



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 3 - V7I3-1790)

Available online at: <https://www.ijariit.com>

## Voice assistant a revolution in technology

Kedar Dhamankar

[kedardhamankar@gmail.com](mailto:kedardhamankar@gmail.com)

MET Institute of Computer Science,  
Mumbai, Maharashtra

Dr. Abhijit Banubakode

[abhijitsiu@gmail.com](mailto:abhijitsiu@gmail.com)

MET Institute of Computer Science,  
Mumbai, Maharashtra

### ABSTRACT

*Virtual assistant is basically a software agent that can perform tasks or services for an individual based on commands or questions. Traditionally if you have to get some information over the internet you would need a desktop/laptop/mobile device of some sort with an internet connection, manually go to the WEB portal and search for what you need to get result which is a little time consuming. Voice based internet access uses rapidly advancing speech recognition technology to give users anytime, anywhere communication and access to the web through the Human Voice over a mobile or home phone. The objective of this research paper is to describe the various technology factors that are making voice assistant the next big opportunity to the various approaches service providers and developers of voice portal solutions can follow to maximize this exciting technology.*

**Keywords**— *Virtual Assistant, Voice Recognition Technology, Alexa, Siri, IoT*

### 1. INTRODUCTION

In general voice assistant is the technology which helps to connect with user without any hard device which reacts to a command received from user and gives back the relevant information as the response to his inquiry. For performing voice assistant we need most important to have a good Voice User Interface (VUI) which helps to makes spoken human interaction with computers possible, using speech recognition to understand spoken commands and answer questions, and typically text to speech to play a reply. Voice user interfaces have been added to automobiles, home automation systems, computer operating systems, home appliances like washing machines and microwave ovens, and television remote controls. They are the primary way of interacting with virtual assistant on smartphones and smart speakers. A voice command device (VCD) is a device controlled with a voice user interface. Text information can be easily converted into a speech response, along with that it can perform the activity given in form of command by the user for e.g., open chrome etc.

Earlier when voice assistant technology got introduced it was in the form of toy, or a voice assisted calculator but later on the evolution happened and the technology took charge in mobile devices, in cars, in smart appliances because of which human efforts got saved. In telecommunication the Interactive voice response (IVR) helps user to have interaction with the host system via either a telephone keypad, or speech recognition after which services can be inquired about through the IVR dialogue. It can generate respond with pre-recorded or dynamically generated audio to further direct users on how to proceed.

### 2. LITERATURE SURVEY

Though we all are familiar to the voice recognition technology or smart voice assistants in last 10 years but the process behind that was started as follows:

- 1911 - Computer speech recognition seems like it has been five to 10 years away for decades. Indeed, the electronics and computer industries have been chasing the goal of voice-directed computers for nearly 100 years, when a simple wooden toy dog called Radio Rex. When his owners called for "Rex!" it responded to children and adults by shooting out of a doghouse (at least some times).
- 1952 – Bell labs brought the product called “Audrey” included Automatic Digit Recognition machine. It occupied huge relay rack, with high power consumption and had streams of cable and exhibited the maintenance problems with complex vacuum-tube circuit. Later on, Audrey could recognize the sound with digit zero-nine with 90% accuracy uttered by its developer HK Davis.
- 1961 – Another tool enabled with digital speech recognition IBM Shoebox a voice-activated calculator though it was initially had market launch in 1961, but presented to the general public during 1962 Seattle World’s Fair. The shoebox was the advanced in speech recognition technology. It could recognize 16 words and 0-9 number with simple arithmetic operations such as addition, subtraction, total through voice command.

- 1970 - Carnegie Mellon University (CMU) in Pittsburgh, Pennsylvania with support of the United States Department of Defense and DARPA agency with a five-year speech recognition program for achieving minimum 1000 words, along with CMU Stanford Research Institute took part. In the result it mastered about 1000 words along with understanding sentences, grammar structure, pre-programmed vocabulary pronunciation for reducing speech recognition errors.
- 1990 – IBM, Philips and Lemout & Hauspie kept digital speech recognition as a feature of their personal computers. Later on, in 1994 first smartphone IBM Simon laid the foundation for virtual assistant which we know today.
- 1997 – Dragon’s Naturally Speaking software introduced recognize and transcribe human speech without pauses at the rate of 100 words per minute and till today it gets used by doctors from US and UK for their medical records documentation.

