

Human Helping Hand Using Advance Technology

Abhishek Kumar
Computer Science and Engineering
University of Engineering and
Management
Jaipur, India
sinhaabhishek3021@gmail.com

Prem Kumar Vishwakarma
Computer Science and Engineering
University of Engineering and
Management
Jaipur, India
prem Kumar609461@gmail.com

Ratul Dey
Computer Science and Engineering
University of Engineering and
Management
Jaipur, India
ratulcs92@gmail.com

Jyoti Khandelwal
Computer Science and Engineering
University of Engineering and
Management
Jaipur, India
jyoti.khandelwal19@gmail.com

Prajanand Kumar
Computer Science and Engineering
University of Engineering and
Management
Jaipur, India
kprajanandk@gmail.com

Abstract— Today the technological evolution is increasing day by day. If individuals are still trying to engage utilizing numerous input devices in this evolutionary period, then it's not worth it. As a result, we created Voice Assistant (VA) in Python to lower the barriers between end users and VA operating systems and to enable users to perform any sort of command on their devices without using a keyboard. The primary aim of VA is to decrease the need of input devices like as keyboards and mic; by lowering the number of input devices, space and expense are reduced. VA are easy to use in latest devices. Nowadays, there are millions of devices that include them in households and some other places. The work represented in this paper is describing the functioning of VA are described, its main process i.e., give some voice command (VC) and get output in text format or on the screen.

Keywords—Voice Assistant, Python, Personal Desktop Virtual Assistant.

I. INTRODUCTION

In the period of technology advancement, VA is not able for user sustain since it's easier to access apps and services on various devices like Microsoft Cortana, Google assistant, Amazon Alexa, etc. They are very famous and mostly used in laptops, and mobile platforms. Virtual Assistant (VLA) is basically design to simulate how a human would behave as a conversational partner that means act as a socio- conversational agent in virtual worlds [1].

The main motivation for creating a VLA is to provide useful information such as weather, facts from Wikipedia, setting an alarm, watching movies on television streaming from Netflix, making a shopping list or scheduling your day's work in the proper way. By using only your simple VC

VA is also used in a dialogue system to help with practical things like customer service, getting information, and navigating [2].

VA has a lengthy history of substantial additions with multiple waves. The usage of virtual assistants (VA) for dictation, search, and VC has become a regular function on smartphones and wearable devices [3].

Complement or replace customer service with humans; that is, instead of hiring humans and wasting extra resources on them, their cost and training on them. We can replace them with a VA who will do the exact same job

with better speed and work with no extra cost or extra resources. VA has several interesting capabilities such as:

- Send Messages.
- Play music from streaming music services.
- Answer to questions asked by users.
- Provide information about weather.

The capabilities of VA are increasing day by day like Amazon and Google have provided platforms for developers in structure to extend their assistants capabilities [4].

This paper has been organized in the following way, section 2 deals with some previous work used in VA Section 3 show the proposed model with the techniques that improve the recognition. Then, section 4 shows the proposed algorithm elaborately. Section 6 shows the result of the VA in the form of screenshots that how can we operate our system without using keyboard or mouse. Section 7 show the conclusion part and future scope of this system.

II. RELATED WORK

There are many voices assistant is which can run on many platforms or for many devices like for Android (Alexa), Apple (Siri), Microsoft (Cortana), Google Home, Amazon Alexa. Every firm that develops a VA has its own unique set of development processes and strategies, but their goal is the same to provide some assistants that are virtually accessible at a cheaper cost and that can also be incorporated into a variety of platforms, such as smart speakers [5].

These company may have their second motive of creating these VA is to make a Socio-Conversational (SC) agent in virtual world or as a conversational partner, they conduct a conversation via auditory or textual methods (Chatbot). During their research, they think that by making these types of VA as a SC agent that can be used as a dialog system for various practical purposes like customer service, information acquisition and how a human would behave as a conversational partner [6]. They clarified that speaking is the most common method of communication used by humans in general. Because this is the most effective method, through which people would be able to communicate with robots using to voice detection, such as

the Hidden Markov Model (HMM) and Dynamic Time Wrapping (DTW) [7].

