

# Voice Based Personal Assistant

Prachi<sup>1</sup>, Abhishek Kumar Singh<sup>2</sup>, Mohd Akmam<sup>3</sup>

SCSE, Galgotias University, Greater Noida, India<sup>1</sup>

SCSE, Galgotias University, Greater Noida, India<sup>2</sup>

SCSE, Galgotias University, Greater Noida, India<sup>3</sup>

**Abstract** This research paper is for solving real world problems with the help of advanced AI (Artificial Intelligence) and ML (Machine Learning) techniques. The voice based personal assistant technology attracts most of the people around the world and people do interaction in many ways like via mobile phone, laptop, computer, etc. This project is mainly focusing on broader parts of windows services or applications which can be automated for making system life easy. The sole purpose of this product is to give more comfort of the user with the help of Automated features of speech recognition. The Most well-known utility of the iPhone is “SIRI” which enables a quick consumer to talk with the quick consumer cellular with speech and it additionally answers to the voice instructions of the consumer. Same form of function exists likewise via way of means of Google that is “Google Voice Assistant” that's used in Android Phones.

**Keywords:** Artificial Intelligence, Android, Optimization, SIRI, Google Voice Search

## I. INTRODUCTION

Internet Connectivity. Intelligent Virtual Assistant it's far a utility that offers facts for example: the voice of the and logical records to offer assist through noting inquiries in everyday dialect in making guidelines and acting activities. Inside the writing the People decide on giving voice instructions in preference to typing them due to the fact they need to store their time. As the time is transferring humans are becoming clever and era is getting advanced. The basic requirements of a user with the voice based personal assistant is work in a relaxed way. Mobile Technology is being well-known for the Customer Experience, due to the fact it's far very smooth to get right of entry to the programs and offerings from everywhere to your Geographic-location. Blackberry, Apple, Windows, Android etc. are diverse well-known & normally used in Mobile OS. All the OS gives masses of programs & offerings for customers. For example, the Contact Applications is used for to shop the touch information of the consumer's touch & additionally facilitates the consumer to attach a name or send an SMS to some other character the use of the contents saved on this utility. We can get comparable kinds of programs all over the international thru Apple Store, Play Store, etc. Nowadays the event of Artificial Intelligence system is that area where any natural-human Interaction can occur by the means of voice, through gestures or facial expression or with eyes scanning etc. And this area is gaining quality. One among the most important study shows that humans are wide spreading in this area it is now no longer need to learn anything to speak with the machines machines are made in this way that they can learns to speak to anybody, exploring anyone's actions and tries to be a good assistant [1]. All those capabilities provide start to diverse varieties of sensors or functionalities to be carried out with inside the cellular devices. The Most known feature of the iPhone is “SIRI” that helps in the consumer to communicate with the quiet consumer to cellular with voice and it additionally responds to the voice instructions of the consumer. User can ask anything from SIRI or can make her do any task they want on the basis of Speech Recognition, Artificial Intelligence and Machine Learning has played a huge role in these types of devices even windows based “cortona” or Android's phone “Google Assistant” responds same as the Apple's SIRI developers have developed these applications or web applications so that any user can make their tasks to do easily and it also saves user's time. Voice based applications or web applications works efficiently on the basis of speech recognition. A review demonstrated that 98% users use Siri more frequently while 30% other users of Siri uses it on daily basis but there are 70% users who uses Siri once in a while. Same sort of utility is likewise advanced via way of means of Google that is “Google Voice Assistant” that's used in Android Phones. But this Application normally works with Internet Connections. But our Proposed System has capability to paintings with and without time period IVA is applied conversely with terms, for example: Conversational Agents, Virtual Personal Assistants, Personal Digital Assistants, Voice-Enabled Assistants or Voice Activated Personal Assistants, to provide some illustrations. Intelligent Virtual Assistant be a part of communicate affirmation, lingo understanding, exchange organization, tongue age and communicate affiliation to reply to customer's request and sales. Voice engaged IVAs like Siri, Google Assistant, Microsoft Cortana and Amazon Alexa are via way of means of and big open on cutting part mobileular phones, and continuously in homes (e.g. Amazon Alexa and Apple Siri) and automobiles (e.g. Google Assistant mixture with Hyundai). The market for IPAs is foreseen to attain four sixty one billion via way of means of the mid 2020s. Airtificial

Intelligence has made our life easier. In this paper we have developed about the web application that we made that works as a virtual assistant that responds on voice commands and has API to perform various tasks like searching weather or surfacing google or configuration of emails or printers.

