Voice Machine Interfacing Technology

By

Mahmudul Hasan (Prince)
ID: 083-19-989

Rongon Deb Raju
ID: 083-19-983

Sumsul Arifin
ID: 093-19-1148

This Report Presented in Partial Fulfillment of the Requirements of the Degree of Bachelor of Science in Electronics and Telecommunication Engineering.

Supervised by

Dr. A.K.M. Fazlul Haque,

Associate Professor

Department Of

Electronics and Telecommunication Engineering

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

AUGUST, 2012

APPROVAL

This Project titled **"Voice Machine Interfacing Technology"** submitted By Mahmudul Hasan Prince, Rangon Deb Raju and Sumsul Arifin to the Department of Electronics and Telecommunication Engineering (ETE), Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Electronics and Telecommunication Engineering and approved as to its style and contents. The presentation has been held 27th August 2012.

BOARD OF EXAMINERS

Dr. Fayzur Rahman
Professor and Head
Department of,
Electronics and Telecommunication Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Dr. A.K.M. Fazlul Haque
Associate Professor
Department of,
Electronics and Telecommunication Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Md. Mirza Golam Rashed
Assistant Professor
Department of,
Electronics and Telecommunication Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Subrata Kumar Aditya
Professor and Chairman
Department of Applied Physics,
Electronics and Communication Engineering
University of Dhaka.

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. A.K.M. Fazlul Haque,** Associate Professor, Department Of Electronics and Telecommunication Engineering, Daffodil International University, Dhaka. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised By:

Dr. A.K.M. Fazlul Haque
Associate Professor
Department Of,
Electronics and Telecommunication Engineering
Daffodil International University

(Signature)

Mahmudul Hasan (Prince)
ID: 083-19-989
Department Of,
Electronics and Telecommunication Engineering
Daffodil International University

(Signature)

Rongon Deb Raju
ID: 083-19-983
Department Of,
Electronics and Telecommunication Engineering
Daffodil International University

(Signature)

Sumsul Arifin
ID: 093-19-1148
Department Of,
Electronics and Telecommunication Engineering
Daffodil International University

(Signature)

ACKNOWLEDGMENTS

While working on this project we have received many invaluable help form a large number of people. We would therefore like to take this opportunity to express our deepest gratitude to everyone has helped us.

First of all we would like to express our cordial gratefulness to Almighty ALLAH for HIS kindness, for which we successfully completed our project with in time.

We fell grateful to express our boundless honor and respect to our supervisor, **Dr. A.K.M. Fazlul Haque**, Associate Professor, Department Of Electronics and Telecommunication Engineering, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of signal processing influenced us to carry out of this project. His endless patient help, friendly support, great enthusiasm and extensive knowledge, which have guided us throughout our work and showed the path of achievement.

We would like to express our heartiest gratitude to **Dr. Fayzur Rhaman**, Professor and head, Department of Electronics and telecommunication Engineering, foe his kind help to finish our thesis and also to other faculty members, the staffs of the ETE Department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

And at last but not the least we must acknowledge with due respect the constant support and patience of our family members for completing this project.

ABASTRACT

This report aims to design and implement the voice recognition and machine interfacing for security. Here the working procedure has been divided into two parts: Software and Hardware part. In order to the system operational, MATLAB and ARDUINO microcontroller have been used for controlling software as well as a steeper motor. The proposed system identifies the authorized and unauthorized voice signal automatically. The projected algorithms have been used to compare the voice from the authorized person as input with the reference voice in data for finding accuracy. Finally, it is found that the accuracy based signal provides the faster speaker identification process.

TABLE OF CONTENT

List of Figures

List of Tables

List of Graphs

Chapter 1
Introduction

1.1. Backgrounds

The voice is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. The goal of this project is to interface between voice and machine. By using voice we can control a machine. We divide it into two phases, Software and Hardware part. In software part, we use Speech Recognition technique to develop for speech input to machine based on major advanced in statically modeling of speech. Here we develop speech recognition technique in Matlab software [1]. At the highest level, all speaker recognition systems contain two main modules feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker [3]. Feature matching involves the actual procedure to identify the unknown speaker by extracted features from his/her voice input with the ones from a set of known speakers. All Recognition systems have to serve two different phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, each registered speaker has to provide samples of their speech so that the system can build or train a reference model for that speaker. In case of speaker verification systems, in addition, a speaker-specific threshold is also computed from the training with stored reference model(s) and recognition decision is made. Speech recognition is a difficult task and it is still an active research area.

Automatic speech recognition works based on the premise that a person’s speech exhibits characteristics that are unique to the speaker. However this task has been challenged by the highly variant of input speech signals. The principle source of variance is the speaker himself. Speech signals in training and testing sessions can be greatly different due to many facts such as people voice change with time, health conditions (e.g. the speaker has a cold), speaking rates, etc. There are also other factors, beyond speaker variability, that present a challenge to speech recognition technology.

In hardware part we used microcontroller, steeper motor, USB cable, power supply etc. Here we used Arduino Microcontroller, to connect it with computer by a USB cable, to rotating steeper motor we use DC power 7.6 V. In our project we just showed only one implementation. Here we showed door controlling system by using steeper motor.

So, if we adjust our software and hardware part, we see that when our speech or voice is matched with the authorized person’s or training phase data then software will send a signal to Arduino microcontroller through the USB cable, Arduino also generate a signal to controlling the steeper motor. Actually this is the basic of the project, each module is discuss detail in later sections.

1.2. Aim of the Research Work

This paper aims to design and implement the voice recognition and machine interfacing for security. Here we use MATLAB (FFT) as software and for hardware interfacing we use ARDUINO microcontroller for controlling a steeper motor. Actually it’s a system which identify authorized and unauthorized voice signal automatically. Basically, this system consists of two parts which are training and recognition.. The task can also be described as distinguishing the speech from the authorized person and unauthorized person. On the other hand, the detection of the voice is a reliable prediction based on hypothesis of the environment [2]. A robust hypothesis can prevent the increasing level of noise deteriorating the recognition of the speech and guarantee the accuracy of the decision. Voice activity detection (VAD) is designed to divide the speech signal into speech segments and non-speech segments.

Automatic speech recognition is therefore an engineering compromise between the ideal, i.e. a complete model of the human, and the practical, i.e. the tools that science and technology provide and that costs allow [3].

1.3. Organization of the Thesis

This thesis provides an overview of speech recognition for data security using the Fast Fourier Transform (FFT) with the help of the MATLAB software. This section is intended to give a short overview of the thesis, by describing the outline of each chapter.

- **Chapter 2.** Introduction about speech recognition, basic speech signal and automatic speaker recognition system is discussed
- **Chapter 3.** Briefly discussed about Software and Hardware development
- **Chapter 4**. Simulation and Results are shown
- **Chapter 5.** Conclusions and some ideas concerning further work in this field.

Chapter 2

Basic Acoustics Speech Signal

And

Speech Recognition

As relevant background to the field of speech recognition, this chapter intends to discuss how the speech signal is produced and perceived by human beings. This is an essential subject that has to be considered before one can pursue and decide which approach to use for speech recognition.

2.1 The Speech Signal

Human communication is to be seen as a comprehensive diagram of the process from speech production to speech perception between the talker and listener, See Figure 2.1.

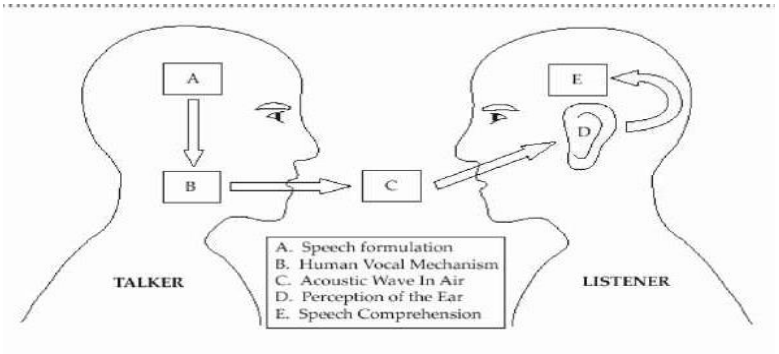


Fig.2.1 Schematic Diagram of the Speech Production/Perception Process

Five different elements, A. Speech formulation, B. Human vocal mechanism, C. Acoustic air, D. Perception of the ear, E. Speech comprehension. The first element (A. Speech formulation) is associated with the formulation of the speech signal in the talker's mind. This formulation is used by the human vocal mechanism (B. Human vocal mechanism) to produce the actual speech waveform.

During this transfer the acoustic wave can be affected by external sources, for example noise, resulting in a more complex waveform. When the wave reaches the listener's hearing system (the ears) the listener percepts the waveform (D. Perception of the ear) and the listener's mind (E. Speech comprehension) starts processing this waveform to comprehend its content so the listener understands what the talker is trying to tell him. One issue with speech recognition is to "simulate" how the listener process the speech produced by the talker. There are several actions taking place in the listeners head and hearing system during the process of speech signals. The perception process can be seen as the inverse of the speech production process. The basic theoretical unit for describing how to bring linguistic meaning to the formed speech, in the mind, is called phonemes. Phonemes can be grouped based on the properties of either the time waveform or frequency characteristics and classified in different sounds produced by the human vocal tract. Speech is:

- Time-varying signal,
- Well-structured communication process,
- Depends on known physical movements,
- Composed of known, distinct units (phonemes),

- Is different for every speaker,
- May be fast, slow, or varying in speed,
- May have high pitch, low pitch, or be whispered,
- Has widely-varying types of environmental noise,
- May not have distinct boundaries between units (phonemes),
- Has an unlimited number of words.

2.2 Speech Production

To be able to understand how the production of speech is performed one need to know how the human's vocal mechanism is constructed, see Figure 2.2. The most important parts of the human vocal mechanism are the vocal tract together with nasal cavity, which begins at the velum. The velum is a trapdoor-like mechanism that is used to formulate nasal sounds when needed [2]. When the velum is lowered, the nasal cavity is coupled together with the vocal tract to formulate the desired speech signal. The cross-sectional area of the vocal tract is limited by the tongue, lips, jaw and velum and varies from 0-20 cm2.

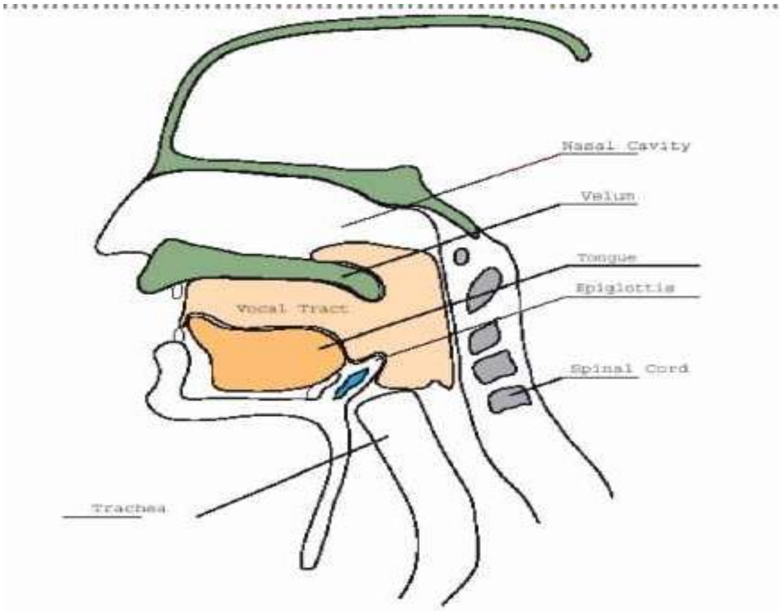


Fig.2.2. Human Vocal Mechanism

2.3 Properties of Human Voice

One of the most important parameter of sound is its frequency. The sounds are discriminated from each other by the help of their frequencies. When the frequency of a sound increases, the sound gets high-pitched and irritating. When the frequency of a sound decreases, the sound gets deepen. Sound waves are the waves that occur from vibration of the materials. The highest value of the frequency that a human can produce is about 10 kHz. And the lowest value is about 70 Hz. These are the maximum and minimum values. This frequency interval changes for every person. And the magnitude of a sound is expressed in decibel (dB). A normal human speech has a frequency interval of 100Hz - 3200Hz and its magnitude is in the range of 30 dB - 90 dB. A human ear can perceive sounds in the frequency range between 16 Hz and 20 kHz. And a frequency change of 0.5 % is the sensitivity of a human ear.

