Chapter 1: Introduction

1.1 Introduction

A virtual assistant is an application that can understand voice commands and complete tasks for a user. For many able-bodied folk, virtual assistants are just another cool way for technology to make us feel like Tony Stark. But for many individuals with disabilities, it's a complete game-changer.From placing phone calls and texting their caretakers to unlocking front doors and ordering groceries; virtual assistants are making important steps towards accessible UIs for the masses.

1.2 Motivation

- Nearly 20% people of the world are suffering from various disabilities;many of them are blind(940 million people with some degree of vision loss) or unable to use their hands(7.7 billion) effectively.They can share information with people by operating computer through voice input.
- Our project is capable to recognize the speech and convert the input audio into text; it also enables a user to perform operations such as open calculator, WordPad, notepad and log off computer.
- For language learning, speech recognition can be useful for learning a second language.Speech recognition is windows software application in which a user gives a predefined voice instruction to a system through microphone, the system understand this command and execute the required function.It facilitates the user to run windows through your voice without use of keyboard, mouse, or any buttons.In simple words it operates an operating system through your voice.Clara is a software agent that assists individuals by conducting particular tasks with the ability of processing natural language.

1.3 Problem Definition

Speech recognition is windows software application in which a user gives a predefined voice instruction to a system through microphone, the system understand this command and execute the required function.It facilitates the user to run windows through your voice without use of keyboard, mouse, or any buttons.In simple words it operates an operating system through your voice.Clara is a software agent that assists individuals by conducting particular tasks with the ability of processing natural language.

1.4 Existing Systems features and lucana

Existing System includes:

- Open and close files
- Setting reminders,alarm.
- Find Basic Information Directly in the Start Menu.

1.5 Lacuna of the existing systems

- The existing system is compatible only with windows 10.
- Existing system doesn’t have default startup.

1.6 Relevance of the Project

Towards the end of the twentieth century, speech recognition systems had found a broad range of use in computerized games and toys, control of different instruments, data collection, and dictation. The feature also proved to be of much help among those who could not obtain keypads and among those with certain disabilities.One of the most notable advantages of speech recognition technology includes the dictation ability it provides. With the help of the technology users can easily control devices and create documents by speaking.Speech recognition technology also makes invaluable contributions to organizations. Businesses which provide customer services benefit from the technology in order to improve self-service in a way that enriches customer experience and reduces organizational costs.It delivers a great customer experience

while improving self-service system's containment rate encourages natural, human-like conversations that create more satisfying self-service interactions with customers and automates what touchtone cannot by collecting dynamic data such as names and addresses enables organizations save agents for more important tasks

1.7 Methodology employed for development

Waterfall model:

In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

Phases of waterfall model:

1. Requirement Gathering and analysis
2. System Design
3. Implementation
4. Integration and Testing
5. Deployment of system
6. Maintenance

All these phases are cascaded to each other in which progress is seen as flowing steadily downwards (like a waterfall) through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name "Waterfall Model". In this model, phases do not overlap.

Chapter 2: Literature Survey (with citation of references)

Chapter 3: Requirement Gathering

3.1 Definition of requirement gathering

requirements gathering iis the practice of researching and discovering the requirements of a system from users, customers, and other stakeholders.The practice is also sometimes referred to as "requirement elicitation".

3.2Functional Requirements

- The application is able to recognize a user's verbal commands without the need for an internet connection.
- The application is capable of giving updates on factors like time, weather and traffic.
- The application allows the user to access other applications using verbal commands.
- Admin is deciding the no of applications to be opened
- The stored data should be platform independent to be sent to a central repository.
- The information of each data upload like the user's voice, the time of upload should be captured
- The central repository should be independent of any type application; ie it should only collect information in a platform independent format.

3.3 Non-Functional Requirements

- A dialogue must provide the possible options of service whenever it perceives that the user will not take the initiative.
- General and explicit instructions on what the system can and cannot do and how interact with it may be provided in a spoken or textual introduction.
- If possible create another sample application which uses the same system but works independent of this application.

3.4.Constraints

- **Time :**

  The voice command should be recognized within minimum time and response should be generated in reasonable time.

- **Correctness:**

  The software should respond according to the commands.(eg. command=open notepad. Expected Output=Notepad should start.)

- **Noise:**

  The software should perform properly even in a noisy environment.

3.5.Various Hardware, Software , Technology and tools available

Technologies to be Used

- Microsoft Visual Studio
- C-Sharp [C#]

