**PERSONAL VOICE ASSISTANT**

A Project Report

Submitted in the partial fulfillment of the requirements for the award of the degree of

# Bachelor of Technology in

Department of Computer Science and Engineering

By

2010030055 GANDE SAITEJA

2010030236 MD. ADNAN

2010030272 GILLA SAMANTH

2010030402 K. TRIBHUVAN

Under the esteemed guidance of

# Dr. ARPITA GUPTA



Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

**DECLARATION**

The Project Report entitled “Personal voice assistant” is a record of bonafide work of Mr. GandeSaiteja (2010030055),Mr. Gilla Samanth (2010030272), Mr. Md. Adnan (2010030236) and Mr. K. Tribhuvan (2010030402) submitted in partial fulfillment for the award of B.Tech in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/University/Institute.

Mr. GANDE SAITEJA – 2010030055

Mr. GILLA SAMANTH – 2010030272

Mr. MD. ADNAN – 2010030236

MR. K. TRIBHUVAN - 2010030402

**CERTIFICATE**

This is to certify that the Project Report entitled “Personal voice assistant” is being submitted by Mr. GANDE SAITEJA bearing Regd. No. 2010030055, Mr. GILLA SAMANTH bearing Regd. No. 20100300272, Mr. MD. ADNAN bearing Regd. No. 2010030236 and Mr. K. TRIBHUVAN bearing Regd. No. 2010030402 submitted in partial fulfillment for the award of B.Tech in Computer Science and Engineering to the K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other department/ University/ Institute.

## Signature of the Supervisor

Dr. Arpita Gupta

(Assistant Professor)

## Signature of the HOD Signature of the External Examination

**ACKNOWLEDGEMENT**

It is great pleasure for us to express our gratitude to our honorable President

**Sri. Koneru Satyanarayana**, for giving the opportunity and platform with facilities in accomplishing the project-based laboratory report.

We express our sincere gratitude to our Principal **Dr. L. Koteswara Rao** for his administration towards our academic growth.

We express our sincere thanks to HOD-CSE **Dr. M Chiranjeevi** for his leadership and constant motivation provided in successful completion of our academic semester. I record it as my privilege to deeply thank for providing us the efficient faculty and facilities to make our ideas into reality.

We express immense gratitude to our guide and Associate Professor **Dr. Arpita Gupta** for her novel association of ideas, encouragement, appreciation and intellectual zeal which motivated us to venture this project successfully.

Finally, it is pleased to acknowledge the indebtedness to all those who devoted themselves directly or indirectly to make this project report success.

**ABSTRACT**

In the Modern Era of fast moving technology we can do things which we never thought we could do before but, to achieve and accomplish these thoughts there is a need for a platform which can automate all our tasks with ease and comfort. Thus we need to develop a Personal Assistant having brilliant powers of deduction and the ability to interact with the surroundings just by one of the materialistic form of human interaction i.e. HUMAN VOICE. The Hardware device captures the audio request through microphone and processes the request so that the device can respond to the individual using in-built speaker module. For Example, if you ask the device ’what’s the Time?’ using its built-in skills, it looks up the time respectively and then returns the response to the customer through connected speaker.

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| Chapter No. | Title | | Page No. |
| 1. | Introduction | | 7 |
| 1.1.  2. | Advantages  Problem Statement | | 7  8 |
| 3. | Methodology | | 9 |
| 4. | Proposed System | | 10 |
| 5. | Techniques | | 11 |
| 6. | Literature Survey | | 12 |
| 7. | System Design & Implementation | | 13 |
| 7.1.  7.2.  8.  9.  9.1.  9.2.  10.  11.  12.  13. | Data Flow Diagram    Flow chart  Model  System Requirements  Software requirements  Hardware requirements  Implementation  Result  Conclusion & Future work  References | | 14  15  16  17  17  17  18-19  20-21  22  23 |
|  | |
|  |

**1. INTRODUCTION**

A voice assistant is a digital assistant that uses voice recognition, language processing algorithms, and voice synthesis to listen to specific voice commands and return relevant information or perform specific functions as requested by the user.

