MULTI LANGUAGE VIRTUAL VOICE ASSISTANT

1. Tule Pavan Kumar
   B.Tech CSE 4th Year
   The ICFAI University
   Raipur, Chhattisgarh, INDIA
   Pavankumartule143@gmail.com

2. Pedalapu Vamsi Krishna
   B.Tech CSE 4th Year
   The ICFAI University
   Raipur,Chhattisgarh,INDIA
   pedalapuvamsikrishna@gmail.com

3.Vamsi Reddy Gayam
   B.Tech CSE 4th year
   The ICFAI University
   Raipur,Chhattisgarh,INDIA
   vamsireddy709@gmail.com

4. Dr.Palak keshwani
   Assistant Professor
   The ICFAI University
   Raipur,Chhattisgarh,INDIA
   palakeshwani@gmail.com

5.Dr.RameshYadav
AssistantProfessor
The ICFAI University
Raipur,Chhattisgarh,INDIA
rameshkumaryadav@iuraipur.edu

ABSTRACT— In this modern era, day to day life became smarter and interlinked with technology. We already know some voice assistants like Google Siri ,Alexa. etc. Here, we have proposed our voice assistance system. It can act as a basic medical prescriber, daily schedule reminder, note writer, calculator and a search tool and it can be easier to send emails without typing any word, searching on Google without opening the browser, and performing many other daily tasks like playing music, opening your favorite IDE with the help of a single voice command. In the current scenario, advancement in technologies is such that they can perform any task with same effectiveness or can say more effectively than us. By making this project, we realized that the concept of AI in every field is decreasing human effort and saving time. This project works on voice input and gives output through voice and displays the text on the screen. The main agenda of our voice assistant is to make people smart and give instant and computed results. The voice assistant takes the voice input through our microphone (Bluetooth and wired microphone) and it converts our voice into computer understandable language and gives the required solutions and answers which are asked by the user. This assistant connects with the world wide web to provide results that the user has questioned. Natural Language Processing algorithm helps computer machines to engage in communication using natural human language in many forms.

Keywords—Speech Recognition, API, Artificial Intelligence, Desktop Assistant, Python, Text to Speech, Virtual Assistant, Voice Recognition

I.INTRODUCTION

Artificial Intelligence when used with machines, it shows us the capability of thinking like humans. In this, a computer system is designed in such a way that typically requires interaction from human. As we know Python is an emerging language so it becomes easy to write a script for Voice Assistant in Python. The instructions for the assistant can be handled as per the requirement of user. Speech recognition is the Alexa, Siri, etc. In Python there is an API called Speech Recognition which allows us to convert speech into text. It was an interesting task to make my own assistant. It became easier to send emails without typing any word, searching on Google without opening the browser, and performing many other daily tasks like playing music, opening your favorite IDE with the help of a single voice command. In the current scenario, advancement in technologies is such that they can perform any task with same effectiveness or can say more effectively than us. By making this project, I realized that the concept of AI in every field is decreasing human effort and saving time.

 As the voice assistant is using Artificial Intelligence hence the result that it is providing are highly accurate and efficient. The assistant can help to reduce human effort and consumes time while performing any task, they removed the concept of typing completely and behave as another individual to whom we are talking and asking to perform task. But I added the concept of typing completely and behave as another individual to whom we are talking and asking to perform task. The assistant is no less than a human assistant but we can say that this is more effective and efficient to perform any task. The libraries and packages used to make this assistant focuses on the time complexities and reduces time.

The functionalities include, It can send emails, It can read PDF, It can send text on WhatsApp, It can open command prompt, your favourite IDE, notepad etc., It can play music, It can do Wikipedia searches for you, It can open websites like Google, YouTube, etc., in a web browser, It can give weather forecast, It can give desktop reminders of your choice. It can have some basic conversation.