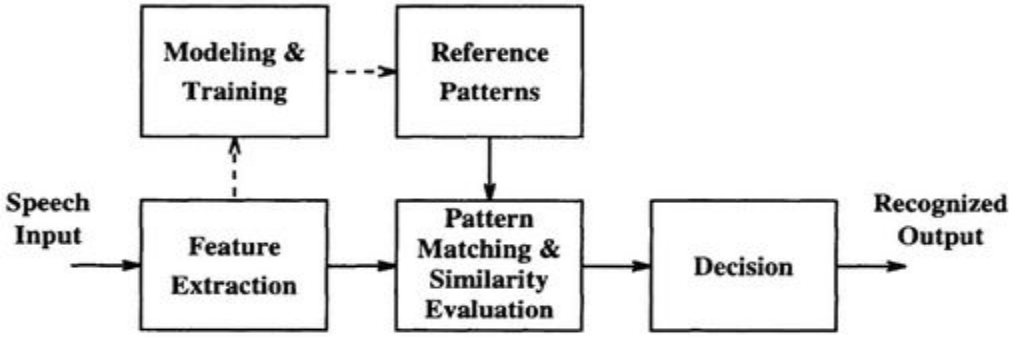
Tools to be Used

- Sound Card
- Microphones/Headphones

Chapter 4: Proposed Design

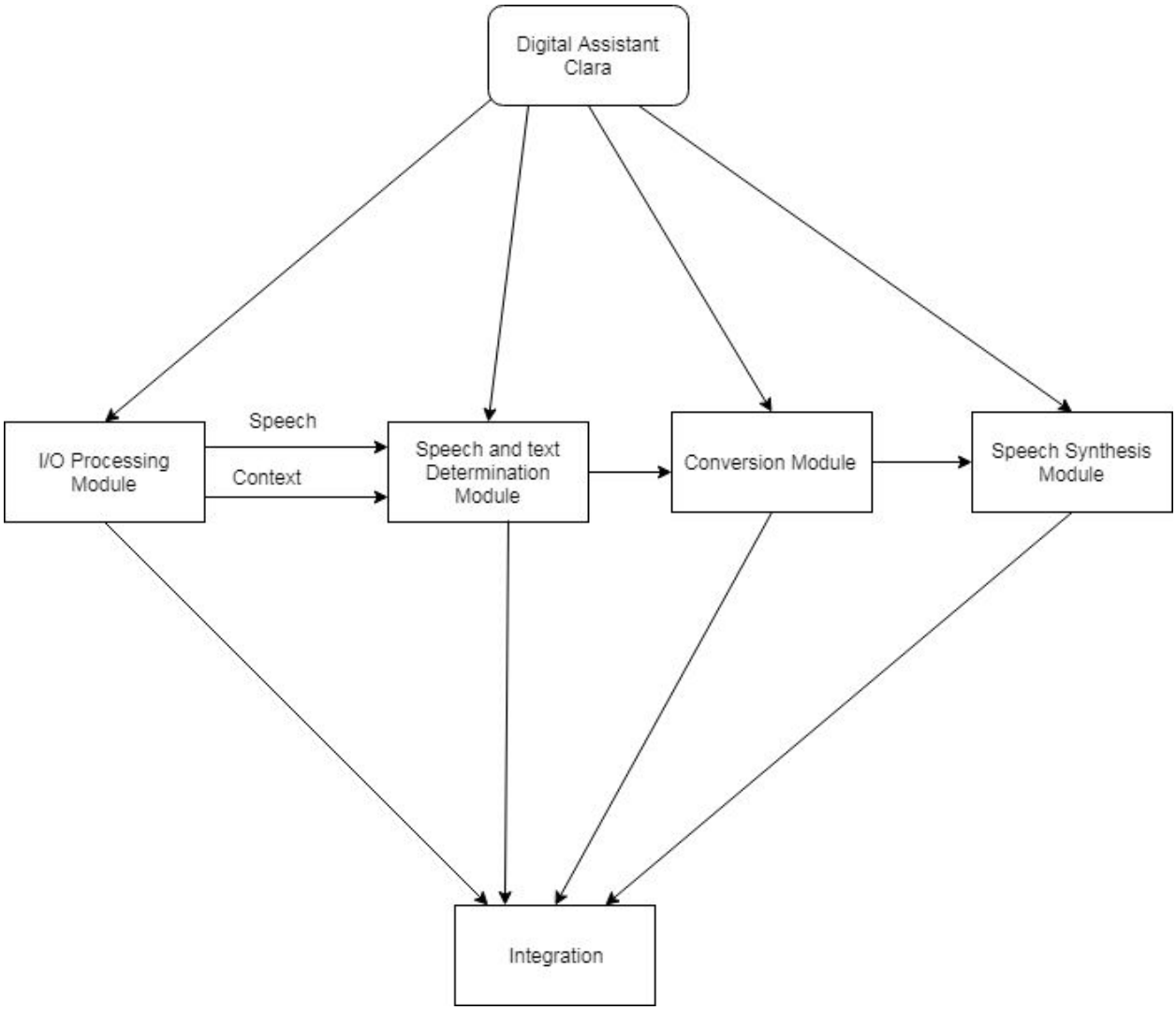
4.1 Block diagram of the system

Block-diagram



The fig is a block diagram which depicts the framework we follow in the design of recognizer .The speech signal is analyzed and the feature is extracted for comparison with the stored reference in the pattern matching block .The reference pattern which characterize individual speech classes to be recognized are obtained by training procedure using a large set of known examples. A decision scheme determines the word or class of the unknown input based on the matching scores or evaluated similarity measures. Methods to be discussed for dealing with adverse conditions may take place in various blocks in the diagram.

