

The Evolution Of Voice Portal And Virtual Assistants

¹ Ashim Jana, ² Dr. Harshali Patil

¹ Student, MET Institute of computer Science, Bandra (W), Mumbai, India.

² Associate Professor, MET Institute of Computer Science, Bandra (W), Mumbai, India.

Abstract

Voice portals and Voice based internet access gives users access to the web through the Human Voice over a mobile or home phone anytime, from anywhere using rapidly evolving speech recognition technology. The objective of this research paper is to describe the various technology factors that are making voice portal and virtual assistants the next big opportunity on the web, as well as the various approaches service providers and developers of virtual assistants can follow to utilize this exciting technology. Also discussing the practical application and limitations of the same.

Traditionally if you have to get some information over the internet you would need a physical device like desktop, mobile phone or other capable device of some sort with an internet connection, manually go to the website or web portal and search for what you need to get result. This approach can sometimes be very inefficient as it is a little time consuming. Basically voice portal can be defined as "speech enabled access to Web based information".

With rapidly increasing number of smart devices users such as smart phones, smart home assistant etc. voice portals and voice assistants have become the latest trend in broadening the reach of services for all businesses.

Key words: Virtual Assistant, Speech Recognition, Voice Commands, VUI, Google Assistant, Alexa, Siri.

1. INTRODUCTION

From a technical point of view, Voice portals are the same as web portals but the information are accessed through the medium of voice that is through voice commands and voice responses. Generally speaking a voice portal is like an access point for any type of information or services found on the Internet. A mobile user with a mobile phone might dial in to a voice portal Web site and request information using voice. Using voice portals text information can be easily converted into speech responses. It has many benefits including low risk and low cost.

Why Voice Assistance?

Voice assistance gives user hands-free access to various functions as they only interact with voice. Voice assistance can be used to remove language barrier while interacting with information on the web.

Voice assistance is providing a huge opportunity to various service providers around the globe to extend the reach of their services to a wider range of customers. With the latest advancements in speech recognition technologies by various organizations voice enabled services are the next big industry opportunity.

Some advantages of using voice activated services:

Audio input does not interfere with visual tasks, such as driving a car. It is entirely hands free and eyes free technology. It allows for easy incorporation of sound-based media, such as radio broadcasts, music, and voice-mail messages into your services. Speech enabled access of information or services need very minimal interaction with any input hardware hence making it ideal to perform along with multiple tasks. Easy way to access internet for the visually impaired. For organisations conversational chatbots can be configured to reduce labour cost of answering calls.

2. LITERARY REVIEW

George Terzopoulos and Maya Satratzemi discusses how Virtual Assistants communicate with users in natural language using AI and cloud computing. Their study shows how Voice Assistants are incorporated in our day-to-day life and the future scope for this technology in field of education [1].

Matthew B Hoy introduces the most popular voice assistants that are embedded into smart phones such as Siri, Alexa and Cortana. Discussing the most common features and basic workings of these Voice Assistants. Also discussing how librarians can provide their services through these Voice Assistants [2].

Eric Hal Schwartz discussed how voice assistance technology has evolved in the last decade and has integrated into our daily lives. From talking about the very first voice assistant to the very latest assistant that was launched in this decade. Discussing the features of each assistant [3].

Diana Ramos discusses the impact voice assistants have on our lives by taking in consideration various surveys conducted in The USA. Predicting that the amount of users of virtual assistants by the end of 2021 would be at least 1.8 billion and how the market for voice assistants has grown over the years [4].

3. HISTORY OF VOICE ASSISTANTS

Voice recognition assistants such as Google Assistant and Alexa have become household names as a lot of tasks such as creating a reminder for any particular date, adding items to shopping list can be done just with a simple voice command. Although voice recognition technologies have been introduced since the 1960s some of the more advanced voice or virtual assistants have been introduced only in the last decade. Following is a brief history of voice recognition technology.

