International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

Smart Assistant System using Speech Recognition

Mohd Abid Hundekari1, Shreya Vitalapuram[2], Lakshmi Chennakesavalu[3], Samarth Mudaliar[4], Prof.S.S.Kulkarni [5]

UG Student, Dept. of Information Technology, Sinhgad Academy of Engineering, Pune, Maharashtra, India[1,2,3,4]

Professor, Dept. of Information Technology, Sinhgad Academy of Engineering, Pune, Maharashtra, India[5]

ABSTRACT: This proposed system provides an application based on ML that uses instant message and voice to create two way communication between human and our machine. This is the desired technology in the 21st century. Our goal is to provide a voice control intelligent system that gives ability to control our machine for its operation. Our system is for enabling the visually impaired people to use appliances or machine. Existing technologies in these fields are dependent on displays and keyboard which are very expensive and scarce. As we know, navigating through any site or webpage users click generates huge amount of clickstream logs. Our major aim is to reduce clicks and minimum use of keyboard to give input to the machine. Our sys-tem is hands free yet accurate way to communicate with the application. System compatibility problems that are generated in existing technologies will overcome in this system. Our pro-posed system is capable of providing maximum functionalities without internet connectivity which is necessary in existing systems like Google SIRI, Google Voice Search and Microsoft Cortana which are not open source software. This intelligent system would provide the blind and physically challenged people access to computer without click of buttons. In future It will soon be used in small and big devices like washing machine too. As per the implementation details we are going to use Google engine for voice to text feature. The text obtained from Google engine can be further processed for assistant. This system works on the principle of data mining and semantic analysis in which for one voice input it gives one or more matching results and the data dictionary recognizes the best matched results and performs the particular task. For this we are going to use xml parsers to store predefined commands. Our main aim is to add new features to Microsoft cortana and remove some of its draw backs.

KEYWORDS: Microsoft cortana, xml parsers, Google engine, Voice Search

I. INTRODUCTION

Nowadays the speech to text technology is getting expanded. The main attraction of any system is reducing human labour, e orts, time and errors. Speech recognition is a technology by which a computer or any machine could interpret human voice and compute the said task. This is the important application of the AI that uses instant message or voice commands to create two way communication. There are lots of systems that allow us to convert human speech to text format. This text that we fetch from the speech to text engine can be used for assistant. It will simplify the use cases and also the text can be used as commands. Our systems also works on the same principle of getting the text from the speech to the text engine and processing the text as commands. The voice assistant systems allows user to control the computer using voice commands. Like Open cmd, write to the notepad send or read emails, get the weather report,set reminders etc. Our project is about providing an application based on AI that uses voice commands to create communication between humans and our machine. This technology has become more desirable in today's world. Software's available in the market are machine specific which has some system compatibility issues which will be solved in our system.

Nowadays mobile technology is being very famous for the user experience, because it is very easy to access the applications and services from anywhere of your geo location. Same kind of application is also developed by Google that is voice Search which is used for android phones, but this mostly works with internet connection which is its major drawback. As there are lots of assistant app available for the PC but they are purchasable also we need to subscribe for the plan for main features to be used, but we decided to implement a system using java language so that it can work on all the systems and to make it open source available to all the users. The software's available in the market are machine specific like windows software won't work on Linux but our system will provide both the interface so that it will work on windows as well as Linux. We are motivated for this project so that physically disabled people can experience today's technologies and can operate pc and perform maximum operations on it without having to click by using voice technology.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

II. LITERATURE SURVEY

[1] The work proposed by Lavin Jalan-Tittled "Speech recognition based learning system" analyses that English is internationally accepted language with various accents in different parts of the world. The study implements a dictionary based on speech recognition using isolated character to provide exact meaning of spoken words with high accuracy. It can be viewed as a unique approach in varied domain of human computer interaction [7]. A real-time speech-based learning/training system distributed between client and server and incorporating speech recognition and linguistic processing for recognizing a spoken question and to provide an answer to end user in learning or training environment implemented on internet, is disclosed. They are acoustic phonetic method, Pattern Recognition method, and artificial intelligence method. Pattern Recognition has mainly two steps involved in it. It involves Pattern comparison and Pattern Training. Two approaches for pattern recognition are stochastic approach and template approach phonetic. This method is based on process of locating sounds and foundation of acoustic phonetic approach. The third method for decoding is artificial intelligence. It can be understood as a combination of pattern recognition approach and acoustic phonetic approach. This includes designing of recognition algorithms, demonstration of speech units and appropriate units. Among all methods of speech recognition, artificial intelligence is most efficient method.