4.2 Modular design of the system

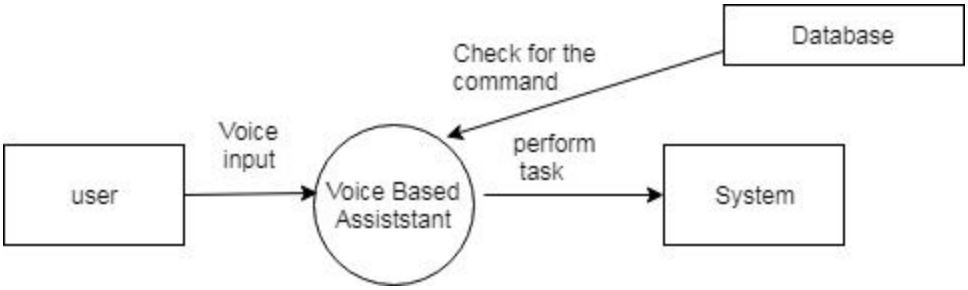


4.3 Detailed Design

DFD - level 0

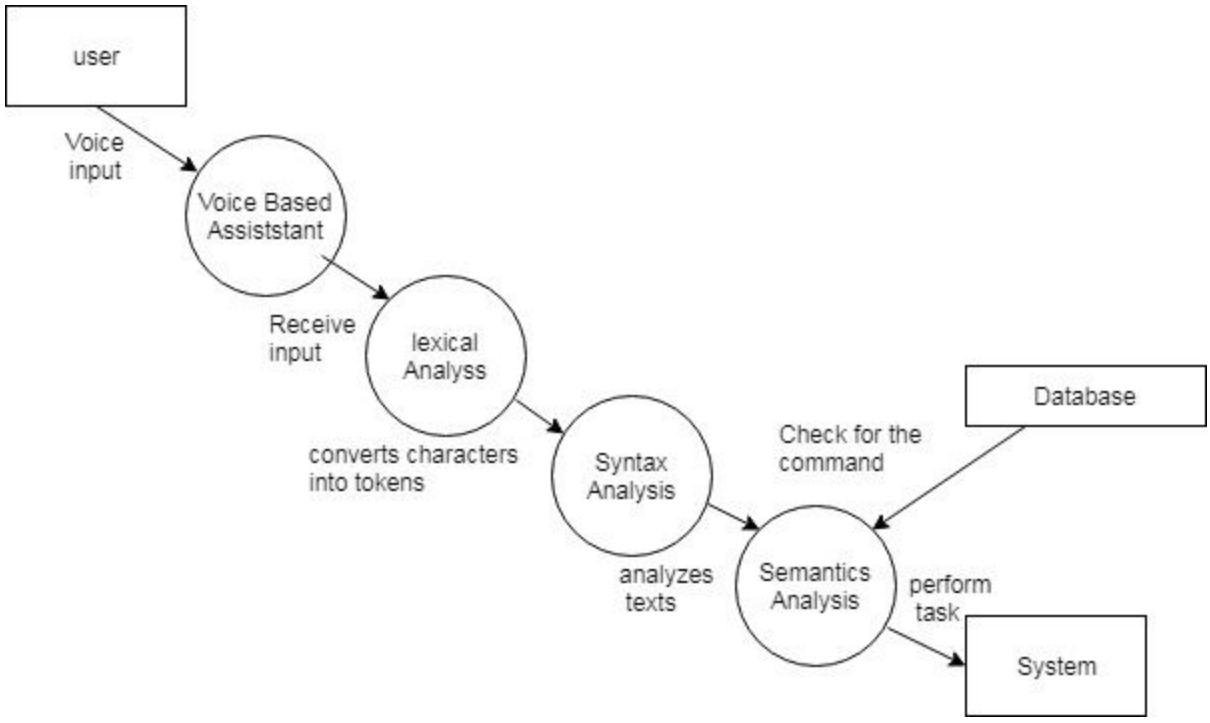
User gives the voice input to the voice based assistant , then system performs the task according to the user input.

DFD-level 1



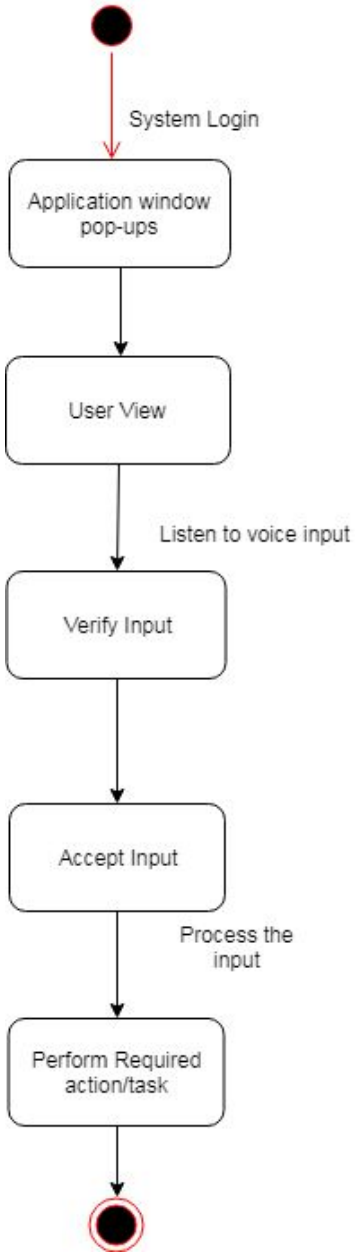
User gives the input voice to the voice based assistant then assistant check that command in the database. If user given command exists in database then assistant will perform the task.