They are working on the next generation of virtual assistants, which will be able to multitask and interpret many input types at once, including audio, images, videos, touch sensors, and so on. In addition, there are certain approaches that are attempting to apply VA in the creation of self-driving automobiles [8]. These individuals do study in the medical area to see how they can best serve Veterans Affairs patients. According to the report, 66% of Americans make use of some kind of prescription medication; yet, around 50% of those customers do not take their medication as prescribed. Patients may schedule timely reminders to ensure they don't forget to take their medication at the prescribed time. By incorporating VLA into healthcare systems, patients would have the ability to converse with their digital assistants in order to either make an appointment with their favourite doctor or discuss their concerns about their health with that physician [9]. These individuals conducted study and descriptions of the language, which explained how voice is the primary interface for the language and how it is the best form of communication. Speech signal was translated into analogue and digital wave structure so that a machine could understand it [10]. This provided the interface between humans and machines. They talk about a genetic algorithm and NLP solicitations of Indian languages at work [11].

III. PROPOSED MODEL

We all have our own part of the creative; therefore, it makes sense that any VA firm's R&D team would use its own special techniques and perspectives while creating its products. Some helpers can execute a wide variety of activities properly and as the user desires, whereas others can only accomplish a subset of those jobs as accurately as possible.

Fig. 1, represents the process of VA. In this diagram, it shows how a VA system can capture the audio through users, cancelling the background noise, implement voice recognition algorithm and uses an API by the help of python.

A. Audio Capture

Words or Sentence spoken by the user act as an input spoken to the system (VA). After taking the input in a form of words and sentences, it will then try to clear the background noise with the help of Noise Cancellation (NC) or Wiener Filtering method. By this way, our system can capture the audio and try to filter that VC by using wiener filtering method. Then, after converted into text via the use of an online voice recognition technology. The user who wants to give their input as their VC into the system may get a text from a specified bulk that is arranged on the network server located in the data canter by accessing the microphone and reading the text that is temporarily stored in the system.

B. Noise Removal Method

Noise is something that is undesirable sound, loud or disruptive to hearing. It can also be used to describe an unwanted electronic noise signal that makes a hissing sound. Background noise is also known as ambient noise, it is any extraneous sound, which can be heard while listening to or monitoring other sounds. noise is coming from various resources such as electronic devices . Like mixer, drilling machine. Unwanted talks from the people around us, Sounds of animal Unwanted sound from the construction side

Most of the noise removal algorithms are subtracting those bonds from the original signal. Many of noise removal methods use static filters such as lowpass, high pass, and bandpass filters. These filters are used to designed for a specific parameter to isolate the dominant signal. Sometimes, these filters are ineffective in some conditions like when the properties of the background noise overlap with the clean signal to isolated. We have an isolated signal

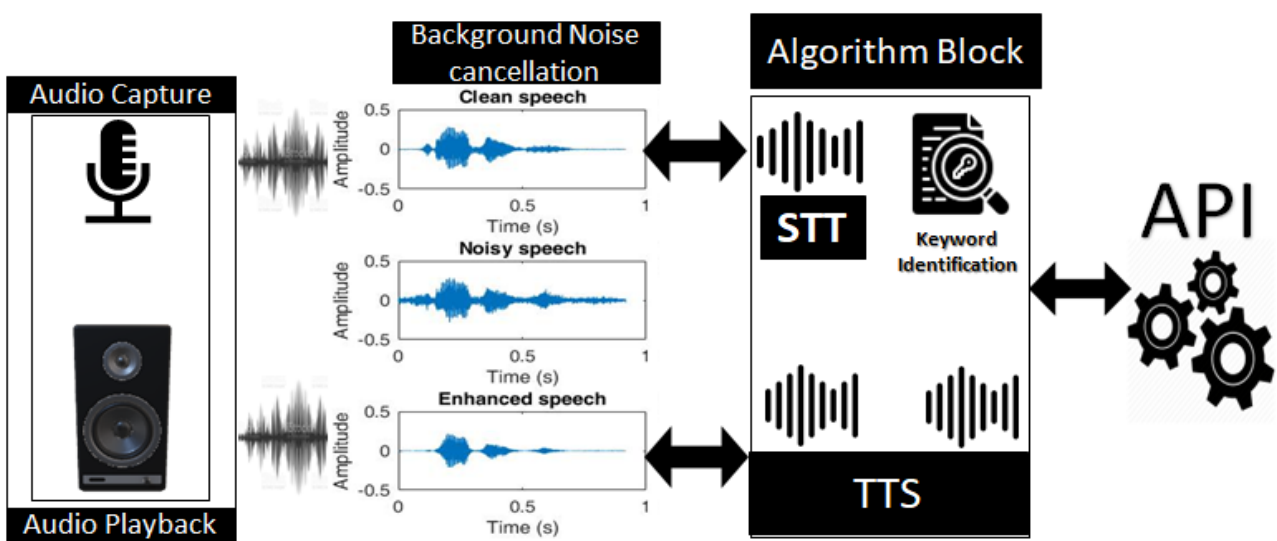


Fig. 1. Process of converting Speech to Text (STT) and Text to Speech(TTS) and noise cancellation process [12].