The IDE used in this project is PyCharm. All the python files were created in PyCharm and all the necessary packages were easily installable in this IDE. For this project following modules and libraries were used i.e., pyttsx3, Speech Recognition, Datetime, Wikipedia, Smtplib, pywhatkit, pyjokes, pyPDF2, pyautogui, pyQt etc. We have created a live web front end design for interacting with the SANDY as it gives a design and interesting look while having the conversation. With the advancement SANDY can perform any task with same effectiveness or can say more effectively than us. By making this project, I realized that the concept of AI in every field is decreasing human effort and saving time. Functionalities of this project include, it can send emails, it can read PDF, It can send text on WhatsApp, It can open command prompt, your

favourite IDE, notepad etc., It can play music, It can do Wikipedia searches for you, It can open websites like Google, YouTube, etc., in a web browser, It can give weather forecast, It can give desktop reminders of your choice. It can have some basic conversation

Now the basic question arises in mind that how it is an AI? The virtual assistant that I have created is like if it is not an A.I, but it is the output of a bundle of the statement. But fundamentally, the mail purpose of A.I machines is that it can perform human tasks with the same efficiency or even more efficiently than humans. It is a fact that my virtual assistant is not a very good example of A.I., but it is an A.I.

II.LITERATURE SURVEY

Bassam A, Raja N. et.al written about statement and speech which is most significant. In the communication between human and machine arrangement was done through analog signal, which is converted by speech signal to digital wave. This technology is massively utilized, it has limitless uses and permit machines to reply appropriately and consistently to user voices, also offers useful and appreciated facilities. Speech Recognition System (SRS) is rising gradually and has indefinite applications. The research has revealed the summary of the procedure; it is a simple model [1]. B. S. Atal and L. R. Rabiner et al, explained regarding speech analysis, and result is regularly completed in combination with pitch analysis. The research described a pattern recognition technique for determining whether a given slice of a speech signal should be categorized as voiced speech, unvoiced speech, or silence, depending on dimensions finished on signal. The main restriction of the technique is the requirement for exercise the algorithm on exact set of dimensions picked, and for the specific recording circumstances [2]. V. Radha and C. Vimala et al, explained that most general mode of communication among human beings is speech. As this is the utmost technique, human beings would identical to utilize speech to interrelate with machines too. Because of this, autonomous speech identification has got a lot of reputation. Most techniques for speech recognition be like Dynamic Time Warping (DTW), HMM. For the feature mining of speech Mel Frequency Cepstrum Coefficients (MFCC) has been utilized which offers a group of characteristic vectors of speech waveform. Prior study has exposed MFCC to be more precise and real than rest characteristic mining approaches in the speech recognition. The effort has been completed on MATLAB and investigational outcomes depict that system is capable of identifying words at satisfactorily great accuracy [3]. T. Schultz and A. Waiel et al, explained about the spreading of speech technology products around the globe, the immovability to novel destination languages turns out to be a useful concern. As a significance, the research emphases on the query of how to port huge vocabulary incessant speech recognition (LVCSR) systems in a fast and well-organized manner. More particularly the research needs to evaluate acoustic models for a novel destination language by means of speech information from different source languages, but only restricted data from the destination language identification outcomes using language-dependent, independent and language-adaptive acoustic models are described and deliberated in the framework of Global Phone project which examines LVCSR methods in 15 languages.[4]. J. B. Allen et al described about the Language that is the utmost significant means of communication and speech is its major interface. The interface for human to machine, speech signal was converted into analog and digital wave shape as a machine understood. A technology enormously utilized and has limitless applications. Speech technologies permit machines to react appropriately and consistently to human speeches and offers valuable and appreciated services. The research provides a summary of the speech identification procedure, its basic model, and its application, techniques and also describes reasonable research of several techniques that are utilized for speech recognition system. SRS is enhancing gradually and has infinite applications. [5] Mughals Bapat, Pushpak Bhattacharyya et al, described a morphological analyzer for most of the NLP solicitations of Indian Languages. [11] During the work they described and estimated the morphological analyzer for Marathi language. They started by planning a to some extent homomorphism "boos trappable" encryption technique that functions during the function f is the techniques individual decryption function. The research showed a great accuracy for Marathi that adventures consistency in inflectional standards in engaging the Finite State Systems for demonstrating language in a sophisticated way. Grouping of post positions and the growth of FSA is one of significant assistances since Marathi have difficult morphtactics [6]. G. Muhammad, M. N. Huda et al, presented a model ASR for Bangla digit. Although Bangla is among the mostly spoken languages around the globe, some of the few works of Bangla ASR can be identified in the collected works, particularly Bangla accented in Bangladeshi. During this research, the quantity is gathered from publics in Bangladesh. Mel-frequency coefficients (MFCCs) dependent characteristics and hidden Markov model (HMM) dependent classifiers are utilized for identification. Dialectical variance make happen a part of performance deprivation. In situation of gender-based trials, female spoken digits had greater accuracy rates than those by male spoken digits [7]. Sean R Eddy et al operated on Hidden Markov models which are a common statistical designing approach for 'linear' issues like sequences or time series and have been extensively utilized in speech identification requests for twenty years. Inside the HMM formalism, it is probable to relate formal, completely probabilistic techniques to profiles and gapped structure arrangements.[12] Profiles based on Hidden Markov model have fixed most of the concerns related with typical profile analysis. HMMs offer a steady theory for notching insertions and deletions, and a constant structure for joining structural and sequence data. HMM based numerous sequence arrangements is quickly refining. Homolog recognition based on HMM is previously adequately influential for HMM techniques to relate satisfactorily to much more difficult threading techniques for protein reverse fold [8].

