MUHAMMAD ZAID RAFIQUE
01-134132-138

Speech Recognition and Emotion Sensor System

Bachelor of Science in Computer Science

Supervisor: Dr Tamim Ahmed Khan

Department of Computer Science
Bahria University, Islamabad

15/5/2017

Certificate

We accept the work contained in the report titled "Speech Recognition and Emotion Sensor System" , written by M.Zaid Rafique as a confirmation to the required standards for the partial fulfilment of the degree of Bachelor of Science in Computer Science.

Approved by . . . :

Supervisor: Dr Tamim Ahmed Khan ()

_____

Internal Examiner: Dr. Asfand e Yar ()

_____

External Examiner: Name of the External Examiner ()

_____

Project Coordinator: Dr. Arif ur Rahman ()

_____

Head of the Department: Dr. Faisal Bashir ()

_____

May, 2017

Abstract

Speech is a basic unit for communication or humans. In this decade computer technology is drastically evolving. In this regard, review of existing work on emotional speech processing is useful for carrying out further research. So to make communication easier between human and computer, Speech Recognition provides a great role. Speech recognition is a technique in which a human speaks to computer in his/her comfortable language and computer is such an intelligent to understand his/her spoken words and respond accordingly. Currently there are number of systems which meet this domain of speech recognition and emotion detection. Keeping in mind these constraints we are trying to develop a system which cannot only resolve these issues but also provide simple and usable interface to public as well. We are aiming to provide some basic professional modules on one platform which are Speech Recognition, Emotion Prediction and separating two Classes of any voice input. Our system shall be a good example of proving credibility of voice signature as it shall offer a very unique service which is very rarely offered in public domain. It shall be able to separate multiple voices which are enclosed in one audio signal. This means we can hear our chosen voice at one time. Speaking style classification is improved by using the modulation spectrum in combination with standard pitch and frequency variation.

Acknowledgments

We are indebted to Almighty Allah, Lord of the Universe and His Holy Prophet (PBUH) whose blessings enabled us to perceive and pursuit higher ideas in life and gave us the strength to complete this project.

We extend our deepest gratitude to our project Supervisor, Dr. Tamim Ahmed Khan, who remained a source of inspiration and motivation for us through the course of the project. .

MUHAMMAD ZAID RAFIQUE
Islamabad, Pakistan

May , 2017

"We think someone else, someone smarter than us,
someone more capable, someone with more resources will solve that problem.
But there isn't anyone else."

Regina Dugan

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Speech Recognition and Emotion Sensor system shall be able to process and differentiate multiple voices (min 2) inputs. These inputs shall then be converted into text and compared to the database for Emotion predictability. Every voice input shall be recorded and stored in database separately. Voice input shall be taken through a MIC. The System shall be able to convert that input into Standard English language. The spoken words and phrases shall also be saved into the database for further use. Classification of voice inputs shall be the next task of the system in which different classes shall be separated according to some key features which include frequency and pitch. Emotion predictability through comparison with spoken words is not a very efficient but still a significant approach. This System shall be able to detect emotion of any user based on what he/she has spoken. Usually emotion detection is applied in signal processing through different techniques. Some those are measuring pitch and frequency. Even those technologies are not efficient enough. This System is a test case as this is very first of its kind. The only module among three of this system which is previously applied is Speech Recognition. Classification of voice inputs technique is not a new technique but due to sensitivity and limited use, this is not found in public domain. These kinds of technologies are generally possessed by Law Enforcements Institutions who do not reveal their technologies so often. So here all the three unique modules are combined for achieving maximum and efficient results. The system shall produce the result in minimum time consumption.

1.1 Project Background

The work that is being done in the 21st century involves the use of technology that will be beneficial for future works. Voice Recognition and Emotion Sensing is currently an emerging technology in the field of Computer Science. The leaders in this innovation are Korea and Japan followed by USA and then European countries Germany, Italy, Sweden and Denmark etc .This System shall provide the interaction between the human and the computer so such technology-based interface will be able to do a better job than the

commonly used interface devices and can also be used for future exploits to extend their function capabilities.

Technologies in industry are growing significantly in this field last two decades. As a result of this, more and more Human Voice Recognized products are being developed which are far more advanced than the previous ones.

At the same time, every telecom and Smartphone manufacturing company is adapting this technology proving its significance and hence has a vast window opportunity in the field. Same goes for the public demand as general public is also preferring and moving towards easy interaction with their electronic devices, especially for the ones that involves high risk work task. Due to technologies are getting higher, the development fees also increase with it. Thus, the products that equipped with high technologies are getting higher in price as well.

1.2 Objective

"To design a Speech Recognition and Emotion Sensor system for better technological processing for monitoring, surveillance and security purposes."

1.3 Problem Description

It is impossible for any human being to process and differentiate multiple voice inputs especially when voice inputs are forged or not so accurate. Non availability of libraries and processing time are the core problems in Speech Recognition systems. Separating multiple inputs w.r.t their pitches and prediction of Emotions are also the problems which shall be countered in this application. Voice Recognition and Classification of Voice classes are rarely used in general public domain as it has very limitations and to some extent restriction as well. This technology is no doubt very efficient and has a lot of potential as well but as far as the usage is concerned, only big companies are able to afford it. So as noted, there is a demand in building this technology that has a low production cost and perform efficiently in a time sensitive manner which would otherwise be quite impossible for a human to perform these tasks [1]. Pakistan lacks in the innovation of this kind of technologies, can enhance and use this as test case for improvement of research and development methodologies.

1.4 Project Scope

Voice communication between people all around the world is very much necessary and understandable. With the passage of time technology has evolved a lot and it has impacted the communication systems as well. Speech Recognition systems are not new to human

race but it hasn't progressed as much as other systems. There are various reasons behind this but the most prominent is unpredictable human behavior. This application shall be able to resolve all above mentioned issues to the approximated level. Voice Activity shall be detected and Emotions shall be extracted from spoken phrases and sentences. It is also important to mention that separation of voice signals is no doubt not usually required in general public usage but still it has a lot importance in law enforcement agencies for security purposes and in some specific companies who manufacture and build sensitive security facilities

1.5 Feasibility Study

With above defined scope, would you be able to meet your project schedule? Do mention following aspects:

1.5.1 Risks Involved

Initially it shall be difficult to resolve all issues like pitch definition and receive multiple voice activity at a time.

1.5.2 Resource Requirement

SQL server, Microsoft Visual Studio, MATLAB, MIC

Chapter 2

Literature Review

The Speech Recognition which is often referred to as Automatic Speech Recognition Or Computer Speech recognition converts spoken words to text. The term "Voice Recognition" sometimes referred to as Speech Recognition where the Recognition System is trained to a particular language. There Emotion Detection and Recognition from speech is a recent field of research that is closely related to Sentiment Analysis. Sentiment Analysis aims to detect positive, neutral or negative feelings from text or speech, Whereas Emotion Analysis aims to detect and recognize types of feelings through the expressions of texts and through variations of audio signals such as anger, disgust, fear, happiness, sadness and surprise. The plan is keeping in mind its commercial achievement and given shape in the form of drawing and art. While in the study of these drawings, careful considerations must be taken of the availability of money, resources and in materials required for successful target of the new conceived idea into reality. While designing any machine components, it is necessary to have good background knowledge of vast subjects such as Mechanics, Mathematics, the specifications of materials used.

2.1 Existing Applications

Currently there is no System (not that we could find) that offers this functionality which we are designing but there are number of systems which meet this domain of speech recognition and emotion detection individually [2]. Speech is a basic unit for communication or humans. In this decade computer technology is drastically evolving. In this regard, review of existing work on emotional speech processing is useful for carrying out further research. So to make communication easier between human and computer, Speech Recognition provides a great role. Speech recognition is a technique in which a human speaks to computer in his/her comfortable language and computer is such an intelligent to understand his/her spoken words and respond accordingly. Speech recognition technology helps to convert recognized and spoken language into text, image or any event stirring by computers and other computerized devices such as Smart Technologies and robotics. Speech recognition

system can be developed for the grammatical structure and some statistical model can be used to improve word predication, but still there is a problem that how much world knowledge of speaking and encyclopedia can be modeled? Of course, we cannot model the world knowledge. So we cannot measure computer system up to human comprehensive. Currently there are number of systems which meet this domain of speech recognition and emotion detection. Some of the existing systems are as following: [2] [3]

- Cortana (Microsoft corp.)
- Siri (Apple.inc)
- IATROS (Linux)
- CMU SPHINX
- Sochi (phys.org)
- Speech Technol
- Speechnotes
- Braina
- Lilyspeech
- Cue-me
- Fusion Speech

2.1.1 Drawbacks

Although these systems are efficient in their particular environment but none of these are developed in altogether modules which we are implementing. These systems are developed according to some specific accents or even languages [2]. Intelligence agencies might have the efficient version of these types o systems for surveillance purposes but those are not in public domain so we do not have much knowledge about that [3].