of both noisy speech and background noise but we can't subtract background noise to get clear speech because of macroscopic distance between the microphones both of have different environments. Another reason is that noise caused by various number of factors including electrostatic charges within hardware components and some vibrations in the environment. Wiener filter uses properties of two signals to get a clear speech and it check an error which is mean squared error is calculated and minimized to produce the clean speech.

To get the filtered speech or clean speech, we use wiener filtering equation which is

$$X(f) = S(f) + A(f). \quad (1)$$

Where.

$X(f)$ is noisy speech signal.

$S(f)$ is original speech signal.

$A(f)$ is additive speech signal.

$S(f)$ can be extracted from the noisy signal by multiplying the noisy speech signal $X(f)$ with the wiener filter function.

$$S(f) = W(f) * X(f) \quad (2)$$

Here, $W(f)$ represent the winner filter in frequency domain.

$$W(f) = \frac{|S(f)|^2}{|S(f)|^2 + |A(f)|^2} \quad (3)$$

In our VA, we remove the background noise using non-gradient method, which we can see in the fig.2. The noisy speech part in which it shows the unwanted frequency signal from the user's surrounding.

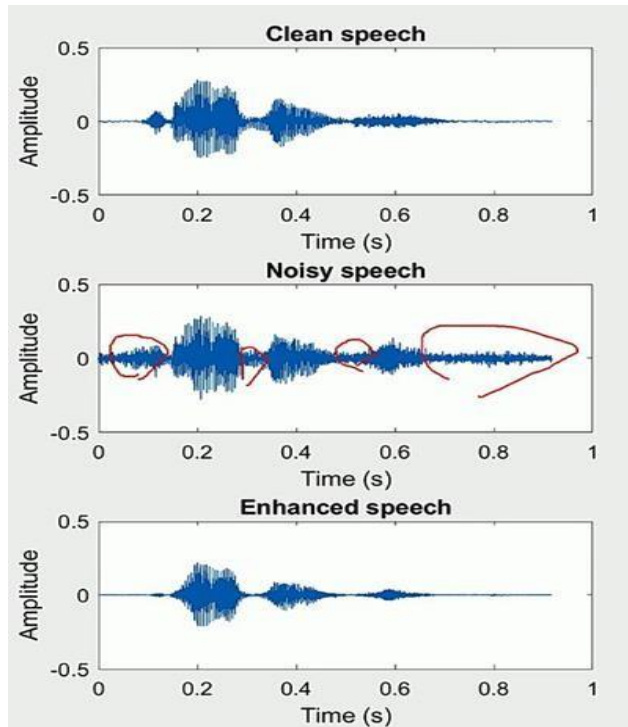


Fig. 2. Checking background noise frequency.

The above diagram shows the background noise coming around the system area or user's surrounding area. Here noise is coming from user surrounding area which is shown by graphically. In Enhanced speech procedure, which shows in the diagram ensures us that we can give input to VA without considering about background noise.

C. NLP (Natural language processing)

NLP is a system that allow VA to interact with human language. It is also known as computational linguistic. It is related with search, logic and knowledge representation. The capability to catch on and process human languages. It is essential in order to break the communication barrier between humans and machines. It consists of many subdomains, such as document classification, machine translation, information extraction, and information representation. NLP application is required when we have to process the text data. Here, the text data mean that the speech spoken by the user is converted into text.

The two appearance of NLP domains that are used in making VA are

Classification. In the condition of conversational systems is to find the intent of the spoken sentence or loosely the action that must be performed.

Information extraction or tagging. It is used to find the useful information from within the sentence, which is needed to complete the task.

Steps used by NLP in VA .

1) Setup the environment:

First of all, we install these libraries :

SpeechRecognition, nltk, gtts and many more.

2) In second step: we use SpeechRecognition to convert speech into text. It gives the ability to listen and understand user voice by transforming the audio signal into text.

