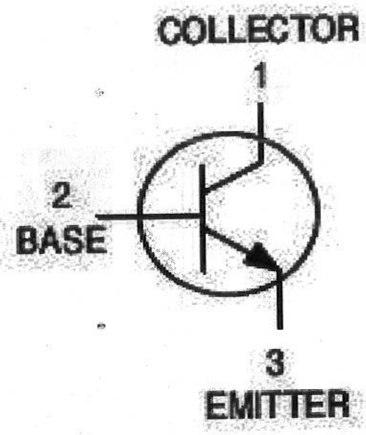


Figure 3.8: A description of 2N2222A NPN Transistor.

Table 3.2: 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

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

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

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

R = Resistance

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

$I_d = Resistor\ desire\ current$

$V_d = Desire\ voltage$

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

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

R = Resistance

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

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

$R_C = Collector\ resistance$

CHAPTER FOUR
IMPLEMENTATION AND RESULTS

4.1 PREAMBLE

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

4.2 DESIGN OF THE SPEECH SYSTEM DATABASE

The speech system database was designed using a sequentially structural method whereby each section leads to another in the whole speech program. The database was done using the MATLAB software (MATLAB 2007). The speech system code was written in MATLAB, has some sections which are listed below and these sections are backed up with their individual code sections too.

- Add a new speech from microphone
- Speech recognition from microphone
- Mel Frequency Cepstral Coefficient (MFCC) Representation
- Database info
- Delete Database
- Exit

The picture below is the graphical user interface (GUI) menu or section code and the execution result

Figure 4.1: The GUI for the speech recognition system

1. **Add new speech from microphone**

This is the first section of the speech recognition system where speech signals or data in the form of speech is added to the database of the system. This can also be said to be the training stage of the whole speech system where the system is trained with different individual's speech signals which are then stored in the speech database, and used as a reference in the recognition stage of the system. In this stage, during the acquisition of the speech signal, some parameters are to be inputted for good acquisition of speech signal, these parameter are; a class number (sound ID) that will be used to differentiate different speech in database and also will be used for recognition; the sampling frequency; bits per sample; duration of recording which is in seconds. There are some recommended values for these parameters listed above, sampling frequency (22050 recommended), sampling bits (8 recommended), and duration of recording in seconds (2 seconds recommended). After recording, the sounds are added to the database of the speech system.

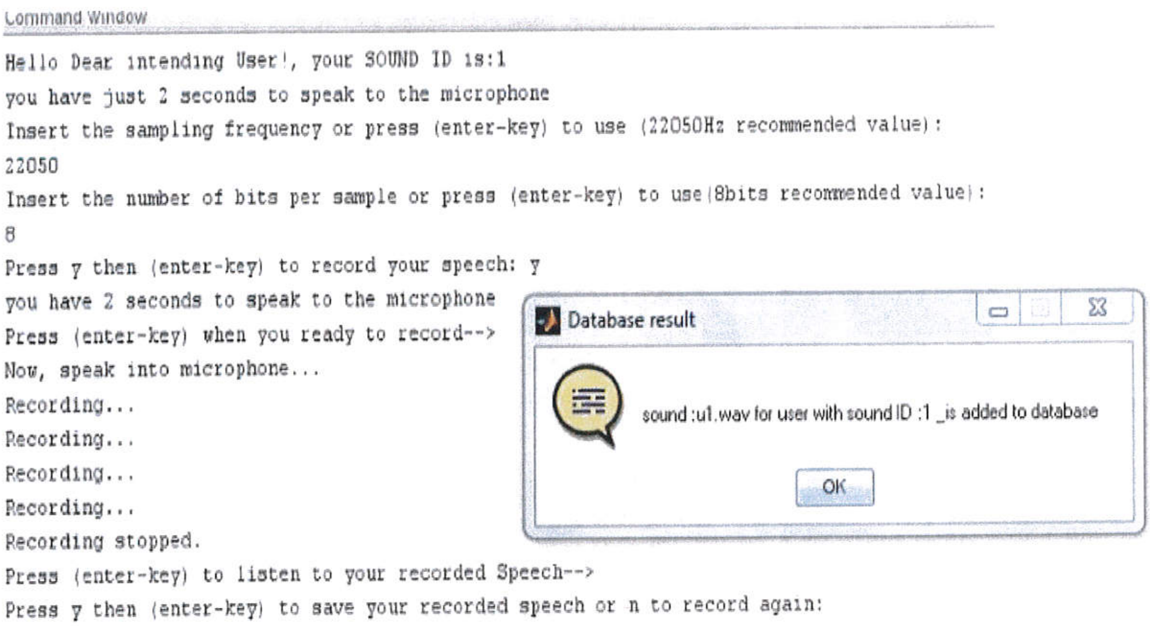
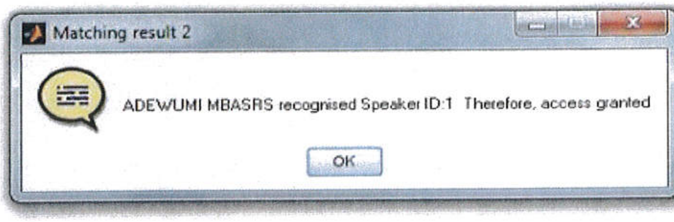


Figure 4.2: Adding speech sample to the database.

2. Speech Recognition from Microphone

This is the second section in the speech system database which is the core section of the whole project. This section controls the flushing of the water closet as it recognizes the sound ID with flush. During the recognition stage of the speech system, speech signals entered into the system are processed and compared to see if there is a match for the signal in the database of the system. The processing of these signals include the use of MFCC for feature extraction and VQLBG (vector quantization using LBG algorithm) for easy recognition of speech signals. In recognition, VQLBG was used instead of HMM, ANN, Gaussian method as it has faster processing and dynamic to changes in the use of different speech signals. VQLBG makes erasing and re-addition of speech signals very easy and faster unlike HMM or ANN which needs must training and is not easy to erase data, they are not as dynamic in change of data during erasing process as the system must be configured to be able to erase speech signals and also reprogram new speech signals countless number of times.


```
Recording...
Recording stopped.
Press enter to listen the recorded Speech-->
Press y to save or n to record again: y
MFCC cofficients computation and VQ codebook training in progress...
```




```
...
...
...
...
Completed.
For User #1 Dist :3.4735
For User #2 Dist :4.6668
For User #3 Dist :4.4148
For User #4 Dist :3.8222
    3.4735

Matching speech:
File:Microphone
Location:Microphone
ADEWUMI MBASRS recognised Speaker ID:1  Therefore, access granted
Opening Port....
```

Figure 4.3: Recognizing the required speech which is flush

3. Mel- Frequency Cepstral Coefficient (MFCC) Representation

MFCC mathematical equation is applied to the speaker speech to generate compressed data of the speech signal with a wave form representation of the speech signal. The zero crossing of the speech signal is done also with the wave form representation and the wave form of the speech signal is done after applying MFCC algorithm.

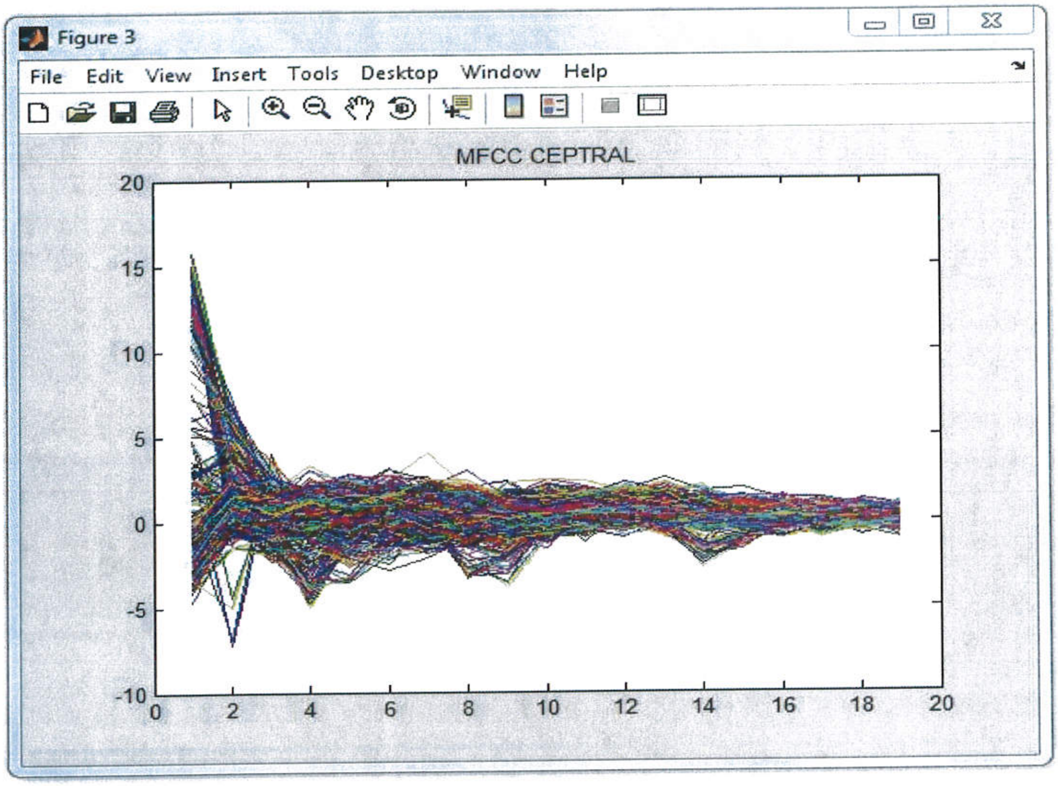


Figure 4.4: MFCC representation of the speech sample.