- ❖ IBM introduced the first speech recognition technology called Shoebox in 1962. It could perform simple arithmetic functions by giving a command over a microphone. The voice inputs were converted into electrical signals by a measuring unit and activated the attached adding machine.
- ❖ In 1971-1976 the harpy program was introduced by Carnegie Mellon University that could recognize about 1000 words.
- ❖ In 1990 Dragon Dictate was launched by Dragon which was the first speech recognition product for normal customers.
- ❖ In 1996-2002 Microsoft introduced Voice recognition for the first time in their products as Clippy and it with their office products as an assistant.
- ❖ In 2007 Windows vista was the first Windows that was launched with built-in speech recognition called the WSR – windows speech recognition which could task such as convert speech to text to perform various tasks.
- ❖ In 2011 Apple launched the first personal virtual assistant called Siri. Siri was first introduced in 2010 as a standalone app but was later acquired by Apple and was integrated with the iPhone.
- ❖ In 2014 Microsoft launched Cortana along with the windows phone as a virtual assistant which was later also integrated with Windows 10 mobile, Windows PCs and XBOX.
- ❖ In 2013 Google launched Google now as a feature of google voice search which allowed the users to search for information on google using voice commands.
- ❖ In 2014 Amazon introduced Alexa which is a virtual assistant along with the Amazon Alexa. It was capable of performing various tasks including accessing information, playing music, setting reminders and could also control smart home devices. In 2015 Amazon launched skill kit which enabled developers to create their own skill thereby expanding the variety of tasks and services that could be provided with Alexa.
- ❖ In 2016 Google launched the personal virtual assistant called Google Assistant which unlike the previous google now was able to hold conversations with the users. The Actions on Google platform also allows developers to create custom actions thereby enabling their service to reach a wider customer base. The Google Home can be connected with any smart devices like air conditioner, fridge and Google assistant paired with Google Home enables users to create a really smart environment to perform daily task with just a simple voice command.
- ❖ In 2017 Baidu launched DuerOS and Alibaba Launched Tmall Genie voice assistant AI. As US companies expand their presence of voice assistants over Asia, Chinese

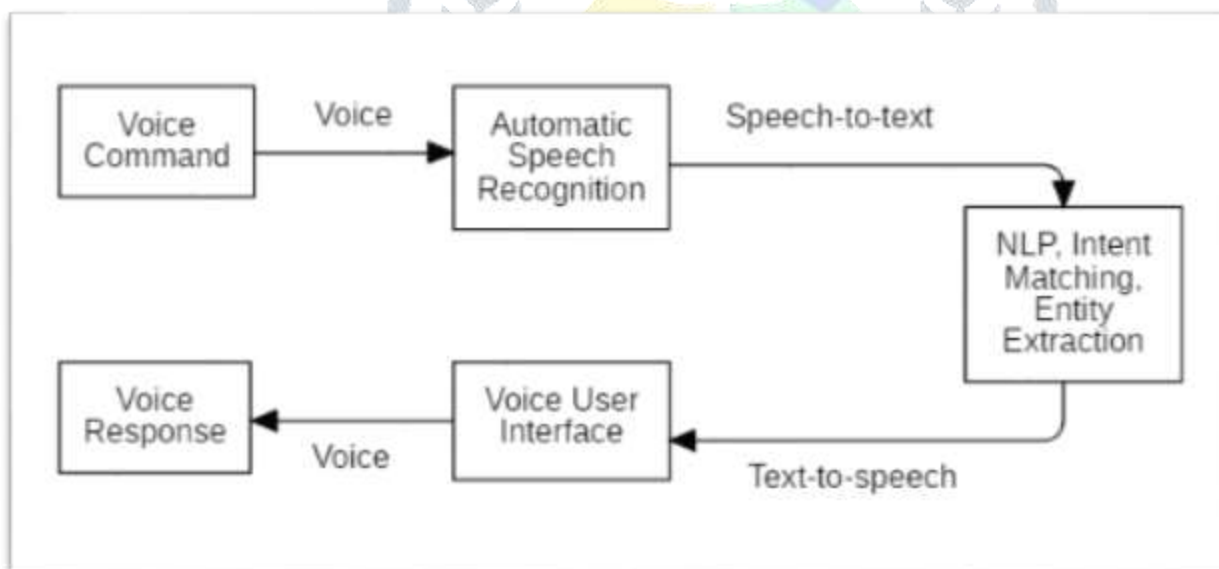
companies Baidu and Alibaba launched their own voice assistants so that the market wouldn't be dominated by Amazon, Google and Apple.