## II. LITERATURE REVIEW

Speech popularity functions a protracted record with numerous waves of important innovations. Speech popularity for dictation, search, and voice instructions has emerge as an average characteristic on smartphones and wearable devices. Design of a compact big vocabulary speech popularity machine that allows you to run effectively on cell devices, as it should be and with low latency. This is accomplished via way of means of using a CTC-primarily based totally LSTM acoustic version which predicts context-impartial telephones and is compressed to a 10th of its unique length using a mixture of SVD-primarily based totally compression and quantization. Quantized deep neural networks (DNNs) and on-the-fly language version rescoring to comprehend real-time overall performance on current smartphones. Let us see few past research papers to study more about the AI and ML.

- A. G. Bohouta, VZ. Kėpuska, "Comparing Voice Based Assistant System (Microsoft API Google API and CMU Sphinx)": it taught about the Goals of voice reputation is to permit herbal communication among human beings and computer systems thru speech, wherein herbal implies similarity to the ways human beings have interaction with every other [9].
- B. In the B. Marr, There some good approaches in Google for Using Deep Learning AI.: Deep learning involves building artificial neural networks technology which used to mimic the organic(living) brains sort and process information. The "deep" in deep learning signifies the use of many layers of neural networks all stacked on top of each other. This data processing configuration is known as a deep neural network, and its complexity means it is able to process data to a more thorough and refined degree than other AI technologies which have come before it. Deep learning is now the trending technology innovation which include the artificial intelligence and nowadays it can be easily seen in many applications like phone apps , desktop apps and for other devices as well.[2]
- C. Huang J., Zhou. M. and Yang D., January. Extracting the chatting bot Knowledge from Online Forums. In IJCAI (Vol. 7): They talked about the Deep Learning the "deep" in Deep Learning signifies the employment of the many layers of the neural networks all stacked-on prime of every different. This processing configuration is thought as a deep neural network its complexness means that it is ready to do the method information to add additional thorough and refined degree than different AI Technologies that has precedes it. It involve s building artificial neural networks which is attempt to copy the organic(living) brains which is sort and process information. [3]
- D. In the Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations between human and artificial intelligence they did a comparison in between human to human online conversations and human to Chatbot conversations: Computers in Human Behavior Comparison of 100 random human IM conversations against 100 random human conversations with Cleverbot, they examined the amount of a written content and the uniqueness of the words which are used, for frequency of profane language, and the usage of standard CMC linguistic features like the shorthand phrases (e.g., "lol") and emoticons (e.g., ":-)", ":-)", ":-). We expected the linguistic profiles of the human-human IM conversations between the consistent with previous studies [10].
- E. "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010: They propose a log-linear modeling framework based on using numerous features which each measure some form of reliability between the fundamental speech and an entire sequence of theorized words. [5].
- F. Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology.: Before discussing any present day bots it is important to learn about the history. When did the idea of artificial intelligence come about? Artificial intelligence exists before computers; in fact, it can be traced back to Greek mythology (Buchanan, 2002, paragraph 1). Whereas the idea of Artificial Intelligence was very old, it has been only since World War Second that taking steps towards making AI a physical reality has been a possibility (Buchanan, 2002, para.2). Even though there have been a great number of important providers to this field, for the purposes of this column we will turn directly to Alan M. Turing and his paper named "Computer Machinery & Intelligence". In this paper work Turing asked the question "Can machines think?" (Turing, 1950). [6]
- G. Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in HCI. In the 5th Conference on Human-Computer-Interaction in South Africa, South Africa (CHISA 2006, ACM SIGHI): In speech synthesis, as in most language technologies, language dependent and language-independent issues need to be considered. This is happen because as we understood the speed optimization issue, this is language independent issue which indicates for a specific design and choices at the algorithmic/data representation level, although the particular characteristics of IJECIERD (International Journal of Electronics, Communication and Instrumentation Engineering Research & Development) certain language which may increase the need in special fine tuning on per language basis. While text normalization and efficient handling of particular language characteristics are language-dependent issues. This paper describes the design issues and implementation approaches for improving the performance of a text to speech synthesis system for screen readers. [7]

- H.** In the Thakur, N., Hiwrale., Selote, S., Shinde, A. and Mahakalkar, N., Artificially Intelligent Voice Based Assistant.: Bots can be created in many ways like by using language like Artificial Intelligence Mark-up Language, it is a language based on XML that allow developers to write the rules for the bot that it needs to follow. Another drawback is writing rules for different scenarios or cases which is very time consuming and it's almost impossible to write rules for every possible scenario. So the important scenarios should be written. [8]
- I.** In the "Accurate and compact, a large vocabulary voice recognition on mobile devices/computers," in inter speech. 2013 ISCA.: they talked about the Many existing catboats consist of conversation management modules to govern the conversation process & Chabot knowledge bases to response to user response. Now Typical implementation of a chatbot's database holds a group of patterns that match user inputs & produce replies [4].
- J.** In the Personal Assistant with Voice Recognition Intelligence witten by Dr. Kshama V. Kulhalli, Dr. Kotrappa Sirbi, Dr. M.S. Sheshagiri and. Abhijit J. Patankar Research Scholar, Computer Science and Engineering: They have created a voice command system for blind users so that their life become easy when doing multiple tasks [17].