3 Database Information

It gives the information on the data (s) stored in the database, it is the section that holds the captured speech sound.

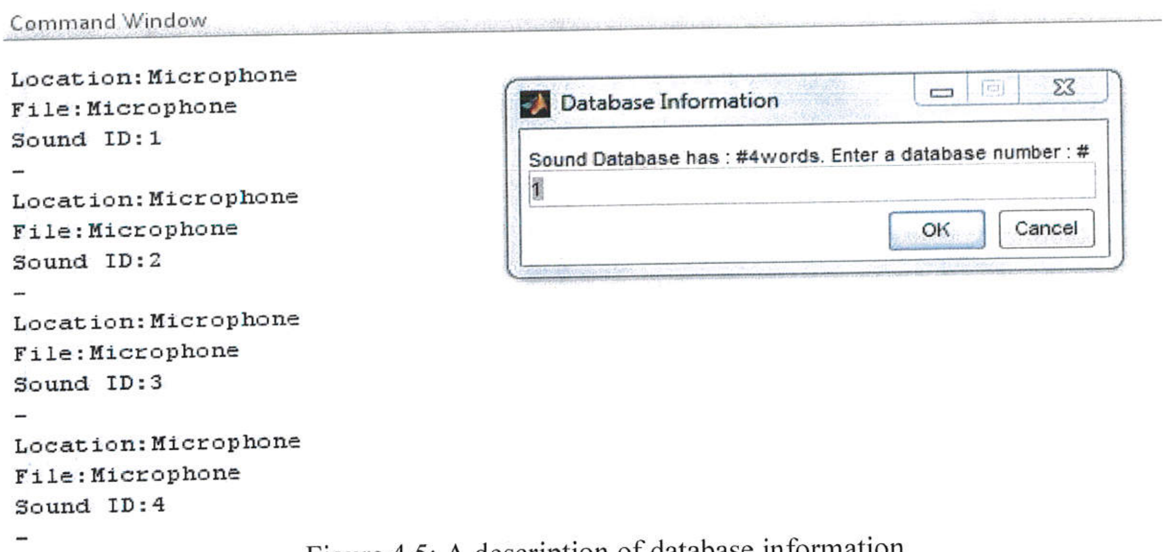


Figure 4.5: A description of database information

4 Delete Database

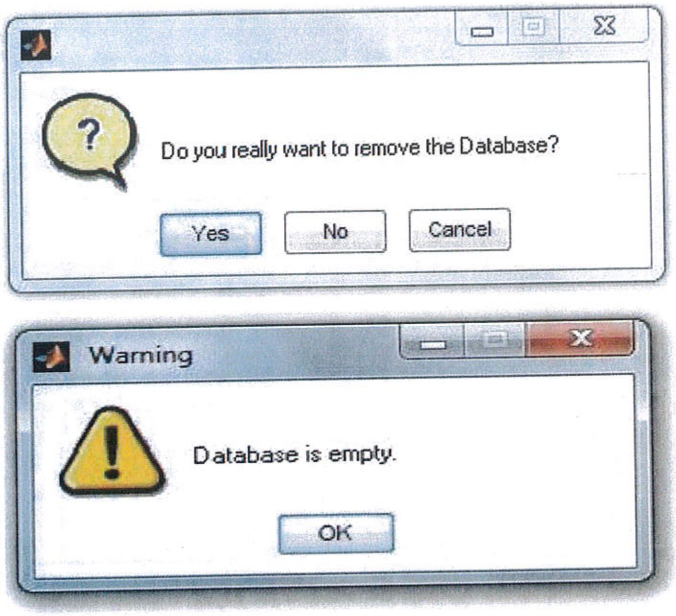


Figure 4.6: Deleting database

It is the section through which speech sound captured can be deleted from the database.

5 **Exit**

Figure 4.7: Exiting the interface

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

4.3 Various Functions used in the Speech Recognition System

There are various metrics used in the achievement of the project in different phase. They are Voice Recording function; MFCC function for feature extraction; and the VQLBG Vector quantization using the Linde-Buzo-Gray algorithm for recognition phase. The code for the function has been specified in the Appendix.

i. **MFCC Function;**

This function is used for accurate extraction of features of a given speech signal "s" having a sampling frequency "fs" and "c" which contains the transformed signal and the function MATLAB code is in the Appendix.

ii. **VQLBG Vector Quantization using the Linde-Buzo-Gray Algorithm;**

This function is used for recognition of features of a speech signal which relates to a stored feature of speech signal in the speech database. It basically checks the distance between parameters of the verifying speech signal features and matches it to a speech signal stored in the database based on the Euclidean distance closeness. The brief MATLAB code is specified in the Appendix as "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 IMPLEMENTATION AND RESULT EVALUATION OF THE SPEECH SYSTEM.

4.4.1 Software Execution and Result Discussion

1. Data acquisition and pre-processing stage

2. Feature extraction and data storage

3. Recognition using VQLBG algorithm

1. Data acquisition and pre-processing

This is the first section of the MATLAB program process, in which data is acquired by the use of analogue microphone using some recommended properties such as; Sound ID (to represent the speech signal), Duration of recording, Sampling frequency (22050Hz recommended) and Number of bits per sample (8bits recommended). Data acquisition is done under a conducive environment such as a silent room to prevent acquisition of unwanted signals called noise that may have effect on the speech system.

After data acquisition, the data is pre-processed by filters and converters which are embedded in the system used. Noise filter which removes unwanted signals from actual signals needed and ADC that converts the signal from analogue form received from microphone to digital process-able signal by computer.

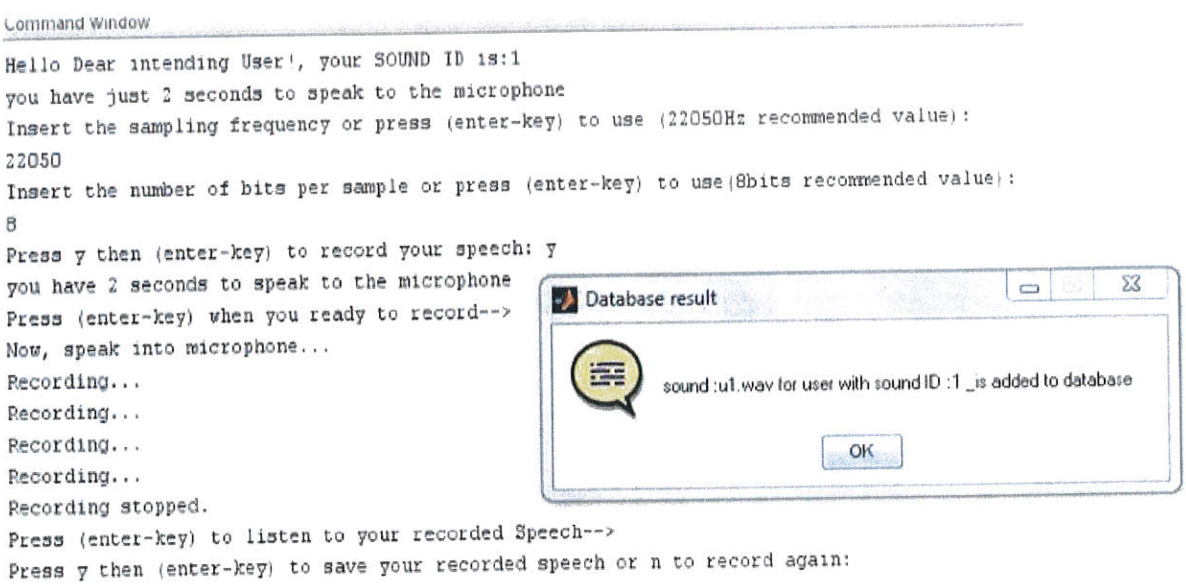


Figure 4.8: Adding speech sample to the database.