- ❖ In 2018 Samsung introduced Bixby which is their version of intelligent virtual assistant. It has similar capabilities to other virtual assistants like Alexa and Google Assistant.

4. Architecture of Voice Portal

The most important aspect of voice portal is VUI (Voice User Interface). VUI is the GUI equivalent for voice commands which provides users with an interface to interact with the internet. VUI are the primary way to interact with virtual assistants and smart speakers. VUI operates with the user of another technology called ASR (Automatic Speech Recognition). ASR works by taking the voice command input and converting it into text for computers to understand. This is also known as Text-to-speech, which allows a software to read out any text as natural speech. An ideal VUI must have multiple fallback strategies in case of failure in speech recognition. Voice portals may also be referred as Interactive Voice Response systems (IVR). IVR systems use touch tone interfaces called DTMF (Dual Tone Multi Frequency) that are touch tones from a telephone or mobile, some voice recognition technology and in some cases even Artificial Intelligence to respond to users via voice. The advantage is that it can be entirely automated and the service will be available 24/7 to the users.

Fig 1. Flow of execution of voice portal



Voice portals can be categorized into two types:

Customer Voice Portal: They are generally used to access common information such as weather, movie timings, sports timings and much more. It provides customers with access to generic information available for the public. AOL by phone, Tellme network are some examples of Customer Voice Portal.

Enterprise Voice Portal: They are always voice portal services provided by an enterprise to extend their services to their customer base. It makes automation of certain queries very simple for an enterprise using voice portals. Avaya, Cisco and Genesys are some examples of enterprise voice portal.

With the development of Personal Virtual Assistants in the last decade, Voice portals can also be accessed now using smart phones or other smart devices such as Amazon Echo and Google Home. Virtual Voice Assistants such as Google Assistant, Amazon Alexa, Samsung Bixby, Apples Siri, and Microsoft Cortana make it easier than ever to access information just with voice commands without the use of a graphical user interface.