Speaker Characteristics,

- Due to the differences in vocal tract length, male, female, and children's speech are different.
- Regional accents are the differences in resonant frequencies, durations, and pitch.
- Individuals have resonant frequency patterns and duration patterns that are unique (allowing us to identify speaker).
- Training on data from one type of speaker automatically "learns" that group or person's characteristics, makes recognition of other speaker types much worse.

2.4 Speech recognition

Speech Recognition is technology that can translate spoken words into text. Some SR systems use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription. Systems that do not use training are called "Speaker Independent" systems [10]. Systems that use training are called "Speaker Dependent" systems. Speech recognition applications include voice user interfaces such as voice dialing (e.g., "Call home"), call routing (e.g., "I would like to make a collect call"), demotic appliance control, search (e.g., find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g., a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).The term voice recognition refers to finding the identity of "who" is speaking, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on specific person's voices or it can be used to authenticate or verify the identity of a speaker as part of a security process.

2.4.1 Automatic Speech Recognition System (ASR)

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the interaction of digital signal processing and natural language processing. Natural language processing is a subfield of artificial intelligence and linguistics. It studies the problems of automated generation and understanding of natural human languages. Natural language generation systems convert information from computer databases into normal-sounding human language, and natural language understanding systems convert samples of human language into more formal representations that are easier for computer programs to manipulate.

2.4.2 Speech coding

It is the compression of speech (into a code) for transmission with speech codec's that use audio signal processing and speech processes techniques. The techniques used are similar to that in audio data compression and audio coding where knowledge in psychoacoustics is used to transmit only data that is relevant to the human auditory system. For example, in narrow band speech coding, only information in the frequency band of 400 Hz to 3500 Hz is transmitted but the reconstructed signal is still adequate for intelligibility.

However, speech coding differs from audio coding in that there is a lot more statistical information available about the properties of speech. In addition, some auditory information which is relevant in audio coding can be unnecessary in the speech coding context. In speech coding, the most important criterion is preservation of intelligibility and "pleasantness" of speech, with a constrained amount of transmitted data.

It should be emphasized that the intelligibility of speech includes, besides the actual literal content, also speaker identity, emotions, intonation, timbre etc. that are all important for perfect intelligibility. The more abstract concept of pleasantness of degraded speech is different property than intelligibility, since it is possible that degraded speech is completely intelligible, but subjectively annoying to the listener.

2.4.3 Speech synthesis

Speech synthesis is the artificial production of human speech. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Synthesized speech can also be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diaphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. The quality of a speech synthesizer is judged by its similarity to the human voice, and by its ability to be understood. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1980s.

2.4.4 Voice analysis

Voice problems that require voice analysis most commonly originate from the vocal cords since it is the sound source and is thus most actively subject to tiring. However, analysis of the vocal cords is physically difficult. The location of the vocal cords effectively prohibits direct measurement of movement. Imaging methods such as x-rays or ultrasounds do not work because the vocal cords are surrounded by cartilage which distorts image quality. Movements in the vocal cords are rapid, fundamental frequencies are usually between 80 and 300 Hz, thus preventing usage of ordinary video. High-speed videos provide an option but in order to see the vocal cords the camera has to be positioned in the throat which makes speaking rather difficult. Most important indirect methods are inverse filtering of sound recordings and electroglottographs (EGG). In inverse filtering methods, the speech sound is recorded outside the mouth and then filtered by a mathematical method to remove the effects of the vocal tract. This method produces an estimate of the waveform of the pressure pulse which again inversely indicates the movements of the vocal cords. The other kind of inverse indication is the

Electroglottographs, which operates with electrodes attached to the subject's throat close to the vocal cords. Changes in conductivity of the throat indicate inversely how large a portion of the vocal cords are touching each other. It thus yields one-dimensional information of the contact area. Neither inverse filtering nor EGG is thus sufficient to completely describe the glottal movement and provide only indirect evidence of that movement.

2.4.5 Accuracy

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

2.4.6 Training

Some speech recognizers have the ability to adapt to a speaker. When the system has this ability, it may allow training to take place. An ASR system is trained by having the speaker repeat standard or common phrases and adjusting its comparison algorithms to match that particular speaker. Training a recognizer usually improves its accuracy.

Training can also be used by speakers that have difficulty speaking, or pronouncing certain words. As long as the speaker can consistently repeat an utterance, ASR systems with training should be able to adapt

2.5 Speech recognition technique

The goal of speech recognition is for a machine to be able to “hear,” understand," and "act upon" spoken information. The earliest speech recognition systems were first attempted in the early 1950s at Bell Laboratories, Davis, Biddulph and Balashek developed an isolated digit recognition system for a single speaker [1]. The goal of automatic speaker recognition is to analyze, extract characterize and recognize information about the speaker identity.

The speaker recognition system may be view working in a four stages:

1. Analysis

2. Feature extraction

3. Modeling

4. Testing

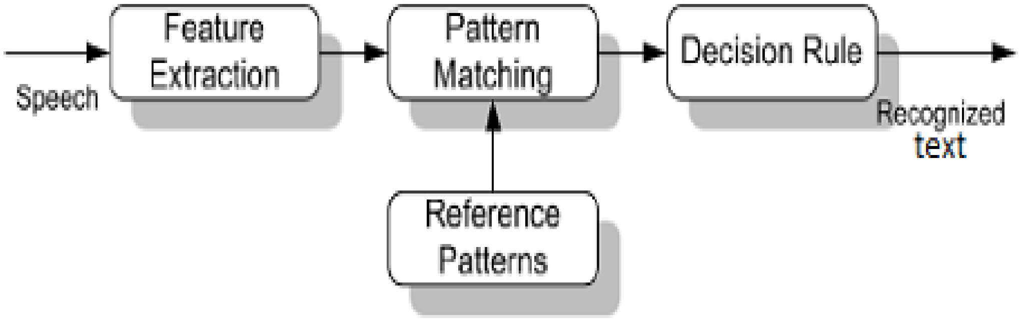


Fig-2.3: Basic working system

2.6 Speech analysis technique

Speech data contain different type of information that shows a speaker identity. This includes speaker specific information due to vocal tract, excitation source and behavior feature. Information about the behavior feature also embedded in signal and that can be used for speaker recognition the speech analysis stage deals with stage with suitable frame size for segmenting speech signal for further analysis and extracting.

2.6.1 Different type of speech recognition techniques:

Research into speech recognition began by reviewing the literature and finding techniques that had previously been used for speech/speaker recognition. It was found that six techniques are commonly used for speech/speaker recognition or have been used for this domain in the past. These were:

- ♦ Dynamic Time Warping (DTW)
- ♦ Hidden Markov Models (HMM)
- ♦ Vector Quantization (VQ)
- ♦ Ergodic-HMM's

♦ Artificial Neural Networks (ANN)

♦ Long-Term Statistics

2.7 Working Procedure

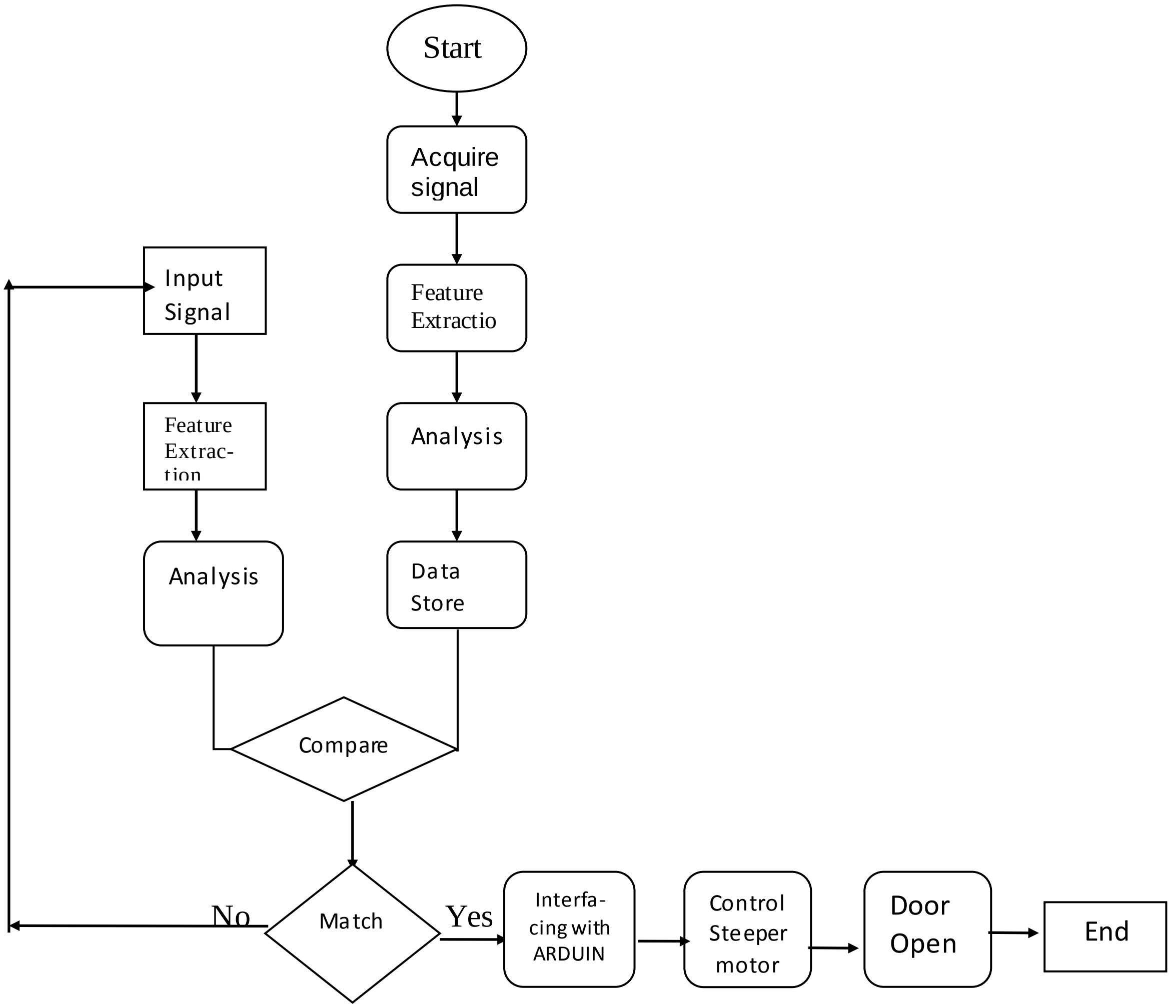


Fig-2.4: Working Procedure

The working procedure is given below:

- At first record a signal from authorized person.
- This signal is acquire signal.
- Then this speech is feature extraction in the feature extraction part.
- Analysis this signal and save this speech.
- Now give a sample speech, it is the input signal.
- This speech is featured extraction and analysis in that section.
- Compare the two speech input & saved signal.
- Then checked the performance of comparing. If voice is matched then a signal is going to the ARDUINO microcontroller and the door is open .if voice doesn't matched, Door remain closed

2.8 Feature Extraction

This stage is often referred as speech processing front end. The main goal of Feature Extraction is to simplify recognition by summarizing the vast amount of speech data without losing the acoustic properties that defines the speech [3].

Obtaining the acoustic characteristics of the speech signal is referred to as Feature Extraction. Feature Extraction is used in both training and recognition phases. It comprise of the following steps:

1. Frame Blocking

2. Windowing

3. FFT (Fast Fourier Transform)

5. Noise reduction

6. Humming Window

2.8.1 Frame Blocking

Investigations show that speech signal characteristics stays stationary in a sufficiently short period of time interval (It is called quasi-stationary). For this reason, speech signals are processed in short time intervals. It is divided into frames with sizes generally between 30 a 100 milliseconds. Each frame overlaps its previous frame by a predefined size. The goal of the overlapping scheme is to smooth the transition from frame to frame.