2. Feature extraction and Data storage

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

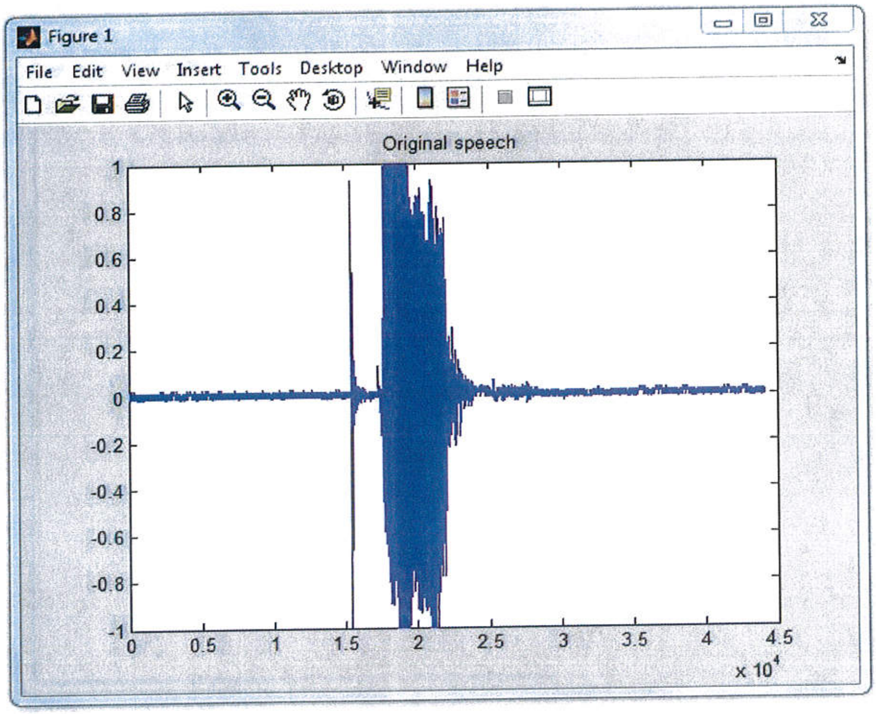


Figure 4.9: Original speech waveform

Figure 4.2 above shows the addition of speech signal into the database, speech signal with representation "Sound ID: 1", sampling frequency of 22050Hz, 8bits per sample, and 2seconds duration of recording. These value are taken constant for subsequent speech signals to be added to the database. The speech used in controlling the system is flush, once the system recognized this speech the flushing operation will be initiated by the hardware and water will be released for flushing.

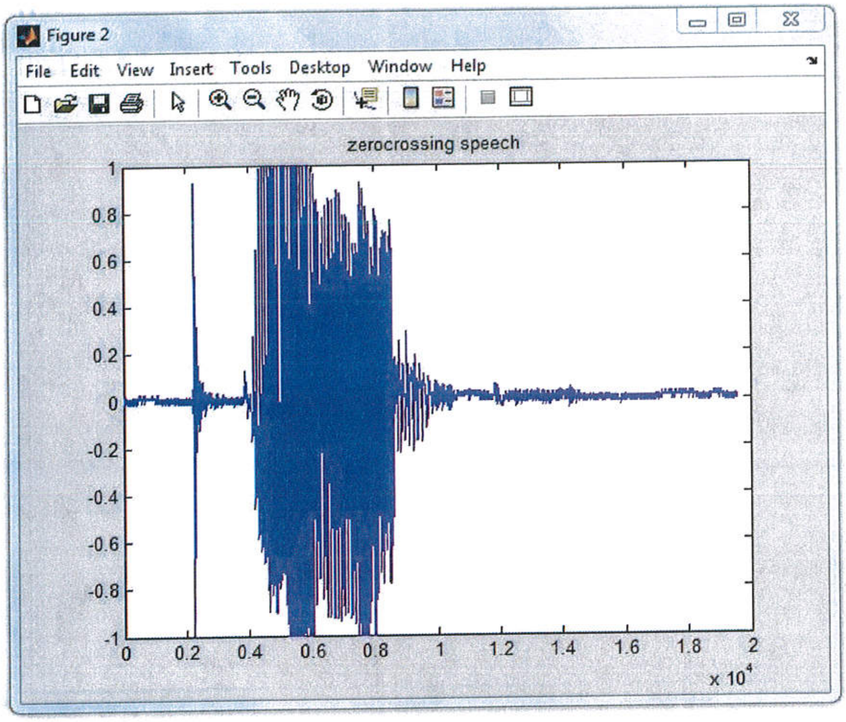


Figure 4.10: Zero crossing speech waveform

The zero crossing was used to remove silence in the speech captured which compensate the work of the MFCC algorithm in the speech processing.

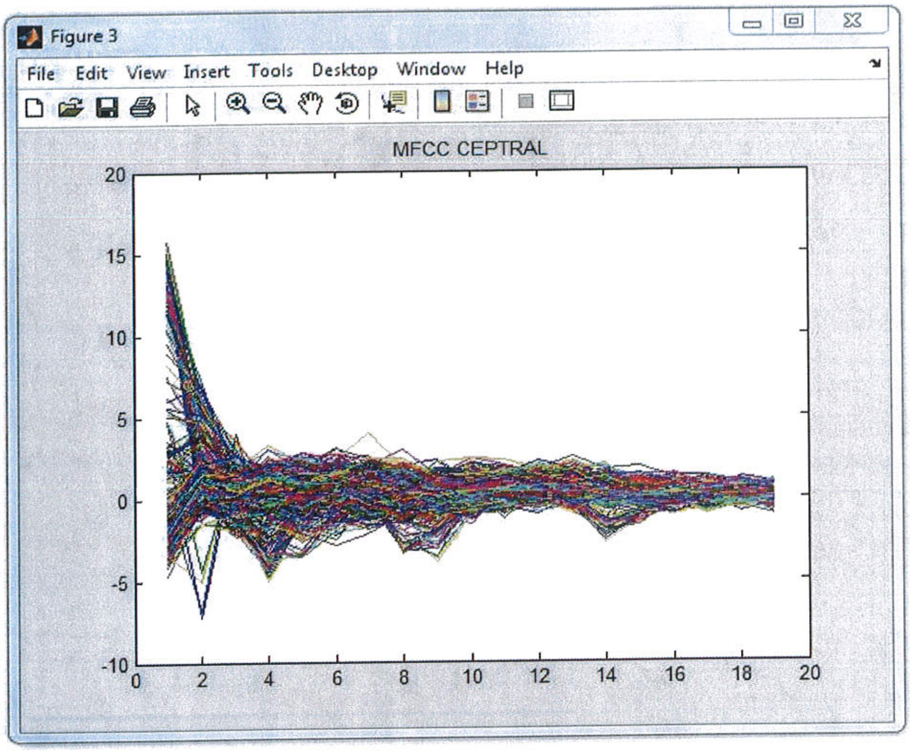


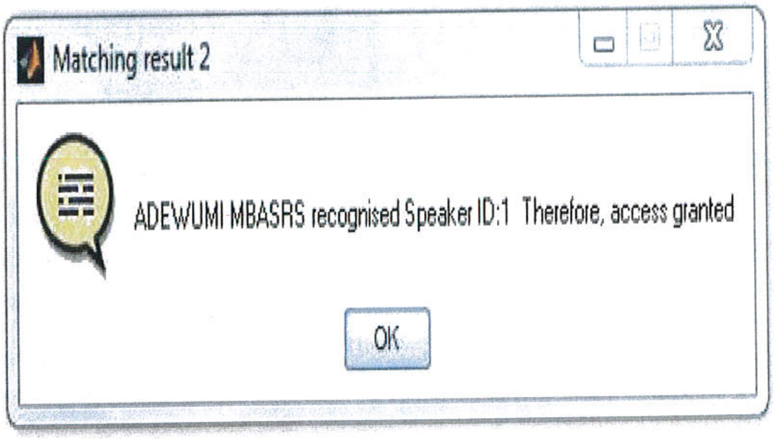
Figure 4.11: MFCC speech waveform

The figure 4.4 above shows the waveform of the speech signal after applying the MFCC algorithm.