DFD-level 2

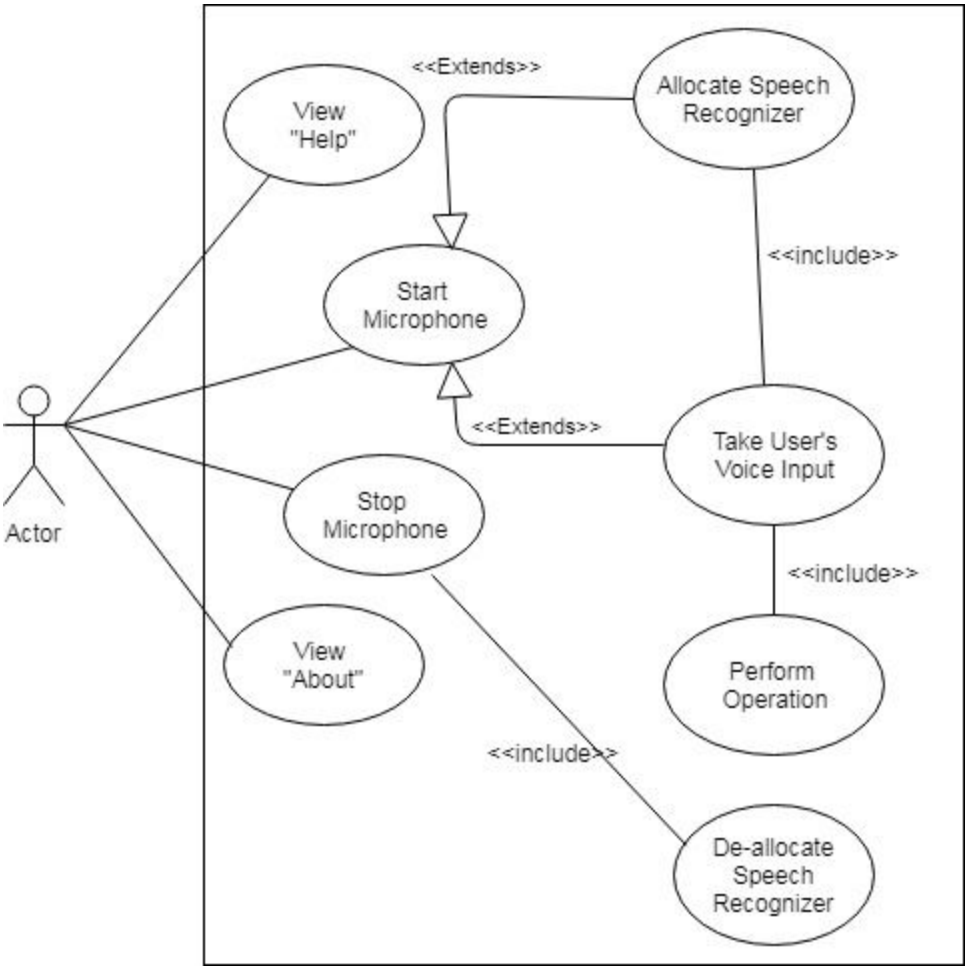


User gives the input voice to the voice based assistant, lexical analysis receives that given input and converts into a sequence of tokens. Syntax Analyzer analysis that tokens and describes it.Semantic Analysis relates syntactic description from the levels of phrases and sentences to their language-independent meanings. And check that command in the database. If user given command exists in database then task will be perform.

State Transition Diagram



Use case diagram



After start the microphone by the user the Speech Recognizer will allocate to takes the input of the user then system will perform the opertaions.

If user stop the microphone then Speech Recognizer will be deallocate.

User can also view the predefined commands of the clara.

Chapter 5: Implementation Details

5.1. Algorithms and flowcharts for the respective modules developed

i) HIDDEN MARKOV MODEL:

(HMM) Is statistical model that output sequence of symbols or quantities. HMM can be trained automatically, simple and feasible to use.In speech recognition, the hidden Markov model would output a sequence of n-dimensional real valued vectors. The vectors would consist of cepstral coefficients, which are obtained by taking a Fourier transform of a short time window o f speech and decorrelating the spectrum using a cosine transform

ii) DYNAMIC TIME WARPING (DTW):

Dynamic time warping is an algorithm for measuring similarity between two sequences that may vary in time or speed. DTW has been applied to video, audio, and graphics –indeed, any data that can be turned into a linear representation can be analyzed with DTW. In general, it is a method that allows a computer to find an optimal match between two given sequences (e.g., time series) with certain restrictions. That is,the sequences are "warped" non-linearly to match each other.This sequence alignment method is often used in the context of hidden Markov models.

iii) NEURAL NETWORKS:

Neural networks emerged as an attractive acoustic modeling approach. Neural networks have been used in many aspects of speech recognition such as phoneme classification,isolated word recognition and speaker adaptation. In contrast to HMMs, neural networks make no assumptions about feature statistical properties and have several qualities making them attractive recognition models for speech recognition . To estimate the probabilities of a speech feature segment, neural

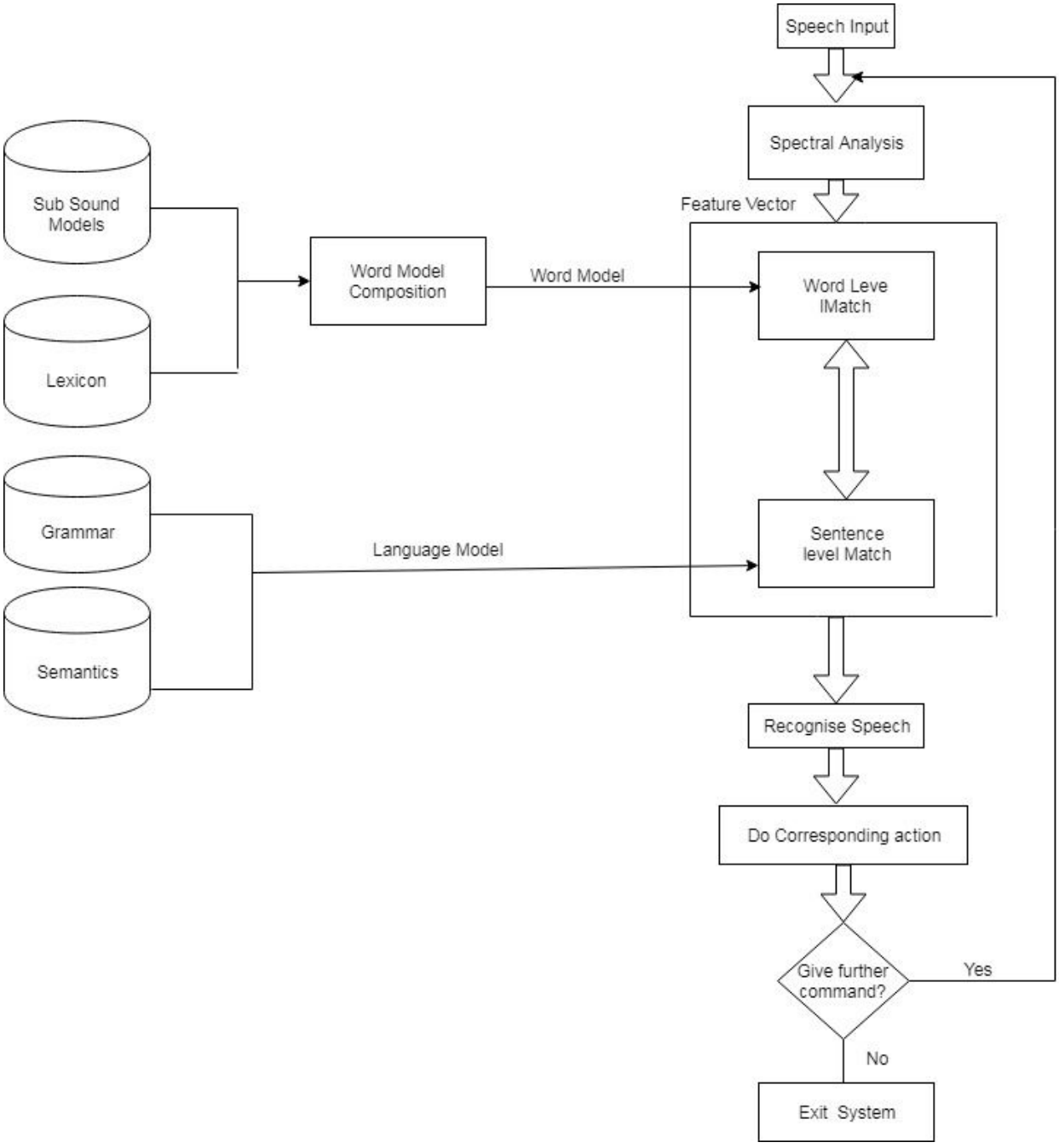
networks allow discriminative training in a natural and efficient way.