2.8.2 Windowing

Windowing is the process of taking a small subset of a larger dataset, for processing and analysis. The second step is to window all frames. This is done in order to eliminate discontinuities at the edges of the frames. If the windowing function is defined as w (n), $0 < n < N-1$ where N is the number of samples in each frame, the resulting signal will be; $y(n) = x(n)w(n)$. Generally hamming windows are used.

2.8.3 Hamming window

In signal processing, a window function is a mathematical that is zero-valued outside of some chosen interval.

For instance, a function that is constant inside the interval and zero elsewhere is called a rectangular window, which describes the shape of its graphical representation. When another function or a signal is multiplied by a window function, the product is also zero-valued outside the interval.

Applications of window functions include spectral analysis, filter design, and beam forming.

In spectrum analysis of naturally occurring audio signals, we nearly always analyze a short segment of a signal, rather than the whole signal. This is the case for a variety of reasons. Perhaps most fundamentally, the ear similarly Fourier analyzes only a short segment of audio signals at a time. Therefore, to perform a spectrum analysis having time- and frequency-resolution comparable to human hearing, we must limit the time-window accordingly.

2.8.4 Fast Fourier Transform (FFT)

A fast Fourier transform (FFT) is an efficient algorithm to compute the discrete Fourier transform (DFT) and it's inverse. There are many distinct FFT algorithms involving a wide range of mathematics, from simple complex-number arithmetic to group theory and number theory.

A DFT decomposes a sequence of values into components of different frequencies. This operation is useful in many fields, but computing it directly from the definition is often too slow to be practical. An FFT is a way to compute the same result more quickly: computing a DFT of N points in the naive way, using the definition, takes $O(N^2)$ arithmetical operations, while an FFT can compute the same result in only $O(N \log N)$ operations. The difference in speed can be substantial, especially for long data sets where N may be in the thousands or millions in practice, the computation time can be reduced by several orders of magnitude in such cases, and the improvement is roughly proportional to $N / \log(N)$. This huge improvement made many DFT based algorithms practical; FFTs are of great importance to a wide variety of applications, from digital signal processing and solving partial differential equations to algorithms for quick multiplication of large integers

. It was found that six techniques are commonly used for speech/speaker recognition or have been used for this domain in the past. These were:

♦ Dynamic Time Warping (DTW)

♦ Hidden Markov Models (HMM)

♦ Vector Quantization (VQ)

♦ Ergodic-HMM's

♦ Artificial Neural Networks (ANN)

♦ Long-Term Statistics

Chapter 3

Software & Hardware Development

In this chapter we will discuss about the hardware and software development, that how we can interfacing software and hardware using Arduino Microcontroller. Here we basically, develop our project based on Matlab software. Stepper motor is controlled by MATLAB codding and operates by Arduino microcontroller, which we are going to discuss briefly in bellow.

3.1 Interface

An interface is a tool and concept that refers to a point of interaction between components and is applicable at the level of both hardware and software. This allows a component, whether a piece of hardware such as a graphics card or a piece of software such as an Internet browser, to function independently while using interfaces to communicate with other components via an input/output system and an associated protocol.

3.1.1 Interfacing with Hardware

Hardware interfaces exist in computing systems between many of the components such as the various buses, storage devices, other I/O devices, etc. A hardware interface is described by the mechanical, electrical and logical signals at the interface and the protocol for sequencing them (sometimes called signaling).Hardware interfaces can be parallel where performance is important or serial where distance is important.

3.2 Types of development

There are two types of work is done in here

- Software development
- Hardware development

3.2.1 Interfacing ARDUINO with MATLAB

Following this process are used to interfacing ARDUINO by MATLAB

First we use COM3 serial port which is indicates below

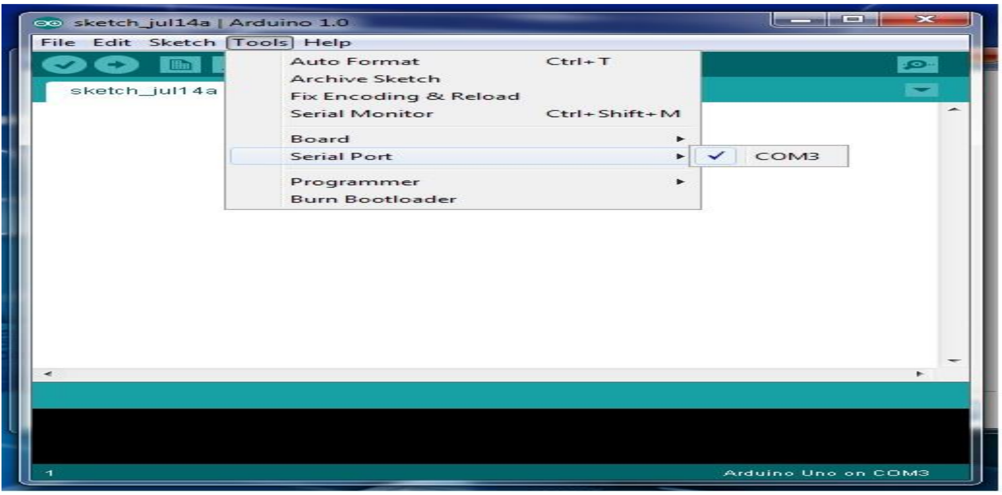
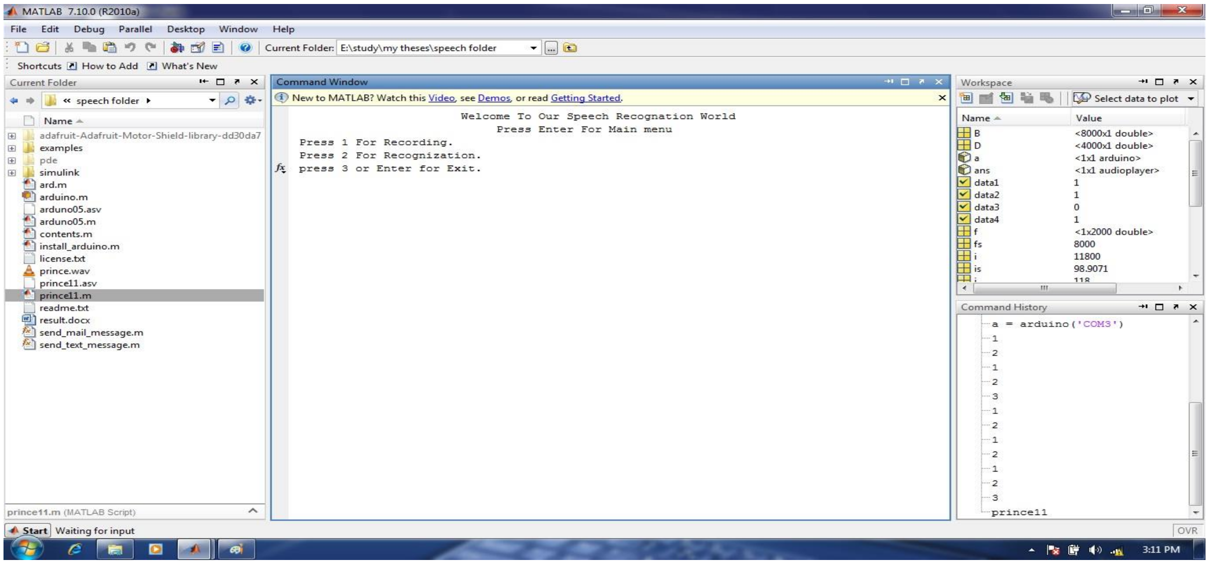


Figure 3.1: Arduino Configuration Procedure

In figure-3.1 we established a connection between pc and Arduino microcontroller through serial port com-3.

This Picture (Figure 3.2) indicates the introducing of speech recognition world. Here, we have three options

- For press 1, Indicates recording,
- For press 2, recognition &
- For press 3 Indicates exit.



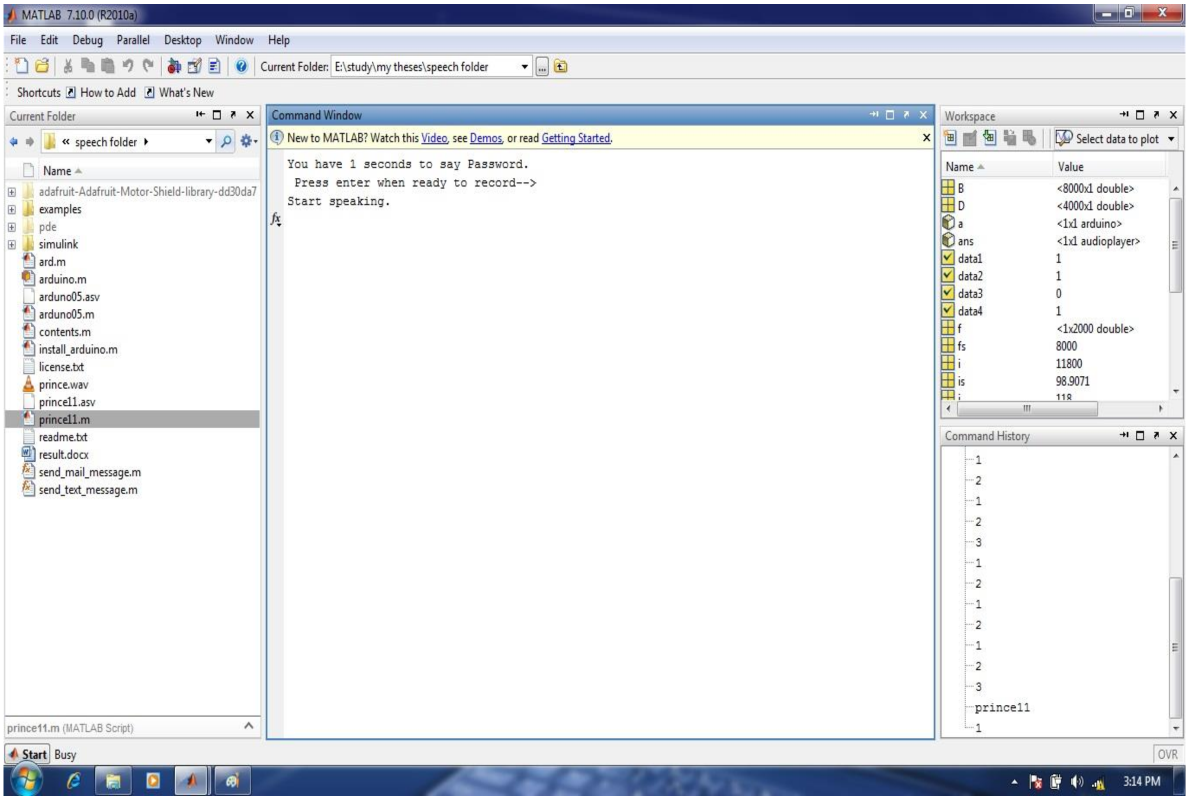
```
                      Welcome To Our Speech Recognation World
                           Press Enter For Main menu
Press 1 For Recording.
Press 2 For Recognization.
press 3 or Enter for Exit.
```

Name	Value
B	<8000x1 double>
D	<4000x1 double>
a	<1x1 arduino>
ans	<1x1 audioplayer>
data1	1
data2	1
data3	0
data4	1
f	<1x2000 double>
fs	8000
i	11800
is	98.9071
j	118

```
a = arduino('COM3')
1
2
1
2
3
1
2
1
2
1
2
3
prince11
```

Figure 3.2: The Starting menu of the program

Now from this pic (Figure 3.3) we press 1 and then press enter for recording speech. Here we stored training signal from speech.