3) After converting speech into text: VA uses different steps of NLP for further process like to get desired output what user wants. The first stage is :

a) Stemming :

In this process, sentence is dividing into its component sentences, usually along a particular mark.

Example : writing : write

b) Tokenization :

In this method, NLP breaks the sentence or text into their constituent words. Those constituent words are known as token.

Example : The sole meaning of life is to serve humanity :

['The', 'sole', 'meaning', 'of', 'life', 'is', 'to', 'serve', 'humanity']

c) Stemming :

It is the process in which tokens are transformed into their base form.

Knowing : Know

This step work same as stemming to convert tokens into their base form but sometime stemming is not able to give accurate result that's why we use lemmatization for more accurate result.

He was taking the ride.

In this step, POS tag will check that base word has what type of parts of speech.

[("Book", "Verb"), ("the", "Det"), ("flight", "Noun")].

Sometimes one word has two or more POS tags for this disadvantage we use Name Entity Recognition to clarify the meaning of that word or for what purpose that is used. Some words used for Company, Location, Name etc.

- ### 5) *Natural Language Generation(NLG)*

It is the process of producing a human language text response based on user VC. In this, NLG can give output to user in the term of human language or also in the form of text format, User can change the voice form to hear the output like if he wants male voice or female voice, they can change it by their need.

It is an abbreviation for Application Programming Interface (API). An API is a piece of software that enables two apps to communicate with one another. In basic terms, it functions as a messenger, delivering your request to the source you are seeking it from and then returning the result to you.

- VA is expected to provide insights into the future of Virtual Personal Assistant for a specific region, locality. The system will keep listening your commands and by given your commands it will recognize it.
- In the Fig.3, you can see the detailed working of VA. If the VA is unable to identify or acquire information from a user command, it will continue to inquire until the need for the ideal solution is satisfied.
- For the user suitability VA can have both male and female voices.
- Playing music, sending emails and messages, searching on Wikipedia, chatting on WhatsApp by your voice command, building lists, or accessing system installed programs, opening anything on the web browser, etc. are some of the services that are enabled by our VLA

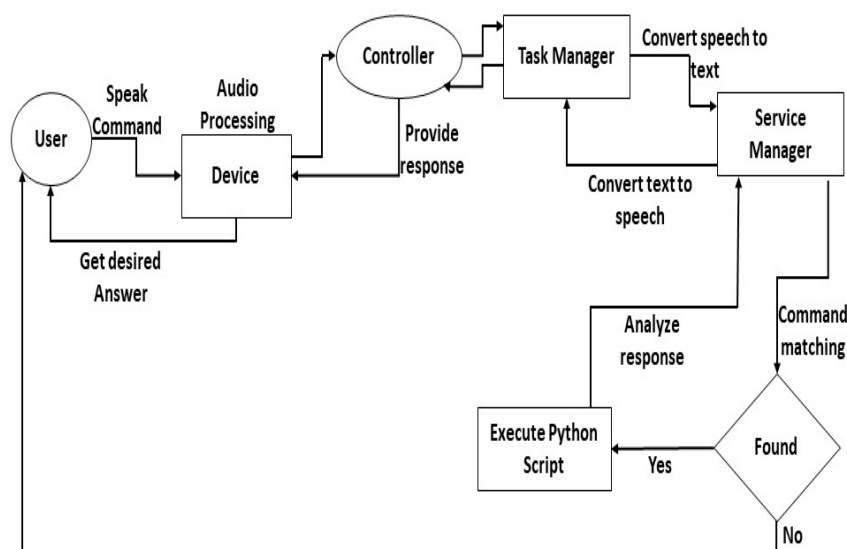


Fig. 3. It shows the detailed process of the VA [12].

In Fig.3. the command that we give to VA is processed by controller to clear the noise. Task manager for converting speech to text and then, service manager check that the command is matching or not. If it is matching then, VA execute python script file else it will move to its starting point. After executing python script file, it will provide our optimal output.

IV. PROPOSED ALGORITHM

Algorithm: Human helping hand proposed model

Input: Voice Command

Output: Response

BEGIN

Step-1: Listen what user want to command and then recognize it.

Step-2: Convert speech to text method.

Step-3: Using non-gradient algorithm, it can purify the given input voice command by clearing the background noise.

Step-4: Store the command in text format and analyse the keyword that VA need to execute.

Step-5: Then, task manager mange the input command. According to service manager, it can match that command is matching with the preferable output or not.