iv) SAMPLING THEORY:

When the microphone records the person's analog speech signal through the computer, so the data quality of the speech signal will directly decide the quality of the speech recognition. And the sampling frequency is one of the decisive factors for the data quality. The analog signal actually consists of a lot of different frequencies' components. Assuming there is only one frequency component in this analog signal, and it has no phase shift. The analog signal x (t) into the discrete-

time signal x (n), which the computer can use to process. Generally, the discrete signal x (n) is always regarded as one signal sequence or a vector

FLOWCHART


Speech Input
Spectral Analysis
Feature Vector
Sub Sound Models
Lexicon
Word Model Composition
Word Model
Word Leve lMatch
Grammar
Semantics
Language Model
Sentence level Match
Recognise Speech
Do Corresponding action
Give further command?
Yes
No
Exit System

5.2. Comparative Analysis with the existing algorithms

Existing System	Proposed System
Only Compatible with windows 2010	Can be installed in any platform
Existing system doesn't have default startup.	We can have Default Startup
It performs advanced commands	It can perform commands given in manual only
More Expensive	Cost Efficient

Chapter 6: Testing

6.1 . Definition of testing

The process or method of finding error/s in a software application or program so that the application functions according to the end user's requirement is called software testing.

Testing Approaches:

1. White Box Testing
2. Black Box Testing
3. Grey Box Testing

White Box Testing: It is also called as Glass Box, Clear Box, Structural Testing. White Box Testing is based on applications internal code structure. In white-box testing, an internal perspective of the system, as well as programming skills, are used to design test cases. This testing is usually done at the unit level.

Black Box Testing: It is also called as Behavioral/Specification-Based/Input-Output Testing. Black Box Testing is a software testing method in which testers evaluate the functionality of the software under test without looking at the internal code structure.

Grey Box Testing: Grey box is the combination of both White Box and Black Box Testing. The tester who works on this type of testing needs to have access to design documents. This helps to create better test cases in this process.

6.2. Types of tests

Static Testing: It is also known as Verification in Software Testing. Verification is a static method of checking documents and files. Verification is the process, to ensure that whether we are building the product right i.e., to verify the requirements which we have and to verify whether we are developing the product accordingly or not.
Activities involved here are Inspections, Reviews, Walkthroughs

Dynamic Testing: It is also known as Validation in Software Testing. Validation is a dynamic process of testing the real product. Validation is the process, whether we are building the right product i.e., to validate the product which we have developed is right or not.
Activities involved in this is Testing the software application

6.3. Type of Testing considered with justification

1. **Unit Testing:**
    - ❖ Unit testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed.
    - ❖ Each module of the project is tested independently before its integration.
    - ❖ For example after giving voice input to open notepad application we tested whether the software opens notepad or not.
2. **Integration Testing:**
    - ❖ Integration testing is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units.
    - ❖ All the individual modules like Homepage, Manual page,Command page,Closing page,speech to text conversion module, text to speech module together.
3. **System Testing:**
    - ❖ System testing is a level of software testing where a complete and integrated software is tested. The purpose of this test is to evaluate the system's compliance with the specified requirements.
    - ❖ After all the modules were tested and integrated successfully we checked for the overall functionality of the system.
4. **Acceptance Testing:**
    - ❖ Acceptance testing is a formal testing with respect to user needs, requirements, and business processes conducted to determine whether or not a system satisfies the acceptance criteria and to enable the user, customers or other authorized entity to determine whether or not to accept the system.

- ❖ At the end we compared the system functionality against the SRS document which was created initially.

6.4 Various test case scenarios considered

Test Case Scenario 1: Check the Homepage Functionality

1. Checked weather the application starts whenever the system is logged in.
2. Checked the position of each button.
3. Checked system behavior when buttons are clicked.

Test Case Scenario 2 : Check the functionality of Manual page.

1. Checked system's response after clicking on help button.
2. Checked system's response after clicking on command button.

Test Case Scenario 3 : Check the functionality of closing page.

1. Checked system's response after giving close command.
2. Checked system's response after clicking on press button.

Test case Scenario 4: Speech to text conversion testing.

1. Checked system's response for text to speech conversion.
2. Checked whether the system is capable of handling noise.
3. Checked whether the system is able to handle homophones or not.

Test case Scenario 5: Text to speech conversion testing

1. Checked system's response after giving text input.
2. Checked whether the system can handle spelling mistakes efficiently.

6.5. Inference drawn from the test

1. The system was not able to detect voice input in too noisy environment.
2. The system was getting confused for some homophones.