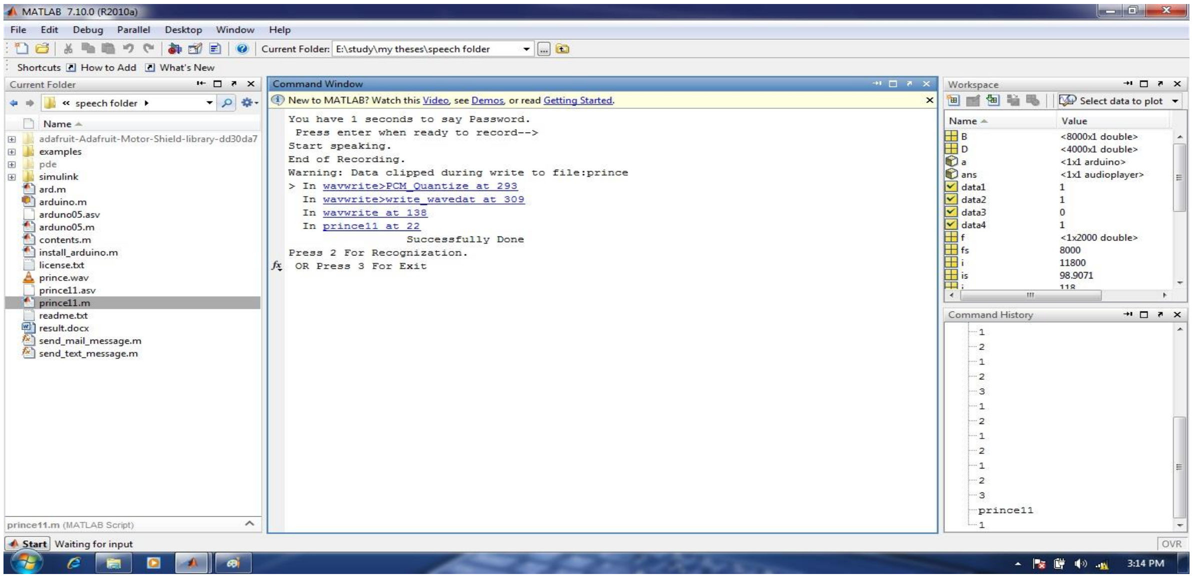


```
You have 1 seconds to say Password.
 Press enter when ready to record-->
Start speaking.
```

Name	Value
B	<8000x1 double>
D	<4000x1 double>
a	<1x1 arduino>
ans	<1x1 audioplayer>
data1	1
data2	1
data3	0
data4	1
f	<1x2000 double>
fs	8000
i	11800
is	98.9071
j	118

Figure 3.3: The Starting menu of the Option 1

After providing the voice, this page (Figure 3.4) is indicates that, our speech or voice recording is successfully done. Then for recognition we need to follow second option.

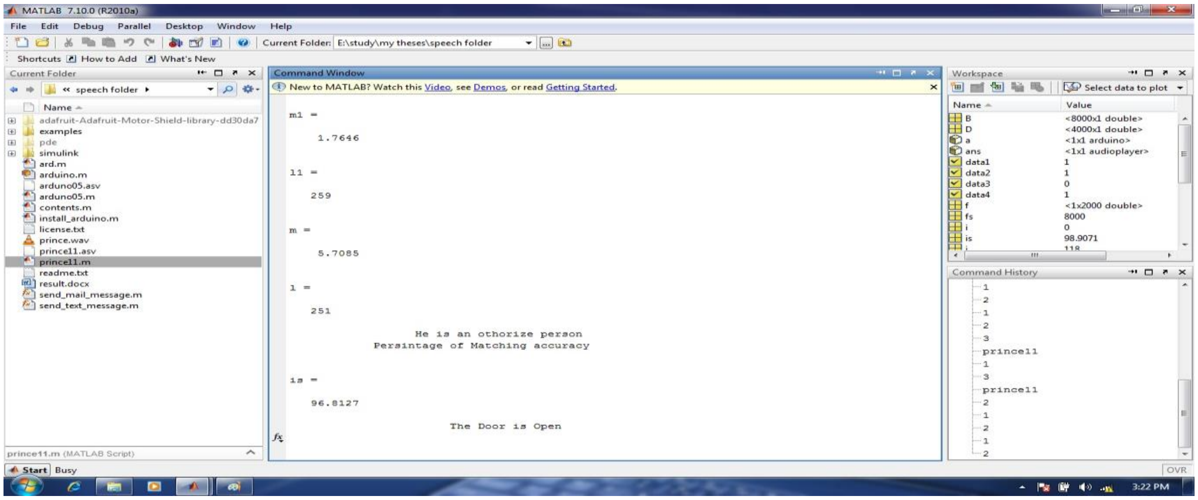


```
You have 1 seconds to say Password.
 Press enter when ready to record-->
Start speaking.
End of Recording.
Warning: Data clipped during write to file:prince
> In wavwrite>PCM_Quantize at 293
  In wavwrite>write_wavedat at 309
  In wavwrite at 138
  In prince11 at 22
                 Successfully Done
Press 2 For Recognization.
 OR Press 3 For Exit
```

Name	Value
B	<8000x1 double>
D	<4000x1 double>
a	<1x1 arduino>
ans	<1x1 audioplayer>
data1	1
data2	1
data3	0
data4	1
f	<1x2000 double>
fs	8000
i	11800
is	98.9071

Figure 3.4: Recording Conformation programme

Figure 3.5 shows that, we recognize our signal through 2nd option. Here we provide our tracing signal from the authorize person. Voice signal is matched then the program is executed and then door will go open.



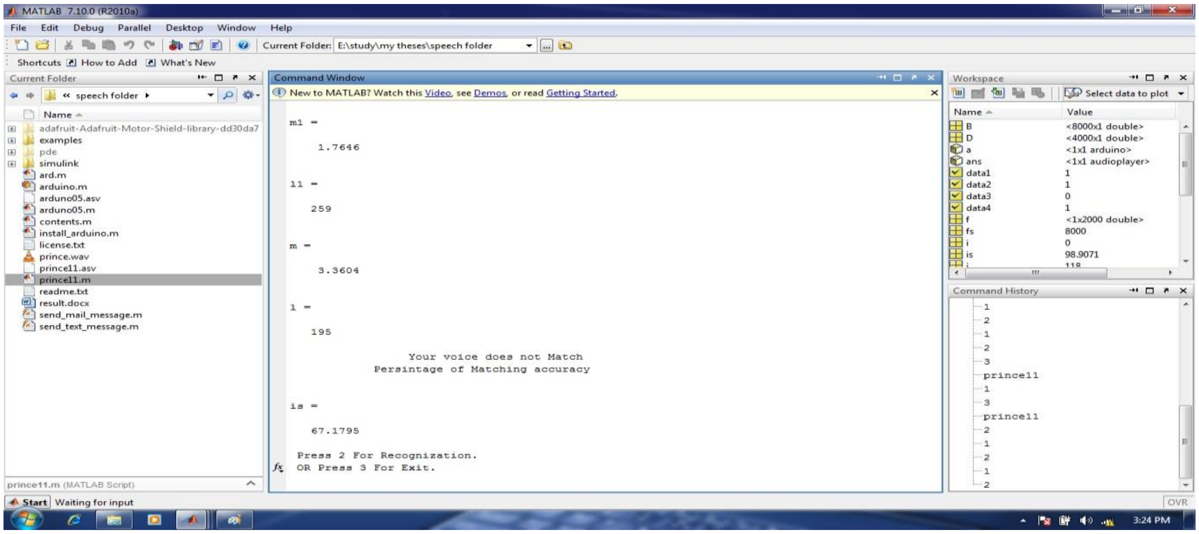
```
m1 =

    1.7646


l1 =

   259


m =

    5.7085


l =

   251

                He is an othorize person
            Persintage of Matching accuracy


is =

   96.8127

                       The Door is Open
```

Name	Value
B	<8000x1 double>
D	<4000x1 double>
a	<1x1 arduino>
ans	<1x1 audioplayer>
data1	1
data2	1
data3	0
data4	1
f	<1x2000 double>
fs	8000
i	0
is	98.9071

Figure 3.5: Authorized final result

Here we see that, the percentage of matching accuracy between the tracing and training signal is 96.82% Which is already crossed to our selected matching threshold level 90.00%.That's why software take this signal as an authorize signal and give the permission to open the door.

If voice does not match, then we can from the figure 3.6



```
m1 =

    1.7646


l1 =

   259


m =

    3.3604


l =

   195

                Your voice does not Match
            Persintage of Matching accuracy


is =

   67.1795

 Press 2 For Recognization.
 OR Press 3 For Exit.
```

Name	Value
B	<8000x1 double>
D	<4000x1 double>
a	<1x1 arduino>
ans	<1x1 audioplayer>
data1	1
data2	1
data3	0
data4	1
f	<1x2000 double>
fs	8000
i	0
is	98.9071
j	118

```
1
2
1
2
3
prince11
1
3
prince11
2
1
2
1
2
```

Figure 3.6: Unauthorized final result

Here we see that, the percentage of matching accuracy between the tracing and training signal is 67.18% Which is already much below to our selected matching threshold level.In this case, software take this signal as an unauthorized signal and don't give the permission to open the door.

3.2.2 Hardware Development:

Below this indicates different types of hardware working.

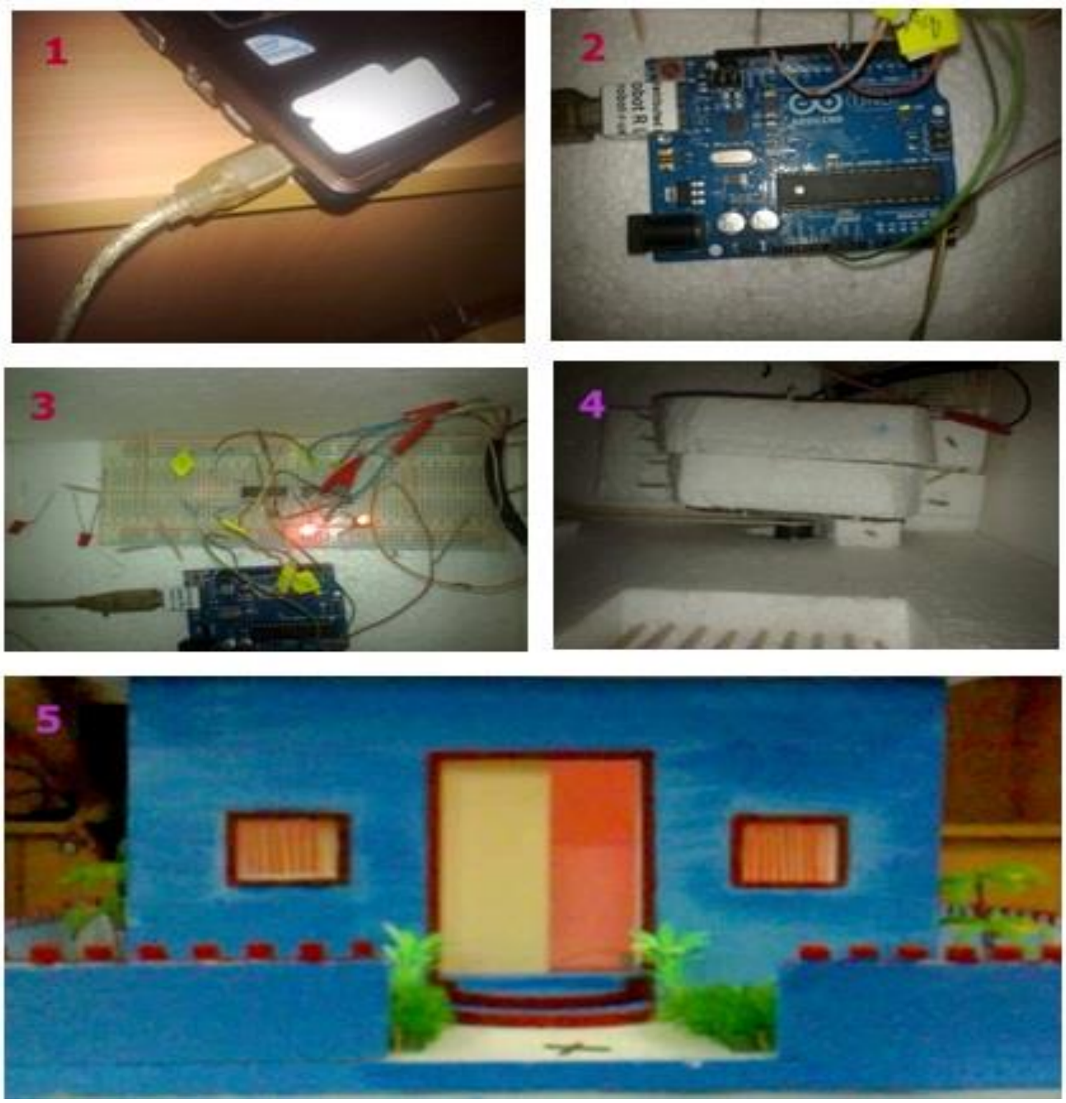


Figure 3.7: Different parts of hardware

From the figure-3.7 we can see sequentially,

1. Here we connect Arduino microcontroller with Laptop through USB cable
2. Arduino Microcontroller is working mode.
3. Our full circuit board, where microcontroller and steeper motor are connected with the circuit board.
4. Steeper motor is working using a ball.