Step-6: If command is not matching, then it again movestowards the starting state.

Where VA can take input and say your command does not recognize.

Step-7: Finally, when command matched it will show user the desired output in voice command or text command or screen form. It depends upon the user that what type of command he/she will give.

END

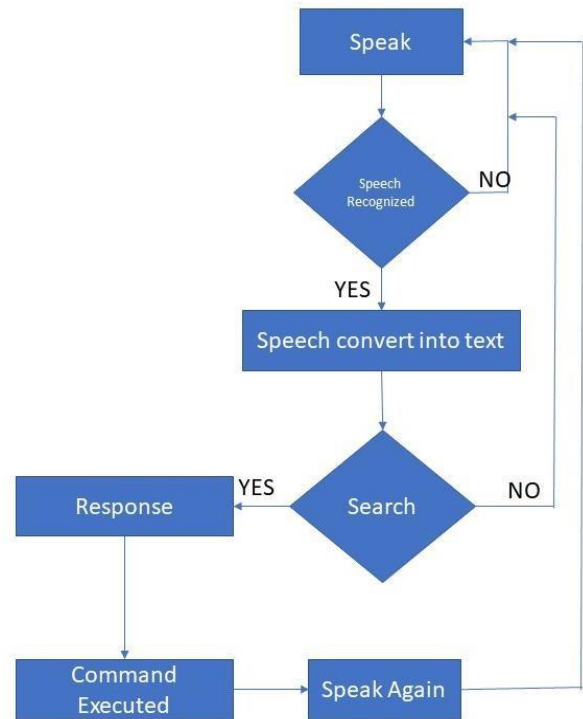


Fig. 4. Detail description about VA using flow chart representation.

Fig.4. Shows the Flow chart of VA and entire process of taking input, recognize process, and action taken to generate the output.

V. RESULT AND DISCUSSION



Fig. 5. Open Calculator Using VA

In Fig.5. The picture shows that we tested our VA by giving command to open calculator. This is not limited to calculator; VA can open any application, which are installed on computer or desktop.



Fig. 6. Open WhatsApp by using VA in which using a simple voice command you can text to anyone. Calculator Using VA

In Fig.6. it is clear that VA is testing for opening WhatsApp. In WhatsApp, you can send any messages by using your voice command that means we don't have to use our keyboard for typing the messages.



Fig. 7. Open Music Player on the request of user. Play that music which user wants or love more.

In Fig.7. it is clear from the above figure that user give command to VA for open the music player. Plays that music which are in the playlist or search that song with their name or singer name to play. You can also interrupt VA during the time of music which plays to change the song.

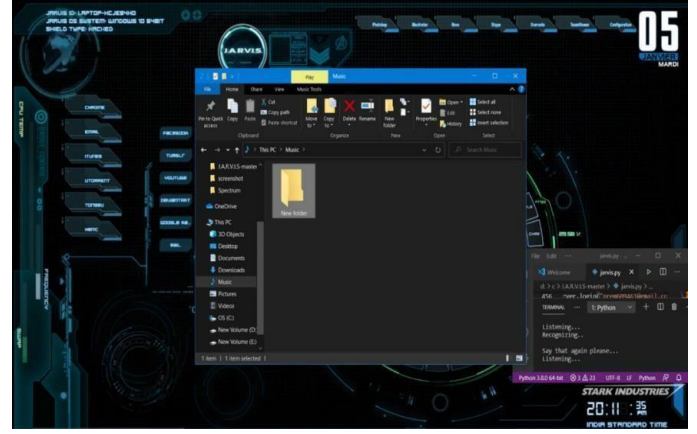


Fig. 8. The process of creating a new folder using VA.

In Fig.8. represents that it is tested for creating a new folder by a simple voice command. For example, if user is on desktop and will say create new folder, it will create a new folder in on desktop and it can delete the folder which user want to delete.

TABLE I. TIME TAKEN BY FOLLOWING COMMANDS

Table No	CONNECTION TIME TAKEN		
	Commands	Connection LAN	Time Taken (In second)
1	Open Calculator	LAN-1	5 Sec
		LAN-2	9 Sec
2	Open WhatsApp	LAN-1	10 Sec
		LAN-2	14 Sec
3	Open New Search Tab	LAN-1	10 Sec
		LAN-2	12 Sec
4	Created New Folder	LAN-1	4 Sec
		LAN-2	7 Sec
5	Open YouTube	LAN-1	9 Sec
		LAN-2	12 Sec
6	Open iemcrp Test	LAN-1	10 Sec
		LAN-2	15 Sec
7	Remove Current Tab	LAN-1	5 Sec
		LAN-2	5 Sec

In the Table I, we check that how much time a VA can take in opening of these commands by using different LAN connection. For opening calculator, it takes time 5sec from connection LAN-1. 9sec from connection LAN-2. As it is we apply same method for different commands.