III.PROPOSED SYSTEM

The proposed system will have the following functionality:

(a) The system will keep listening for commands and the time for listening is variable which can be changed according to user requirements.

(b) If the system is not able to gather information from the user input it will keep asking again to repeat till the desired no. of times.

(c) The system can have both male and female voices according to user requirements.

(d) Features supported in the current version include playing music, emails, texts, search on Wikipedia, or opening system installed applications, opening anything on the web browser, etc.

(e)The system will keep listening for commands and the time for listening is variable which can be changed according to user requirements.

(f) If the system is not able to gather information from the user input it will keep asking again to repeat till the desired no. of times.

(g) The system can have both male and female voices according to user requirements

a) *FUNCTIONALITIES OF THIS PROJECT :*

- It can send emails.
- It can read PDF.
- It can send text on what Sapp.
- 4. It can open command prompt, your favourite IDE, notepad etc.
- 5. It can play music.
- 6. It can do Wikipedia searches for you.
- 7. It can open websites like Google, YouTube, etc., in a web browser.
- 8. It can give weather forecast.
- 9.It can have some basic conversation.

b) *WE GIVEN THE ADVANCE FEATURES OF THIS PROJECT :*

- We have created a live Web front end deign by using Html CSS JavaScript for interacting with the Assistance as it gives a design and interesting look while having the conversation.
- It can switch to your comfortable language like Telugu Hindi etc, by default we given the English language, whenever you given the such commands as switch to Telugu, switch to Hindi etc, that time voice assistant completely switch to your comfortable language that time it can

start recognizing to your comfortable language only, that time it did not recognizing English command, it can only take to your comfortable (Telugu, Hindi) language commands.

- It can automatically typing words, what you're saying.
- It can give desktop reminders of your choice. like check battery percentage, set alarm, check internet speed, check Ip address etc.
- It can translate the words like English to Telugu, Telugu to English etc.
- It can shut down your system and it can sleep your system; it can lock the system.
- It can write a note and save a note
- It can take Photo to you and also its shows to you photo how its takes.
- It can take screen shots.
- It can take screen records.
- It can open files in to your system

PRESENT SYSTEM

We are familiar with many existing voice assistants like Alexa, Siri, Google Assistant, Cortana which uses concept of language processing, and voice recognition. They listens the command given by the user as per their requirements and performs that specific function in a very efficient and effective manner. As these voice assistants are using Artificial Intelligence hence the result that they are providing are highly accurate and efficient. These assistants can help to reduce human effort and consumes time while performing any task, they removed the concept of typing completely and behave as another individual to whom we are talking and asking to perform task. These assistants are no less than a human assistant but we can say that they are more effective and efficient to perform any task. The algorithm used to make these assistant focuses on the time complexities and reduces time. But for using these assistants one should have an account (like Google account for Google assistant, Microsoft account for Cortana) and can use it with internet connection only because these assistants are going to work with internet connectivity. They are integrated with many devices like, phones, laptops, and speakers etc.

and specialized vocabularies are available through third-party publishers.