Fig-3.7

5. For the working of steeper motor, our door is going to open mode.
6. To rotating steeper motor we use dc power supply 7.6 volt

3.3 Overview of Arduino

As relevant background to the field of Arduino microcontroller and steeper motor, this chapter intends to discuss how the Arduino microcontroller and stepper motor work. These are essential devices that have to be considered before one can pursue and decide which approach to use for speech recognition.

3.3.1 About Arduino

Arduino is a tool for making computers that can sense and control more of the physical world than your desktop computer. It's an open-source physical computing platform based on a simple microcontroller board, and a development environment for writing software for the board.

An Arduino board consists of an 8-bit Atmel AVR microcontroller with complementary components to facilitate programming and incorporation into other circuits. An important aspect of the Arduino is the standard way that connectors are exposed, allowing the CPU board to be connected to a variety of interchangeable add-on modules known as shields. Some shields communicate with the Arduino board directly over various pins, but many shields are individually addressable via an I²C serial bus, allowing many shields to be stacked and used in parallel.

3.3.2 Reason for using Arduino

We already know that, Arduino is a popular open-source single-board microcontroller, descendant of the open-source Wiring platform, Arduino also simplifies the process of working with microcontrollers, but it offers some advantage for us over other systems. That’s why we use it in our project.

- **Inexpensive** - Arduino boards are relatively inexpensive compared to other microcontroller platforms.
- **Cross platform** - The Arduino software runs on Windows, Macintosh OSX, and Linux operating systems. Most microcontroller systems are limited to Windows.
- **Open source and extensible software** - The Arduino software and is published as open source tools, available for extension by experienced programmers.
- **Open source and extensible hardware** - The Arduino is based on Atmel's ATMEGA8 and ATMEGA168 microcontrollers. The plans for the modules are published under a Creative Commons license, so experienced circuit designers can make their own version of the module, extending it and improving it.

3.3.3 Matlab supported packages for Arduino

Matlab support package for Arduino allows you to communicate with an Arduino uno or duemilanove over a serial port. It consists of a Matlab api on the host computer and a server program that runs on the Arduino. Together, they allow you to access Arduino analog i/o, digital i/o, and motor shield from the Matlab command line.

Sample usage:

```
%-- connect to the board
a = arduino('COM4')
```

```
%-- specify pin mode
 a.pinMode(4,'input');
 a.pinMode(13,'output');

 %-- digital i/o
 a.digitalRead(4) % read pin 4
 a.digitalWrite(13,0) % write 0 to pin 13

 %-- motor shield
 a.motorRun(4, 'forward') % run motor forward
 a.servoWrite(1, 175); % move servo#1 to 175 deg position
 a.stepperStep(1, 'forward', 'double', 100); % move stepper motor

 %-- close session
 delete (a)
```

3.4 STEPPER MOTORS

A stepper motor is a motor controlled by a series of electromagnetic coils. The center shaft has a series of magnets mounted on it, and the coils surrounding the shaft are alternately given current or not, creating magnetic fields which repulse or attract the magnets on the shaft, causing the motor to rotate.

There are two basic types of stepper motors,

- Unipolar steppers and
- Bipolar steppers.

3.4.1 UNIPOLAR STEPPER MOTORS

The unipolar stepper motor has five or six wires and four coils. The center connections of the coils are tied together and used as the power connection. They are called unipolar steppers because power always comes in on this one pole.

3.4.2 Bipolar stepper motors

The bipolar stepper motor usually has four wires coming out of it. Unlike unipolar steppers, bipolar steppers have no common center connection. They have two independent sets of coils instead. We can distinguish them from unipolar steppers by measuring the resistance between the wires. Here we should find two pairs of wires with equal resistance. If we've got the leads of your meter connected to two wires that are not connected, you should see infinite resistance.

3.4.3 Reason for using Bipolar Stepper motors

We already known about the stepper motors, in our project we used Bipolar Stepper motors. Because we need bipolar stepper motor rotate at both sides. That was our expectation. That's why we use it confidentially.

Chapter 4

Simulation & Results

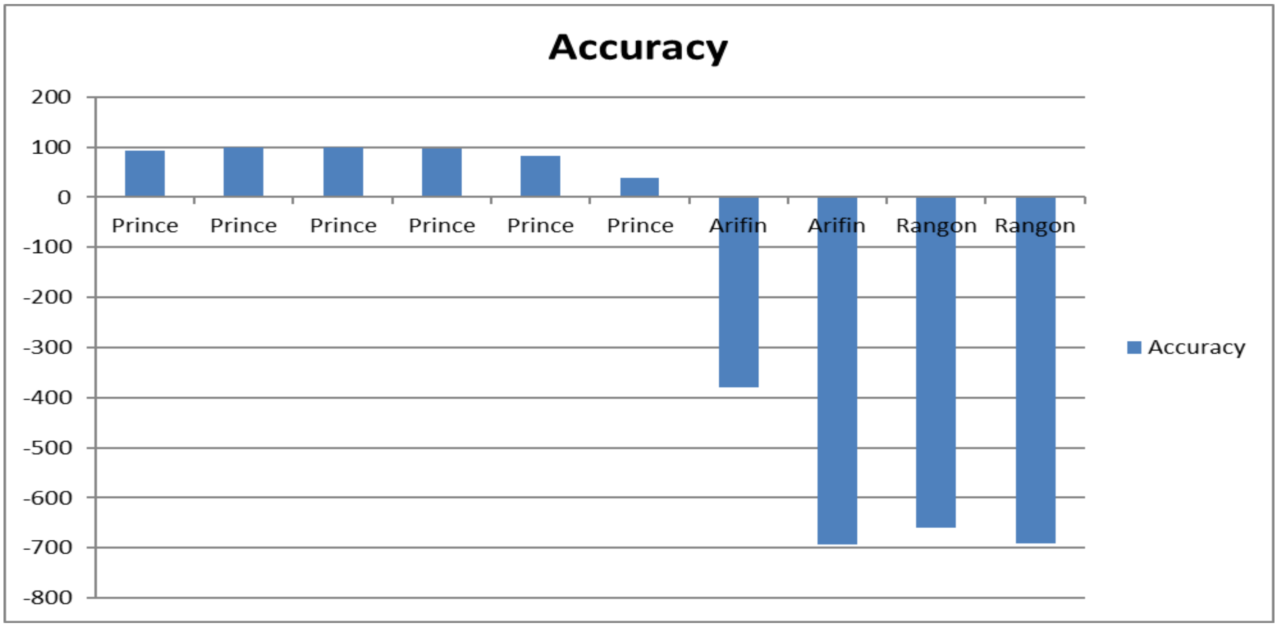
This chapter is represents about performance and analysis of the project. Also different types of table, graph & compares is shown in this chapter.

Performance Analysis

In Tabel-1, Prince's voice signal "Open" is stored as a train signal, different types of test signal is used for authentication. When prince's voice is used as test signal for five times, then for the 1st four times Prince gave same voice "Open", as a result 4th times match the signal with the train signal, but last time for Prince's "Hello" signal, signal didn't match with the train signal, that's why door was still closed. To check systems accuracy again and again is checked by using different signal from different users.

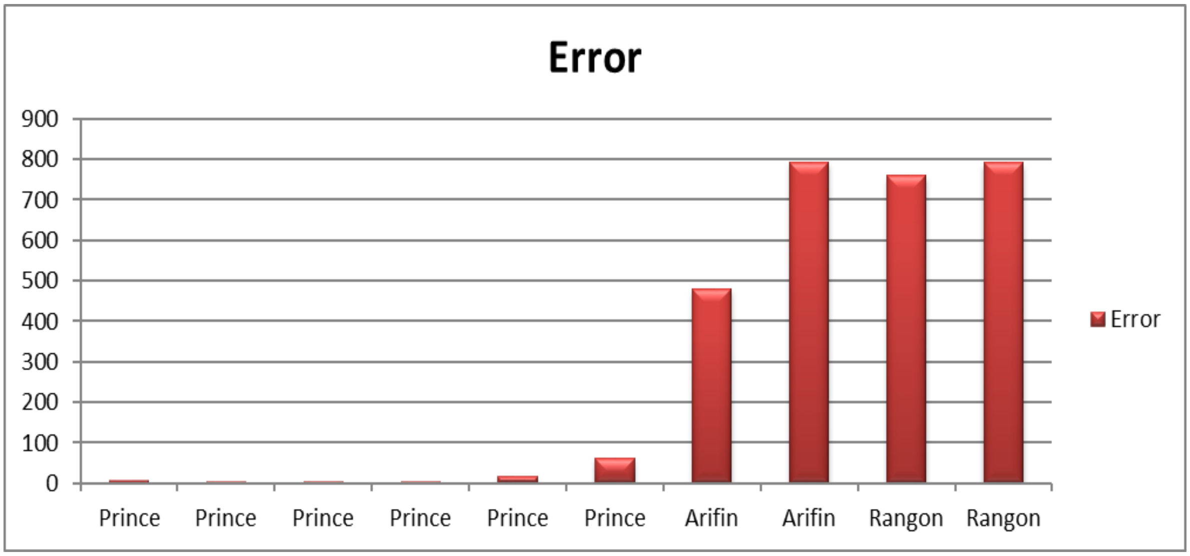
Tabel-1: Train & Test voice of Prince

Name	Input Signal	Train M1	Train L1(freq -peak)	Test Mo	Test Lo	Accur acy	Error	Comment (Door condition)	Result
Prince	Open	23.05	232	25.47	248	93.54	6.46	Open	Match
Prince	Open	23.05	232	21.48	230	99.13	0.87	Open	Match
Prince	Open	23.05	232	16.77	229	98.69	1.31	Open	Match
Prince	Open	23.05	232	15.82	227	97.89	2.11	Open	Match
Prince	Hello	23.05	232	48.55	280	82.85	17.15	Close	Mismatch
Prince	Me	23.05	232	11.53	143	37.76	62.24	Close	Mismatch
Arifin	Ok	23.05	232	8.401	40	-380	480	Close	Mismatch
Arifin	Open	23.05	232	13.01	26	-692.3	792.30	Close	Mismatch
Rangon	Hi	23.05	232	2.26	27	-659	759	Close	Mismatch
Rangon	Open	23.05	232	6.21	26	-692	792	Close	Mismatch



Graph- 1

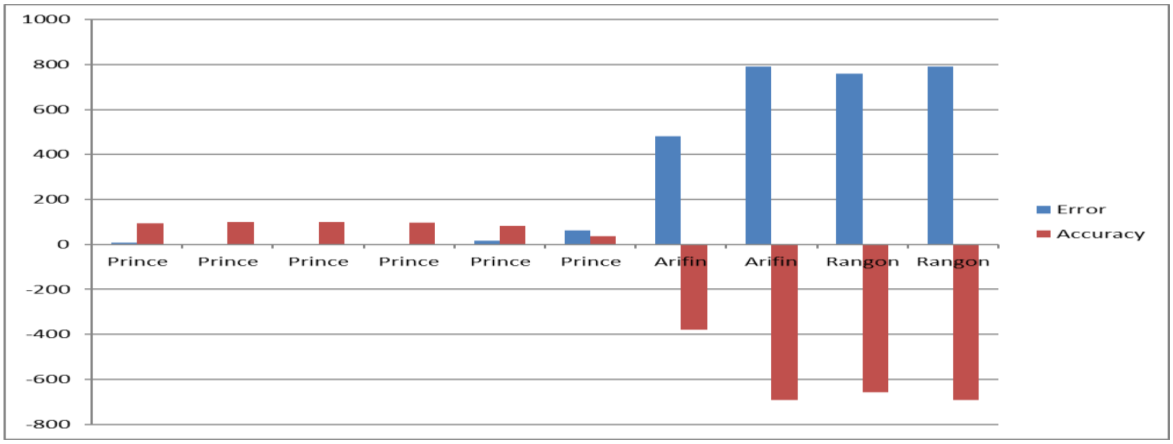
Graph 1 shows the input voice and accuracy rate. Here authorized person gives four times same voice so accuracy rate is approximately 100% , two times other voice so accuracy rate is medium and rest of them are unauthorized person so accuracy rate is below zero.