- Language
- Accent
- Time Consuming
- Reliability
- Manageability

2.2 Proposed System

Keeping in mind these constraints we are trying to develop a system which can not only resolve these issues but also provide simple and usable interface to public as well. The current systems are mostly developed in MATLAB which takes too much time on runtime. We are aiming to provide some basic professional modules on one platform which are Speech Recognition, Emotion Prediction and separating two Classes of any voice input. Separating multiple classes of any audio signal means if any audio signal contains more than one voice inputs, we shall separate these inputs from each other for understandability and clear audio. This system shall have some of the following features:

- Typical features are the pitch, the formants, and the vocal tract cross-section areas.
- Mel-frequency cepstral coefficients [3].
- Phase/frequency modulation is represented by the sub band instantaneous frequency.
- Emotions shall be examined from variations of signal.

Our system shall be a good example of proving credibility of voice signature as it shall offer a very unique service which is very rarely offered in public domain. It shall be able to separate multiple voices which are enclosed in one audio signal. This means we can hear our chosen voice at one time. Speaking style classification is improved by using the modulation spectrum in combination with standard pitch and energy variation.

Chapter 3

Requirement Specifications

The purpose of this document is to describe the external behavior of the system. Requirements Specification defines and describes the operations, interfaces, performance, and quality assurance requirements of the system. The document gives the detailed description of both functional and non-functional requirements. From this document, this system can be designed and developed. The document is developed after number of consultations with the client and considering the complete requirement specifications of the given Project. The clear understanding of the system and its' functionality will allow our team to construct the appropriate system for the end user.

3.1 Functional Requirements

The Functional Requirements of our Final Year Project are as following:

3.1.1 MIC

Audio shall be taken through the Mic as input.

3.1.2 Detect Audio

Application shall detect Audio and represent it in waveform.

3.1.3 Separate Classes

All classes shall be separated w.r.t their features.

3.1.4 Speech to Text

Separated classes of audio shall be converted to text.

3.1.5 Emotion Detection

Emotions shall be predicted on the bases of text and phrases.

3.1.6 Hardware Requirements

The hardware Requirements of our Final Year Project are “MIC” and “LAPTOP/PC”

3.2 Non Functional Requirements

The Non Functional Requirements of our Final Year Project are as following:

3.2.1 Reusable

The application will be maintainable and reusable because it Shall be developed by using C#.

3.2.2 Accessibility

System should be accessible by the authorized user.

3.2.3 Performance

System should respond fast while sending/receiving information to/from the server.

3.2.4 Security

System should be secure in regard of sensitive data and from different viruses as well.

3.2.5 Availability

System shall be available at the particular time when required.

3.2.6 Adaptability

As we know these type of systems are meant to be complex but it shall be much more user friendly so that general public can use it without any hesitation.

3.2.7 Response Time

The response time shall be in seconds.

3.3 User Interface

I have used Buttons in our GUI which are reliably much easy to use and interact. Visibility and Control are always the key features in any GUI which we have encountered carefully.

3.4 Performance Requirement

The Laptop/PC should have the minimum system requirements for the System. The software can even run on a normal processor and the memory system. But to make it work faster the processor and the memory should be compatible with the ones the application is designed for. The system will take few seconds to process each audio and hence generate results.

Chapter 4

Design View

This chapter gives the detailed description of the system "Speech Recognition and Emotion sensing" through diagrams. The high and the low level design of the system is shown in the chapter for the better understanding of the working of the system. The complete overview of the system is explain through the different diagrams like Activity diagram, Flow Chart diagrams, Use Cases and Class diagram show the association between the classes of the system. The design ensure that the system is fullling the required specication.

4.1 Flow Charts

The flow diagram of the system shows the flow of things or people's actions regarding the sequence of movement in any complex activity.

4.1.1 Speech to Text Flow Diagram

The Figure 4.1 shows the mechanism of one of the modules. The figure elaborates how speech to text module works involving database and API

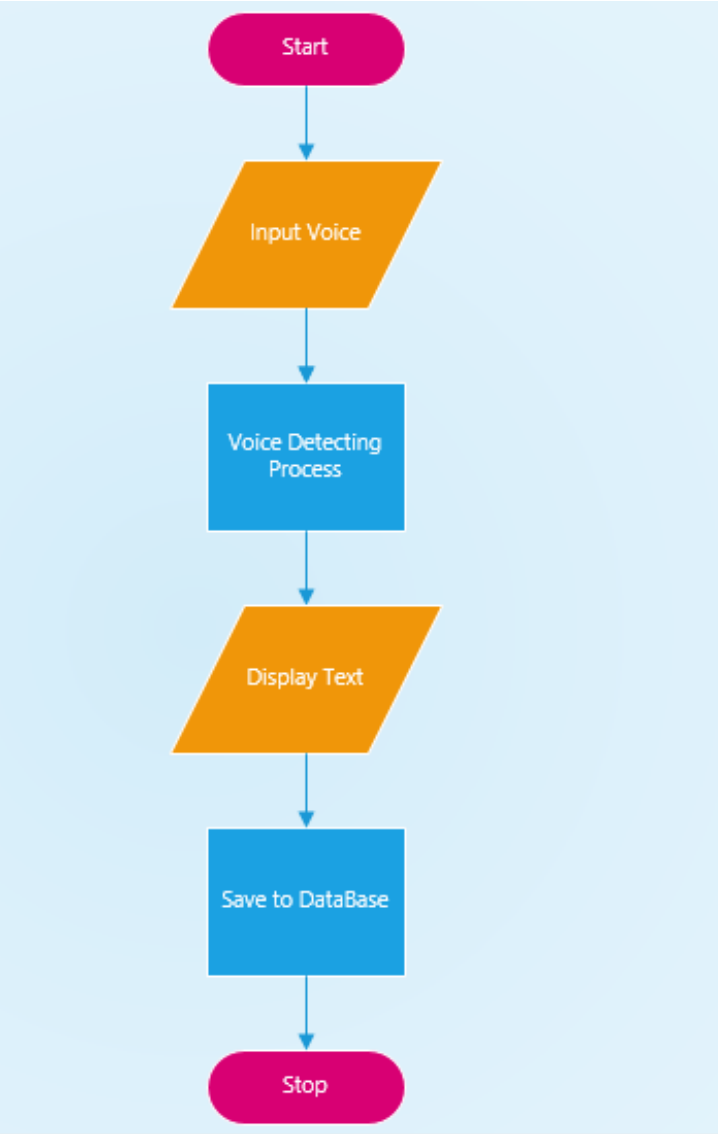

Figure 4.1: Speech to text

4.1.2 Emotion Detection

The Figure 4.2 shows the flow of data in emotion sensing module which takes the voice as an input and then compares with the database. The result is displayed on screen.