3. Recognition using VQLBG algorithm

At this stage the controls for the flushing of the water closet is initialized as the system recognizes the sound ID with flush. During the recognition stage of the speech system, speech signals entered into the system are processed and compared to see if there is a match for the signal in the database of the system. The processing of these signals include the use of MFCC for feature extraction and VQLBG (vector quantization using LBG algorithm) for easy recognition of speech signals.

```
Completed.
For User #1 Dist :0.16892
For User #2 Dist :0.23687
For User #3 Dist :0.23105
For User #4 Dist :0.1784
    0.1689
```




```
Matching speech:
File:Microphone
Location:Microphone
ADEWUMI MBASRS recognised Speaker ID:1  Therefore, access granted
Opening Port....
Complete
```

Figure 4.12: The result after recognizing the sound ID with the speech flush

```
...
Completed.
For User #1 Dist :0.24984
For User #2 Dist :0.27563
For User #3 Dist :0.29918
For User #4 Dist :0.21531
    0.2153

Matching speech:
File:Microphone
Location:Microphone
ADEWUMI MBASRS guest you are Speaker ID:4  Therefore, access denied
```

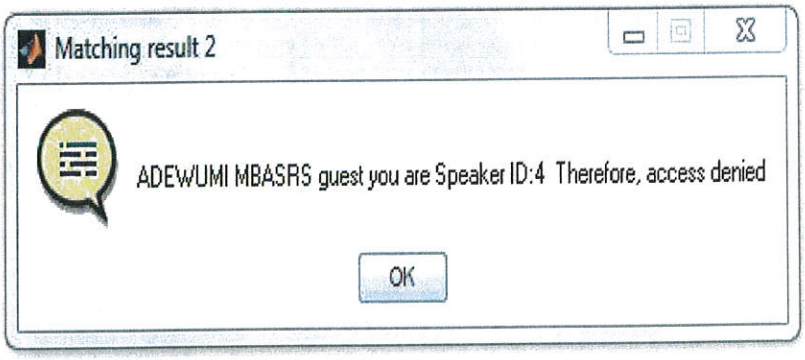


Figure 4.13: The result for not recognizing the sound ID with the speech flush

4.4.2 Parametric Value of the User Speech Signal.

Table 4.1 The MFCC coefficient of speech sample [Flush] of two speakers in a noisy environment

	No of Trials	Speech Sample: [FLUSH]
SPEAKER 1	**1**	0.4125 6.1458 3.1142
	2	0.3654 -1.2448 1.7147
	3	-0.8390 2.3153 4.1758
	4	2.8356 4.6006 1.8996
	5	-5.4210 0.1720 -0.6795
SPEAKER 2	**1**	3.9819 6.6636 3.2317
	2	-2.0607 2.4683 1.2470
	3	3.2699 3.0837 1.2162
	4	3.5252 -1.9557 0.6381
	5	-4.2162 -2.3676 1.1961

The table 4.1 above is the parametric value of the word flush used in controlling the system, the speech was collected in an environment with a background noise effect from two different speakers from the above vector features / co-efficient it depicts that the speech signal has been distorted by noise.

Table 4.2 **The MFCC coefficient of speech sample (Flush) of two speakers in a noiseless environment**

	No of Trials	Speech Sample: [FLUSH]
SPEAKER 1	1	-3.1827 3.5463 -1.9316
	2	-5.3972 -0.4870 -2.2592
	3	-5.5585 1.5663 -1.1051
	4	-5.4482 -4.8212 2.4374
	5	-4.0506 -3.8500 -8.3575
SPEAKER 2	1	-3.2616 -2.1540 -2.1315
	2	-3.8374 -0.2546 -2.1356
	3	-3.0832 3.3290 -2.1789
	4	-2.6759 -3.3360 -4.3766
	5	-3.9052 -1.7874 2.5250

The table 4.2 above is the parametric (vector co-efficient) representation for the word flush used in controlling the system, the speech was collected in a noiseless environment from two different speakers taking the variation in the two environment and speaker gives us the ability to analyze the speech signal and made it clear that there will be instability in the speech signal from the noisy environment compare to noiseless environment with little or no noise distortion.

Table 4.3 **Euclidean distance of speakers to speech sample**

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: [FLUSH]			
	Distance to User-1	**Distance to User-2**	**Distance to User-3**	**Distance to User-4**
Speaker-1	**0.16892**	**0.23687**	**0.23105**	**0.1784**
Speaker-2	**0.1571**	0.19536	0.16174	0.15938
Speaker-3	**0.19072**	0.26319	0.22305	0.2138
Speaker-4	**0.17816**	0.25686	0.21152	0.20253
Speaker-5	**0.19497**	0.24652	0.19603	0.20217

The table 4.3 above is the Euclidian distance for the speech sample for flush in relation with other speech sample in the database by calculating their Euclidian distance and matched it with the speech with flush which has the lowest distance, the figure highlighted in the table 4.3 has the Euclidean distance closed to the word speech in the database compared to other speech sample in database, the speech was collected from three different speakers. The system respond to the ech sample flush which has priority for the controlling of the hardware system.

le 4.4 **Coefficients of Speech Sample using MFCC, LPC and PLP**

ature Extraction Techniques	Speech Sample: [FLUSH]
MFCC	-5.5163 -0.5720 4.4282
LPC	6.2154 -0.0546 0.1966
PLP	-0.6153 -0.3913 -0.2641

le 4.4 above gives the vector coefficient of the speech signal (Flush) using Mel Frequency stral Coefficients feature extraction technique, Linear Predictive Coding and Perceptual Linear ictive Analysis.