[2] The work proposed by Dr.KshamaKulhalli [et.al] Titled \Personal assistant with voice recognition intelligence" in which they proposed Nowadays mobile technology is being very famous for the user experience , because it is very easy to access the applications and services from anywhere of your geo location. Same kind of application is also developed by Google that is voice Search which is used for android phones, but this mostly works with internet connection which is its major drawback. A personal assistant system uses context to intelligently make contacts and for provisioning an address book. The system uses history information to make predictions of whether calls will be successful when no preference is specified by user.

[3] The work proposed by Prince Bose [et.al] Titled" Digital assistant for the blind" implements a voice based intelligent system for the blind and can benefit from using the technology to convey words and then hear the computer recite them as well as use a computer by commanding with their voice, instead of having to look at the screen and keyboard [5]. This paper showed possibility of user guide interaction with intelligent assistant .We expected that if we measure the users' feedback by doing the users research might be good to show the e effectiveness of the suggestion. Additionally, we will extend this methodology with multi- modality-based suggestion system.

[4] The work Proposed by RoshanJadhav [et.al] Titled \Speech synthesis for generation of artificial personality" developed a learning model for voice based intelligent system that is security. Voice recognition can become effective solution to security problems of different types including telephone based crime, frauds and many others. Biometric voice recognition software are actively used in banking and e-commerce sectors.[1]The speech synthesis is a process which artificially produces the speech for text to speech synthesis. In this paper speech synthesis is shown where core component is construction of intonation pattern called as rules. These rules are set up in format of frequency consisting of group, vowel group and natural consonant group.

[5] The work proposed by Md .Almin[et-al] Titled \Design of intelligent home assistant" developed voice based system as talking calculator in which user will get arithmetic calculations in the form of voice.[1] The answer of the calculation will get converted into an audio signal and software will give answer to the user in the form of audio along with text. Basic mathematic operation like addition, subtraction, multiplication, division are performed by the speaker independent trained machine system. Numbers from 0-9 are also frequent words such as minus, plus, multiply, divide are recognized by the machine [1]. Microphone is used as input where speech is detected and recognized and is sent for the calculation there after respective output is produced. Words and numbers are trained to the machine. There are various states involved in the training where some of them are clear, reset, erase memory etc. Testing and validation is done after the completion of training of machine.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

[6] The work proposed by Li Deng [et-al] Titled "Machine Learning Paradigms For Speech Recognition" describes the following :

Here is a remote station where feature extraction is located. From the input speech given by the user the frame including input voice is extracted by feature extraction and then is sent to the central processing station. The process that takes place inside central station is that the word is decoded and the valid output is given. Machine learning is very important in the field of acoustic recognition system. Sequence of data is converted into words made by sequence which automatically done by speech recognition system. Frames which are also called as sub-process From Machine Learning point of view there are number of processes required for converting ASR. Lately the method for converting speech to text for sending as an email message also text is displayed on the screen .When the caller sends audio message le txt extension is generated and is sent to the receivers address. In windows there is speech recognition system namedcortana. It has a feature for opening le .Drawback of this feature is that cortana cannot open le with extension. e.g.: .xml,.pdf,.doc Suppose the end user wants to open a document le and if the end user speaks open .doc file cortana will show how to open the le with steps but it itself will not open

III. METHODOLOGY USED IN EXISTING SYSTEM

This project provides an environment for developing adaptive speech recognition system .It provides various modules like signal processing module,word recognition design module,validation of the module and adaptor of module.Signal processing module includes most common technique used in speech processing.Word recognition design module provides architecture for adaptive speech recognition system.Graphical user interface is designed to allow interaction between different modules. As per the implementation details we are going to use Google engine for voice to text feature.The text obtained from Google engine can be further processed for assistant.This system works on the principle of data mining and semantic analysis in which for one voice input it gives one or more matching results and the data dictionary recognizes the best matched results and performs the particular task. For this we are going to use xml parsers to store predefined commands. For sending an email GoogleAPIS are used which is a set of application program interface developed by Google which allow communication with Google services amd their integration with other services.Marrytts library is used for text to speech conversion.Voice synthesizers in which every word is broken down down into pieces are used for audio conversion.