Chapter 7: Result Analysis

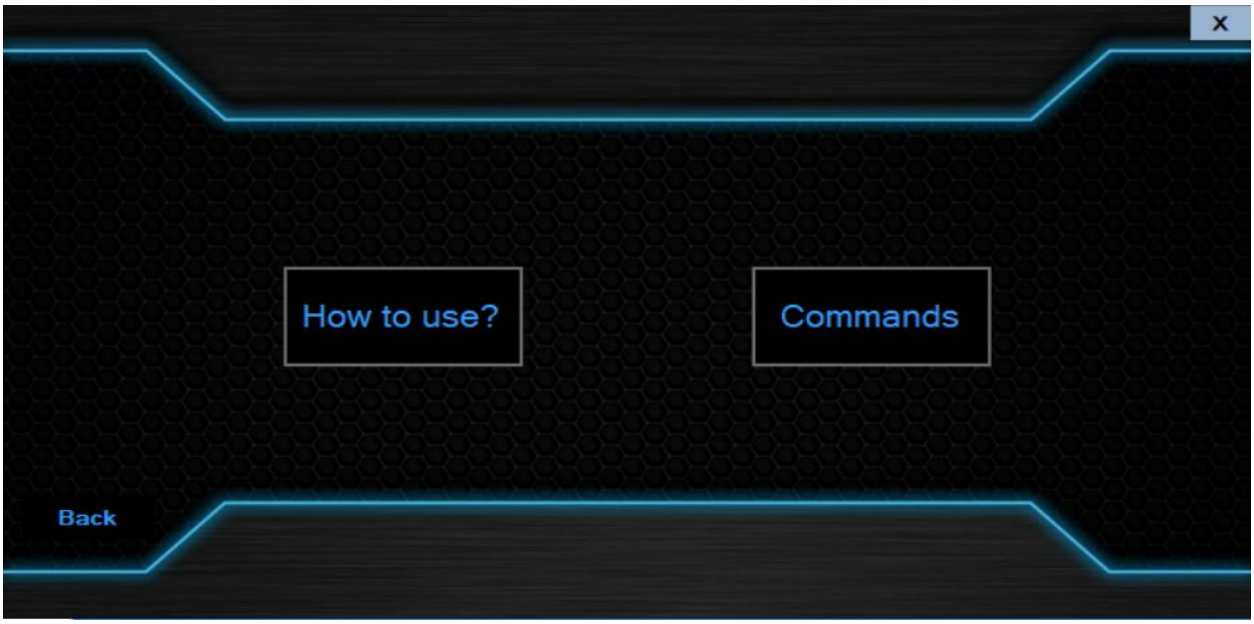
7.1. Parameters considered

a. Noise : Noise is an important parameter as the application was tested to perform well in a noisy environment.

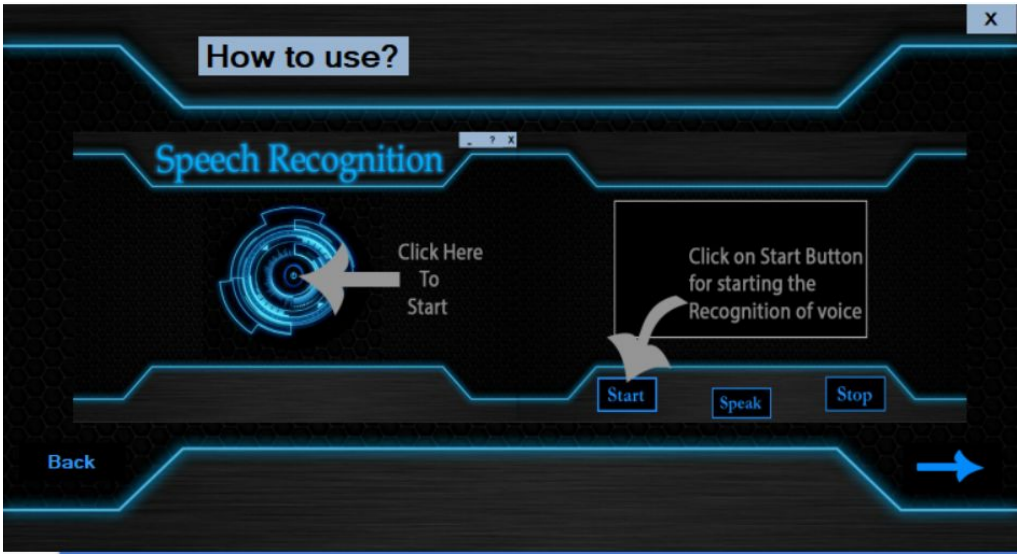
b. Time : The application should respond within minimum time period so time is an important parameter.

7.2. Screenshots of User Interface (UI) for the respective module ---with its explanation

This is the Home Page for Clara The Digital Assistant



This the Help/Manual page ,here you can see two buttons 1.How to use? 2.Commands

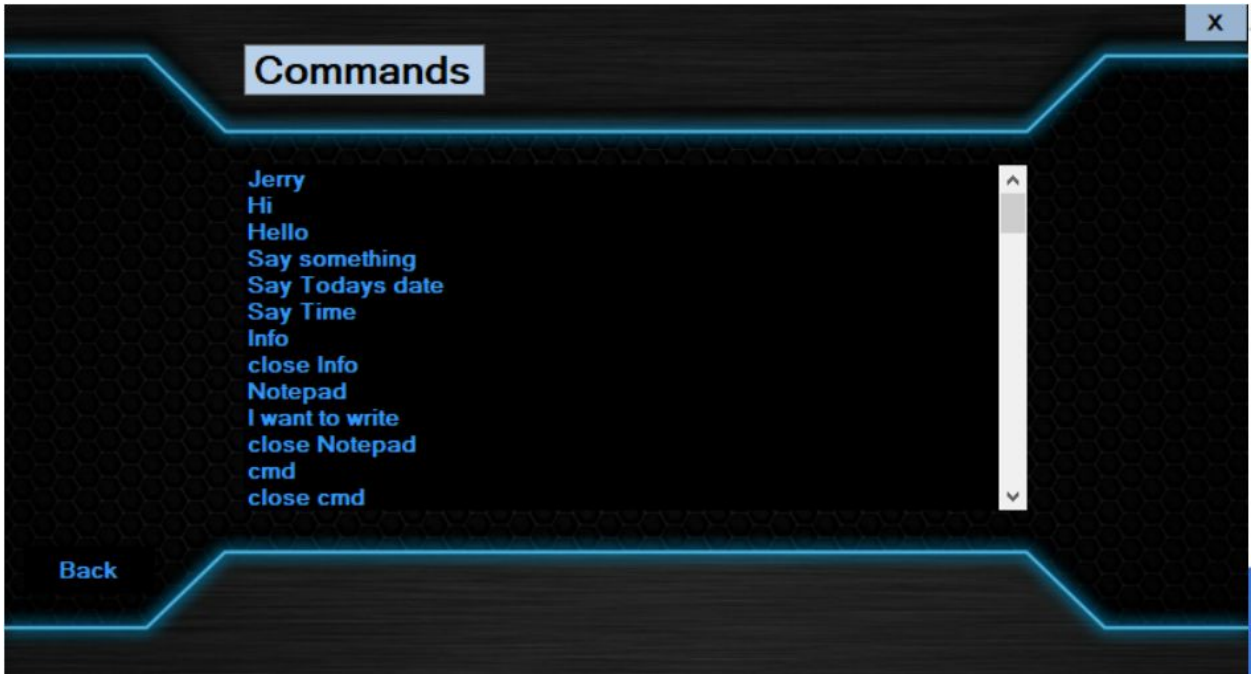


If you click on the How to use button then system will show you how the clara works. Here you can see three buttons 1. Start 2.Speak 3.Stop

After click on start button you just have to type anything then system will speak that word.