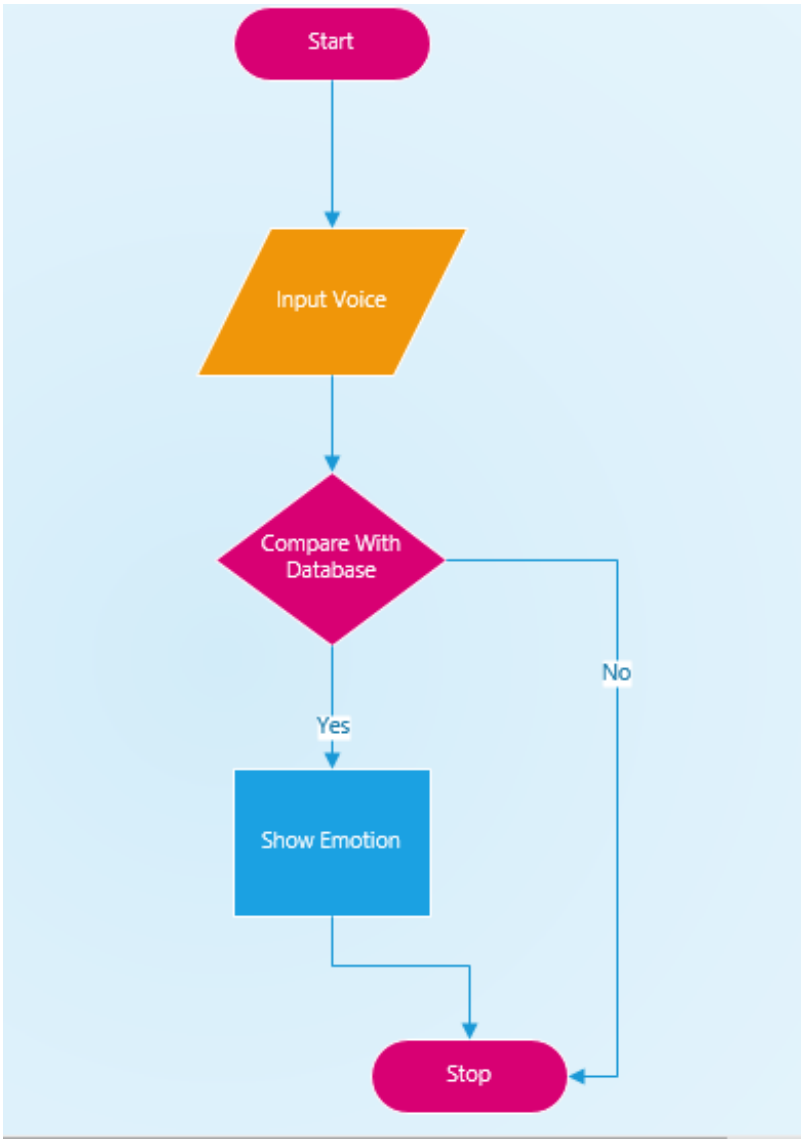

Figure 4.2: Emotion Detection

4.1.3 Classification of Voice Flow Diagram

Here is the flowchart in Figure 4.3 of the third module which is separation of voice classes in any given mixed signal. Features are extracted and processed which are then analyzed. The result is in separation of classes.

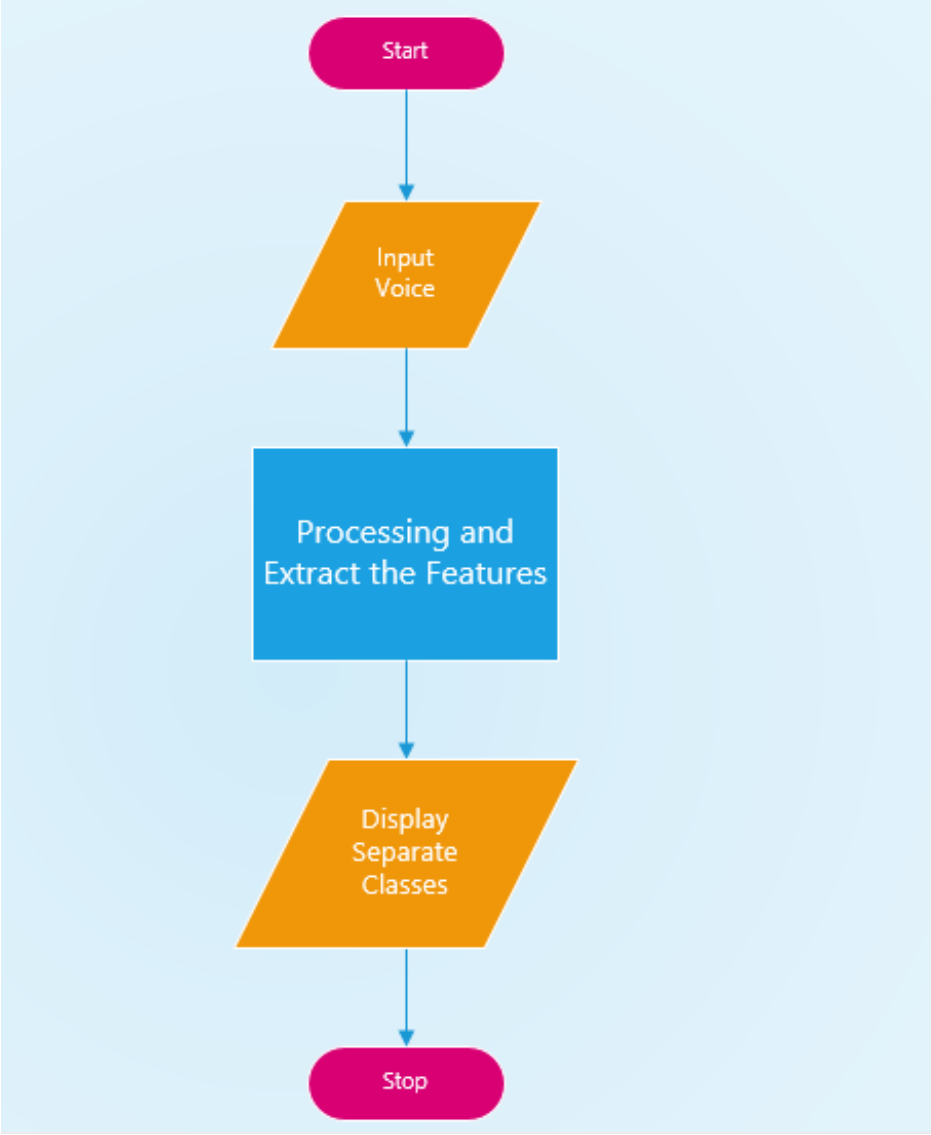

Figure 4.3: Features Extraction

4.2 Deployment Diagram

The below deployment diagram in Figure 4.4 shows the behavior of system when it is finally deployed for the user. All the modules and their working mechanism is shown in this diagram. All of the three modules are combined on a single platform which leads to the complete system.