IV. METHODOLOGY OF VIRTUAL ASSISTANT USING PYTHON
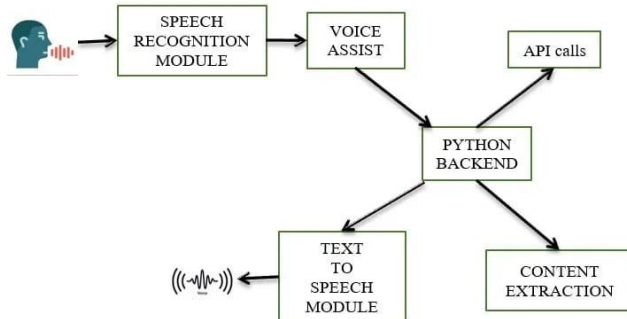
- ### Speech Recognition module: The system uses Google's online speech recognition system for converting speech input to text. The speech input Users can obtain texts from the special corpora organized on the computer network server at the information centre from the microphone is temporarily stored in the system which is then sent to Google cloud for speech recognition. The equivalent text is then received and fed to the central processor.

- ### Python Backend: The python backend gets the output from the speech recognition module and then identifies whether the command or the speech output is an API Call and Context Extraction. The output is then sent back to the python backend to give the required output to the user.

- ### API Calls: API stands for Application Programming Interface. An API is a software intermediary that allows two applications to talk to each other. In other words, an API is a messenger that delivers your request to the provider that you're requesting it from and then delivers the response back to you.

- ### Content Extraction: Context extraction (CE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. In most cases, this activity concerns processing human language texts using natural language processing (NLP). Recent activities in multimedia document processing like automatic annotation and content extraction out of images/audio/video could be seen as context extraction TEST RESULTS.

- ### Text-to-speech module: Text-to-Speech (TTS) refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. TTS engines with different languages, dialects

SYSTEM DESIGN

The data flow for SANDY is as follow:

- START

Live Web front end design for interaction will appear on screen.

- INPUT

Live Web front end design for interaction will appear on screen.

- PERFORM

It will perform the required task for the user like opening notepad, searching on browser, sending mails, playing songs etc.

- EXIT

It keeps on asking for the command from user until the user say "Quit". Once the user says "Quit", it exits

DATA FLOW FOR SANDY

The system is designed using the concept of Artificial Intelligence and with the help of necessary packages of Python. Python provides many libraries and packages to perform the tasks, for example pyPDF2 can be used to read PDF. The details of these packages are mentioned in Software Details of this report.

The data in this project is nothing but user input, whatever the user says, the assistant performs the task accordingly. The user input is nothing specific but the list of tasks which a user wants to get performed in human language i.e., English, Telugu.

PROPOSED TO PLAN OF WORK

The work started with analyzing the audio commands given by the user through the microphone. This can be anything like getting any information, operating a computer's internal files, etc. This is an empirical qualitative study, based on reading above mentioned literature and testing their examples. Tests are made by programming according to books and online resources, with the explicit goal to find best practices and a more advanced understanding of Voice Assistant.
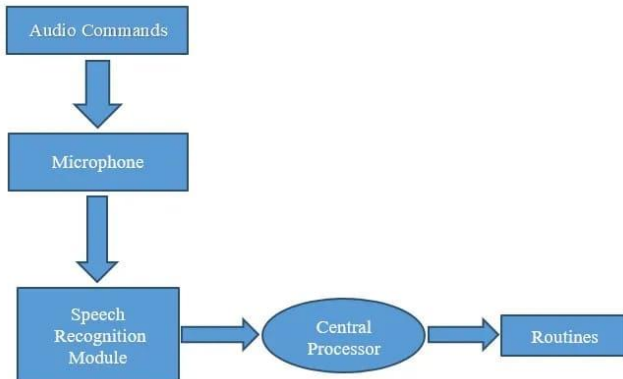
shows the workflow of the basic process of the voice assistant. Speech recognition is used to convert the speech input to text. This text is then fed to the central processor which determines the nature of the command and calls the relevant script for execution.