After click on the Speak button you just have to give the commands to the clara then after recognizing users command clara will perform it.

After click on the Stop button we can exit our system.



If you click on commands button you can see the list of commands which are define in the system.

Chapter 8: Conclusion

8.1 Limitations

1. The software is not able recognize voice input in a noisy environment.
2. The software is not able to recognize some homophones.

8.2 Conclusion

The developed system matches almost all the functionalities mentioned in requirement specification document.This software provides the most user friendly graphical user interface.It minimizes lots of efforts of the user of typing and mouse strokes.

8.3 Future Scope

1. The software can be used to write a whole document by giving voice input.
2. Given a specific name of a song or a movie it will directly play it.

9. References

9.1. Newspaper articles referred

- Speech recognition is tech's next giant leap, says Google

  https://www.theguardian.com/technology/2018/sep/24/speech-recognition-tech-google

  Abstact:

  The next big leap in technology will be advances in voice services, according to Google's head of search, Ben Gomes, who says that a better understanding of common language is crucial to the future of the internet

.

- **Speech, the Experience Game-Changer**

  https://news.microsoft.com/2010/08/03/speech-the-experience-game-changer/

  Abstact:

  The growth of connected devices, from automobiles to your mobile phone, coupled with the increase in data consumption is signaling the beginning of a broad shift in technology toward an era of more integrated, natural experiences driven by speech, touch and gesture.At the 2010 SpeechTEK Conference in New York, Zig Serafin, general manager of the Speech Group at Microsoft, delivered a keynote address describing Microsoft's vision for speech and natural user interfaces (NUIs). Serafin demonstrated the latest in speech recognition technology that has been designed into upcoming Microsoft products. These products promise to deliver more elegant and accessible interfaces, allowing users to utilize their voices and, in some cases, their bodies to perform actions and access information.

- **Spread the Word: Speech Recognition Is the "New Touch" in Computing**

  https://news.microsoft.com/2009/10/28/spread-the-word-speech-recognition-is-the-new-touch-in-computing/

  Abstract:

  Keyboards and mice still are the dominant methods for working with a PC or laptop. But big leaps in speech-recognition technology mean that talking to a computer may soon be as natural as using a mouse.Leading Microsoft's charge to that audible future is Zig Serafin, general manager of the Speech at Microsoft group. Serafin says his team's goal is simply to create the world's most advanced speech platform, one that spans cloud-based voice services, mobile phones and world-class servers for enterprise customers. "Voice is the new touch," says

Serafin. “It’s the natural evolution from keyboards and touch screens. Today, speech is rapidly becoming an expected part of our everyday experience across a variety of devices. Bill Gates articulated this vision a decade ago, and we’re seeing it happen today.”

9.2. Book articles referred

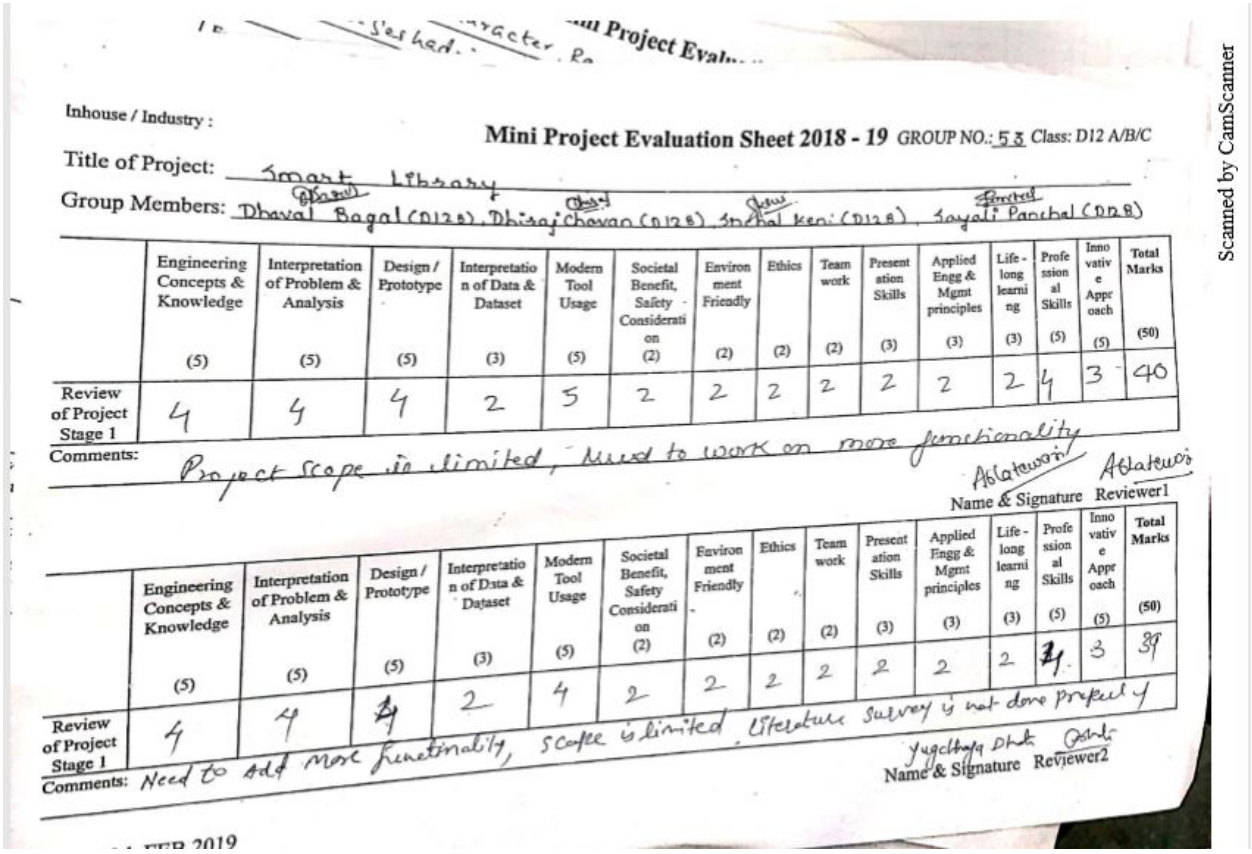
- Acoustic Phonetics by Kenneth N. Stevens
- Data-Driven Techniques in Speech Synthesis [COLI 28:4] by Thierry Dutoit