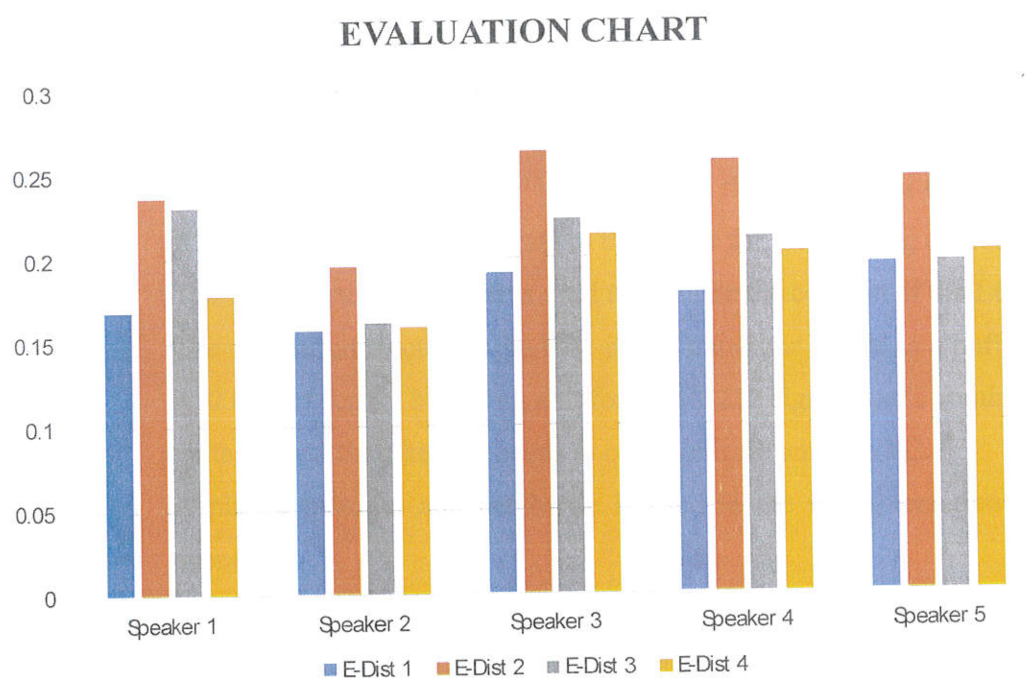


Figure 4.14: Evaluation chart for the Euclidian distance

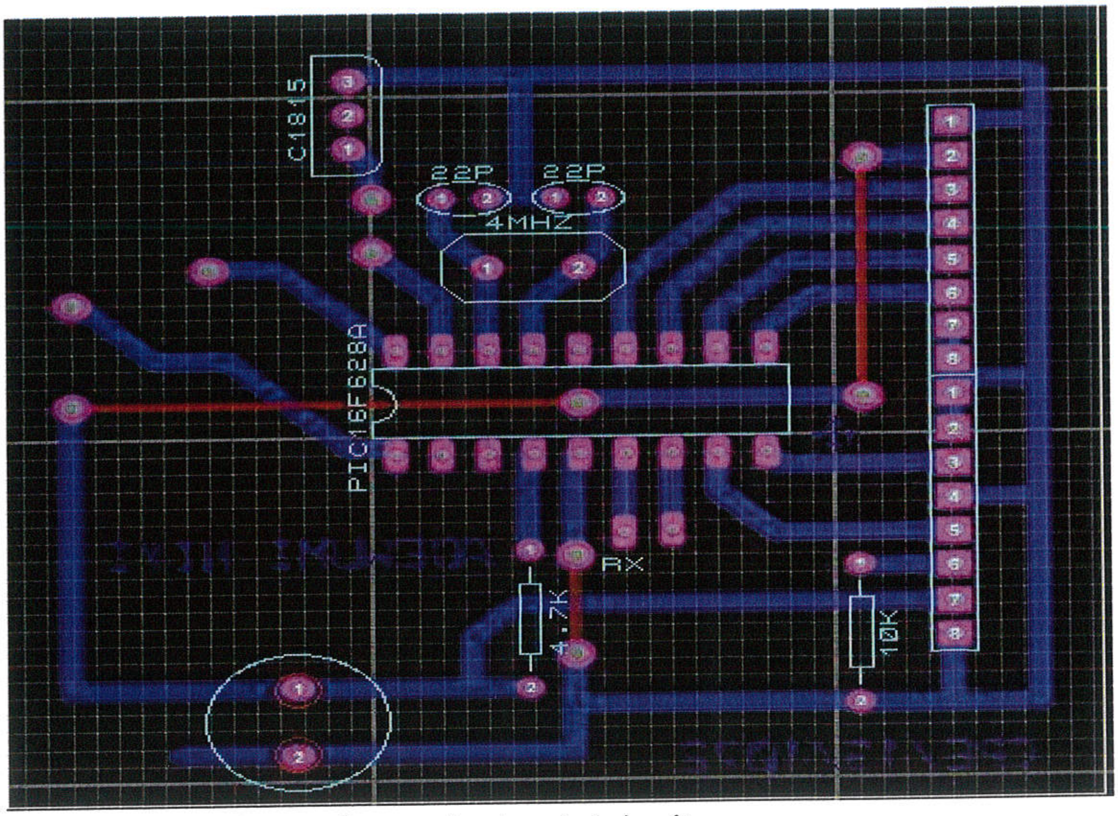


Figure 4.15: PCB layout of water closet control circuit

The PCB layout diagram of the water closet control circuit is shown above. A 5V DC is supplied into the circuit by a Computer System through 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 of the PIC16F628A is connected to the 2N2222A transistor which controls the DC motor in the opening and closing of the submersible water pump for flushing. 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.

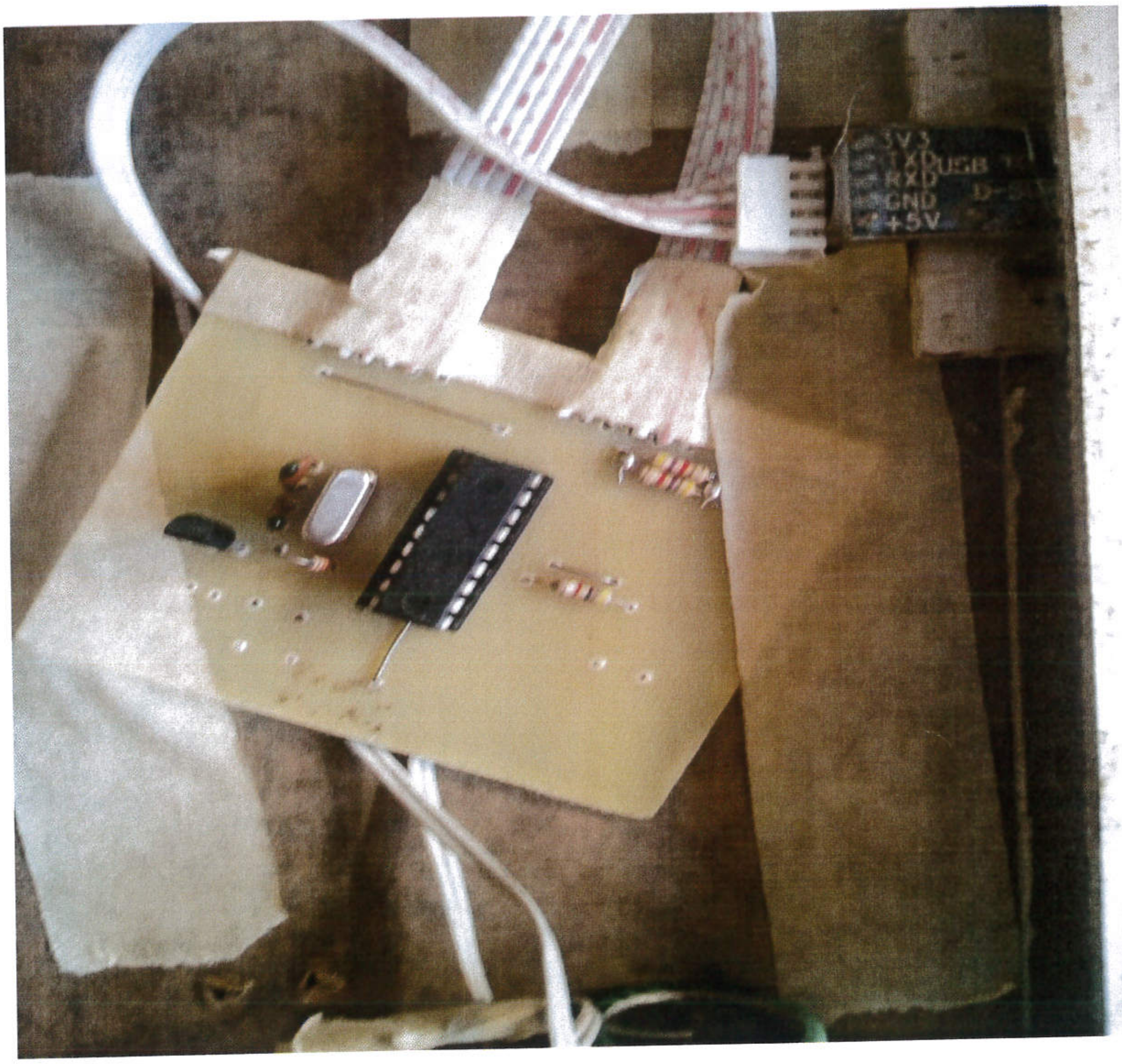


Figure 4.16: Internal arrangement of the prototype/model water closet circuit and components.

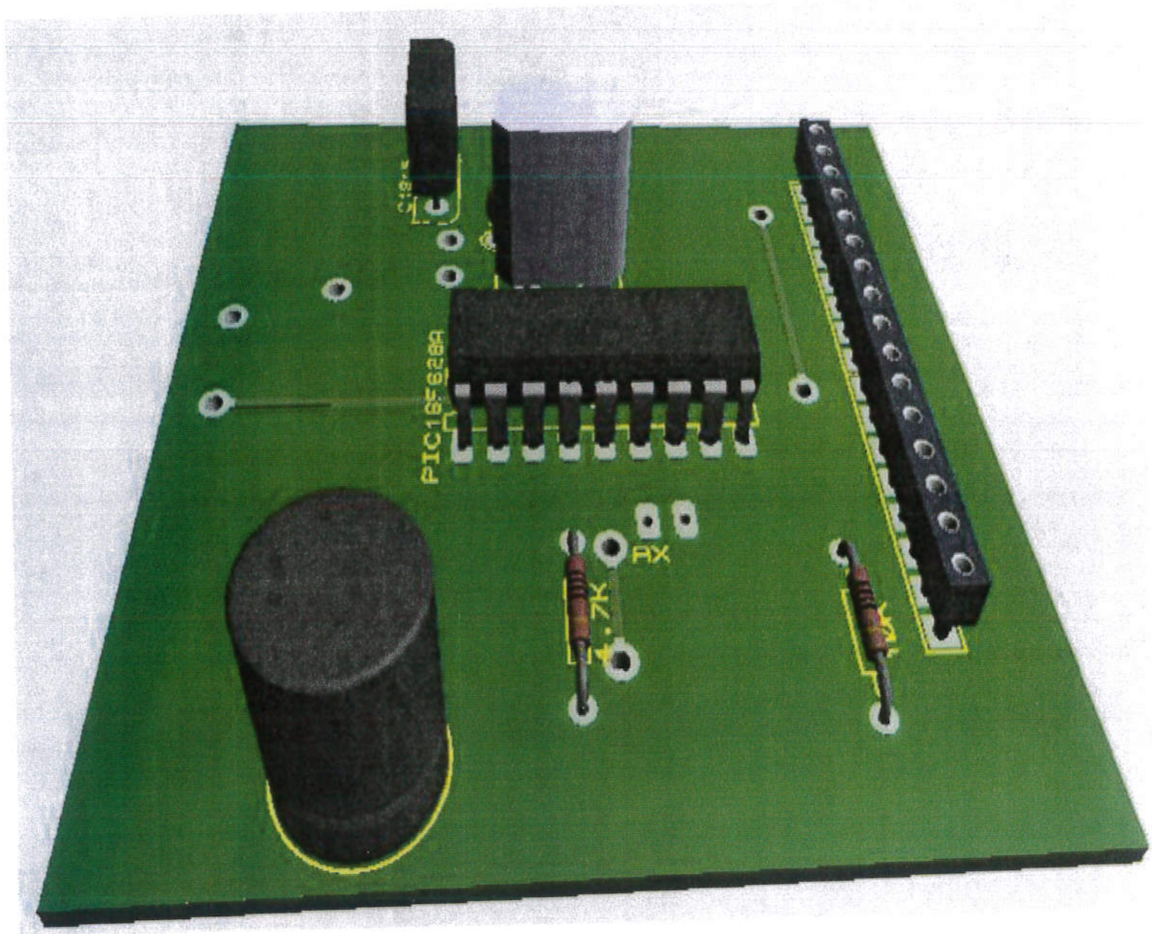


Figure 4.17: 3D Visualization of water closet circuit (Top View)