Virtual Assistants are software programs that help you ease your day to day tasks, such as showing weather report, creating reminders, making shopping lists etc. They can take commands via text (online chat bots) or by voice. Voice based intelligent assistants need an invoking word or wake word to activate the listener, followed by the command.

This system is designed to be used efficiently on desktops. Personal assistant software improves user productivity by managing routine tasks of the user and by providing information from online sources to the user.

In fact, voice searches tend to be far faster than written searches. We are capable of speaking around 150 words per minute at the same rate as we can write 40 words per minute.

**1.1) Advantages**

Voice Assistant allows you to gain the perks of high-end technology and its functionalities. Our proposed application points to many advantages: 1. Our proposed application provides security to the user as it can authenticate the authorized user using Face Recognition technique. 2. The face recognition technology make the system secure and robust for the user as this does not required any input from the user through keyboard or mouse. 3. The application provides flexibility to the user as it can send email just listening the command given by the user.

**2. PROBLEM STATEMENT**

We already have multiple virtual assistants. But we hardly use it. These systems can understand English phrases, but they fail to recognize in our accent. Our way of pronunciation is way distinct` from theirs. Also, they are easy to use on mobile devices than desktop systems. There is need of a virtual assistant that can understand in Indian accent and work on desktop system.

Artificial Intelligence personal assistants have become plentiful over the last few years. Applications such as Siri, Bixby, Ok Google and Cortana make mobile device users’ daily routines that much easier. You may be asking yourself how these functions. Well, the assistants receive external data (such as movement, voice, light, GPS readings, visually defined markers, etc.) via the hardware’s sensors for further processing - and take it from there to function accordingly. Not too long ago, building an AI assistant was a small component of developers’ capacities; however, nowadays, it is quite a realistic objective even for novice programmers. To create a simple personal AI assistant, one simply needs dedicated software and around an hour of working time. The voice assistant is design to make the work easier of the user. As user can give command to them without making visual access to the screen. The biggest disadvantage of this system is that confidential data can be accessed by unauthorized user so the privacy can be breached. Due to this, the confidentiality, integrity and availability (CIA) of user data is affected.

**3. METHODOLOGY**

The overall system design consists of following phases:

(a) Data collection in the form of speech.

(b) Voice analysis and conversion to text

(c) Data storage and processing

(d) Generating speech from the processed text output

In first phase, the data is collected in the form of speech and stored as an input for the next phase for processing. In second phase, the input voice is continuously processed and converted to text using STT. In next phase the converted text is analysed and processed using Python Script to identify the response to be taken against the command. Finally once the response is identified, output is generated from simple text to speech conversion using TTS.

**4. PROPOSED SYSTEM**

The proposed system will provide following features:

1) It always keeps listing for its name and wakes up to response upon calling with the assigned functionality.

2) It keeps learning the sequence of questions asked to it related to its context which it remembers for the future. So when the same context is mentioned, it starts a conversation with you asking relevant questions.

3) Performing Arithmetic calculations based on voice commands and giving back the computed solution through a voice.

4) Searching Internet based on user’s voice input and giving back the reply through a voice with further interactive questions by machine.

5) Other features such as playing music, setting an alarm, checking weather conditions of device’s location. Setting reminders, spell-correct, etc can be performed by an input from user’s voice.

**5. TECHNIQUES**

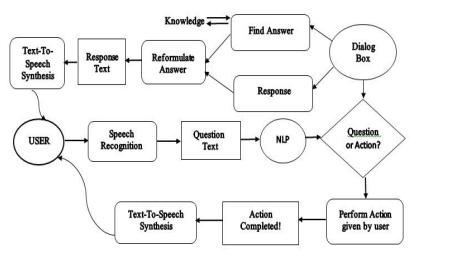
Conventional Speech Enchantement (SE) Technique:

It uses the parameters required for noise suppression are derived based on statistical models and are estimated from the noisy observations. The aim of a conventional speech enhancement system is to suppress the noise in a noisy speech signal. For robust speech recognition, such a system is used as a preprocessor to a speech recognizer Since it produces a clean speech signal, no changes in the recognition system are necessary to make it robust. A number of speech enhancement techniques have been reported in the literature [32]. They include spectral subtraction, Wiener and Kalman filtering , MMSE estimation, comb filtering ,subspace methods and phase spectrum compensation. The technique that has been used most for this purpose is spectral subtraction, in which the power spectrum of clean speech Pxx(m, f) is estimated by explicitly subtracting the noise power spectrum Pww(f) from the noisy speech power spectrum Pyy(m, f) using Equation (6.8). This requires information about the noise power spectrum, which can be estimated from the nonspeech frames detected by voice activity detection (VAD). However, it is not always possible to detect the nonspeech frames correctly, which affects the estimation of the noise power spectrum and may result in poor speech enhancement. Hence, a practical spectral subtraction scheme has the form.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S. No. | Authors | Title | Publishing | Technique | Dataset | Pros | Cons |
| 1 | Dr. Kshama V.Kulhalli,  Dr. Kotra Abhijit,  J. Patankar | Personal Assistant with Voice Recognition Intelligence Techniques | 10 April 2020-Published by Research Trend | CTC based LSTM acoustic model .  SVD based compression quantization.  PARI the wellknown Technique . | It consists of 14956 voices with different languages | Responds Faster than the online Voice Search Appication. | Performance decreases with increase in dataset. |
| 2 | Dr.Subhash Mani Kaushal,  Dr. Megha Mishra. | A Review on Voice Assistance using Python | April 2021-Published by IOP Publishing Ltd. | Node MCU chips using Arduino IDE. python script. | NA | Responds faster  Better for physically Abled | Couldn't determine the speed of responding. |
| 3 | Dr.Abhay Dekate.  Dr.Chaitanya Kulkarni.,  Dr.Rohan Killedar. | A Literature Review on Study of Voice Controlled Personal Assistant Device. | December 2021. | Firebase cloud system architecture.  Audio Capture and Audio Playback | Many machine voices  And in many languages | can control IoT applications.  Lowers interacting with multiple sub-Systems. | Should upgrade so that there will be an improvement in terms of speed and accuracy |
| 4 | Parul Sharma , Yash Paul Singh Berwal , Wiqas Ghai | Voice Recognition Intelligence in there preferable language | 14November 2019 | Audio Capture and Audio Playback | 17,929  DNN Sounds with compressed background noise. | Responds faster Responds faster. Simple working Structure. | Performance decreases with increase in dataset. Without internet can be used . |

**6. LITERATURE SURVEY**

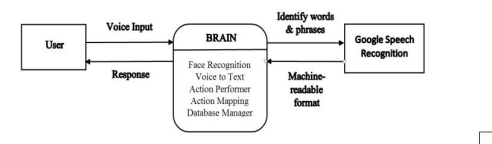
**7. SYSTEM DESIGN AND IMPLEMENTATION**

****

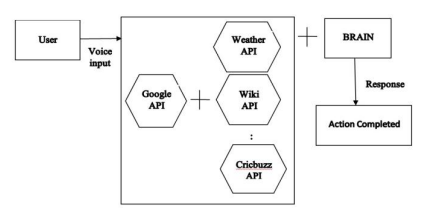
**FIG:(7.1)**

The proposed model of the voice assistant is as shown in the above figure 1. The model consists of user input through microphone to accept commands from the user. These commands are then go through Speech Recognition, it is the ability of a machine or program to identify words and phrases in spoken languages and convert them to a machine-readable format. DFD is a graphical representation which provides information flow between input and output data.