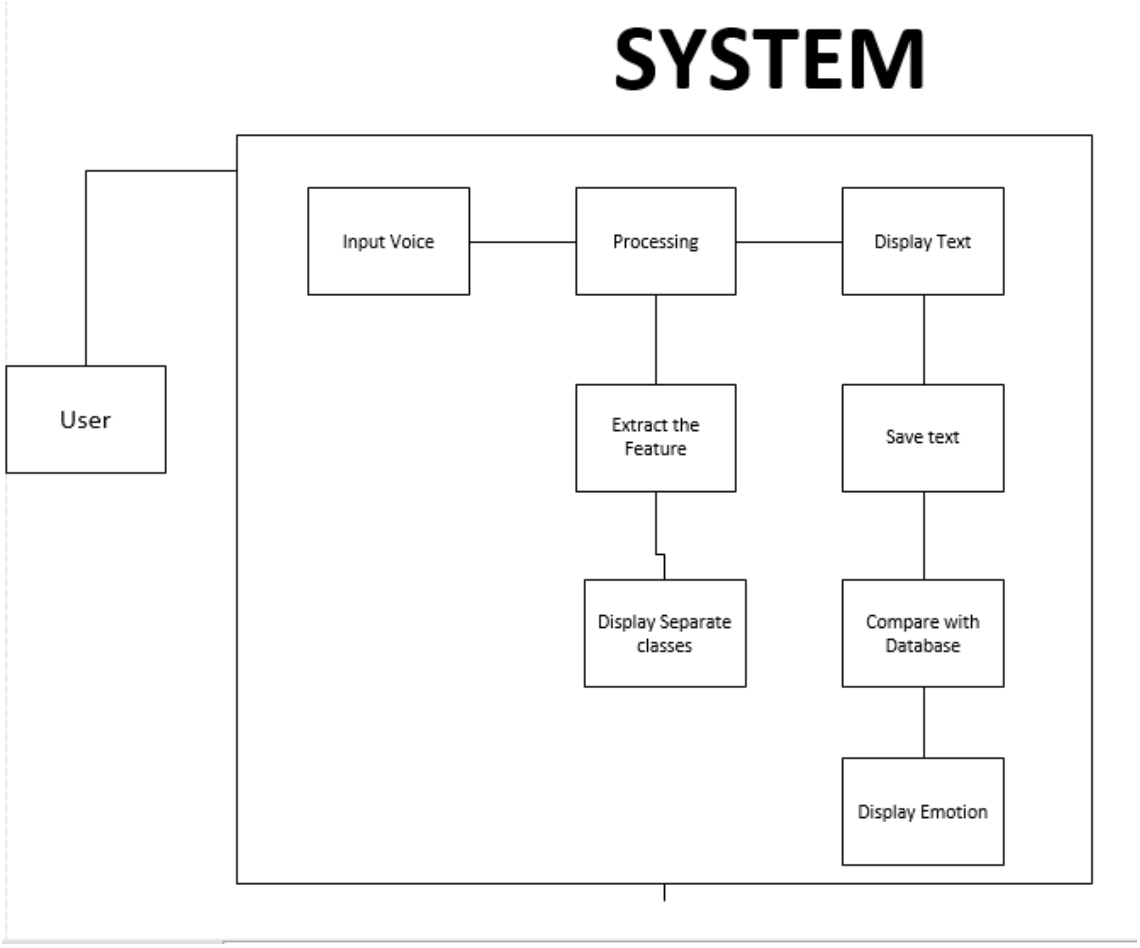

Figure 4.4: Deployment Diagram

4.3 Sequence Diagrams

Sequence diagrams is and interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arrange in time sequence and tells how user interact with the system.

4.3.1 Sequence Diagram of Speech to Text

The Figure 4.5 shows the sequence of the speech to text module. The user interacts with system with his voice as an input. That voice input is converted into Text and saved into the database.

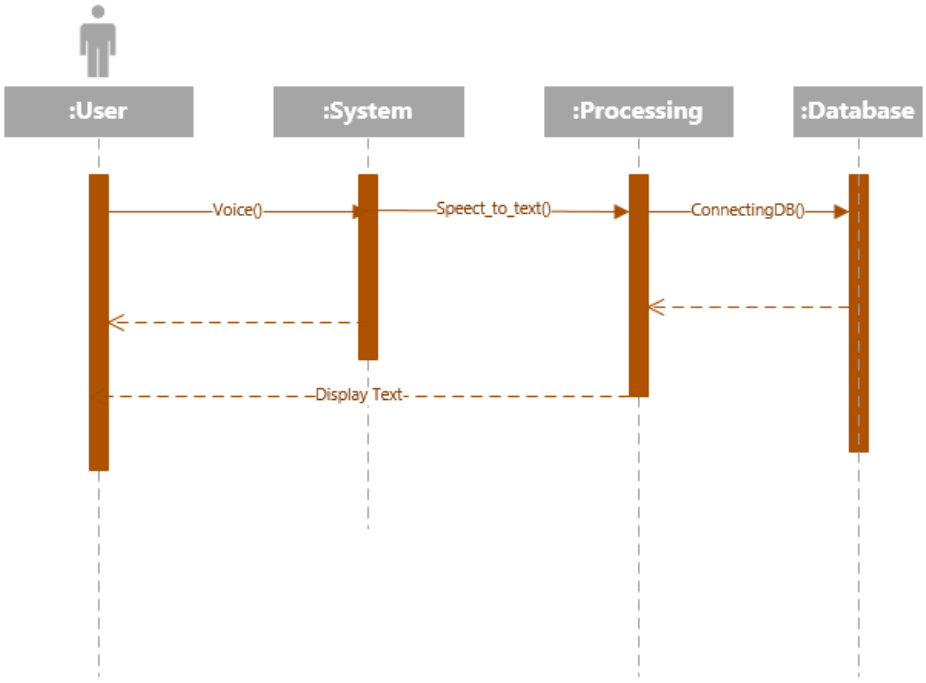

Figure 4.5: Sequence Diagram of Speech to Text

4.3.2 Sequence Diagram of Emotion

In the below Figure 4.6 , the sequence of another module emotion predictability is shown. This involves the texts which user has already spoken and stored into the database are compared with the keywords, phrases and sentences which are classified into different emotions. These specific words, phrases and sentences are stored in database w.r.t its particular emotion.

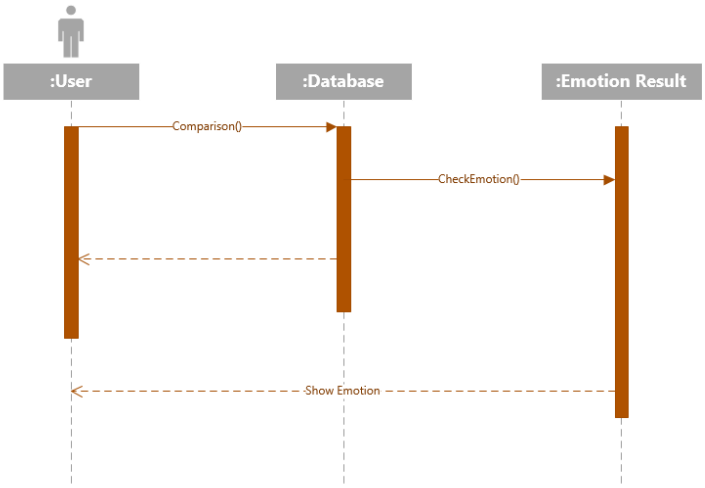

Figure 4.6: Sequence Diagram of Emotion Detection

4.3.3 Sequence Diagram of Voice Classification

The Figure 4.7 shows the sequence of last module which is separation of voice classes. Voice input is taken through a Mic and this contained signal is processed and analysed according to its features. The extraction of features leads to separation of voice signal.

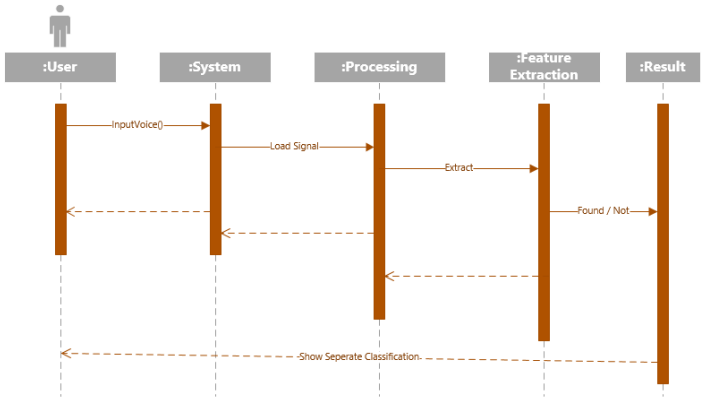

Figure 4.7: Sequence Diagram of Feature Extraction

4.4 Use Case

A use case is a list of actions or event steps, typically defining the interactions between an actor and system, to achieve a goal. The actor can be a human or any other external system.

4.4.1 Use Case of Speech to Text

The following Figure 4.8 shows the usage of speech to text module individually.