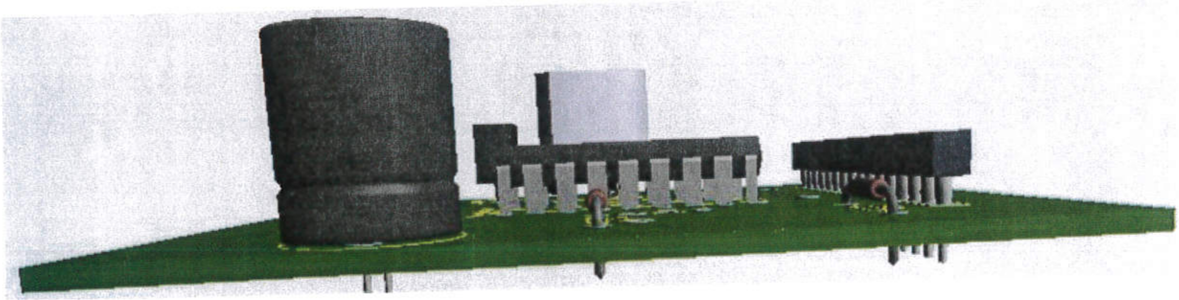


Figure 4.18: 3D Visualization of water closet circuit (Side View)

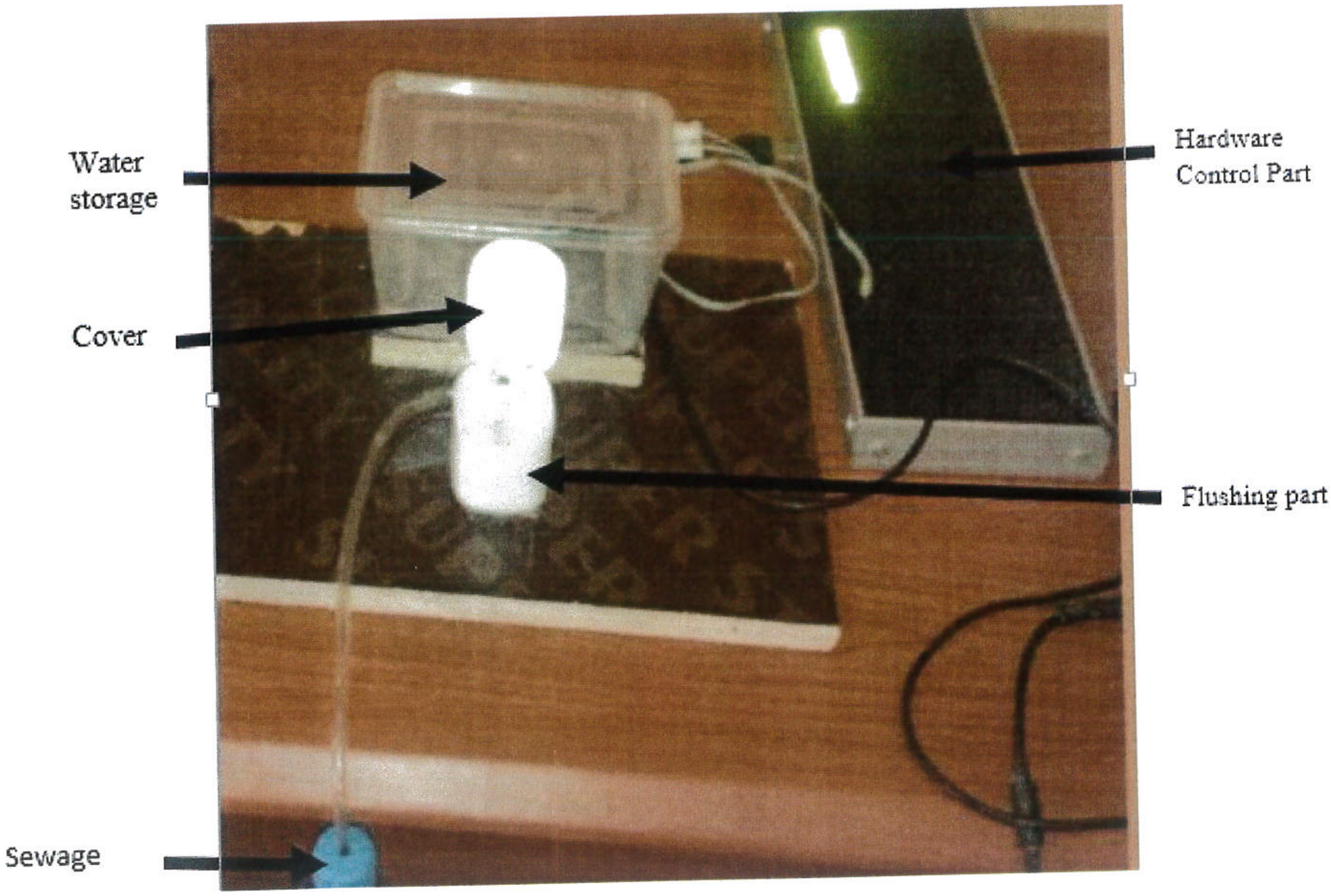


Figure 4.19: The prototype/model of the water closet controlled by speech recognition.

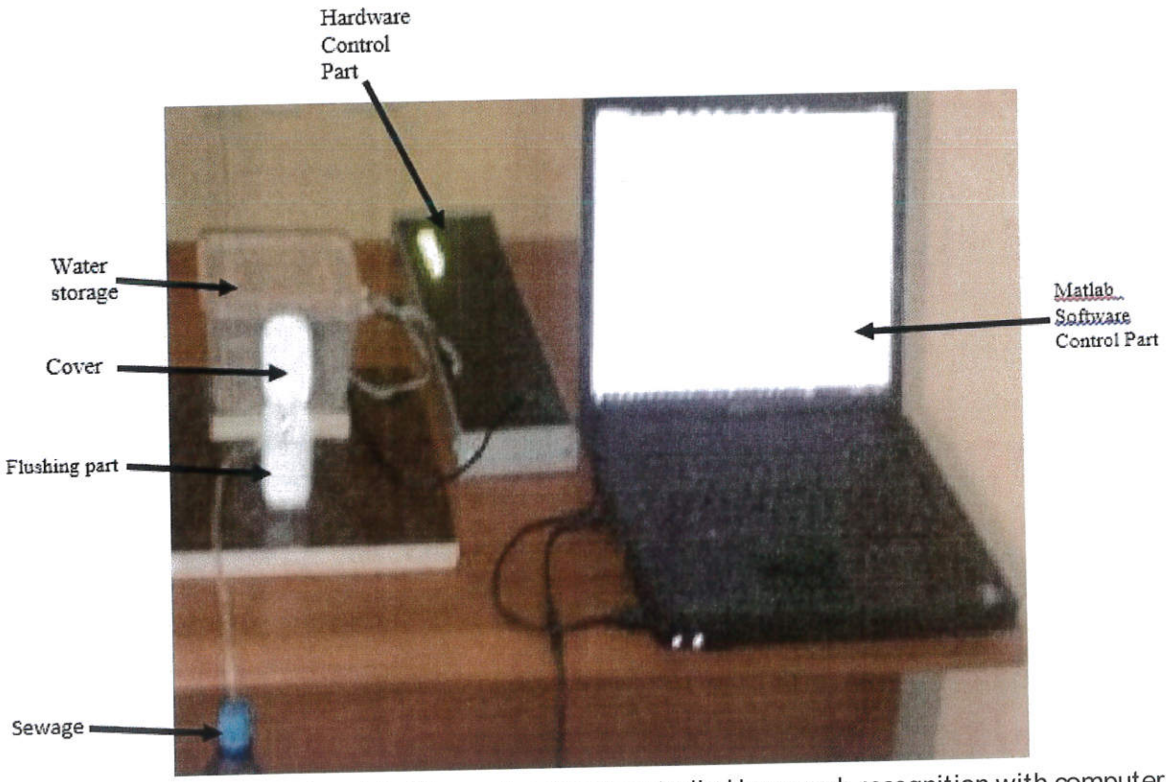


Figure 4.20: The prototype/ model of the water closet controlled by speech recognition with computer system.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATIONS

5.1 CONCLUSION

The system developed which is "speech recognition water closet system" is based on recognizing a known speech which is flush from a given set of registered speech in the system database. This project was accomplished using MATLAB software in creating the speech recognition system. In the first step of developing the speech recognition model, speech data acquisition and pre-processing was done, followed by the feature extraction using Mel Frequency Cepstral Coefficients. Next on recognition phase, we went for feature mapping using the vector Quantization using LBG algorithm. The results obtained using MFCC and VQ are appreciable. The code books were generated using LBG algorithm which optimizes the quantization process. Euclidean distance was measured and matched to the closest speech signal in the speech database. Finally, the functionality of the project was implemented and tested using a model water closet system built, which has PIC16F628A microcontroller that receives access signal from the speech recognition system design on MATLAB and controls the 2N2222A transistor which controls the flushing of the water closet system.