The last 10 years made massive transformation in thinking of people about voice assistant technology, it’s becoming as part of life

- 2010 – Siri which we know now as a pure voice assistant for apple but earlier Siri was published as a mobile application and later on in 2010 apple saw the potential in Siri and took over the company along with its creation. Apple introduced the Siri as its voice assistant in iPhone 4S and from then it’s started with capable product, current ecosystem of smart speakers and that’s how Siri became first voice assistant for huge number of people. Now it’s been a decade with the abilities are changed and expanded but till apple keeping control over the functionality part with keeping a limited third- party developers.
- 2011 – In 2000s only Google started testing voice search technology but in 2011 google.com launched a little microphone button for searching through the voice though the voice search was restricted only to google chrome but suddenly a huge market had chance to try out and by applying voice technology to the massive power to its search engine marked a big step to the industry it would be years before google assistant launched in the market.
- 2013 – For the first time at Microsoft BUILD Developer Conference in San Francisco Cortana was first demonstrated with the key ingredient “makeover” for the future operating system for windows phones and windows. Later on, it expanded to the platform’s windows 10, windows phones, Xbox.
- 2014 – Amazon announced a smart speaker with its own voice assistant named for the Library of Alexandria till that time Amazon’s participation in e-commerce and cloud computing were overshadowed its attempts for building consumer technology. Alexa is most powerful voice assistant around now, jump ahead of Siri or Cortana. Amazon aggressively pushing the voice assistant into a developing array of smart speakers and other products from wearables to cars. Alexa marketplace may be the most important contribution to voice assistant space. There are more than 100,000 skills on the market.
- 2016 – Google Assistant is primarily available for smart phones and smart homes. The google assistant can engage in two-way conversations unlike the company’s previous virtual assistant. It challenges Alex at least in the U.S. though google home and google assistant offers many of the same features like Alexa but integrations into Google’s larger in tech ecosystem. It started creating a smart home network by launching Nest. While not avoiding the other voice assistant aspects, Google Assistant an universal part of people’s lives made comprehensive smart home.
- 2018 – Though Samsung introduced their voice assistant S Voice application in 2012 with Galaxy S III, later discontinued in June 2020. In 2017 Samsung announced that Bixby is coming in line of family hub 2.0 refrigerator making non-mobile product which included voice assistant. Bixby comes with types such as “Bixby Voice”, “Bixby Vision” a reality camera can identify object real-time and offer the user to purchase online translating text, read QR codes and recognize, “Bixby Home” for information weather, fitness activity and buttons for controlling their smart home gadgets, “Bixby Touch” which makes recommendations based on intelligent recognition and can access services like translation, online shopping, media by touching the screen. It’s still in development phase hasn’t released publicly yet[17].

### **3. METHODOLOGY**

#### **3.1 Architecture**

- (a) Takes input in form of speech
- (b) Analyze the voice and convert into text
- (c) Data Processing
- (d) Process the text output and converts into speech

In the first phase the data is collected from the command given by the user in the form of speech and the collected data will be stored as an input for next phase. In the second phase the input data gets converted into text format using Speech-To-Text (STT). Then in next phase the text is analyzed and looks into the script to identify the correct response for the command. Finally, the output is generated in text form then, it gets converted into speech using Text-to-speech (TTS), once the response is identified [18].

#### **3.2 System Elements**

Basically, there are two main elements while using voice assistant technology either smart devices or functionality which includes activities related to smartphones

- Smart Devices: If any smart device needs to perform an activity like controlling light in home, control television in home or office, control the temperature of the room, prepare food in Microwave, lock the doors etc. for this an AI device is required like Amazon’s Echo, Echo Dot, Google nest. For this transaction the AI Device and the smart device should be on WIFI network.
- Firebase: Android applications helps the user to share his personalized data. The data gets processed through a Built in API which gets stored to firebase cloud storage and with the help of firebase cloud messaging service sends the message to the target device through a token and can achieve the activities needs to perform through phones like reminders, open camera, to do list or open an application which has voice assistant feature.

#### **3.3 Data Flow**

- Device Initialization: By calling the name for the device can initialize it like for Google assistant its “OK Google”, for Siri its “Hey Siri”, for Alexa its “Alexa” etc. and after that can pass the command for the activity needs to be done.

- **Converting:** After receiving the command in the form of speech can convert it to text format using Speech-To-Text (STT).
- **Analysis:** Go through the command and matching them with web service or cloud server.
- **Command Execution:** Once the command matches can look into the script and execute it and sends the output in the text format and convert it into speech using Text-To-Speech (TTS).

For activities which needs to be perform through the smartphones like open reminders for that after searching into script for the exact match the token which is generated by firebase cloud messaging (FCM) through its SDK. Then with the help of this token once the command is given by user for an activity that comes in form of message with body to the target device and then it performs the activity.