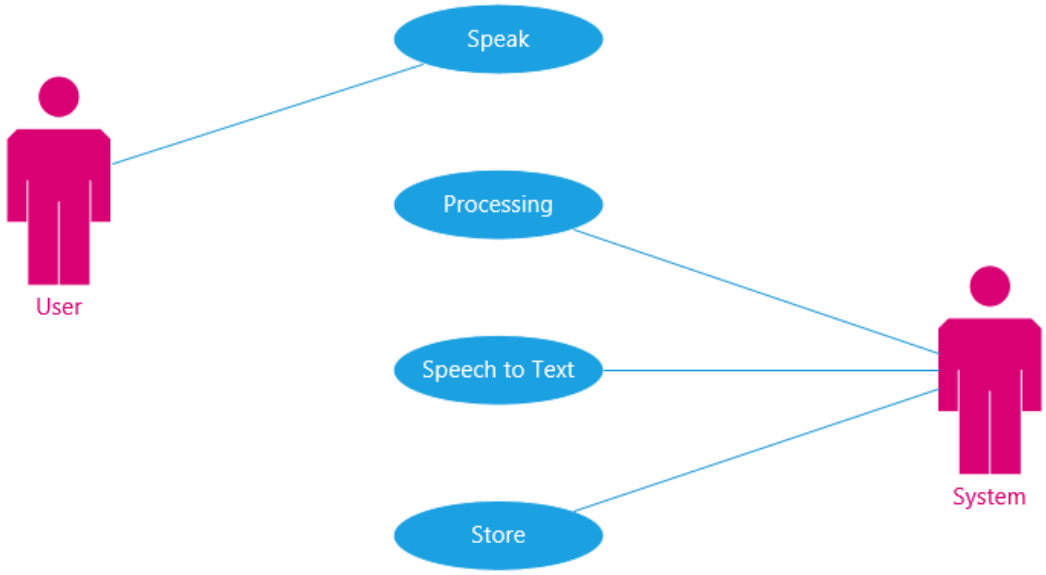

Figure 4.8: Use Case of Speech to Text

Detailed description of Use case 1

Table 4.1: Use Case of Speech to Text and Save Text

Use Case ID:	1.0
Use Case Name:	Convert to Text and save to database
Actors:	Any Person
Description:	Actor will Speak and system detect the voice and convert it to the wave form and then Convert into text form and save the text in database.
Preconditions:	Running application
Priority:	High

4.4.2 Use Case of Emotion

In this following Figure 4.9 , the use case of emotion detection module is shown.

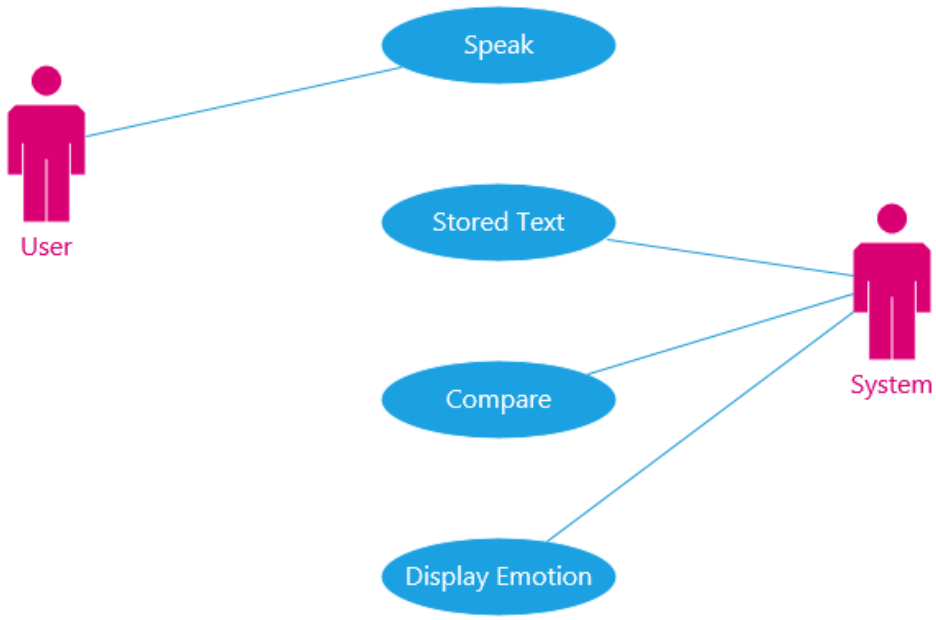

Figure 4.9: Use Case of Emotion Detection

Detailed description of Use case 2

Table 4.2: Use Case of Emotion Detection

Use Case ID:	2.0
Use Case Name:	Emotion
Actors:	Any Person
Description:	Actor will Speak and train the system to tell the emotion of a person. System will compare with the database and then display the emotion of user.
Preconditions:	Running application
Priority:	High

4.4.3 Use Case of Voice Classification

In this following Figure 4.10, the use case of separation of voice classes module is shown.

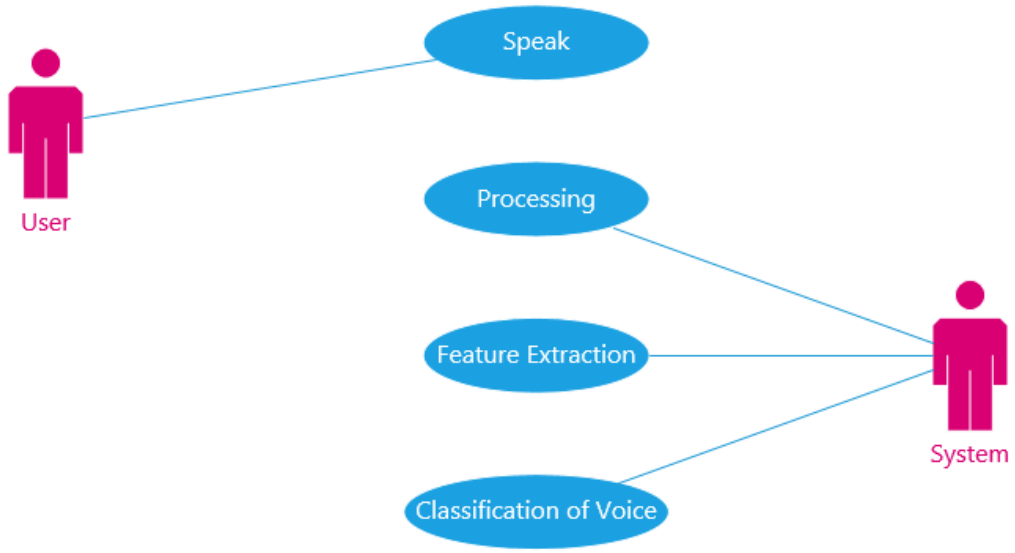

Figure 4.10: Use Case of Feature Extraction

Detailed description of Use case 3

Table 4.3: Use Case of Feature Extraction

Use Case ID:	3.0
Use Case Name:	Feature Extraction
Actors:	Any Person
Description:	Actor will Speak and system detect the voice and convert it to the wave form and then extract the features of that wave with respect to their pitch and frequency and at the end display the classification of each class.
Preconditions:	Running application
Expectation:	Wrong Input Conflicts about:- • Emotion • Voice Features
Priority:	High

4.5 Activity Diagram

Activity Diagram can be shown in Figure 4.11.

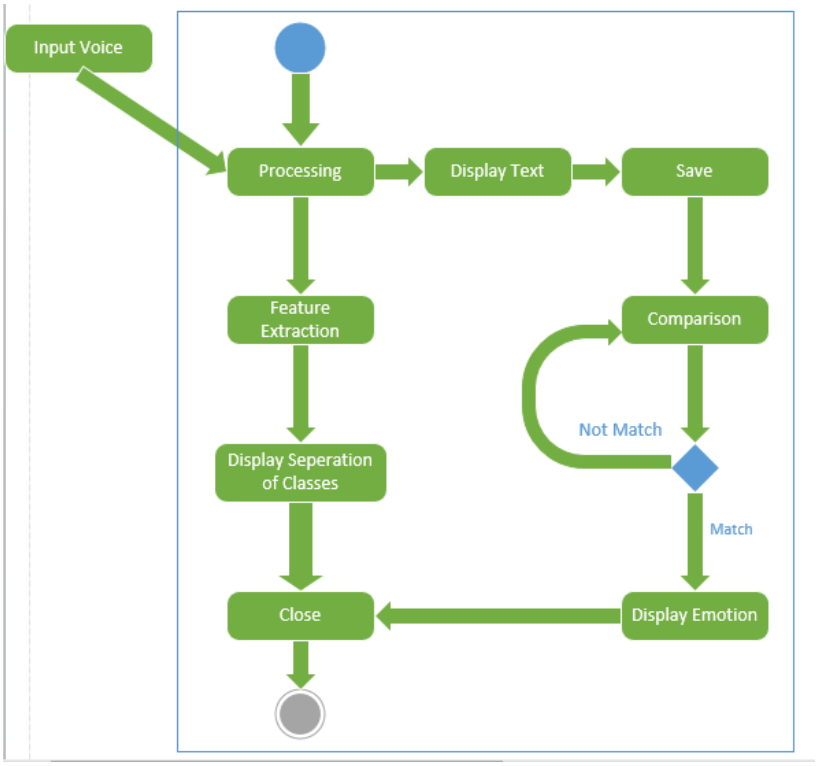

Figure 4.11: Activity Diagram

4.6 Design Constraints

Large vocabulary and infinite language. Linguistic patterns are hard to characterize and solution must understand and represent what is invariant.