Algorithms

Our System uses 3 main algorithms Filtering text, Feature extraction and Template matching algorithms.

1. Filtering Text Algorithm:

This Algorithm is mainly for filtering the adult or bad words. When the user gives input to the system the system produces multiple results then this filtering algorithm helps to filter out the unwanted words and then display the accurate result to the user.

Steps:

1. Get speech spoken by user.
2. Extract text from the speech.
3. Register listeners.
4. Compare text received by speech with the predefined unwanted words.
5. If any match found send not valid message to user.
6. Stop.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

2. Feature extraction Algorithm:

This Algorithm is mainly used for converting speech to text.This is based on linguistic model and grammar. In the classification problem the speech feature extraction is used for reducing the dimension of the input vector, while maintaining the perceptive power of the signal. Feature extraction is a special form of dataset and it results in extraction of specific features. Features carry the characteristics of the useful information regarding speech. Feature design and selection is the main challenging problem in the speech recognition system development for specific application. For speech identification and verification development, the number of training and testing vector are needed for the classification. The problem grows with the dimension of the given input.

Steps:

1. Get the speech spoken by user.
2. Extract signal from speech.
3. Remove noise from speech as much as possible.
4. Amplify signal so that the signal strength will get increase.
5. Pass speech to signal to grammar engine.
6. Grammar engine parses signal and returns text to user.
7. Stop.

3.Template Matching Algorithm:

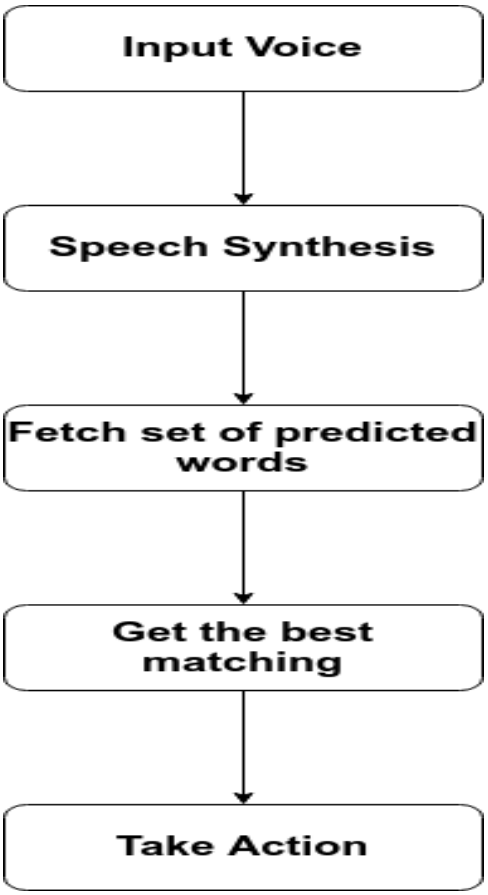
The goal of TML is too accurately and efficiently convert a speech signal into a text message transcription of the spoken words. This process should be independent of- The device used to record the speech (i.e., the microphone) the speaker's characteristics (i.e., age, gender, accent) the acoustic environment (i.e., quiet vs. noisy rooms, outdoors)

Steps:

1. Start speech engine
2. Get the speech from the user
3. Extract feature using feature extraction algorithm.
4. Compare feature with linguistic model and accent specified by speech library.
5. If it matches with the template of grammar return text to user.
6. Stop

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||



There many existing systems to control machine using voice commands like cortana on windows, Jarvis on Linux etc but they all are dependent on the quality of hardware in order to predict the proper words otherwise they may interpret words wrongly. Cortana cannot recognize symbolic letter alphanumeric. Cortana also does not have a feature where the administrative user can change his/her password or user id or username. Cortana shows the user steps on how to change the password or username but itself does not change the password or username. Another drawback of cortana is that there are specific sentences and words to perform a particular task. If the end user speaks those specified words or sentences then only the respective work will be performed otherwise error message is displayed on the screen.