### III. CLASSIFICATION OF SPEECH RECOGNITIONS SYSTEM

Speech popularity structures may be labeled in numerous ways via means of describing the sort of speech utterance, sort of speaker version and sort of vocabulary that they have got the cap potential to recognize. The demanding situations are in short defined as following:

#### A. The types of speech utterance:

Speech recognition are classified consistent with what sort of utterance they need ability to acknowledge. They are classified as:

- Isolated word: They are some phrase recognizer commonly calls for every vocable to own quiet (loss of an audio signal) on every aspect of the pattern window. It accepts an unmarried phrase at a time.
- Connected word: it is nearly like a remoted word; however, it permits separate utterances to "run-together" which includes a minimal pause in among them.
- Continuous Speech: It lets in the customers to talk obviously and in parallel the computer will decide the content.
- Spontaneous Speech: it is the speech that's herbal sounding and isn't always rehearsed.

#### B. Types of speaker model:

The Speech recognition system is divided into the two main categories:

- Speaker dependent models: These systems are designed for a single selected speaker. They are easier to develop and more accurate but they are not flexible.
- Speaker independent models: These systems are designed for a diversity of speakers. These systems are difficult to develop and fewer accurate but they are flexible.

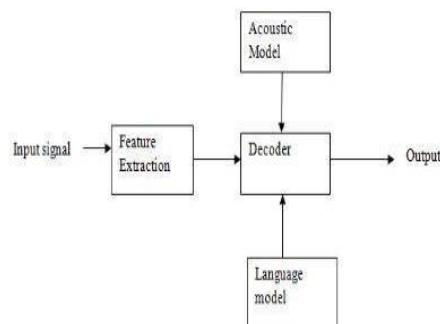
#### C. Types of vocabulary:

The vocabulary size of the speech recognition systems affects the processing requirements, accurateness and density of the system. In voice recognition system(dot): voice-to-text .The types of vocabularies can be classified as follows:

- Small vocabulary includes solo letters. Like: a, b, c, d and etc.
- Medium vocabulary means two or three letter words. Like: The, who, why, will etc.
- Large vocabulary means more letter words. Like: voice, speech, machine, weather etc.

### IV. OVERVIEW OF THE VOICE RECOGNITION SYSTEM

1) **Fig 1: Data Conversion from Audio to Text**

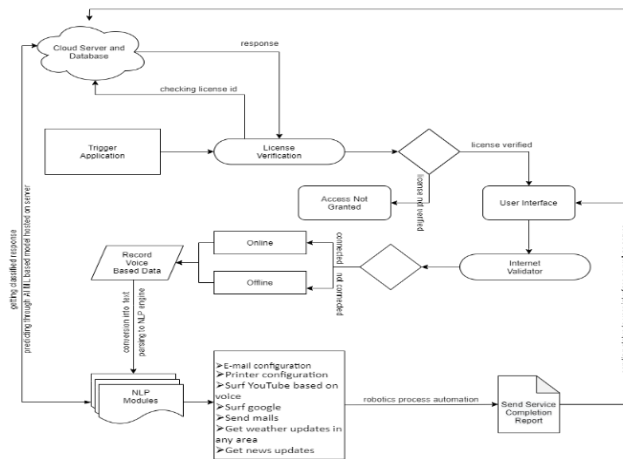


**Input signal:** Voice enter via means of the user. Feature Extraction, it should keep beneficial information of the sign, deduct redundant and undesirable information, display much less variation from one talking surroundings to another, arise typically and certainly in speech [16]. Acoustic model-it includes statistical representations of every distinct sounds that makes up a word. Decoder-it's going to decode the enter sign after function extraction and could display the preferred output. Language model-it assigns an opportunity to a chain of phrases with the aid of an opportunity distribution. Output-interpreted textual content is given by means of the computer. See (fig 1) The key part of the project is to recognize speech using Mel-frequency cepstrum Coefficients and Vector Quantization techniques. The feature excavating will be done using Mel Frequency Campestral Coefficients(MFCC). The steps of Mel Frequency Campestral Coefficients(MFCC) are as follows: -