Chapter 5

System Implementation

This chapter describes the implementation details of the SRESS (Speech Recognition and Emotion Sensing System). Multiple algorithms are used in the systems which are briefly discussed. Furthermore this chapter also describes tools and technologies used in the implementation. System implementation uses the structure shaped throughout architectural design and the consequences of system analysis to construct system fundamentals that meet the stakeholder necessities and system requirements established in the initial life cycle stages. These system fundamentals are then assimilated to form intermediate aggregates and finally the comprehensive system.

5.1 Software Architecture

The general architecture of the application has been stated in the following. Proposed project is a desktop base application that will provide three modules to the user. Considering the constraints of the system user will have to interact with the system. User can select any of the three modules to execute. Three modules include "Speech to Text", "Emotion Sensing" and "Classification of Voice Input".

5.2 Development/Environment Languages Used

Following are the languages and platforms which are used for the development of system:

- MATLAB
- .NET FRAMEWORK
- SQL SERVER

5.2.1 MATLAB

Data is being input through a .wav file that can contain voices of one or more individual. The .wav data is converted into a binary data and the combined voice is sampled into binary data. The time domain samples are converted to frequency domain using FFT function shown in Figure 5.1. Once the FFT function is used the data is now converted into imaginary and real part shown in Figure 5.2, But the imaginary and real part are plotted for analysis [4] [3]. Both the imaginary and real values will be required to reconstruct the signal in later stages. The magnitude plot of the FFT function is analysed and the fundamental frequency components are drawn out, Pitch consideration is also very important so the magnitude of the frequency is also analysed. After establishing the frequency present in the system of the two voices shown in Figure 5.4 without edges of signal, the next step is to filter out the voices. For separation of the voices we used band pass and high pass filter. We determine the fundamental frequency of the voice pattern and use the filter to separate the two frequency components [1]. After filtering out the frequency components, the values are passed through the IFFT function to re generate the signal that contains the separate voices shown in Figure 5.3 and Figure 5.5