Graph- 2

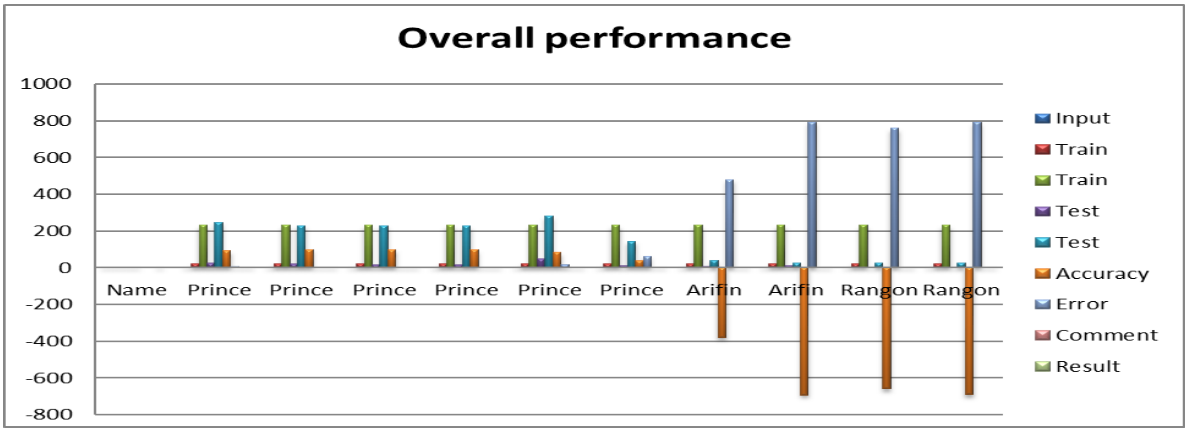
Graph-2 shows the input voice and error rate. Here authorized person gives four times same voice so error rate is very low , two times other voice so error rate is high and rest of them are unauthorized person so error rate is very high near eight hundred.

Error and Accuracy



Graph- 3

Graph-3 shows the input voice and error rate, here the error rate is very low but authorized person gives other voice as input then accuracy is lower than the error. Matching accuracy and error rate of the signal, which is in graph-3. This speech is taken from the medium noisy area.



Graph -4

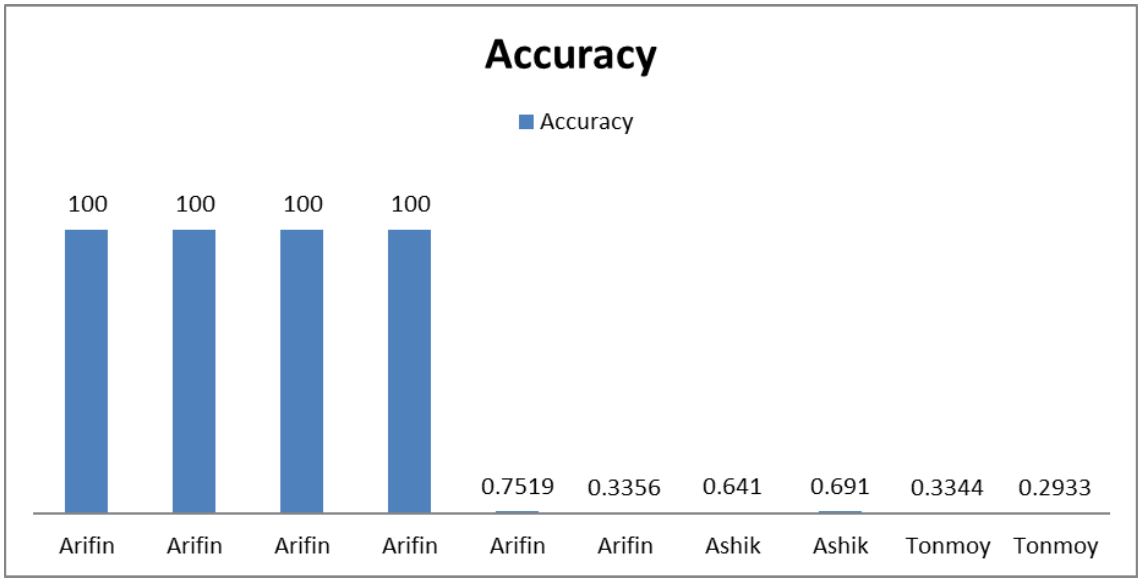
Above graph-4 represent overall performance figure of the train and test voice of prince. First color means Fourier co-efficient of train signal, second color means peak frequency of train signal 3rd and 4th color means test signal. 5th and 6th color means accuracy and error.

From the table (Tabel-2), here stored Arifin's voice signal as a train signal, for this train signal different types of test signal is used for authentication. Arifin's voice is used as test signal for six times, for the 1st four times Arifin gave same voice signal "Open" as a result 4th times match the signal with the train signal, but last two time for Arifin's "Hello" signal, signal didn't match with the train signal, that's why door was still closed. To check systems

Tabel-2: Train & Test voice of Arifin

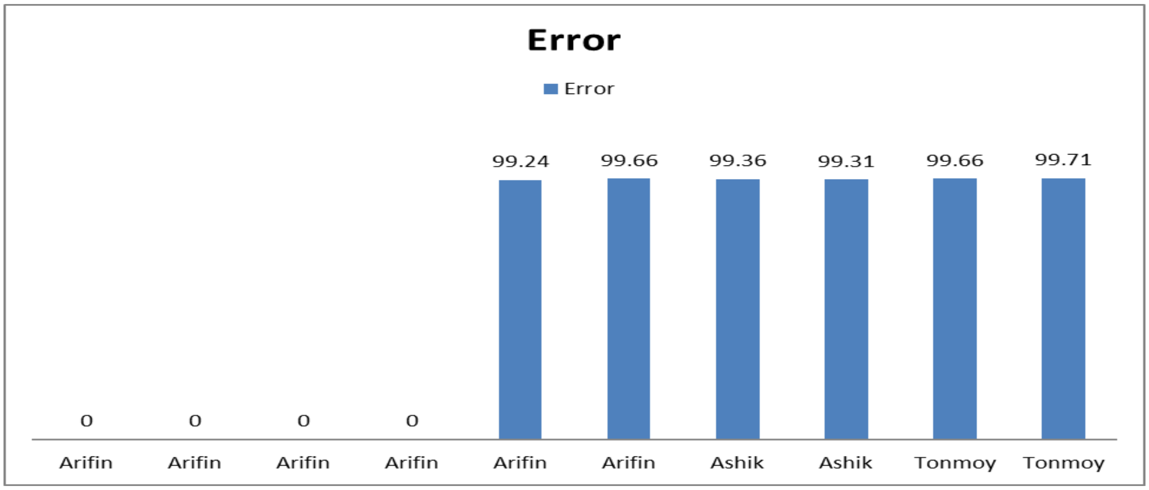
Name	Input Signal	Train M1	Train L1(freq-peak)	Test Mo	Test Lo	Accuracy	Error	Comment (Door condition)	Result
Arifin	Open	0.0588	1	.0632	1	100	0	Open	Match
Arifin	Open	0.0588	1	.0574	1	100	0	Open	Match
Arifin	Open	0.0588	1	.0771	1	100	0	Open	Match
Arifin	Open	0.0588	1	.0662	1	100	0	Open	Match
Arifin	Hello	0.0588	1	.2418	133	.7519	99.24	Close	Mismatch
Arifin	Hello	0.0588	1	.7870	298	.3356	99.66	Close	Mismatch
Ashik	ok	0.0588	1	.0662	156	.6410	99.36	Close	Mismatch
Ashik	ok	0.0588	1	.0652	180	.6910	99.31	Close	Mismatch
Tonmoy	Over	0.0588	1	.6121	299	.3344	99.66	Close	Mismatch
Tonmoy	Open	0.0588	1	.7115	341	.2933	99.71	Close	Mismatch

accuracy again Asik & Tonmoy are gave different signal. But every times software represent same mismatch result.



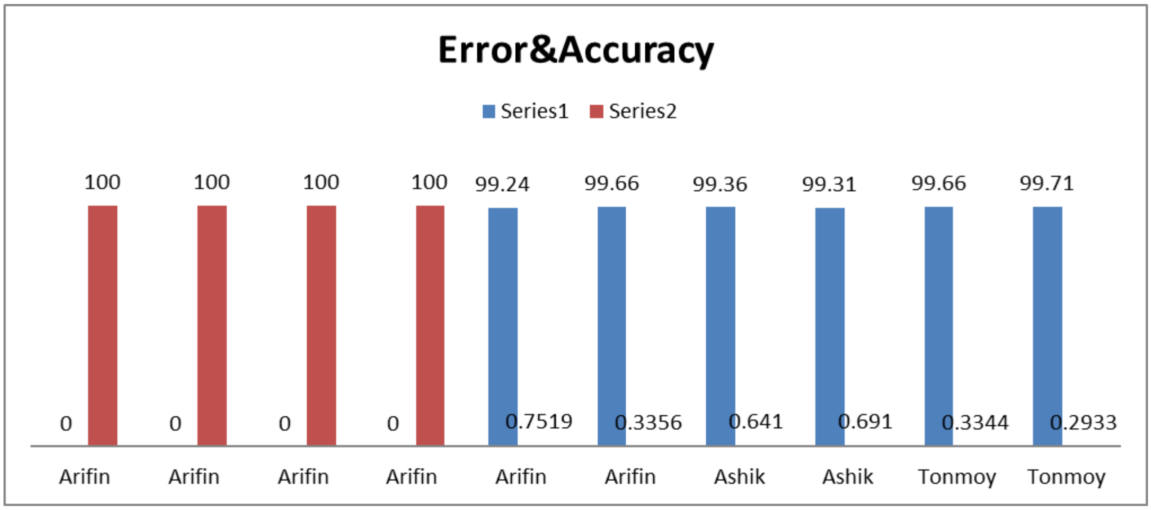
Graph-5

The results of the performance of the voice recognition for data security are in graph 5, 6, 7. This performance is measured by the ratio of accuracy, error of the recognized speech.



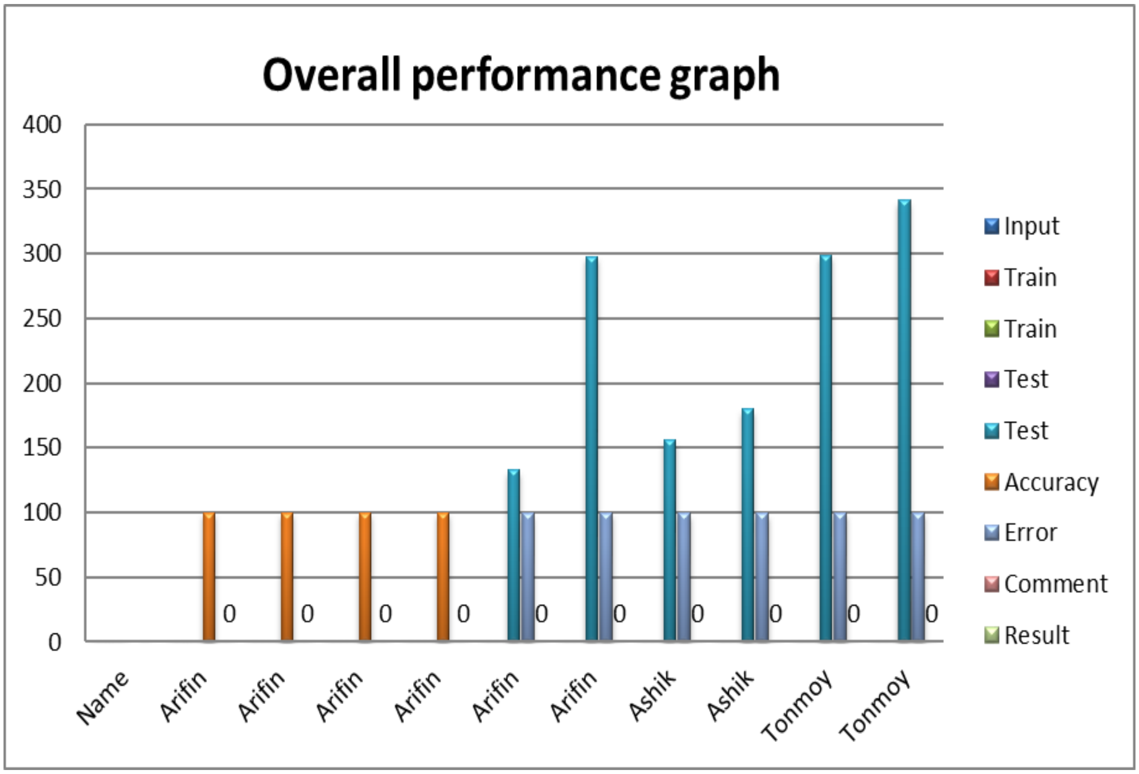
Graph-6

Graph-6 shows the input voice and error rate, the error rate is very low. Authorized person other voice error is very high such 99.25, 99.66, 99.31.Then matches the accuracy and error of the signal that is plot in the graph -3. This voice is taken in the less noisy area.