5.2 PROBLEMS ENCOUNTERED

Diverse problems were encountered during the project implementation. The problem include coding the system, implementing the project, modelling the hardware part and getting the exact components required for the system to work posed a great challenge . The major problem encountered are as follows:

1. The processing of the speech signal and its representation was challenging at the onset of the project execution.
2. There was problem in getting the exact component required in the implementation of the hardware part which result to the use of an alternative provision.
3. The construction of the flushing aspect of the system takes a lot of time and energy in which the parts was modelled several times just to suit the purpose of the project work.

4. There was problem in representing the different speech samples in the speech system database with characters like words such as exact names of the speaker. This problem was solved by using numbers to represent each speech sample.

5.2 RECOMMENDATIONS

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

1. The whole project can be made standalone by implementing it on a very high speed microcontroller or using a raspberry pie interfaced with some necessary component and an object oriented programming language can be used in the coding like python or the like of it which would cost more to achieve.
2. A system can be developed which the speech samples to be stored in the database will be represented with an exact name instead of sample ID.
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 system can be improved on by making its control to be speaker dependent, this will allow control over the use of the system.

REFERENCES

Abdulla, W. H., Chow, D., & Sin, G. (2003). Cross-words reference template for DTW-based speech recognition systems. In *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region* (Vol. 4, pp. 1576-1579). IEEE.

Akshay Mewada, Ayush Mishra, Manoj Gupta, Rahul Dash, Nilofer Mulla (2016), a research paper " Voice Controlled Home Automation" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016 ISSN: 2277 128X.

Anusuya M.A., Katti S.K. (2009) "Speech Recognition by Machine: A Review" *(IJCSIS)* International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009 Department of Computer Science and Engineering Sri Jaya chamarajendra College of Engineering Mysore, India.

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

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

C.S.Myers and L.R.Rabiner (1981), A Level Building Dynaimc Time Warping Algorithm for Connected Word Recognition, IEEE Trans. Acoustics, Speech Signal Proc., ASSP-29:284 297.

C.H.Lee and L.R.Rabiner (2004), A Frame Synchronous Network Search Algorithm for Connected Word Recognition , IEEE Trans. Acoustics, Speech, Signal Proc., 37(11):1649 1658.

Emerson Juarez, Son Le, Eva Okhankhuele Steve Rudenko (2004), a project titled, "Voice Activated Door Control System" at the University of Houston College of Technology.

Furui, S., Ichiba, T., Shinozaki, T., Whittaker, E. W., & Iwano, K. (2005). Cluster-based modeling for ubiquitous speech recognition. *Interspeech*, 2865-2868.

Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications, 10*(3), 16-24.

Ghai, W., & Singh, N. (2012). Literature review on automatic speech recognition. *International Journal of Computer Applications, 41*(8).

Gin-der Wu and Ying Lei (2008), "A register array based low power FFT processor for speech recognition. " *Journal of Information Science and Engineering.*

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

Huan Fang, Xin Liu, Ingo Sander, (2007). Paper named as " SOPC-based Voiceprint Identification System. "

Hermansky H , Hanson B. A., and H. Wakita, (1985) "Perceptually based linear predictive analysis of speech," *Proc. IEEE Int. Conf. on Acoustic, speech, and Signal Processing,"* pp. 509 512.

Juang, B. H., & Rabiner, L. R. (2005). Automatic speech recognition–A brief history of the technology development. *Encyclopedia of Language and Linguistics.*

King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., & Wester, M. (2007). Speech production knowledge in automatic speech recognition. *The Journal of the Acoustical Society of America, 121*(2), 723-742.

Kharka Bahadur Rai, Jeetendra Thakur, Nirmal Rai (2015) , "Voice Controlled Wheel Chair Using Arduino" International Journal of Science, Technology & Management, Volume No.04, Issue No. 06, June 2015, ISSN (online): 2394-1537

Klevans, R. L., & Rodman, R. D. (1997). *Voice recognition.* Artech House, Inc.

Koustav Chakraborty, Asmita Talele and Savitha Upadhya (2014) "Voice Recognition Using MFCC Algorithm" International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 10.

Lawrence Rabiner, Biing Hwang Juang, Fundamental of Speech Recognition, Copyright 1993 by AT&T.

Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi (2010), "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques" Journal of Computing, Volume 2, Issue 3, ISSN 2151-9617.

Muhammed Zahit ÖZDEM_RCAN (2008), "Robot Control with Voice Command" Yildiz Technical University Faculty of Electric and Electronics Department Of Computer Engineering.

Morales, N., Hansen, J. H., & Toledano, D. T. (2005, March). MFCC Compensation for Improved Recognition of Filtered and Band-Limited Speech. In *ICASSP (1)* (pp. 521-524).

Picone J. W., "*Signal modelling technique in speech recognition*," *Proc. Of the IEEE,* vol. 81, no.9, pp.1215-1247, Sep. 1993.

Rabiner L R and Junang B H (1986) "an Introduction to Hidden Markov Models". In *IEEE ASSP Magazine 4–16*. January 1986.

R.K.Moore (1994), twenty things we still don t know about speech, Proc.CRIM/ FORWISS Workshop on Progress and Prospects of speech Research and Technology.

Santosh K.Gaikwad and Pravin Yannawar (2010), A Review, International Journal of Computer Applications *A* Review on Speech Recognition Technique Volume 10– No.3.

Shanthi Therese Chelpa Lingam, International Journal of Scientific Engineering and Technology (ISSN: 2277-1581) a Review of Feature Extraction Techniques in Automatic Speech Recognition, Volume No.2, Issue No.6, pp : 479-484 1 June 2013.

Shikha Gupta, Mr.Amit Pathak, Mr.Achal Saraf (2014) "a study on speech recognition system" International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 8.

Shaikh Naziya S. *and* Deshmukh R.R. (2016) "Speech Recognition System" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 4, Ver. II, PP 01-09

Shreya Narang and Divya Gupta (2015) CSE Dept., Amity University, Noida, India, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015, pg. 107 114

Urmila Shrawankar and Dr. Vilas Thakare, (2010), "*Techniques for Feature Extraction in Speech Recognition System: A Comparative Study",* International Journal of Computer Applications in Engineering, Technology and Sciences (IJCAETS), ISSN 0974-3596, pp. 412-418

Weibel A, (1989), Phoneme recognition using time-delay neural networks , IEEE Trans. Acoustics, Speech, Signal Proc.,37,pp.393-404.

Wiqas Ghai, Navdeep Singh (2012), " Automatic Speech Recognition", International Journal of Computer Applications (0975 – 8887, Volume 41– No.8