#### **4. IMPORTANCE OF VOICE ASSISTANT**

- **Multiple formative Skills:** There are multiple skills available on the platform of voice assistant. Skill is nothing but a functionality is enabled by the user to make the process smooth. For e.g., if u want to view the traffic status you can easily get the response in the form of voice, ordering the favorite pizza by enabling the skill i.e. over the command the pizza will arrive. Apart from this there are very interesting things for all age groups for children there are some skills for storytelling, rhymes, for those who plays instruments like tuner skills are there etc. like from information which you asked or information related your bank accounts till the current affairs happening around the skills makes your life easy[19],[20].
- **Dependence on screen reduced:** This can be a major important factor in the voice assistant platform. Every time for checking the news, for checking traffic, for checking flights in next hour or few from a specific airport, listening music etc. for all this stuff earlier we had to take a device in hand which might be a smart phone or laptop but because of this platform all the work became just on voice which saved our time and as well as it is less gadget used as the conversation happens between AI device and user.
- **Becoming more popular other than homes and smart phones:** Not only with the actions being performed smart phones and AI devices but also voice assistants is capturing all the components in the market. Now a days even cars voice assistants, in smart TV also now a days it helps to give command through voice to search something etc., as well as in home appliances like fan, microwave, lights, AC, door locks are now able to use with voice assistant.
- **VUI is developing:** Voice User Interface (VUI) is developing so rapidly, before few years there were problems to find the VUI right. The automated customer center was one of the components who faced this issue. But earlier the experience is more becoming like a human conversation than a robotic conversation. This clearly shows that VUI is improving to make the devices in a manner in which it will fit in human society.

#### **5. SECURITY CONCERNS**

Though the voice assistants are highly popular but there are some facts about its privacy and the security aspect are there like for classroom environment or for educational purpose its not fully considerable. Based on several studies since devices listen all the time so the microphone is all the time stays on so they can respond to users at any point of time. Lau et al. (2018a) took interview of 17 users who use smart speakers and 17 who don't use smart speakers to find out their views on adopting new technology and about their privacy and security concerns According to many non-users these devices are not useful in all the companies and those who users they are less concern about privacy and company should safeguard their personal data which is not interesting to others. Some people have kept it in home based on accessibility. Most of the children loves using voice assistant many of them purchased before to get adopted to the new technology but while children are using it so it should be a separate account for them with the ability to mute the device with audio commands. There should be an incognito mode to preventing the data collection by the companies.

The perception of the user towards privacy along with its usage studied by Lau et al. (2018b). Users reported that even though they are aware to erase the audio logs of Google and Amazon devices or physically mute the device but haven't used it so far. And also, the user suggested that there should be private browsing option when surfing over the internet in the case of smart speakers.

Privacy is majorly considered by Hoy (2018) is that anyone can use or access the voice activated device can give command to the device and gather any personal information which can be linked with the account, services associated with device and may ask to perform any task which may related to any smart device or smart phones.

At the point when the data affectability and person's security concerns were high, text was pretty much as significant as voice to initiate human-like impression of the voice aide. Furthermore, discoveries propose that voice collaborations raised the sensations of having a social discussion between the client and the voice associate and this prompted positive assessment toward the voice colleague. Be that as it may, this happened just when data being asked was less delicate and keeping in mind that people revealed low degrees of security concerns.

#### **6. CONCLUSIONS**

Voice assistant have the ability to update the existing education system with the immersive learning technologies. Virtual system which enhancing the reality along with voice assistant can provide new learning experiences. In this paper, research regarding alliance of AI voice assistant in most of the scenarios is presented. Research on this topic is limited as voice assistants and smart speakers are gaining popularity very immensely. In coming years, the smart speakers and voice assistant will enter the everyday life of household and will be at center of interest. After some developments can be used efficiently in the learning process as currently there are challenges, one of that can be the language issue it does not speak all languages. In addition, they do not have proper security and privacy filters which can be used in home or can be used by children.

They will use the natural language to process and can be add on with AI techniques to accomplish smart assistant which can control

IoT applications and solve the queries given by user using web searches and it can be designed in a way which will minimize human efforts to contact with many other subsystems which can save the time doing it manually by which the system can make human life comfortable. This system is designed to have interaction between other subsystems intelligently and control devices which includes IoT devices, or getting news from internet, providing information, getting personalized data which saved previously on system etc.

Adding alarms, calendar events, reminders such type of events should allow by android application for adding the data by user over the voice. By this the software will help to ease of access to various other platforms and devices. The phases for the system will be collecting the data in the form of voice, conversion to text using Speech to text (STT), storing the data and processing, generating the speech from the processed output text using Text to speech (TTS). The generated data at every phase can be used find pattern of queries and later on suggest user. This may be the major base for AI machines which learns and understands user.