But the complexities don't stop there. Even with hundreds of hours of input, other factors can play a huge role in whether or not the software can understand you. Background noise can easily throw a speech recognition device off track. This is because it does not inherently have the ability to distinguish the ambient sounds it "hears" of a dog barking or a helicopter flying overhead, from your voice. Engineers have to program that ability into the device; they conduct data collection of these ambient sounds and "tell" the device to filter them out. Another factor is the way humans naturally shift the pitch of their voice to accommodate for noisy environments; speech recognition systems can be sensitive to these pitch changes.

SOFTWARE DETAILS:-

THE IDE USED IN THIS PROJECT IS PYCHARM. ALL THE PYTHON FILES WERE CREATED IN PYCHARM AND ALL THE NECESSARY PACKAGES WERE EASILY INSTALLABLE IN THIS IDE. FOR THIS PROJECT FOLLOWING MODULES AND LIBRARIES WERE USED I.E. PYTTSX3, SPEECHRECOGNITION, DATETIME, WIKIPEDIA, PYWHATKIT, PYJOKES, PYPDF2, PYAUTOGUI, ETC. I HAVE CREATED A LIVE WEB FOR INTERACTING WITH THE SANDY AS IT GIVES A DESIGN AND INTERESTING LOOK WHILE HAVING THE CONVERSATION.

IMPLEMENTATION OF WORK DETALIS:-

SANDY a desktop assistant is a voice assistant that can perform many daily tasks of desktop like playing music, opening your favorite IDE with the help of a single voice command. SANDY is different from other traditional voice assistants in terms that it is specific to desktop and user does not need to make account

to use this, it does not require any internet connection while getting the instructions to perform any specific task.

REAL LIFE APPLICATION

- Saves time: SANDY is a desktop voice assistant which works on the voice command offered to it, it can do voice searching, voice-activated device control and can let us complete a set of tasks.
- Conversational interaction: It makes it easier to complete any task as it automatically do it by using the essential module or libraries of Python, in a conversational interaction way. Hence any user when instruct any task to it, they feel like giving task to a human assistant because of the conversational interaction for giving input and getting the desired output in the form of task done.
- Reactive nature: The desktop assistant is reactive which means it know human language very well and understand the context that is provided by the user and gives response in the same way, i.e. human understandable language, English. So user finds its reaction in an informed and smart way.
- Reactive nature: The main application of it can be its multitasking ability. It can ask for continuous instruction one after other until the user "QUIT" it.
- No Trigger phase: It asks for the instruction and listen the response that is given by user without needing any trigger phase and then only executes the task.

DATA IMPLEMENTATION AND PROGRAM EXECUTION

As the first step, install all the necessary packages and libraries. The command used to install the libraries is "pip install" and then import it. The necessary packages included are as follows:

LIBRARIES AND PACKAGES

- pyttsx3: It is a python library which converts text to speech.
- Speech Recognition: It is a python module which converts speech to text
- pywhatkit: It is python library to send WhatsApp message at a particular time with some additional features.
- Datetime: This library provides us the actual date and time.