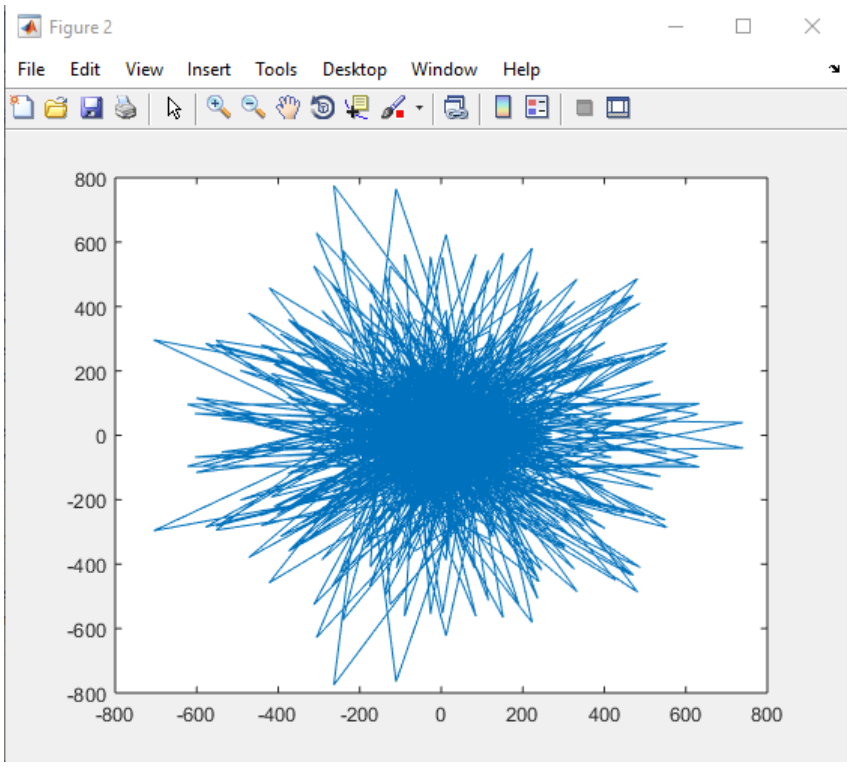

Figure 5.1: FFT of Signal

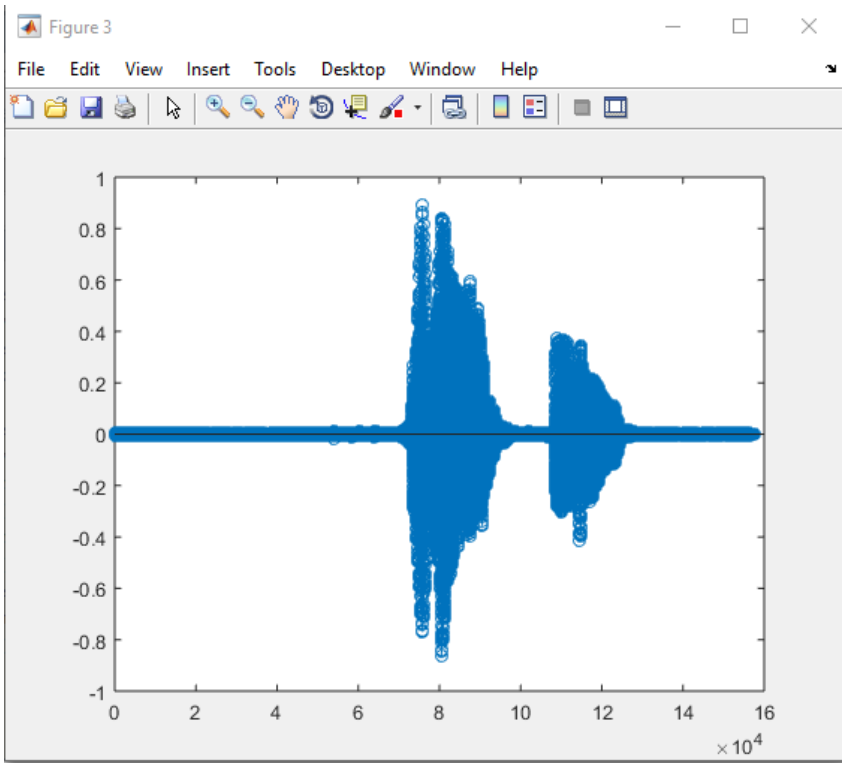

Figure 5.2: Full Signal

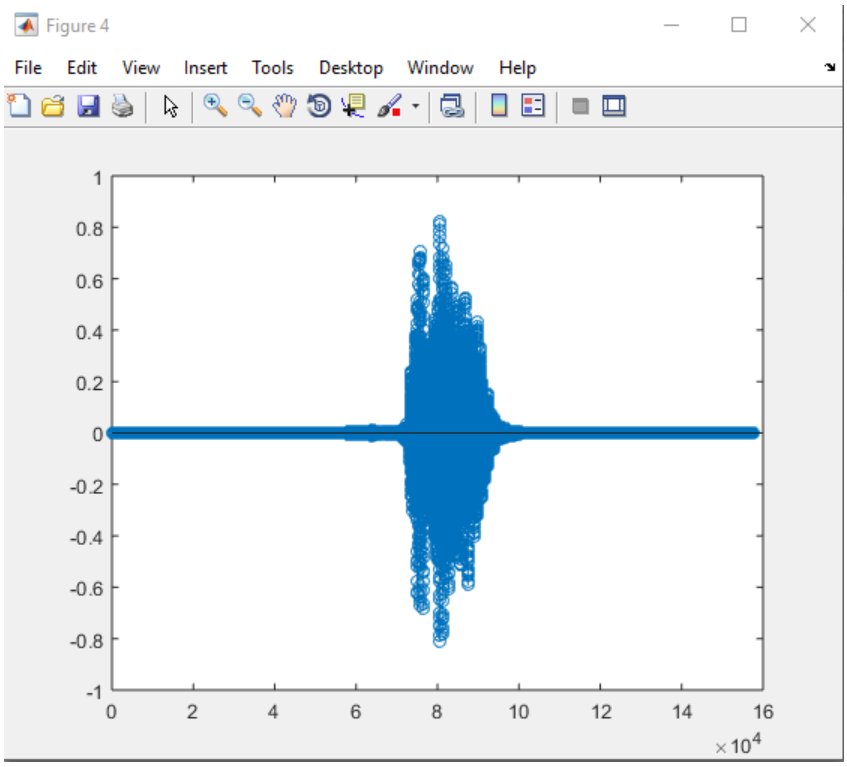

Figure 5.3: Separated Signal

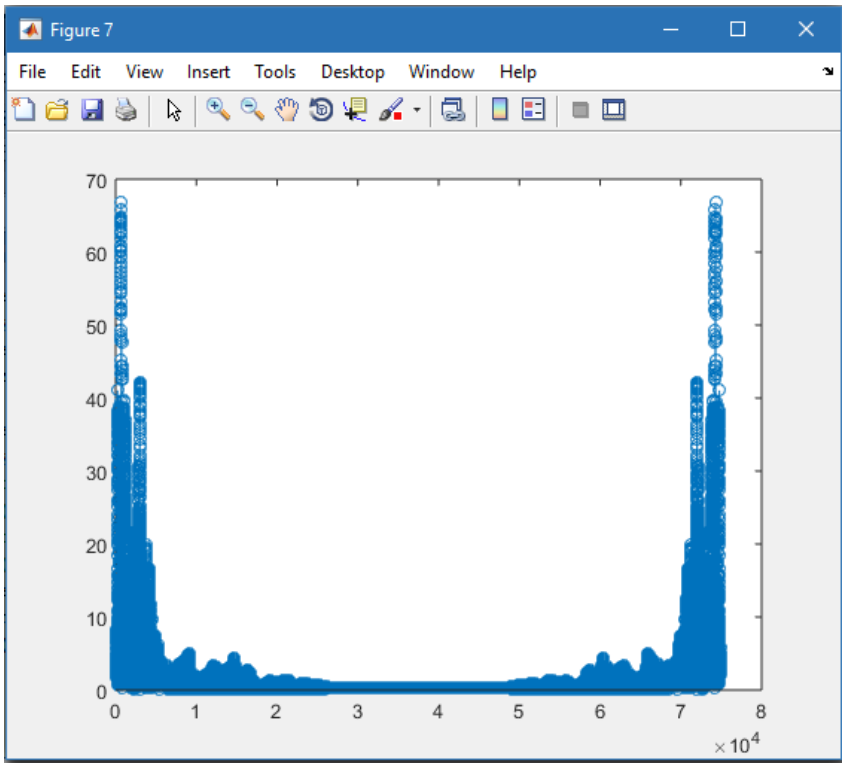

Figure 5.4: Full Signal Without Edges

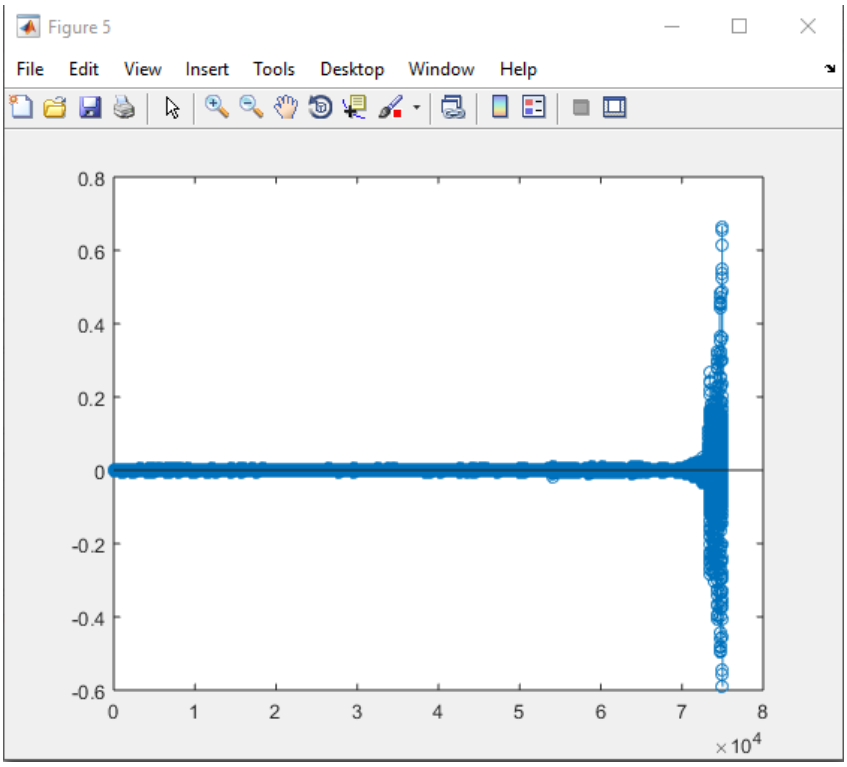

Figure 5.5: Separated another Signal

5.2.2 .NET FRAMEWORK

Microsoft Visual Studio is a consolidated change environment (IDE) from Microsoft. It is used to make PC programs for Microsoft Windows, and also destinations, web systems and organizations. Visual Studio uses Microsoft programming change stages, for instance, Windows Programming interface, Windows Outlines, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can make both nearby code and directed code. Visual Studio supports particular programming. Worked in dialects consolidate C, C and C/CLI (by method for Visual C), VB.NET (by method for Visual Fundamental .NET), C# (by method for Visual C#), and F# (as of Visual Studio 2010[7]). Support for various dialects, for instance, Python, Ruby, Node.js, and M among others is open by method for tongue organizations presented autonomously. It also bolsters XML/XSLT, HTML/XHTML, JavaScript and CSS. Java (and J) were supported some time recently.

The Below Figure Figure 5.6 show the main interface of our application.

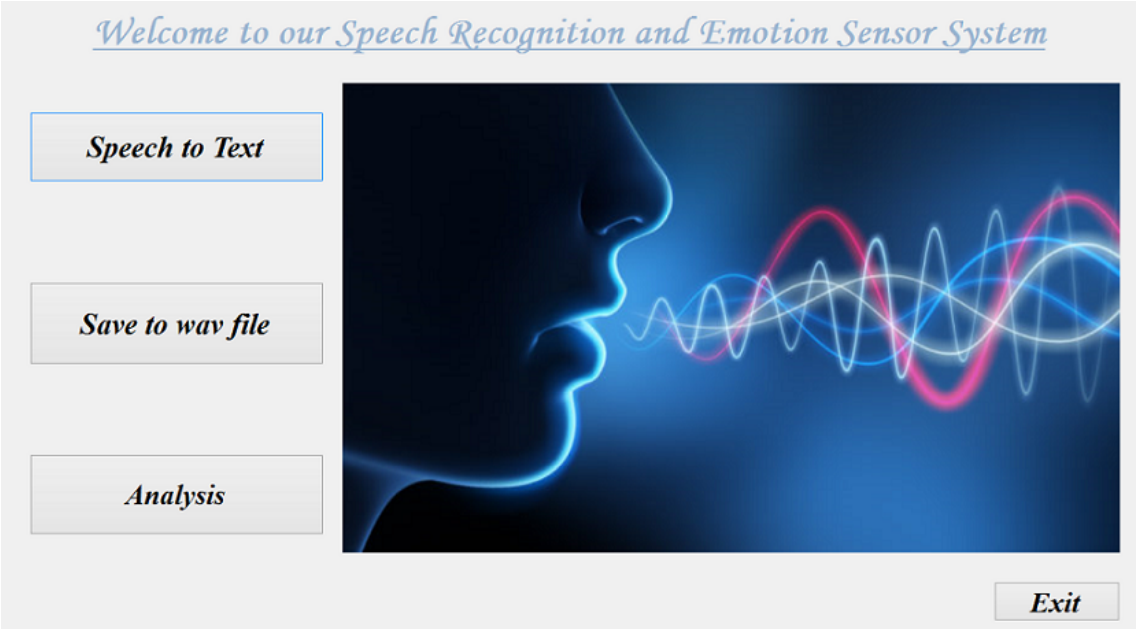

Figure 5.6: Main Form

5.2.3 SQL SERVER

SQL Server is a well-known database provided by Microsoft, the reason for using SQL Server in system is to setup keywords and phrases for emotion predictability of users so no one other than authentic emotions are detected. Our database also contains the phrases spoken by users via microphone. Both SQL Server and C Sharp are provided by Microsoft That's why there is no compatibility problem and don't need to use third party tools for database interaction. Our database include "5" emotion classes which are flooded with respective keywords of that emotion.