**7.1) DATA FLOW DIAGRAM**

** FIG:(7.2)**

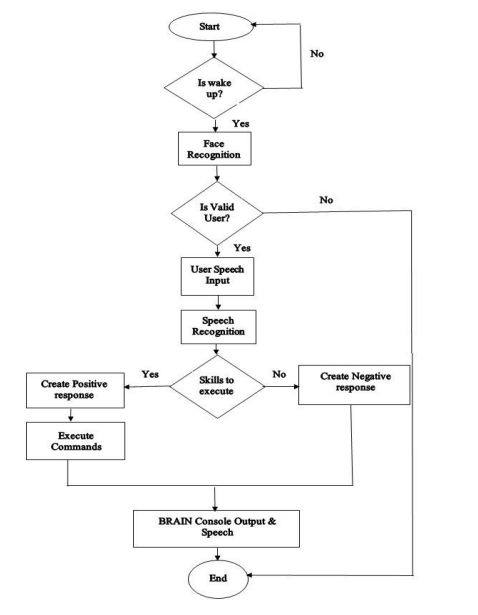
The user gives the input in the form of voice; this voice command is recognized by the application. Then it will check whether it is the authorized user, then action is performed as per the command given by the user. Command given is compared as a form of action and question and responsed with the dialog box or search through the knowledge base.

****

**FIG:(7.3)**

Input is given by user in the form of voice. GoogleVoiceAPI will convert this voice data in text form and then the action is performed by the voice assistant according to the command given by the user by comparing with the dialog box and knowledge base.

**7.2) FLOW CHART**

****

**FIG:(7.4)**

**8. MODEL**

**HIDDEN MORKOV MODEL**

Hidden Markov Models (HMM) are mainly used for general-purpose speech recognition systems. In general, the speech signals are observed as a stationary signal whose amplitude and frequency remains constant. It takes a very short time scale for a speech to be estimated as a static process. In HMM, it is possible to train the data set automatically, which makes it easy computationally, and hence it is extensively used. The HMM would generate a sequence of n-dimensional real-valued vectors every ten milliseconds (with ‘n’ being a small integer value such as 5 or 10) in speech recognition. The first coefficient of a Fourier transform of a small part of the speech is extracted and decorrelated with the cosine transform to calculate the cepstral coefficient vectors. These HMM states can be used to normalize different recordings and the speaker conditions using the cepstral normalization method. Vocal tract length normalization (VLTN) can be used to further normalize male-female or other speaker criteria. To capture the speech dynamics, linking and linear discriminant analysis (LDA) based projects can be used, which is followed by either the heteroscedastic LDA step or global semi-tied covariance transform method. To optimize the classification-related measure of the training data, many systems employ discriminative training strategies, which can be used to avoid the purely statistical approach to estimate the HMM parameter. The HMM is used to convert speech features into HMM parameters and calculate all speech samples’ likelihood. Recognition of this likelihood of speech samples is used to recognize the spoken words. The extracted features based on the parameters such as acoustic models, pronunciation dictionary, and the language model for which speech recognition is required.

**9. SYSTEM REQUIREMENTS**

**9.1) Software requirements:**

The major software requirements of the project are as follows:

Language : Python

Operating system : Windows 10

Tools : Jupyter Notebook, Kaggle Datasets

**9.2) Hardware requirements:**

The hardware requirements that map towards the software are as follows:

RAM : 4.00 GB

Processor : Intel(R) Core(TM) i5-4210U CPU @ 1.70GHz 1.70 GHz

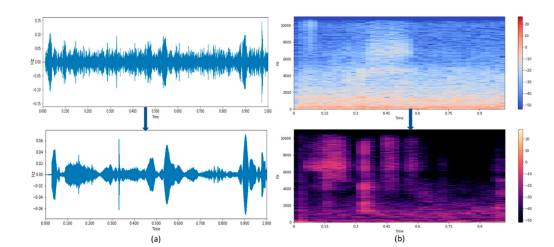
**10. IMPLEMENTATION**

**CODE**

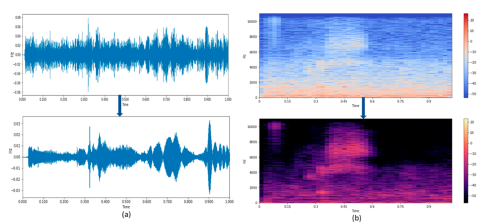
1. import speech\_recognition as sr
2. from gtts import gTTS
3. from playsound import playsound
4. import os
5. import datetime as dt
6. import pywhatkit as pk
7. import webbrows
8. listener= sr.Recognizer
9. def speak(cmd):
10. tts = gTTS(cmd, lang="te")
11. tts.save("audio.mp3")
12. playsound("audio.mp3")
13. os.remove("audio.mp3")
14. va\_name="సాయి"
15. speak(" నమస్కారము నేను మీ పర్సనల్ వాయిస్ అసిస్టెంట్ మీకు ఏ విధంగా సహాయ పడగలను")
16. def take\_cmd(check):
17. command=""
18. try:
19. with sr.Microphone() as source:
20. print("listening")
21. audio=listener.listen(source)
22. if check:
23. command = listener.recognize\_google(audio, language="te"
24. )if va\_name in command:
25. command=command.replace("సాయి","")
26. print(command)
27. # speak(command)
28. else:
29. command=""
30. else:
31. command = listener.recognize\_google(audio, language="en-US")
32. except:
33. print("check your mic")
34. return command
35. while True:
36. final\_cmd=take\_cmd(True)
37. if final\_cmd!="":
38. if " టైం" in final\_cmd
39. current\_time=dt.datetime.now().strftime("%I:%M %p"
40. speak(current\_time)
41. if "యూట్యూబ్" in final\_cmd:
42. speak(" ఏ వీడియో ప్లే చేయాలో చెప్పండి")
43. final\_cmd=take\_cmd(False)
44. pk.playonyt(final\_cmd)
45. speak("ఆనందించండి ")
46. if "గూగుల్" in final\_cmd:
47. speak("ఏమి వెతకాలో చెప్పండి")
48. final\_cmd=take\_cmd(False)
49. pk.search(final\_cmd)

**11. RESULT**

One of the numerous issues that automatic speech recognition systems face is processing spontaneous speech. Spontaneous speech is characterized as utterances that comprise Sensors 2021, 21, 7025 9 of 15 well-formed phrases similar to those found in written texts. Disfluencies (complete pauses, repetitions, false starts, and so on) are the main characteristics of this type of speech, and numerous studies have concentrated on detecting and correcting them. The waveform and the spectrogram results produced by the ISE algorithm are shown in show a speech signal’s time waveform and spectrogram, respectively. The noisy waveform is getting transferred to noise reduced waveform signal using the ISE algorithm.