9.3. Research Papers Referred

- **Research on Speech Recognition Technology and Its Application**
  https://ieeexplore.ieee.org/abstract/document/6188154
  Abstract:
  Speech recognition is a kind of technology that is using computer to transfer the voice signal to an associated text or command by identification and understand. The paper depicts the speech recognition system and the main techniques of speech recognition, and makes a preliminary exploration for its application in various fields.
  Published in: 2012 International Conference on Computer Science and Electronics Engineering

10. Appendix

Project review sheet 1



Inhouse / Industry :

Mini Project Evaluation Sheet 2018 - 19 GROUP NO.: 53 Class: D12 A/B/C

Title of Project: Smart Library

Group Members: Dhaval Bagal (D128), Dhiraj Chavan (D128), Snehal Keni (D128), Sayali Panchal (D128)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	5	2	2	2	2	2	2	2	4	3	40

Comments: Project scope is limited, Need to work on more functionality

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2	2	2	2	2	4	3	39

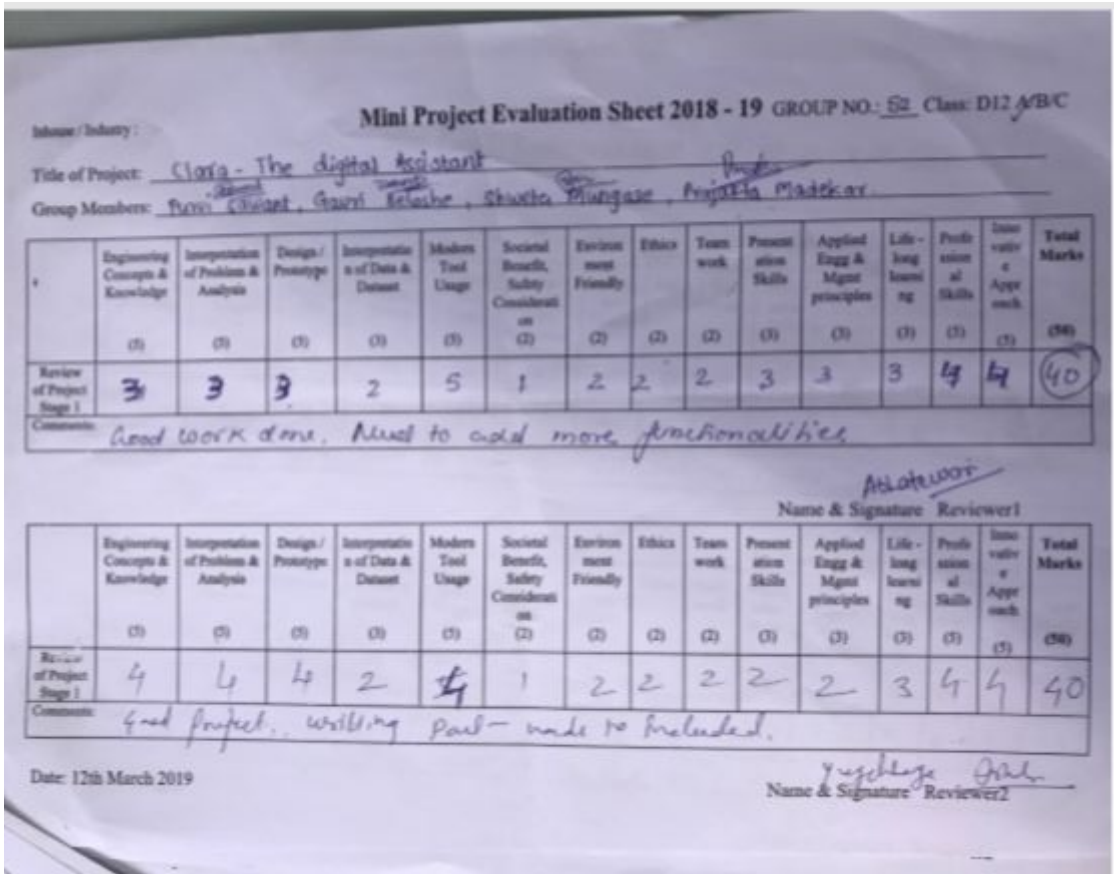
Comments: Need to Add more functionality, Scope is limited, Literature Survey is not done properly

Name & Signature Reviewer2

FEB 2019

Scanned by CamScanner

Project review sheet 2



Inhouse / Industry :

Mini Project Evaluation Sheet 2018 - 19 GROUP NO.: 52 Class: D12 A/B/C

Title of Project: Clara - The digital Assistant

Group Members: Purvi Chhajed, Gauri Kelashe, Shweta Mungase, Prajakta Mahekar

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	3	3	2	5	1	2	2	2	3	3	3	4	4	40

Comments: Good work done. Need to add more functionalities

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	1	2	2	2	2	2	3	4	4	40

Comments: Good project. Working part - made to be included.

Date: 12th March 2019

Name & Signature Reviewer2