IV. METHODOLOGY USED IN PROPOSED SYSTEM

In this project the physically disabled people would be able to operate the computers through their voice commands. We intend to develop a software that would translate speech into text and text into speech and further it would execute all possible commands given by the user. Some of the systems like Microsoft has its own dictation software which is not accessible in other software's like notepad so we intend to develop universally useful software for dictation and reading the text documents. This text documents can also be converted into an audio _le so that the user can further use it. Even some existing systems do not support composing mail through voice our system would support composing mails and even navigating through a browser using voice. Our assistant system would also implement a voice based music system for the better user's experience. This system helps to replace the human efforts of typing with voice commands.Security issues would also be resolved as there is no need of syncing contacts or personal information in our system.The project is carried out by following the software development lifecycle phases. We have used agile methodology. Agile model believes that every project needs to be handled differently and the existing methods need to be tailored to best suit the project requirements. Agile method is all about iterative planning, making it very easy to adapt when some requirements change. Iterative approach is taken and working software build is delivered after each iteration. Each build is incremental in terms of features. This desires of man to create a machine resembling itself and having to interact with people in natural language is now become a reality using AI and voice recognition systems. The more the advances in technologies more are the areas to deploy them. These areas of applications might be beneficial or

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

sometimes might even prove to be disadvantageous for the domain of people using it. Project organization is a process which provides arrangement and decisions about the realization and the project. The structuring, organizing and configuration is included in it. Itdevelops a software for physically disabled people to ease their work of operating the pc and benefit from the new evolving technologies. Windows 10 provides to organize Speech Recognition to help users control a system with voice. The features allows to open app, programs, open _les makes a hand fee type in document .All this runs with the help of microphone, either built in or external.

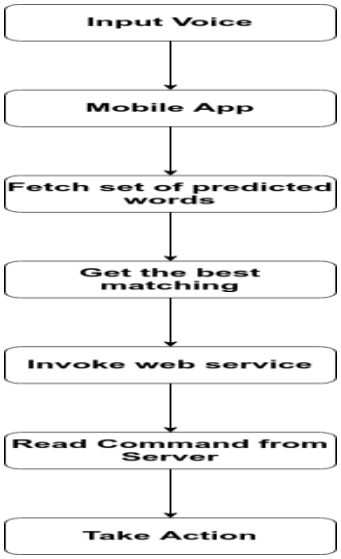


Fig. System Architecture

We are going to develop following modules:

A. Android Speech to text engine

This is the main feature provided by android OS where we can directly convert speech into text by using Voice Recognizer thread.

B. PHP web server

This web server is used to send commands from mobile device to machine where we want to take actions.

C. Virtual Key set:

The main use of this library is to handle the actions based on the command received from mobile device.

D. Login:

In this system, the first step in action is to login. Hence, the user has to login to the system by inserting the username and password. Inserting the valid username and password will further lead to next process of the system. Invalid data inserted can show wrong results. The output of correct login is shown through speech output saying “Welcome to smart assistant how can I help you” where this shows the text to speech module. After the login screen is done, the valid one later open up the home screen with different modules shown on screen.

E. File Writing Module:

In this module we use vinterface database which is used to communicate with files. We have vcommand table that consists of timestamps. The command with large timestamp is considered as recent one. Whatever we speak goes to vcommand table and then the recent command is displayed using buffered reader. In File writing mode it plays an important role as it takes speech input and displays the same on the editor screen. On saying save command it saves the file for further references.