Graph-7

Graph 7 shows that authorized person voice error is zero (0) and accuracy is 100%.Sky colour means authorize person but other voice error is 99.24% and accuracy is shown 0.7519%.


Overall performance graph
400
350
300
250
200
150
100
50
0
0
0
0
0
0
0
0
0
0
0
Name
Arifin
Arifin
Arifin
Arifin
Arifin
Arifin
Ashik
Ashik
Tonmoy
Tonmoy
Input
Train
Train
Test
Test
Accuracy
Error
Comment
Result

Graph-8

Graph-8 shows the overall performance figure of the train and test voice of Arifin. First colour means Fourier co-efficient of train signal, second colour means peak frequency of train signal 3rd and four colour means test signal 5th and 6th colour means accuracy and error.

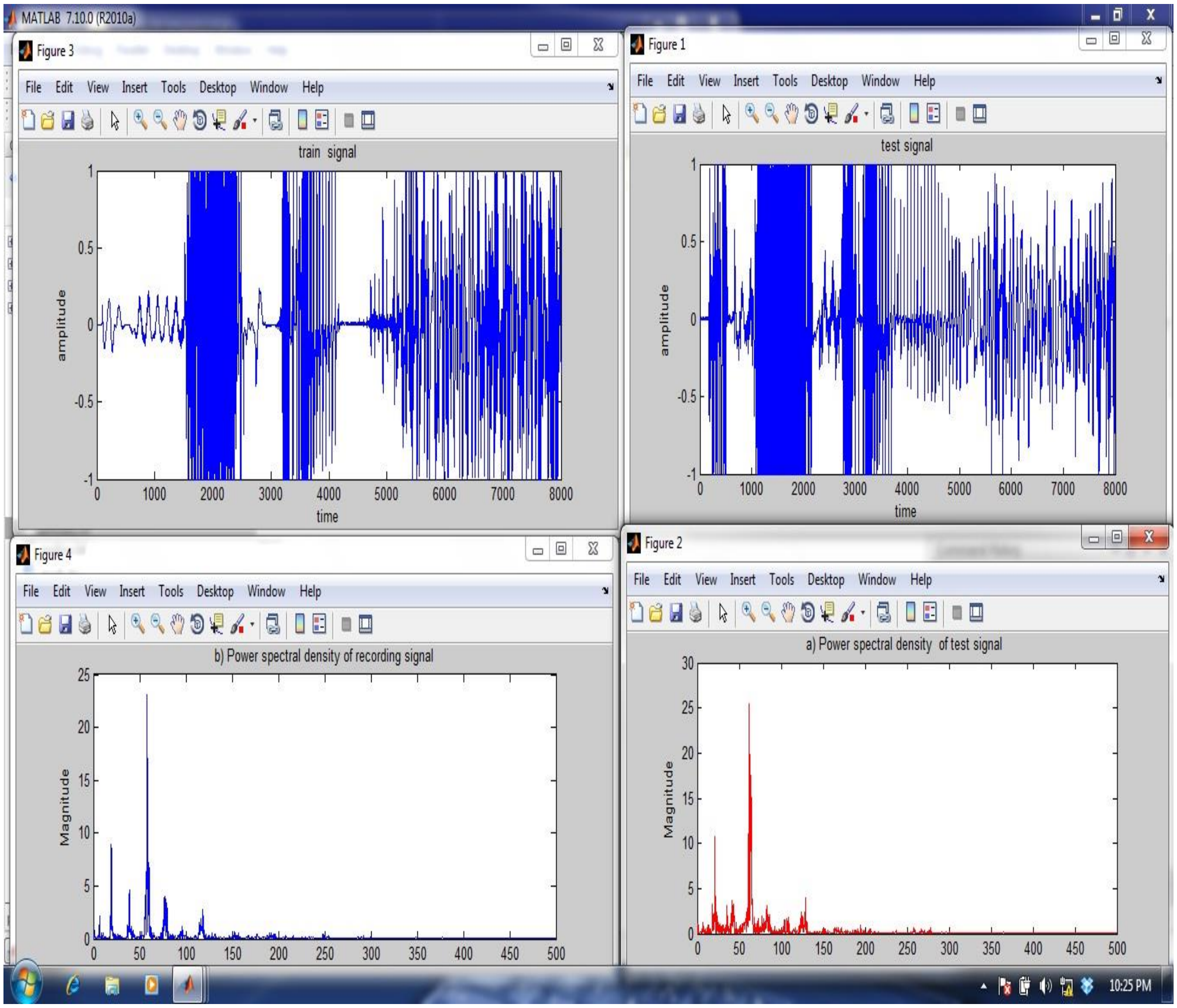


Figure-4.1: Authorized person signal

Above two part of this signal is a train and test part of authorized person. Below two part of this signal represent power spectral density of train and test part.

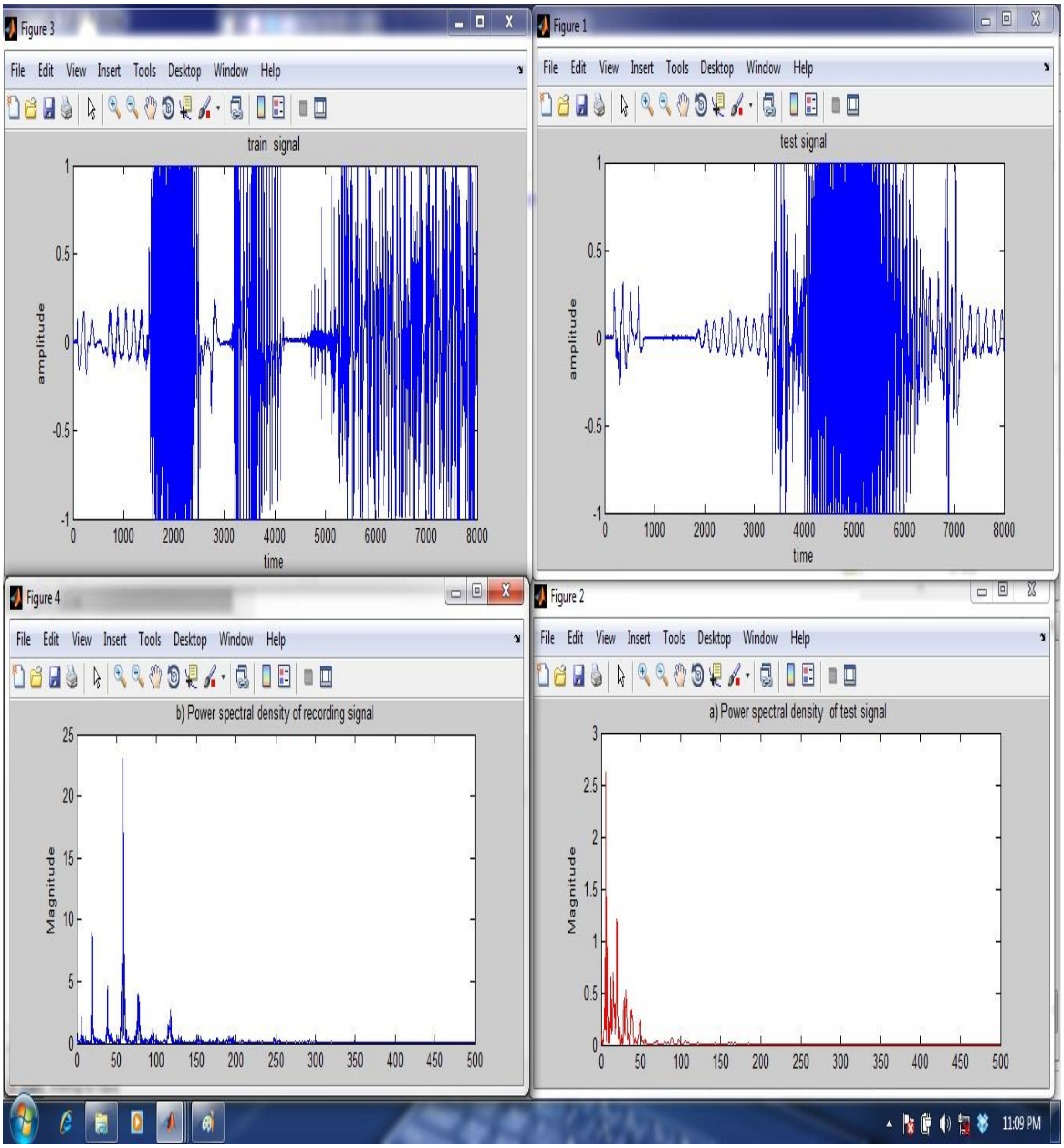
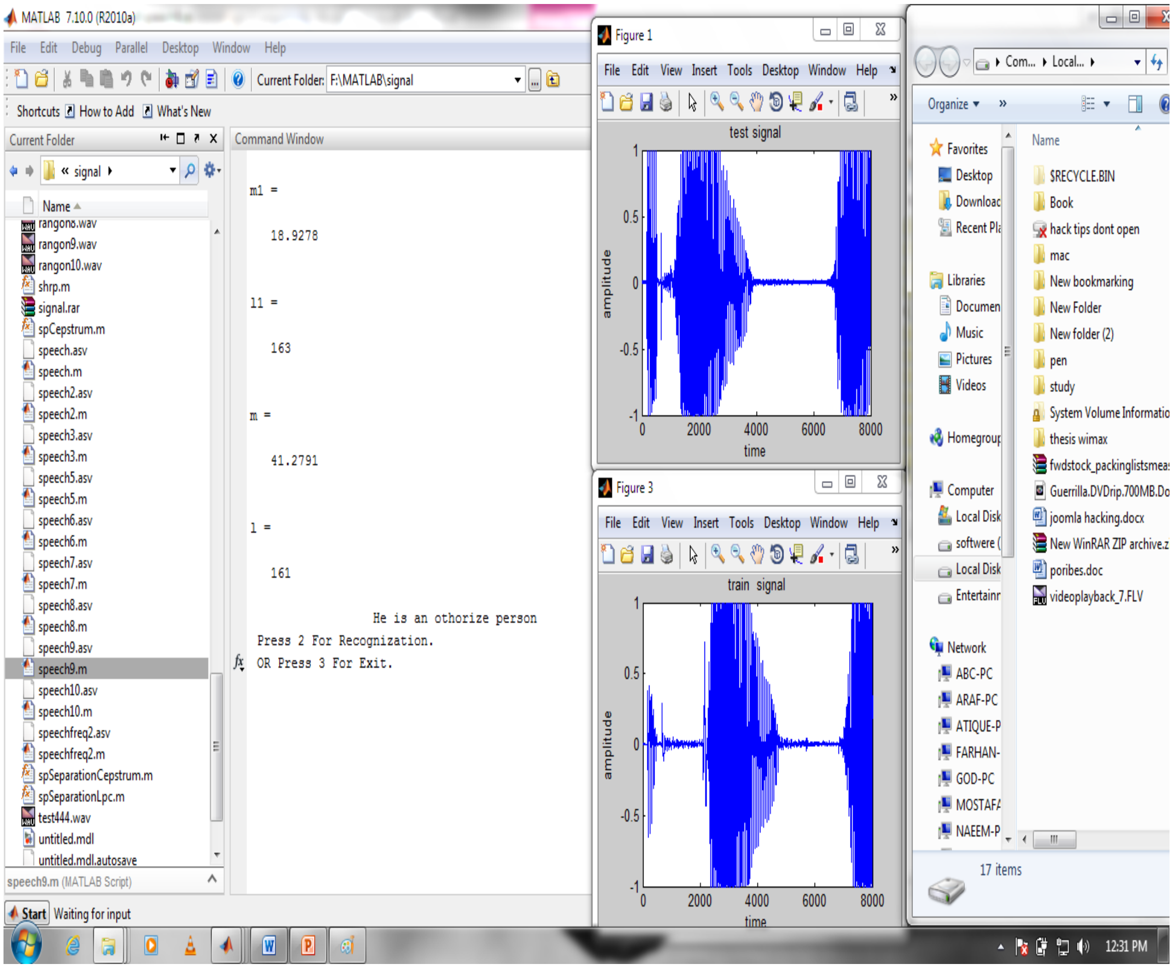