VI. CONCLUSION AND FUTURE WORK

In this paper, our aim is to reduce the barrier between human and machine interaction. By our proposed algorithm, we used analog audio to convert speech into English Text. then it used NLP to convert user command as an instruction and then our system perform operation on this instruction and show some output on the screen. It uses python that has

many libraries, which are easy in making this system, and it has many advantages of its own. Therefore, with our paperwork, we mostly try to reduce the communication barrier between human and machine.

It also uses the concept of a non-gradient algorithm to hear the pure voice or actual command of the user. By reading this paper, you already know that our system is fully dependent on the internet and that it only works on laptops or desktops. For the future work, our approach is to make it available for mobile phones, i.e., for Android or Apple OS, and trying to make it offline.

REFERENCES

- [1] B. Ravenet, M. Ochs, and C. Pelachaud, "Architecture of a socio-conversational agent in Virtual Worlds," 2014 IEEE International Conference on Image Processing (ICIP), 2014.
- [2] D. J. Patil, A. Shewale, E. Bhushan, A. Fernandes, and R. Khartadkar, "A voice based assistant using google dialogflow and machine learning," International Journal of Scientific Research in Science and Technology, pp. 06–17, 2021.
- [3] Dr. K V. Kulhalli, Dr.K Sirbi, Mr. A J. Patankar.: Personal Assistant with Voice Recognition Intelligence. International Journal of Engineering Research and Technology. ISSN 0974-3154 Volume 10,Number 1 (2017). International Research Publication House.
- [4] G. TERZOPOULOS and M. SATRATZEMI, "Voice assistants and smart speakers in Everyday Life and in Education," Informatics in Education, pp. 473–490, 2020.
- [5] V. Kepuska and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018.
- [6] H. K. J. Zohu, and M. Yang, "Clustering of thread posts in online discussion forums," Extracting Chatbot Knowledge from Online Discussion Forums, vol. 7, no. 1, 2007. (Vol. 7, (pp. 423–428).
- [7] v radha and c vimala, "Speech recognition for agglutinative languages," A review on speech recognition challenges and approaches, vol. 2, no. 1, pp. 423–428, 2012.
- [8] G. Lugano, "Virtual assistants and self-driving cars," 2017 15th International Conference on ITS Telecommunications (ITST), 2017.,Warsaw, Poland.
- [9] v radha and c vimala, "Speech recognition for agglutinative languages," A review on speech recognition challenges and approaches, vol. 2, no. 1, pp. 423–428, 2012.
- [10] k v kulhalli, a j patankar, and k sirbi, "Desktop Voice assistant with speech recognition intelligence (DVAbot)," International Journal for Research in Engineering Application & Management, vol. 10, no. 1, pp. 301–305, 2020.
- [11] S. Srivastava and S. Prakash, "Security enhancement of IOT based smart home using hybrid technique," Communications in Computer and Information Science, vol. 6, pp. 543–558, 2020.
- [12] Ravikumar N R, Prateek C, Sathvik Bhadhar, Rahul Kumar, Mayura D Tapkire. "Virtual Voice Assistant" International Research Journal of Engineering and Technology(IRJET) Volume : 07 Issue : 04 Apr 2020.
- [13] R. Sarikaya, "The technology behind Personal Digital assistants: An overview of the system architecture and Key Components," IEEE Signal Processing Magazine, vol. 34, no. 1, pp. 67–81, 2017.
- [14] S. K and P. A. J, "Personal assistant with voice recognition intelligence," 2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON), vol. 10, no. 1, pp. 416–419, 2017.
- [15] A. S. Tulshan and S. N. Dhage, "Survey on virtual assistant: Google assistant, Siri, Cortana, Alexa," Communications in Computer and Information Science, pp. 190–201, 2019..
- [16] M. Chkroun and A. Azaria, "Lia: A virtual assistant that can be taught new commands by speech," International Journal of Human–Computer Interaction, vol. 35, no. 17, pp. 1596–1607, 2019.