F. Calculator Module:

Calculator is designed to make arithmetic calculations using voice recognition. Operations supported in this system are addition, multiplication, division, subtraction. An additional operation present is square root. Clear function is used to

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

erase the previous performed result. Result function is used to display result. Working of calculator: First the user speaks first number then the operation which he/she wants to perform followed by second number. When the user says result, output is displayed on the screen. On saying clear the result is erased.

G. Reader module:

This module is another featured of Smart Assistant System. Because this module is not only the absolute user need, it's a need of today's generation for every class from kids to old age people. Now the task of 'Read out' Button. So if the file is not selected or path is empty or file is not in '.txt' format, then we show Option Pane to show message to "PLEASE SELECT A TEXT FILE". Because we are using Lexar class that only will process normal text as lexical analysis. We need lexical analysis for stopping the speaker at full stop or comma for a while. Other than '.txt' format files are formatted text, which can include images too, so that Lexar class would not process the text. On closing again we get back to previous UI of Welcome () class. On command "Document to Audio" the window will launch of convertor, Everything in this window is same as to "Read Out Document", But the major difference in only that 'Read out Document' just reading the text file for current instance, and this is store the file in '.wav' format, so user can listen it anywhere anytime. It is the convertor of text file to audio file. The application of this feature is to convert text books into audible books. The user once use the Editor to save some task in text file, later on he/she can convertor that text file to audible file which he/she can listen any time. Working of this application is same as to 'Read out Document'. Just select the file using File Dialog class's method of opening file. But additional to that we also getting saving location of audio file from user by Class JFileChooser object's method showSaveDialog ().

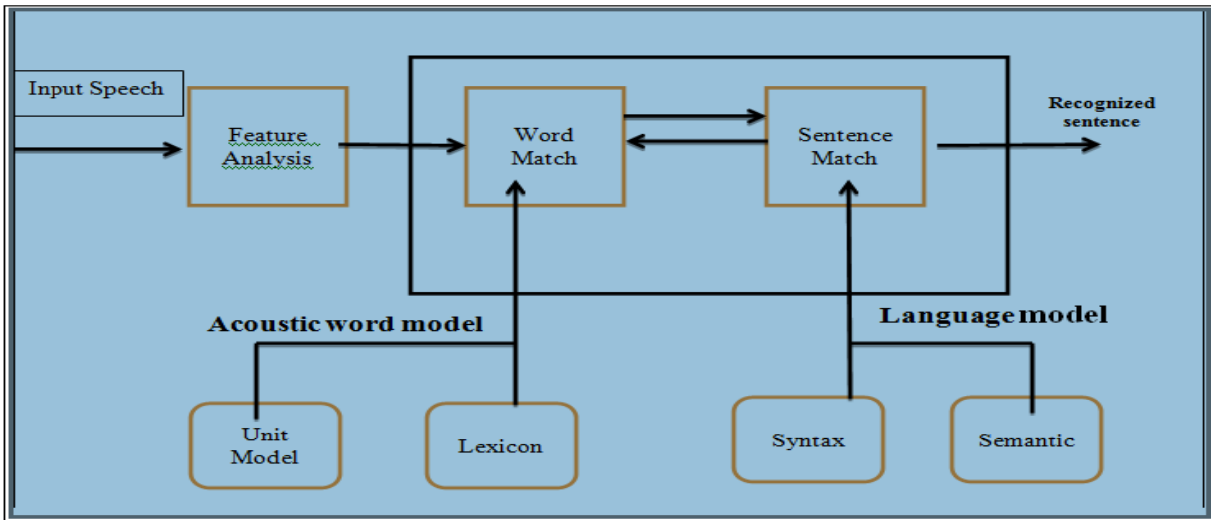
H. Command Mode:

Command processor is used to open different utilities provided by OS. But first user have to enter manually on which command what utility user want to open. The UI for command processor and adding new command to system are as follow: As Editor, this module also using all same packages imports like InputStreamReader, BufferReader, URL, URLConnection and swing. The first figure shows command prompt of Smart Assistant System, where user have to give command over voice. For some general common use, we have given built-in commands. On closing of Command Screen by 'close' command, it will self-dispose and launch the same Home Screen, from where it was originated.