- Wikipedia: It is a python module for searching anything on Wikipedia.
- Smtplib: Simple mail transfer protocol that allows us to send mails and to route mails between mail servers.
- Pyjokes: It is a python library which contains lots of interesting jokes in it.
- Web browser: It provides interface for displaying web-based documents to users.
- Pyautogui: It is a python library for graphical user interface.
- OS:- It represents Operating System related functionality.
- Play Sound: The play sound module is the simplest module to use for playing sound. This module works on both Python 2 and Python 3, and is tested to play wav and mp3 files only. It contains only one method, named play sound
- Translate: Python google trans is a module to translate text. It uses the Google Translate Ajax API to detect languages and translate text. Python google trans list languages
- GTTS: gTTS (Google Text-to-Speech) is a Python library and CLI tool to interface with Google Translate text-to-speech API. We will import the gTTS library from the gtts module which can be used for speech translation. The text variable is a string used to store the user's input. The text can be replaced by anything of your choice within the quotes.
- Requests: The requests module allows you to send HTTP requests using Python. The HTTP request returns a Response Object with all the response data (content, encoding, status, etc).
- Capture: This module can be take photos and it can be used to capture the screen
- BS4soup: The BS4 stands for beautiful Soup The beautiful Soup is a Python library which is used for pulling out data of the HTML & XML files using the Python program. The beautiful Soup library was created basically for the purpose of web scraping
- Subprocess: Subprocess has a method call () which can be used to start a program. Save process output (stdout) We can get the output of a program and store it in a string directly using check output.

SYSTEM TESTING:-

The system testing is done on fully integrated system to check whether the requirements are matching or not. The system testing for SANDY desktop assistant focuses on the following four parameters

FUNCTIONALITY:-

In this we check the functionality of the system whether the system performs the task which it was intended to do. To check the functionality each function was checked and run, if it is able to execute the required task correctly then the system passes in that particular functionality test. For example, to check whether SANDY can search on Google or not, as we can see in the given picture, user said "Open Google", then Sandy asked," What should I search on Google?" then user said, "What is Python", sandy open Google and searched for the required

Output
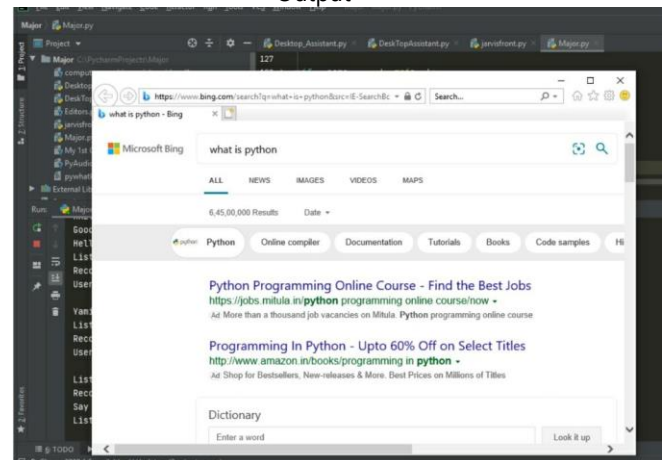
A. USEBILITY

Usability of a system is checked by measuring the easiness of the software and how user friendly it is for the user to use, how it responses to each query that is being asked by the user. It makes it easier to complete any task as it automatically do it by using the essential module or libraries of Python, in a conversational interaction way. Hence any user when instruct any task to it, they feel like giving task to a human assistant because of the conversational interaction for giving input and getting the desired output in the form of task done. The desktop assistant is reactive which means it know human language very well and understand the context that is provided by the user and gives response in the same way, i.e., human understandable language, English. So, user finds its reaction in an informed and smart way. The main application of it can be its multitasking ability. It can ask for continuous instruction one after other until the user "QUIT" it. It asks for the instruction and listen the response that is given by user without needing any trigger phase and then only executes the task.

B. SECURITY

The security testing mainly focuses on vulnerabilities and risks. As SANDY is a local desktop application, hence there is no risk

of data breaching through remote access. The software is dedicated to a specific system so when the user logs in, it will be activated.

C. STABILITY

Stability of a system depends upon the output of the system, if the output is bounded and specific to the bounded input then the system is said to be stable. If the system works on all the poles of functionality, then it is stable.