Thus, on the basis of literature review can see the immense transformation has happened till the existing system, already the technology has taken important part in humans' life. There are still lot of changes to be covered up in the automation world, skills of the device can help to build voice-controlled device new generations and bring the sustaining change in the automation world.

Along with that few things need to see across like privacy default settings, improvements to muting, audio log features as well as security layers for correct voice recognitions and providing offline capabilities. Amazon Alexa and Google Assistant were in consideration in the majority of paper. There are some emerging privacy and security concerns, which can do affect user's adoption to Voice Assistant or any particular device. Along with that both the concerns laps in third party software's reviewed by manufactures.

## 7. REFERENCES

- [1] Krazit, Tom. "Google finding its voice". *CNET*. Retrieved 23 October 2020.
- [2] Carlos Paras, "Complete guide to voice assistants – Most Popular Digital Helpers"
- [3] Levy, Karyne (April 2, 2014). "Microsoft Has Its Own Version Of Siri, A Voice Assistant Called 'Cortana"
- [4] [https://www.researchgate.net/publication/292210788\\_A\\_multi-agent\\_conversational\\_system\\_with\\_heterogeneous\\_data\\_sources\\_access](https://www.researchgate.net/publication/292210788_A_multi-agent_conversational_system_with_heterogeneous_data_sources_access)
- [5] Maria Navarro, Juan L. Castro "A multi-agent conversational system with heterogeneous data sources access" Article in Expert Systems with Applications · January 2016 DOI: 10.1016/j.eswa.2016.01.033
- [6] "The Decade of Voice Assistant Revolution" ERIC HAL SCHWARTZ December 31, 2019
- [7] The future is AI, and Google just showed Apple how it's done Published October 5, 2016, Retrieved July 5, 2018
- [8] Karner, Jen (29 March 2017). "What is Bixby and is it like Google Assistant?". *Android Central*. Retrieved 30 March 2017.
- [9] Vincent, James (10 February 2017). "So, what is Samsung's Bixby AI assistant really made of?". *The Verge*. Retrieved 5 April 2017.
- [10] Voice Assistants and Smart Speakers in Everyday Life and in Education George TERZOPOULOS, Maya SATRATZEMI Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece, November 2019
- [11] <https://itchronicles.com/technology/the-pros-and-cons-of-speech-recognition-and-virtual-assistants/> The pros and Cons of Speech Recognition and Virtual Assistants, William Goddard, September 18, 2018
- [12] <https://blogs.oracle.com/marketingcloud/advantages-and-disadvantages-of-voice-assistants-for-marketers> - August 2020, Advantages and Disadvantages of Voice Assistants for Marketers – John Rampton
- [13] Lau, J., Zimmerman, B., Schaub, F. (2018a). Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 102. DOI: 10.1145/3274371
- [14] Lau, J., Zimmerman, B., Schaub, F. (2018b). "Alexa, Stop Recording": Mismatches between Smart Speaker Privacy Controls and User Needs. Poster at the 14th Symposium on Usable Privacy and Security (SOUPS 2018).
- [15] Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. Medical reference services quarterly, 37(1), 81–88. DOI: 10.1080/02763869.2018.1404391
- [16] Study of Voice Controlled Personal Assistant Device, December 2016 ,ISSN: 2231-2803
- [17] Abhijit Banubakode "A Review paper on Multi keyword rank search on encrypted cloud data", International Organization of Scientific Research (IOSR) Journal of Computer Engineering, Pp 28-32, 2015, e-ISSN2278-0661, P-ISSN 2278-8727
- [18] Abhijit Banubakode "Scalable Distributed Processing of K nearest Neighbor Queries over Moving Objects", International Journal of Innovative Research in Science, Engineering and Technology , Vol. 5, Issue3, January 2016, ISSN: 2347-6710.
- [19] Abhijit Banubakode "Hierarchical Clustering Algorithm for Improved Pattern Classification" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue2, May 2016, ISSN: 2320-9798.
- [20] Abhijit Banubakode "Query Optimization in the Object-oriented Database Using Equi-join", Advances in Computational Sciences and Technology, vol.4, no.1, 2011, pp. 83-94.
- [21] Abhijit Banubakode "Query Optimization in Object-oriented Database Management Systems: A short review", International Journal on Computer Science and Engineering Technology Subject Category: Computer Science and Engineering vol. 1, no.1, pp.1-6, 2011.