5.3 System Methodology

Speech Recognition and Emotion Sensing System is comprised of three modules. All these modules are set on independent and single platforms which means these can be run independently as well as combined on a platform as. Speech to text module is developed in C#. This module is developed in MATLAB and called in C using open cv. Voice input is taken from a microphone and converted into text using an API [2]. Separation of Voice Classes is the second module which separates the different voice classes w.r.t their certain features. Different algorithms are also applied in this feature extraction technique. For third module Database is created which not only stores spoken words from user but also the specific keywords for 5 different emotions. The spoken words are compared with the specific allotted keywords for specific emotion and thus the Emotion is detected. All of the three modules are joined in such a way that the purpose of this system can be achieved fulfilling all the standard functional and non-functional requirements.

5.4 System Components

This System is divided into three modules which are also components of the system. All of the components have their unique functionality. Speech Recognition module is for conversion of spoken words into text whereas Classifier module separates the different voice classes. The third and the last component detect the emotion of the user.

5.5 Processing Logic/Algorithms

System implementation has been roughly divided into three major phases keeping in consideration the scope of the proposed application .The details of implementation methodology and processing logic are summarized in this section.

5.5.1 Speech to Text

- This module accepts voice input through microphone and converts that input into text using “API”.
- This converted text is stored in database (SQL SERVER).
- This module is implemented in Microsoft Visual Studio independently.

5.5.2 Classification of Speech Signals

This module classifies the mix signal into separate voice classes by applying different algorithms like “Fast Fourier Transform” and “K-Mean”. Fast Fourier Transform and K-Mean algorithms are applied in MATLAB which happens to be a lot easier than in .NET

Framework. Classification is based on voice frequency, pitch and some other features like energy and density. Frequency Band Filters are used to avoid distortion in the signal [5].

5.5.3 Emotion Predictability

Emotion predictability is based on what the user has spoken instead of frequency and pitch variations of speech signals. This is more authentic and useful approach as variations in speech signal can have multiple difficulties. Emotions are detected by comparing already stored text and phrases.

5.5.4 Algorithms

There are multiple algorithms are used in the system which are given as following: [4] [6] [1] [3]

- Fast Fourier Transform
- K-Mean
- Reverse Fast Fourier Transform
- Band Filtration

Chapter 6

System Testing and Evaluation

This section is meant for the validation and verification of the software according to the established and proposed requirements. This phase has significant importance in the development life cycle because history has proven that a tiny bug in the software may result in loss of money and human life due to the negligence in the testing phase. Software testing process ensures that each module of the product is working and providing results according to the requirements. Each of the individual modules of the application is tested individually and complete application is also tested. The main objective of testing is to check whether the developed software meets the required quality standards or not. Testing is also aimed at determining whether the application is providing the desired result.

6.1 Graphic User Interface Testing

GUI testing is aimed at measuring the ease with which the user can be use the system as shown in Table 6.1 . It also ensures the ease of interactivity of the application developed. As shown in the previous sections our system's graphical user interface is user friendly for any non-computer literate personal. The desired actions are achieved through simple button clicks like starting the application, loading the image and generating the report. Proper exception handling is also provided to discourage malicious users.

Table 6.1: TC_01:Testing GUI

Test Case ID	TC_FUNCT_01	
Description	Testing the GUI.	
Intial Condition Expected Result	Task	
Step	Task and Expected Result	Status
1	Open the Application	PASS
2	Graphics up to the mark	PASS
3	Verify that buttons are working	PASS
4	Activation and creation of correct dataset	PASS

6.2 Software Performance Testing

Software performance testing is the process of checking the performance, efficiency and reliability of the system. In general, the performance of the developed system is effective, reliable and efficient.

6.2.1 Load Testing

As our system is developed in MATLAB and Csharp, so for the single platform MATLAB code is called in C by adding libraries and DLL files.

6.2.2 Testing Strategies

Dynamic and static testing are the two procedures used to test the working of software and analyzing the satisfaction of user. Dynamic testing includes the checking of work flow and its rightness, reliability and dependability. Static testing will include the incorporation of necessities of programming in the design.

6.2.3 Component Testing

All of the three modules are tested according to their functionality and environment. Tests and their results are given below. First module's test case which is "Speech to Text" is shown in Table 6.2.

Test Case 1

Table 6.2: TC_02:Speech to text

Test Case ID		TC-1
Description		Test that the voice input is converted into text or not?
Applicable for		Window desktop systems
Requirements		Microphone and System.Speech.Recognition Library
Initial Conditions		English Language
Step	Task and Expected Result	Status
1	System allows Microphone to feed voice input	PASS
2	System is converting into Text	PASS

Second module which separates voice classes is tested and given below in Table 6.3.

Test Case 2

Table 6.3: TC_03:Feature Extraction

Test Case ID		TC-2
Description		Test that the voice classes are separated (Feature Extracted) or not?
Applicable for		Window desktop systems/MATLAB
Requirements		Voice signal contains voices of at least 2 people
Initial Conditions		Human Understandable Language
Step	Task and Expected Result	Status
1	System accepts .wav file	PASS
2	System is separating Voice Classes(Feature Extraction)	PASS

Test Case of third and the last module which detects Emotion is shown in Table 6.4.

Test Case 3

Table 6.4: TC_04:Emotion Detection

Test Case ID		TC-3
Description		Test that System is matching Emotions or not?
Applicable for		Window desktop systems
Requirements		Voice Text is stored in Database
Initial Conditions		Human Understandable Language/Database
Step	Task and Expected Result	Status
1	Match data with database data successfully.	PASS
2	Emotions are detected	PASS

6.3 Usability Testing

Usability testing is aimed at measuring the ease with which the system can be used. The usability testing is carried out by choosing a sample of representative users and providing them the opportunity to use the application. Later, the feedback of the users can be recorded to identify the usability issues and resolve them.

6.4 System Testing

The system as a whole is tested to check its behavior is according to set priorities or not.

Chapter 7

Conclusions

This project was the first attempt to automate a system of this nature. We identified from the commencement that creating a whole outcome would be impossible within the given time frame. We observed the project as a journey where we learned many lessons and increased some visions to the topic which we tried to share in this report. We tried to look at the problem from many points of view which produced some new concepts that could be discovered in the future. We proposed formal frameworks for modeling and examining the system which are by no means comprehensive but could become an initial for examining around this area. More, we have learnt short term project planning and implementation which includes:

- Gathering requirements
- Research about the domain
- System design
- Implementation and basic functional and quality testing.
- GUI design
- Interaction with different tools and technologies. management.

Personally, we would consider this project a success if the concept described in the report can become a beneficial reference for future work on the subject. This System is surely very unique in its nature as we couldn't find any related project like our project.

References

[1] Abdullah I Al-Shoshan. Speech and music classification and separation: a review. *Journal of King Saud University*, 2006. Cited on pp. 2, 26, and 31.

[2] bonding in methane - sp3 hybridisation. url-http://www.chemguide.co.uk/basicorg/bonding/methane.html. (Accessed on 05/24/2017). Cited on pp. 5, 6, and 30.

[3] Yi-Lin Lin and Gang Wei. Speech emotion recognition based on hmm and svm. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, 2005. Cited on pp. 6, 7, 26, and 31.

[4] Mahwish Pervaiz and Tamim Ahmed Khan. Emotion recognition from speech using prosodic and linguistic features. *International Journal of Advanced Computer Science & Applications*, 1:84–90, 2016. Cited on pp. 26 and 31.

[5] Samuel Kim, Panayiotis G Georgiou, Sungbok Lee, and Shrikanth Narayanan. Real-time emotion detection system using speech: Multi-modal fusion of different timescale features. In *Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on*, 2007. Cited on p. 31.

[6] Michael Abernethy. Data mining with weka, part 1: Introduction and regression. *URL= https://www. ibm. com/developerworks/library/os-weka1*, 2010. Cited on p. 31.