- Frame & Block
- Window
- Fast Fourier Transform
- Mel-Scale technique
- Discrete Cosine Transform

## V. METHODOLOGY

Now, we will see the methodology that has been involved. We will take a look at the flow involving all the crucial steps. See (fig 2)



**Fig 2: Working**

**Cloud Server:** - Cloud Server or Database helps in to check the license id of the user which tells the system if the user is a legit user or not. The input of User's speech will converting into texts from the special domain which are specially organized on the pc network server at the knowledge center from the electro-acoustic transducer is briefly hold on within the system which is sent to cloud for recognizing the speech. The equivalent text is then received and sent to the central processor.

**API:** - An API is a technology that allows applications to talk to each other. In other term, an API is like a messenger that delivers any request of the user to the provider who handles the user's request and then delivers the response back to the user.

**Backend:** - The python backend script receives the output from the voice reputation module after which it identifies whether or not the command or the voice output is an API Call, Context Extraction of data, or it's just a System Call. The output is then dispatched lower back to the python backend module to represent the expected output to the user. Speech recognition is the process of converting the voice of a user into text. This is commonly used in voice based personal assistants like Google, Cortana, etc. Python provides an API which called Speech-Recognition to allow programmer to convert the user's voice into text for further processing [13].

**Internet Validator:** - It is used to let the system know if the system is allowed to do online/offline work or both.

**NLP Modules:** - NLP helps in speech recognition and converting into text. TTS conversion which talk about the skill of computers to read text properly or not. A Text to Voice Engine that can able to translates written text to a voice representation, then converts the voice representation to waveforms that is the output which we get as the audio. Text to Voice engines with the support of different languages and vocabularies that are available to programmer with the help of third-party publishers like APIs and Text to Voice python modules.

Natural Language Processing Modules has already classes named for each of the following modules:

- E-mail Configuration, it lets the system to do an e-mail configuration on outlook or Gmail on its own.
- Google assistant excelled in Virtual Reality and Human Interactions by 60% while Siri replied with 40% correctness in Virtual Reality & Human Interactions. There were only few differences between Google Assistant's and Siri's correctness but next assistant i.e. Cortana decreased steadily. Its accuracy answered up to 28%. And Alexa answered up to 7% which shows poor results as compared to among others. Our web application has success rate of 50%.
- Printer Configuration, it lets the system to do printer configuration on outlook on its own.
- Surf YouTube, it lets the system to open any YouTube video we tell them to open. It will showcase the searched page.
- Surf Google, it lets the system to open any search result from google we tell them to open. It will showcase the searched page.
- Weather Report, it will showcase the weather report of any Location we asked of.
- The Date-Time module is imported to support the functionality of the date and time. For example-the user wants to know the current date and time of anywhere. In short this module supports classes to manipulate date and time and perform operations according to it only [11].

## **VI. RESULT AND DISCUSSION**

### **Voice Based Recognition and Related Understanding:**

Virtual Assistant are going to be very often useful in our daily life because of its reach in smart-phone ir-respective of operating system whether its "SIRI," GOOGLE ASSISTANT", and" CORTANA" many of us own at least once. There is main challenge in VBPA (voice based personal assistant) is that people's voices vary and they speak in different languages along with different accent which sometimes create issue to reorganize voice. In voice base recognition, all needed to know is that it was the way regularly the virtual assistants could perceive the words users were saying to found more. A survey have been done. In this survey users tried speech recognition from various gadgets and it also changing level of foundation noise found out that Google and Siri knew their users well according to the graphs. There was a couple of misconceptions, similar to when someone asked Siri that "Will I need an umbrella for next two months can you suggest?" And voice based personal assistant gave user the date for one week later, so it is clear that voice-based assistant not good at understanding. Google-voice-based personal assistant is very good at understanding natural languages, Alexa is good in music related activities and their thing, But Alexa is not comfortable with basic yet simple questions on the other side Cortana is not that good in basic conversation.

### **Human Free Collaboration:**

The issue that our survey found out that for some users, hands free connection is the real utilization condition. We thought that in this specific situation, interrupt ability was a huge burden. They were puzzled when they were requested by the intelligent virtual assistants draw in outwardly with the screen, or to select alternatives by tapping the touch screen as opposed to through utilizing dialog. This appeared to interfere with the without hands involvement of IVAs and was considered especially complicated in circumstances, for example like when driving. We suggest that keeping up dissertation as the fundamental information and yield all through the connection should be a need in the future plans of IPAs with a specific end goal to guarantee that hand free association is supported completely and that events are not caught up by a cooperation methodology move. Hands free technology is the future where people can utilize their maximum time of their lives. Speaking of desktop and mobile applications, the ones whose interface is powered by AI and is processed by Chabots, they have the ability of providing perfect delivery and quick access of the required data, at the call of the users [14] [15].