I. Mail:

Mail Composing is one of the module shown in the home screen of the system. The working of Gmail module includes a) composing function b) sending functions The user composes the required information that needs to be send as email to the person. User needs to send and the first composing function is done. This is done through speech input and the information to be send is saved it. The second function is the sending function. This function takes place after the user composes the data, but not a compulsion to be done composing. The user selects the sender's Id and the word SEND is voiced as input. Thus the email is send to the required sender. And the sender gets notification as "SUCCEESSFULLY SENT" in the senders device.

V. RESULTS AND DISCUSSION



| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

The diagram above shows block diagram of speech recognition system. This block diagram can be made to perform virtually any speech recognition task that has been using since the past years. Some of the examples are Isolated word recognition, connected word recognition, continuous speech recognition etc. In fig below the feature analysis module gives the acoustic feature vectors that are used for characterizing the spectral properties of the time varying speech signal. The word acoustic match module is used for finding out the similarity between the input feature vector sequence and a set of acoustic word models for all the words in the recognition task vocabulary to find out which words were most widely used and spoken. The sentence level match module uses a language model to determine the sequence of words which are used more. Syntactic and semantic rules can be defined for the model eithermanually or statistically. Searching and recognition decisions are made by 502 by taking into consideration all word sequences that are widely used and selecting the one which best suits the found out results as the output . Almost all the modules of the speech recognition systems is been studied over the years. The result of this study has given us the idea and knowledge to design a voice recognition system, that takes input finds all matching patterns and selects the one which best matches the result.

1. Presentation Tier
2. Data Tier
3. Application Tier

The proposed System uses the Three Tier architecture. The 3 tier architecture consists of:

Presentation Tier

The presentation tier is the topmost level of the application. It is a layer which users can access directly.
It is a layer where user can access by inserting username and password to enter into next
Screen which is home screen. This home screen contains the view structure of modules done in the project system.This modules are clicked and then next steps are proceeded. Each module describes each in detail.Calculator module has the arithmetic operations to be performed. The file writing has the functions of save and edit where particular phrases are dictated and is stored up.
Gmail has functions of composing and sending.Command modules has commands like open
Terminal, browser etc .... These all are involved in presentation tier and the exit button makes
User reach back to the home screen again.
The filter module is well viewed when each action is performed as the adults words have to be discarded to make accurate words spoken.

Data tier

This is the second most layer of application. The data tier includes data persistent mechanisms like the database SQL apache.As we know the SQL is used to communicate with the database,where the SQL statements are used to perform tasks are such as update and retrieve data on database.The user data tables will store data of user and whatever is entered when logged in to the system. It is special purpose of programming language to store data.

Application tier

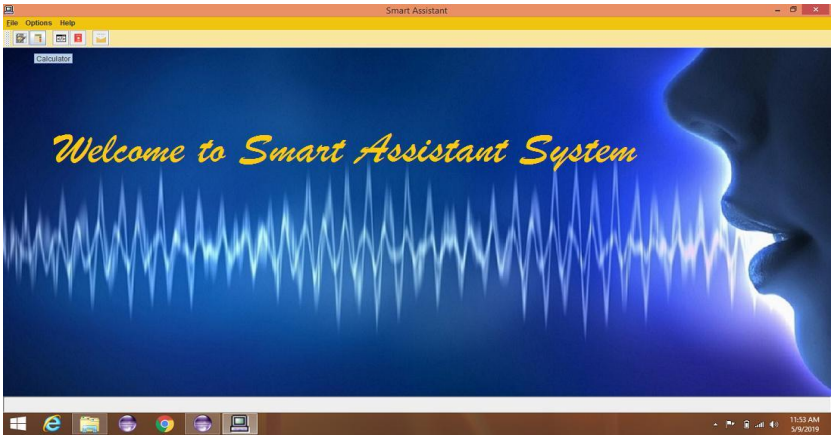
The user has to login to be able to use the Smart Assistant System through the GUI. On Login, the user will be provided a username and Password. This information will be stored in the database. This third tier is the application tier where what operations are performed in this system each are stored in the database system.The username, the file writing para-Phrases, readable documents, the mail that is composed and the sender's userid all this are Stored up.