APPENDIX A

Program Code in MATLAB for speech recognition using MFCC

Speech Recognition

```
%% Project: Speech Recognition and Identification system BY
% ADEWUMI ADENIYI B., SUPERVISED BY ENGR. N.S. OKOMBA.DEPARTMENT OF COMPUTER
ENGINEEERING, FEDERAL
% UNIVERSITY, OYE-EKITI, NIGERIA
%-----------------------------------------------------------------------

%% Main Function Speech Recognition
function []=speechrecognitionmfcc()
% For clear screen
clc;
% min_distance is a variable used to directly set the minimum distance for
% speech recognition.
min_distance=10;
%
% u1, u2, u3, u4, will be used for filenames related purposes reducing
% code redundancy.
char st;
disp('Project: Microcontroller Based Speech Recognition Identification
system');
disp('Electronic Water Closet Control System By ADEWUMI ADENIYI B. ');
disp(' ');
pause(0.5);
disp('LOADING ');


        disp('> ADEWUMI MFCC SPEECH RECOGNITION SYSTEM')
        pause(2);
        chos=0;
        possibility=6;
        while chos~=possibility,
           chos=menu('Speaker Recognition System','Add a new speech from
microphone','Speaker recognition from microphone','MFCC
Representation','Database Info','Delete database','Exit');

%% 10.1 Add a new sound from microphone

             if chos==1
                 clc;

                if (exist('sound_database.dat','file')==2)
                    load('sound_database.dat','-mat');
                        classe = sound_number+1;
message=strcat('Hello Dear intending User!, your SOUND ID
is:',num2str(classe));
                        disp(message);
    message=('The following parameters will be used during recording:');
                    disp(message);
```

```
                message=strcat('Sampling
frequency',num2str(samplingfrequency));
                disp(message);
                message=strcat('Bits per sample',num2str(samplingbits));
                disp(message);
               duration=2; %recording time in 2seconds
                option = 'n';
                option_rec = 'n';
                option_rec = input('Press y to record: ','s');
                if option_rec=='y'
                while option=='n',
                disp('you have 2 seconds to speak to the microphone');
                input('Press enter when ready to record--> ');
                micrecorder =
audiorecorder(samplingfrequency,samplingbits,1);
                disp('Now, speak into microphone...');
                pause(0.5);
                record(micrecorder,duration);


                while (isrecording(micrecorder)==1)
                    disp('Recording...');
                  pause(0.5);
                end
                disp('Recording stopped.');
                y1 = getaudiodata(micrecorder);
                y = getaudiodata(micrecorder, 'uint8');
                input('Press enter to listen the recorded Speech--> ');
                sound(y1,22050);
                pause(2);
           option = input('Press y to save or n to record again: ','s');
                end
                if size(y,2)==2
                    y=y(:,1);
                end
                y = double(y);
                sound_number = classe;
            data{sound_number,1} = y; %push user sample data into stack
             data{sound_number,2} = classe; %push user sound ID into stack
          data{sound_number,3} = 'Microphone'; %connect and save microphone
%signal into stack
                data{sound_number,4} = 'Microphone';
             st=strcat('u',num2str(sound_number)); %concatenate to string
                wavwrite(y1,samplingfrequency,samplingbits,st)
            save('sound_database.dat','data','sound_number','-append');
   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

%% Speech Recognition from microphone

          if chos==2
                         clc;
             if (exist('sound_database.dat','file')==2)
                load('sound_database.dat','-mat');
```

```
Fs = samplingfrequency;
  duration=2;
option = 'n';
option_rec = 'n';
option_rec = input('Press y to capture speech: ','s');
if option_rec=='y'
while option=='n',
 disp('you have 1.5seconds to speak to the microphone');
input('Press enter when ready to record--> ');
micrecorder = audiorecorder(samplingfrequency,samplingbits,1);
 disp('Now, speak into microphone...');
 pause(0.2);
 record(micrecorder,duration);
 while (isrecording(micrecorder)==1)
     disp('Recording...');
     pause(0.5);
 end
 disp('Recording stopped.');
 y1 = getaudiodata(micrecorder, 'uint8');
 y = getaudiodata(micrecorder, 'uint8');
 input('Press enter to listen the recorded Speech--> ');
 sound(y1,22050);
 pause(2);
option = input('Press y to save or n to record again: ','s');
 end
 wavwrite(y1,samplingfrequency,samplingbits,'v');

 if size(y,2)==2
     y=y(:,1);
 end
 y = double(y);
 %----- code for speech recognition -------
disp('MFCC cofficients computation and VQ codebook training in
progress...');
 disp(' ');
 % Number of centroids required
 k =16;
 for ii=1:sound_number
% Compute MFCC cofficients for each sound present in database
     v = mfcc(data{ii,1},Fs);
     % Train VQ codebook
     code{ii} = vqCodeBook(v, k);
     disp('...');
 end
 disp('Completed.');
 % Compute MFCC coefficients for input sound
 v = mfcc(y,Fs);
 % Current distance and sound ID initialization
 distmin = Inf;
 k1 = 0;
 for ii=1:sound_number
d = distance(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 < min_distance
            disp(distmin);
            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==1
            message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
            disp(message);
        msgbox(message,'Matching result 2','help');


    disp('Opening Port....')
SerPIC = serial('COM8');
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','E');pause(0.2)
 fclose(SerPIC);
delete(SerPIC)
clear SerPIC

                    ch3=0;
                  while ch3~=3
                        ch3=menu('Matched result
verification:','Recognized Spkr','Recorded spkr','Exit');

                        if ch3==1
                            st=strcat('u',num2str(speech_id));
                            [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
```

```
elseif speech_id==2
    message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
    disp(message);
    msgbox(message,'Matching result 2','help');
                        ch3=0;

    while ch3~=3
ch3=menu('Matched result verification:','Recognized Spkr','Recorded
spkr','Exit');

        if ch3==1
            st=strcat('u',num2str(speech_id));
            [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
            elseif speech_id==3
    message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
    disp(message);
    msgbox(message,'Matching result 2','help');
                        ch3=0;

    while ch3~=3
ch3=menu('Matched result verification:','Recognized Spkr','Recorded
spkr','Exit');

        if ch3==1
            st=strcat('u',num2str(speech_id));
            [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
            elseif speech_id==4
    message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
    disp(message);
    msgbox(message,'Matching result 2','help');
                        ch3=0;

    while ch3~=3
        ch3=menu('Matched result
verification:','Recognized Spkr','Recorded spkr','Exit');
```

```
                if ch3==1
                    st=strcat('u',num2str(speech_id));
                    [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
    elseif speech_id==5
                message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
                disp(message);
            msgbox(message,'Matching result 2','help');



            ch3=0;
            while ch3~=3
                ch3=menu('Matched result
verification:','Recognized Spkr','Recorded spkr','Exit');

                if ch3==1
                    st=strcat('u',num2str(speech_id));
                    [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
                                    elseif speech_id==6
                message = strcat('Speaker ID:
',num2str(speech_id),'_ is Recognised and Authenticated');
                disp(message);
            msgbox(message,'Matching result 2','help');
                                        ch3=0;

            while ch3~=3
                ch3=menu('Matched result
verification:','Recognized Spkr','Recorded spkr','Exit');

                if ch3==1
                    st=strcat('u',num2str(speech_id));
                    [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
            else
                warndlg('Database is empty.',' Warning ')
        end
    end
%% 10.3 MFCC Representation
    if chos==3
            if (exist('sound_database.dat','file')==2)
                load('sound_database.dat','-mat');
            st=strcat('u',num2str(sound_number));
                        [signal fs nb]=wavread(st);
                        [versig fs nb]=wavread('v');

silencesig=silenceremoval_endpoindet(fs,signal);
                        mfccsig=mfcc(signal,fs);
                         figure
                        plot(signal);
                        title('Original speech')
                         figure
                         plot(silencesig);
                         title('zerocrossing speech')
                         figure
                         plot(mfccsig);
                         title(' MFCC CEPTRAL')


                    else
                    warndlg('Database is empty.',' Warning ')
                end
    end         %-------------------------------------------------------------
-----
%% 10.4 Database Info
                        clc;
            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
        clc;
        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 Great!.from ADENIYI.','Thank You','help');
```

CODE FOR MFCC

```
function c = mfcc(s, fs)
% MFCC Calculate the mel frequencey cepstrum coefficients (MFCC) of a signal
%
% Inputs:
%       s       : speech signal
%       fs      : sample rate in Hz
%
% Outputs:
%       c       : MFCC output, each column contains the MFCC's for one speech
frame

N = 256;                             % frame size
M = 100;                             % inter frame distance
len = length(s);
numberOfFrames = 1 + floor((len - N)/double(M)); % for getting the number of
a frame in the signal
mat = zeros(N, numberOfFrames); % vector of frame vectors

for i=1:numberOfFrames % using loop to calculate the total number of the
frame in the sample
    index = 100*(i-1) + 1; % it saves the frame
    for j=1:N
        mat(j,i) = s(index);
        index = index + 1;
```

APPENDIX B

PROGRAM CODE FOR PIC16F628A 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 = 10

LCDOUT $FE,$80,"    SPEECH    "

LCDOUT $FE,$C0," RECOGNITION  "

pause  4000

high portb.1
```

```
LCDOUT $FE,$80," WATER CLOSET   "
LCDOUT $FE,$C0,"    CONTROL     "
pause  4000
low portb.1
D1:
LCDOUT $FE,$80,"   SYSTEM       "
LCDOUT $FE,$C0,"    READY       "
pause  3000
inicio:
serin portb.0,T2400,valor
if valor == "E" then high porta.0:gosub mss 'E->Encendido
if valor == "A" then low porta.0  'A->Apagado
goto inicio
mss:
LCDOUT $FE,$80,"  FLUSHING      "
LCDOUT $FE,$C0,"    TOILET      "
pause  10000
LCDOUT $FE,$80,"     DONE       "
LCDOUT $FE,$C0,"                "
low porta.0
PAUSE 1000
GOTO D1
END
```

APPENDIX C: BILL OF ENGINEERING MEASUREMENT AND EVALUATION.

S/N	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, 22K Ω	1	50	50
7	Resistor, 1K Ω	1	50	50
8	LCD(16X2)	1	900	900
9	Resistor, 4.7K Ω	1	50	50
10	Resistor, 10K Ω	1	50	50
11	4MHz Crystal Oscillator	1	400	400
12	RS232 Com Cable	1	1,500	1,500
13	Casing	1	3,000	3,000
14	Water Closet Model	1	2,000	2,000
15	Fiber Clad Board	1	1000	1000
16	Soldering Lead	4	100	400
17	Connector/Jumper wire	8	50	400
18	Male and Female USB Cord	1	400	400
19	USB Module	1	450	450
	TOTAL (₦)			**13,300.00 K**