LIVE INTERACTIVE FRONT END WEB DESGIN FOR THIS ASSISTANCE
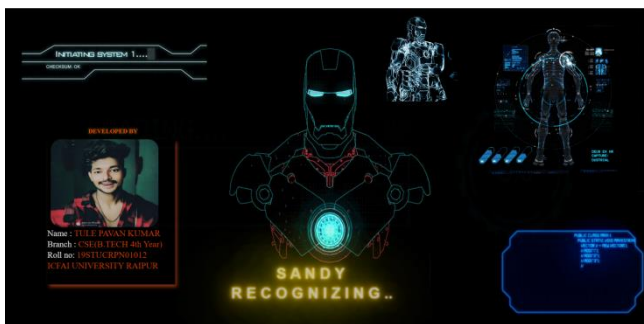
We have created a live web front end design for interacting with the SANDY as it gives a design and interesting look while having the conversation.
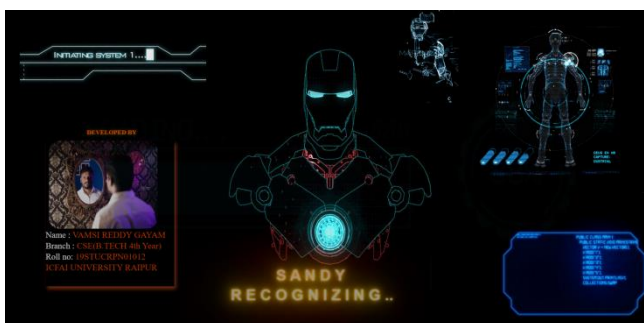
ADVANTAGES

- Easily configured to perform many of regular tasks by simply giving voice commands
- Voice based search that is boon for many like senior citizens who are not comfortable using the keyboard
- Able to write the text through both keyboard and voice input
- Different notead commands such as open save clear etc
- Requires less consumpion of time in writing text
- Lower operation cost
- Open different software, based on voice input

DIS ADVANTAGES

- Not good in Noisy place
- A user has to speak much louder and slower than their normal speed and volume
- Low accuracy

V.CONCLUSION

SANDY is a very helpful voice assistant without any doubt as it saves time of the user by conversational interactions, its effectiveness and efficiency. But while working on this project, there were some limitations encountered and also realized some scope of enhancement in the future which are mentioned below:

VII. SCOPE AND FUTURE WORK

- Add Password protection
- Add Face Recognition
- Make SANDY to learn more on its own and develop a new skill in it.
- SANDY android app can also be developed.
- Make more Sandy voice terminals.
- Voice commands can be encrypted to maintain security.

VIII. REFERENCES

[1] M. Bapat, H. Gune, and P. Bhattacharyya, "A paradigm-based finite state morphological analyzer for marathi," in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.

[2] B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212, 1976.

[3] V.Radha and C. Vimala, "A review on speech recognition challenges and approaches," doaj. org, vol. 2, no. 1, pp. 1–7, 2012.

[4] T. Schultz and A. Waibel, "Languageindependent and language adaptive acoustic modeling for speech recognition", Speech Communication, vol. 35, no. 1, pp. 31–51, 2001.

[5] J. B. Allen, "From lord rayleigh to shannon: How do humans decode speech," in International Conference on Acoustics, Speech and Signal Processing, 2002.

[6] M. Bapat, H. Gune, and P. Bhattacharyya, "A paradigm-based finite state morphological analyzer for marathi," in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, 2010.

[7] G. Muhammad, Y. Alotaibi, M. N. Huda, et al., pronunciation variation for asr: A survey of the "Automatic speech recognition for bangla digits," literature," Speech Communication, vol. 29, no. in Computers and Information Technology, 2009. 2, pp. 225–246, 1999.

[8] S. R. Eddy, "Hidden Markov models," Current opinion in structural biology, vol. 6, no. 3, pp. 361–365, 1996.

[9] excellent style manual for science writers is "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010.

[10] Srivastava S., Prakash S. (2020) Security Enhancement of IoT Based Smart Home Using Hybrid Technique. In: Bhattacharjee A., Borgohain S., Soni B., Verma G., Gao XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_44

[11] S. Srivastava and S. Prakash, "An Analysis of Various IoT Security Techniques: A Review," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 355- 362, doi: 10.1109/ICRITO48877.2020.9198027

[12] Saijshree Srivastava, Surya Vikram Singh, Rudrendra Bahadur Singh, Himanshu Kumar Shukla," Digital Transformation of Healthcare: A blockchain study" International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 5, May 2021