The results of our working project is shown in the below images.It consists of the entire execution of the proposed system.The system successfully implements all the expected output and is providing the desired results Graphical User Interface

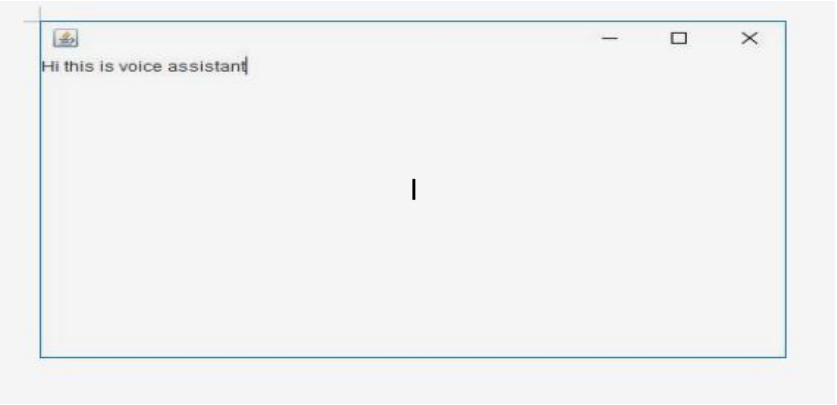
| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