## **VII. FUTURE SCOPE**

In this paper we have talked about Voice based personal assistant "DOT" projects have the most features that can be added in the application. It already performs so many tasks but in the future we can also add more features in it. Waiting time would be minimized. The voice range would be maximized. The results can be seen in the minimum time possible. User can compose mail through his voice without typing anything like user assistant will be able to help the user by writing their long mails. Surfing would also be easier like opening a browser to surf YouTube videos without any typing or to play or pause any YouTube video to make user experience great or to even only searching anything from the google and to download any image would be possible on the basis of user's commands. This "DOT" voice based personal assistant will reduce the typing efforts and make the user experience great, these features will help to compete with the competitors in the market. It is not wrong in saying that Desktop Assistants are progressing. Desktop Assistant are more intelligent. Even there are reports that 80- 90% of businesses will be giving out more enhanced Desktop Assistants by 2030[12].

**VIII.CONCLUSION**

DOT is a web application designed to ease our daily life work by saving out time by doing our work on Voice Commands and has the capability to understand what the user is saying without internet connection.

DOT has various functionalities which makes it fun for the user like to open the YouTube video or to configure the mail or to doing fast google research etc. It only able to work on human voice Text-to-Speech mechanism convert it into commands and gives desired answers to the user on the basis of the user's query that is being asked to make a call or to perform any acts and operations. DOT also gives the same output as other does and along with it gives greeting to the user who says like Hello or Hi so that user will feel more comfort and feels free to relate with the voice based personal assistant. This application is also able to reduce any kind of manual work which is required in the user's life which are hectic or feel bored. The DOT can easily do all the tasks. The entire system works on the user's voice input rather than the text which is a one-step advancing in the world of applications. Hence, DOT is the application which fully operates on the Voice Command of the user.

**IX.REFERENCE**

1. AI Based Voice Assistant Using Python by Deepak Shende, Ria Umahiya, Monika Raghorte, Aishwarya Bhisikar, Anup Bhangre
2. B. Marr: The best Ways Google Uses Deep Learning Artificial Intelligence.
3. Huang J., Zhou. M. and Yang D., January. Extracting Chabot Knowledge from Online Discussion Forums. In IJCAI (Vol. 7).
4. Accurate and compact large vocabulary speech recognition on mobile devices in INTERSPEECH. 2013, ISCA.
5. Speech recognition with flat direct models, IEEE Journal of Selected Topics in Signal Processing, 2010.
6. Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology.
7. Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in HCI (Human-Computer Interaction). In 5th Conference on Human Computer Interaction in South Africa (CHISA 2006, ACM SIGHI).
8. Thakur N, Hiwrale, A., Selote, S., Shinde, A. and Mahakalkar, N., Artificially Intelligent Chabot.
9. Bohouta, VZ Kępuska : "Comparing Speech Recognition Systems (Microsoft API Google API and CMU Sphinx)".
10. Hill , Ford, WR and Farreras , I.G., 2015. "A Real conversation with Artificial Intelligence: A common comparison of conversations human in online and human with Chabot. Computers in Human Behavior.
11. VOICE ACTIVATED DESKTOP ASSISTANT USING PYTHON-Dhiraj Pratap Singh, Deepika Sherawat, Sonia
12. DESKTOP VOICE ASSISTANT Gaurav Agrawal\*1, Harsh Gupta\*2, Divyanshu Jain\*3 , Chinmay Jain\*4 , Prof. Ronak Jain\*5
13. Desktop Assistant Theory Explained," <https://www.geeksforgeeks.org/Desktop-Assistant-theoryexplained/>
14. "Freya Riki, "Future of Desktop Assistant in 2019," <https://yourstory.com/mystory/future-of-Desktop-Assistant-in-2019-8wulieg1yx>,
15. "Desktop Assistant from Wikipedia," [https://en.wikipedia.org/wiki/Desktop\\_Assistant](https://en.wikipedia.org/wiki/Desktop_Assistant),
16. Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018.
17. Personal Assistant with Voice Recognition Intelligence Dr. Kshama V. Kulhalli Vice-Principal & HOD IT D.Y. Patil College of Engineering and Technology, Kolhapur-416006 Dr. Kotrappa Sirbi Department of Computer Science and engineering Belgaum, India Mr. Abhijit J. Patankar Research Scholar, Computer Science and Engineering.