Figure 3
File Edit View Insert Tools Desktop Window Help
train signal
amplitude
time
1
0.5
0
-0.5
-1
0
1000
2000
3000
4000
5000
6000
7000
8000
Figure 1
File Edit View Insert Tools Desktop Window Help
test signal
amplitude
time
Figure 4
File Edit View Insert Tools Desktop Window Help
b) Power spectral density of recording signal
Magnitude
25
20
15
10
5
0
0
50
100
150
200
250
300
350
400
450
500
Figure 2
File Edit View Insert Tools Desktop Window Help
a) Power spectral density of test signal
Magnitude
3
2.5
2
1.5
1
0.5
0
11:09 PM

Fig. 4.2: Authorized train and unauthorized test signal

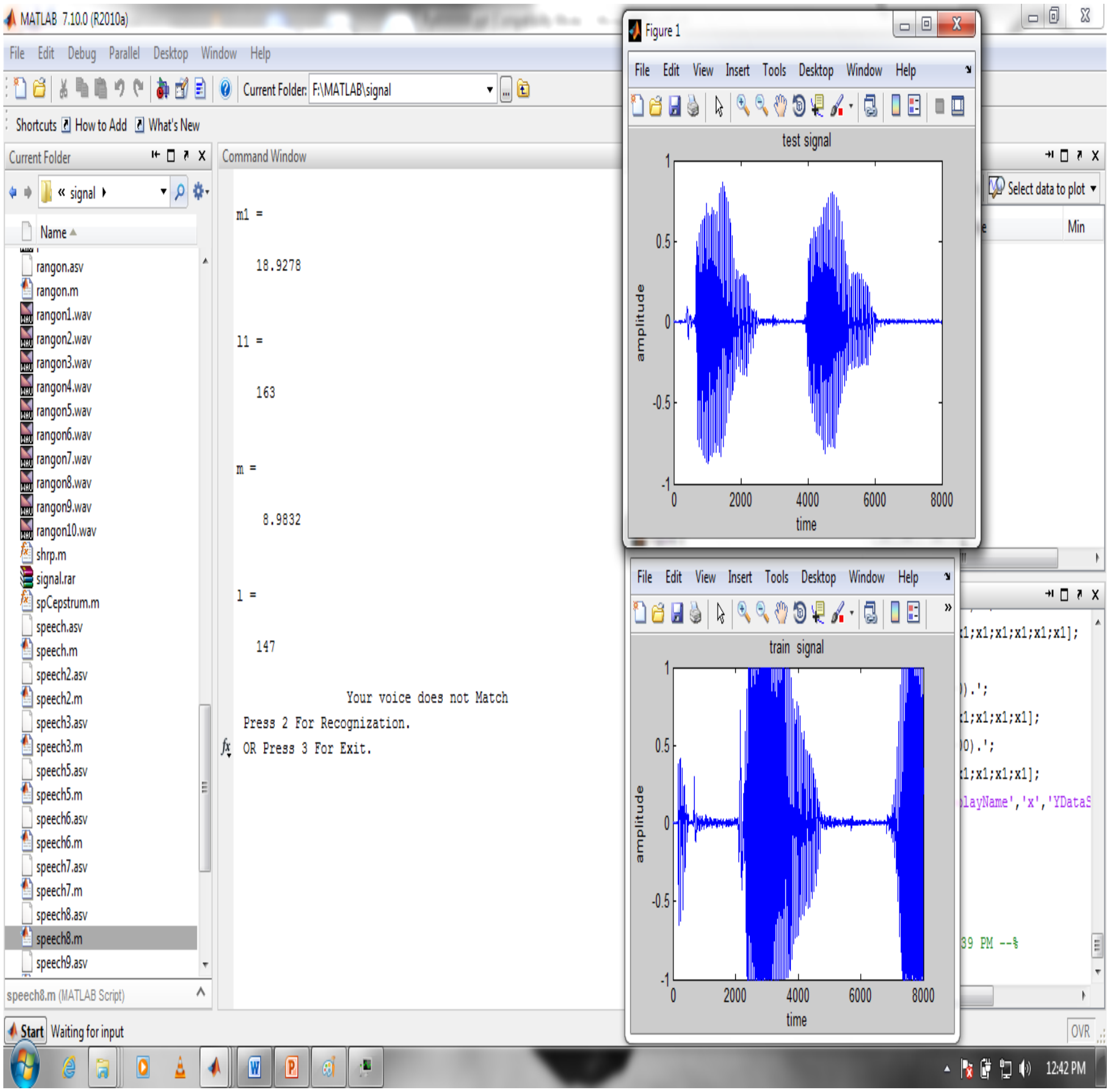
Above figure represents that train part is authorized and test part is unauthorized person.



```
m1 =

    18.9278


l1 =

    163


m =

    41.2791


l =

    161

                He is an othorize person
 Press 2 For Recognization.
 OR Press 3 For Exit.
```

Figure-4.3: Final results of authorized person

This voice is matched, speaker is authorized person. His\her train signal and test signal is same. So door is open.



```
m1 =

    18.9278


l1 =

    163


m =

     8.9832


l =

    147

                    Your voice does not Match
 Press 2 For Recognization.
 OR Press 3 For Exit.
```

Figure-4.4: Final results of unauthorized person

This voice is unmatched, speaker is unauthorized person. His\her train signal and test signal is not same. So door is closed.

Chapter 5

Conclusion

The goal of this thesis was to create Voice Machine Interfacing and apply it to security purpose. The features are extracted of the unknown speech and then it is compared with the stored extracted features for each different speaker in order to identify the unknown speaker. The feature extraction is done by using MATLAB. In this method, the FFT algorithm is used to measure the accuracy. In the recognition stage, individual distortion has been measured based on the difference of accuracy which was used when matching an unauthorized speaker with the authorized speaker’s database.

Finally, it is found that the accuracy based signal provides the faster speaker identification process. Here for the authorize person’s voice signal is matched and door is open.

References

[1] Santosh K. Gaikwad, Bharti W. Gawali, Parvin Yannawar, “A Review on Speech Recognition Technique” Dept. of CS & IT, University of Aurangabad.

[2] Lingfeng Liu, Mahtab Alam, Xi Fu, “Voice activity detection and noise reduction”, Dept. Of Communication Technology, Alborg University, December 2005.

[3] M. Hafidz M. J, S. A. R. Al Haddad, Chee Kyun Ng, “Speech Recognition System for Cerebal Pulsy” Dept. of Communication and System Engineering, University Putra Malaysia.

[4] Jamel Price, “Design of an Automatic Speech Recognition System Using MATLAB”, Dept. of Engineering and Aviation Sciences, University of Maryland Eastern Shore Princess Ann, August 2005.

[5] David Sundermann, “Voice Recognition Matlab Toolbox”, Technical university of Catalonia, Barcelona, Spain.

[6] Kimberly Dawn Vall, “A Methodology of Error Detection: Improving Speech Recognition in Radiology”, School of Computing Science, Simon Fraser University, 2011.

[7] Alexander Seward, “Efficient Method for Automatic Speech Recognition”, Royal Institute of Technology, Stockholm, 2003.

[8] Dr. Joseph Picone, “Fundamental of Speech Recognition”, Dept.of Electrical and Computer Engineering, Mississippi State University.

[9] Paris Smaragdis, Madhusudhana V. S. Shashanka, “A Framework of Secure Speech Recognition”, Boston University.

[10] D. A. Liauw Kie Fa, “Topics in Speech Recognition”, Dept. of Mediamathics, Delft University of Technology, Netherlands.

[11] Jeffrey Adam Bilmes,” Natural Statistical Models for Automatic Speech Recognition” International Computer and Science Institute, October 1999.

Appendix

```
clc;
disp('                              Welcome To Our Speech Recognation World')
```

```
input('                             Press Enter For Main menu ')
prince=input(' Press 1 For Recording. \n Press 2 For Recognization.\n press 3
or Enter for Exit.');
clc;


i=0;
 while(1)
if prince==1;
     clc




% Record your voice for .5 seconds.
input('You have 1 seconds to say Password.\n Press enter when ready to
record--> ');
recObj = audiorecorder;
disp('Start speaking.')
recordblocking(recObj, 1);
disp('End of Recording.');
y = getaudiodata(recObj);
play(recObj);
wavwrite(y,'prince');
    disp('')
    disp('')
    disp('')
disp('                 Successfully Done')
prince=input('Press 2 For Recognization.\n OR Press 3 For Exit')
```

```
clc

elseif prince==2

% Record your voice for 1 seconds.
disp('You have 1 seconds to say your name.')
input(' Press enter when ready to record--> ');
recObj = audiorecorder;
disp('Start speaking.')
recordblocking(recObj, 1);
disp('End of Recording.');
play(recObj);
clc;
plot(B)
title('test signal')
xlabel('time')
ylabel('amplitude')
%wavwrite(B,'prince');
fs=8000;
% perform 8000-point transform
D = fft(B,4000);
x = D.* conj(D) / 4000;
f = 1000*(1:2000)/4000;
figure
%s=x(200:300)
plot(f(1:2000),x(1:2000),'r')
title('a) Power spectral density  of test signal ')
ylabel('Magnitude')
```

```
v=wavread('prince.wav');
figure
plot(v)
title('train  signal')
xlabel('time')
ylabel('amplitude')
D = fft(v,4000);
figure
plot(f(1:2000),x1(1:2000))
title('b) Power spectral density of recording signal')
ylabel('Magnitude')
m1=max(x1)
for j=1:2000
    if x1(j)== m1
     l1=j
    end
  end
m=max(x)
for j=1:2000
    if x(j)==m
        l=j


    end
end


if abs(l-l1)<=4 %&&abs(m-m1)<=3
    disp('                  He is an othorize person')
     p=(abs(l-l1)/l)*100;
    disp('            Persintage of Matching accuracy'),is=100-p;
    disp('                         '),is

    winopen('E:')
else
    disp('')
    disp('')
    disp('')

    disp ('                 Your voice does not Match')
    p=(abs(l-l1)/l)*100;
    disp('            Persintage of Matching accuracy'),is=100-p;
    disp('                         '),is







end
m=max(x);
m1=max(x1);
   prince=input(' Press 2 For Recognization.\n OR Press 3 For Exit.')
i=i+1
if i==4
    clc;
    disp('')
    disp('')
    disp('')
    disp('                                 Password Blocked')
    break
end

%prince==input('                            Press 3 For Exit.\n
OR Press 2 For Recognization. ')



elseif prince==3
    disp('')
    disp('')
    disp('')
    disp('                         Thank You For Your Cooperation')
     break
end
if prince== input('\n\n\n\                 press Enter')
    disp('')
    disp('')
    disp('')
    disp('                         Thank You For Your Cooperation')
     break
 end

 end
```