Login Screen



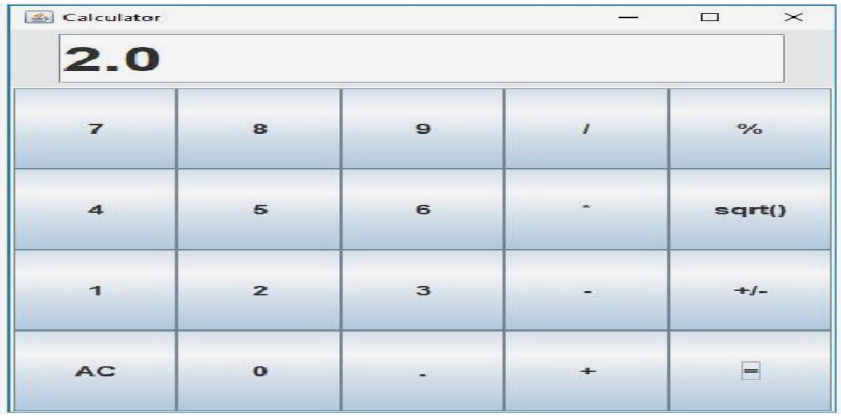
Homescreen



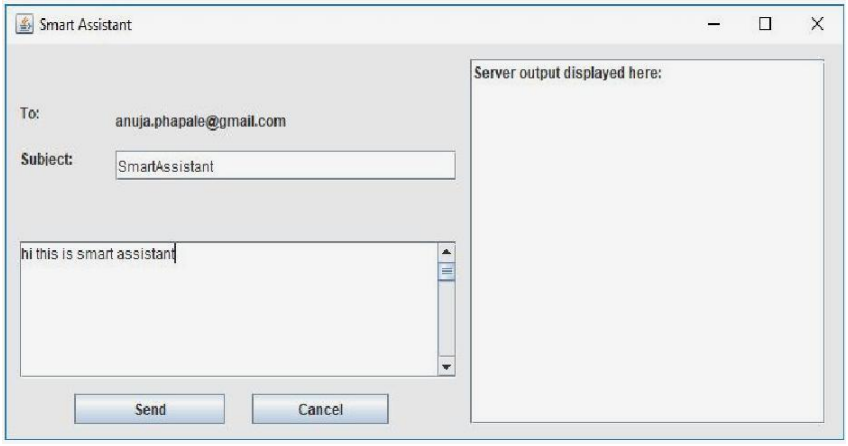
Editor

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

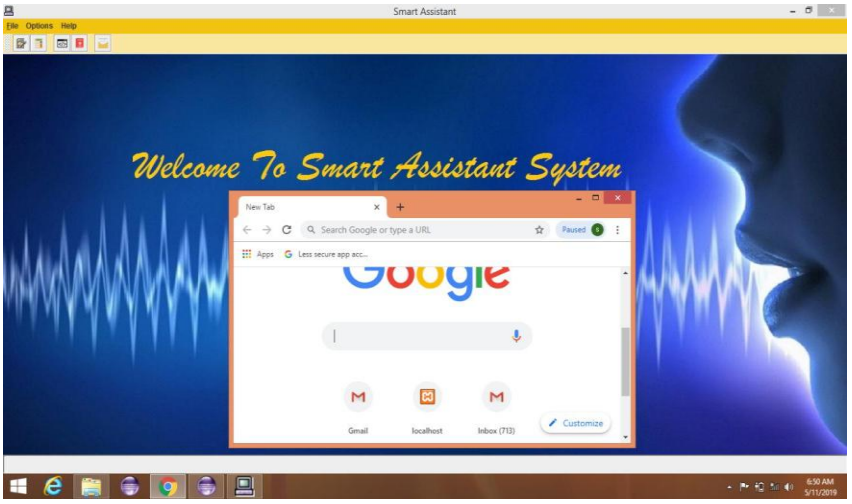
||Volume 9, Issue 6, June 2020||



Calculator performing operations



Mail Composing



Opening Chrome

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

||Volume 9, Issue 6, June 2020||

VI. CONCLUSION

This project is undertaken to develop a software for the physically disabled based on speech
Recognition systems. This included the study of all aspects including technologies, trends in the market based on AI applications, home appliances using voice recognition systems, the areas which are still under development in this fields and to find out a way to provide an efficient and accurate system to the user so that the user can control its device using speech instead of typing and looking at the screen. We have concluded through this project that the strongest desire of man to create computer resembling itself and having ability to interact with people can become a reality. Speech recognition has become the most desirable technology of today's need of mankind. People can interact with their devices to send emails, read weather report, get daily news, voice dialing and other things using speech based systems. In this project we have researched all the areas where this speech based systems have significant use. All the technological aspects of this field is been studied and we have evaluated that our system would have an efficient use for the physically disabled people who can operate devices using speech recognition system and interact and control the device. Software available in the markets based on this technologies have various drawbacks such as they can operate using Bluetooth devices and also they are purchasable so these drawbacks were studied and are resolved in our system. Our system efficiently composes mails,search for a particular _le, writes a documents and saves it, reads out a selected _le and also saves it into audio format, also we have developed a speech calculator that performs maximum operations.Thus we conclude the successful developmentof our project

REFERENCES

[1] Yahya S H Khraisat" speaker independent voice recognition calculator" Department of Electrical and Electronics Engineering Contemporary engineering science vol 5,2012 no -3, 119-125
[2] Li Deng, Fellow, Xiano Li, \Machine Learning Paradigms For Speech Recognition \IEEE TRANSACTION ON AUDIO AND LANGUAGE PROCESSING, VOL 21 MAY 2013, ISBN: 1558-7916
[3] K Naga Abhishek Reddy, "A Comparative study on Speech Recognition Approaches and models" IJCTT-Volume 43 Number 2-Jan 2017.
[4] Khaled M. Alhawati, "Advances In Articial Intelligence" International Journal of com-puter, Vol: 9, No: 6, 2015
[5] Mathew P. Aylett; Alessandro Vinciarelli; Mirjam Westers; \Speech synthesis for the generation of articial personality \IEEE TRANSACTIONS ON AFFECTIVE COM-PUTING 2017 ISSN: 1949-3045
[6] Md .Almin; SyedaZinathAmanl \Design of Intelligent Home Assistant", 2016 7th Inter-national Conference on Intelligent Systems, Modelling and Simulations.
[7] Dr.KshamaKulhalli; AbhijeetJ.Patinas, "Personal Assistant Using Voice Recogni-tion Intelligence, \International Journal of Engineering Research and Technology", ISSN: 0974-3154 Volume 10 number 1(2017)
[8] **Prince Bose; ApurvaMalpthak; UtkarshBansal, "Digital Assistant for the Blind" 2nd International conference for Convergence in Technology (I2CT) 2017**