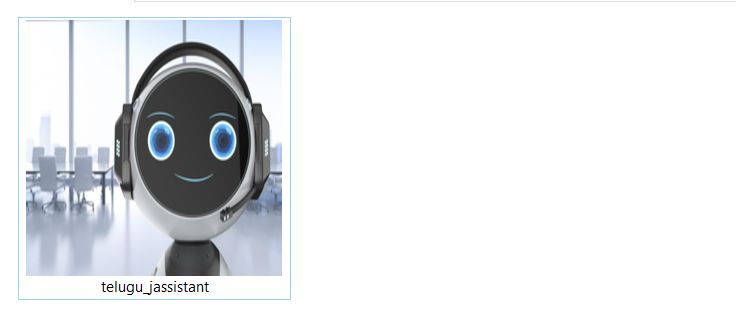
****

**FIG:(11.1)**

****

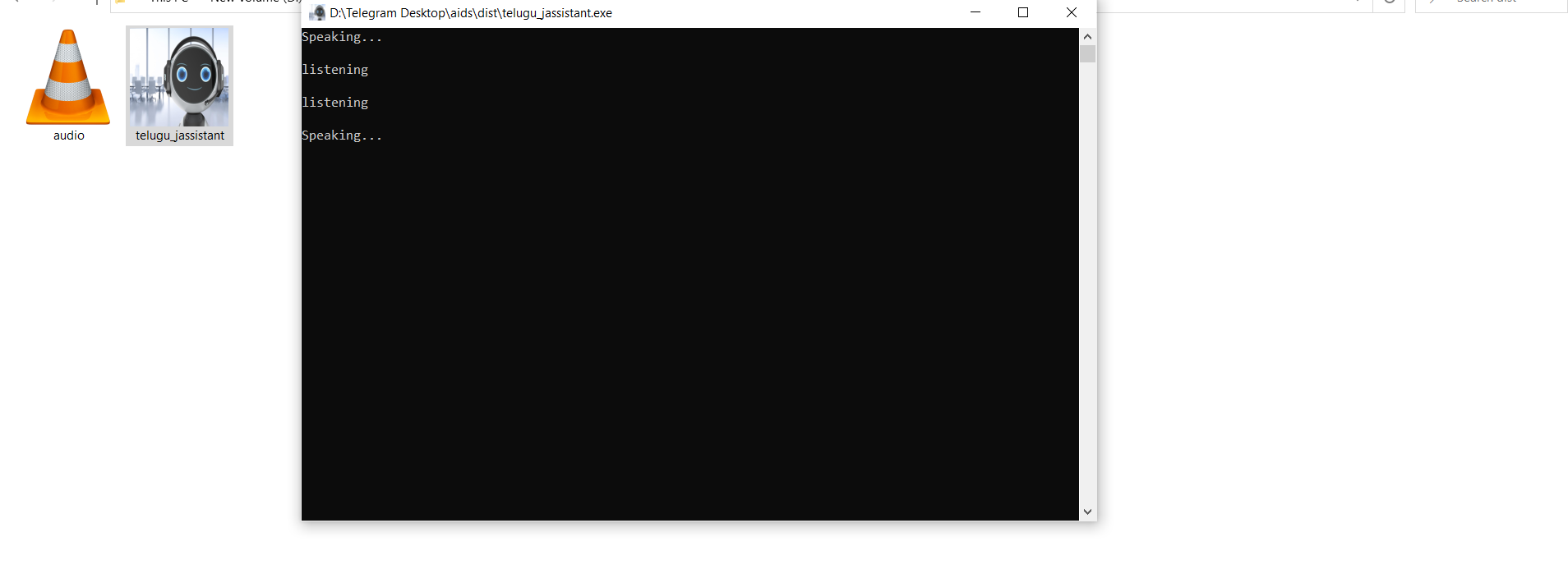
**FIG:(11.2)**

By clicking on the image personal voice assistant will be activated and it will respond and talk back to the user in comfortable language.

****

**FIG:(11.3)**

Here user can see the audio file in which the user commands will be saved and assistant will be asset the user according to user saved commands.

****

**FIG:(11.4)**

**12. CONCLUSION & FUTURE WORK**

Voice Controlled Personal Assistant System will utilize natural language processing and artificial intelligence techniques to create a smart assistant that can perform IoT applications and even solve user queries by using web searches...Using it can reduce human effort required for interaction with many other subsystems, which would otherwise need to be performed manually. Voice Controlled Personal Assistant System will use the Natural language processing and can be integrated with artificial intelligence techniques to achieve a smart assistant that can control IoT applications and even solve user queries using web searches.. It can be designed to minimize the human efforts to interact with many other subsystems, which would otherwise have to be performed manually

**FUTURE WORK**

Using this system as a framework, the system can be expanded to features security. Security is important these days so it can be combined with this system to give more advanced security features. In this, the voice authentication technology can be implemented for more security. More advancement are possible like operating on various tones or accents from different regions that mean it should be able to perform operations on various voice tones and accents.. Further modifications are possible like learning the answer of questions that are not known by the voice assistant and replying whenever next time the same question is put up by the user.

**13. REFERENCE**

* Yash Mittal, Pradhi Toshniwal “A voice-controlled multifunctional Smart Home Automation System”, 2015 India Conference (INDICON), IEEE.
* Prerna Wadikar, Nidhi Sargar, Rahool Rangnekar, Prof.Pankaj Kunekar “Home Automation using Voice Commands in the Hindi Language”, 2020.
* Nagesh Singh Chauhan,“Build Your First Voice Assistant”,March,2019 <https://towardsdatascience.com/build-your-first-voice-assistant>.
* Aditi Bhalerao, Samira Bhilare, Anagha Bondade, Monal Shingade, Aradhana Deshmukh, “Smart Voice Assistant: a universal voice control solution for non-visual access to the Android operating system”, Jan-2019
* Dongmahn SEO, Suhyun KIM, Gyuwon SONG, Seung-gil, "Speech-to-Text-based Life Log System for Smartphones”, IEEE International Conference on Consumer Electronics (ICCE), 2020