Architecture of Virtual Assistant

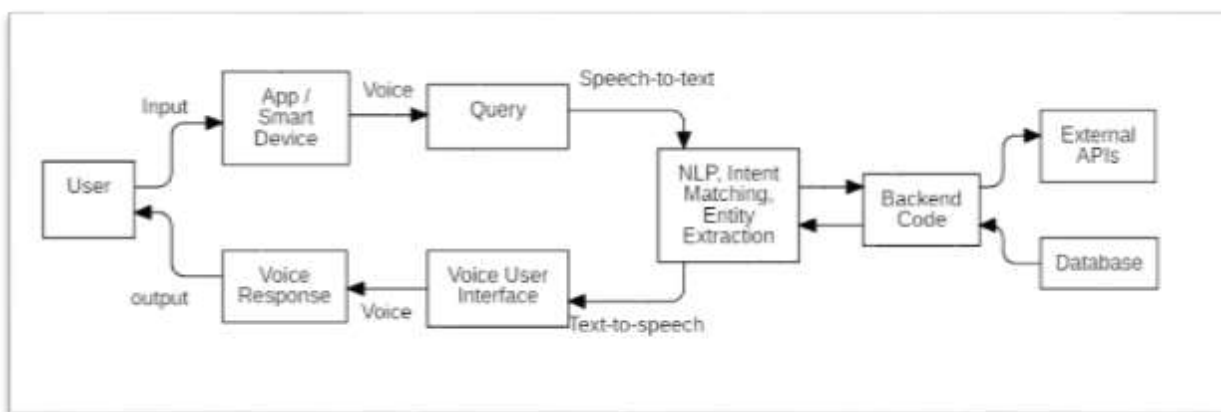


Fig. 2 Flow of execution for Virtual Assistant

The general flow of execution for any virtual assistant like Alexa or Google Assistant starts with the user's voice command which is the input to the VUI. Every virtual assistant provide a console for developers to develop frontend intents like Alexa developer console, Dialogflow, Actions on Google. In the front end after receiving the command the Speech is then converted into text format which can be understood by computers by the Natural Language Processor. The command is then analysed and matched with an intent or a web service. The matched intent will trigger the backend code for the particular command if available which can perform tasks like hitting external APIs or retrieval of data from a database. Once the required data has been fetched successfully the backend code returns to the desired in text format which is again converted back to Speech (Text-to-Speech) in the frontend. This result is spoken out by the Virtual assistant as the output. Android developers can also extend their app with virtual assistants to make the assistant perform their app specific tasks by setting up a backend code and using token generated Firebase Cloud Messaging (FCM) for accessing the same.

5. Developing technologies for voice assistance

There are thousands of Actions available on Voice Assistant platforms that provide various services to the users. The user just needs to trigger these actions to use the service. These services can range from any action like playing nursery rhymes for kids to solving complex mathematical equations and much more depending on the service provided by the provider.

There are three ways to integrate voice assistance into an app

1. Extending an app with already existing voice recognition technologies with the help of development tools such as App Actions by Google which allows the users to directly jump into a specific feature of your app by using built-in intents and handling deep links in the app
2. Developing an intelligent voice assistant or a chatbot that allows the users to hold a conversation with the chatbot and perform various tasks related to the specific skill by using open source services and APIs. Some of the technologies that provide these features for integration with an app are Actions on Google, Dialogflow, Alexa developer console and SiriKit. You also need to setup a fulfillment file to extract various entities and parameters from the voice commands which will be sent to the web service to retrieve appropriate result for the users query.
3. Developing your own Personal Virtual Assistant and embedding it in your application. To do that you need to first setup a speech engine for Speech-to-text (STT) and Text-to-speech (TTS) to greet users and receive inputs. Setting up the commands function which consist of training phrases or general commands that are expected to be received and make decisions accordingly. Setting up a Voice User Interface (VUI) to interact and hold a conversation with the user.

6. Practical uses for Virtual Assistants

Juniper research paper evaluated that there are around 3.25 billion users of voice assistants in February, 2019[5]. Voice Assistant technologies allows users to get the most out of their devices. It is quite a unique user experience as users able to perform tasks by verbally communicating with their devices. A lot of private companies have started providing services that are supported by voice assistants. But where does Virtual Assistants stand in regards to practical use in day to day life? Here are a few practical applications of virtual assistants:

- ❖ **Task Automation:** Virtual Assistants can be used to automate process in medical labs where everyone is required to gloves and bodysuits to prevent contamination. It can also be used to set reminders for future events or appointments. Nowadays a lot of household devices have become smart such as smart TV, smart AC, smart lighting, smart Fridge and much more. All these devices can be easily controlled with a virtual assistant.
- ❖ **Reduce Screen time:** Virtual Assistants allows the users to perform a huge variety of tasks hands-free and without even interacting with a device screen. Some of the tasks that can be performed are making a call to someone, opening apps on your phone, read out messages and emails verbally to instead of reading them, finding local shops and businesses of your interest, speaking out messages or emails and converting it to text instead of manually typing it, creating and maintaining a to-do list, stream music,

get sports scores, get weather updates and many more such tasks can be performed without even a need for a physical device.

- ❖ **Provide Aid to the Visually Impaired:** As Virtual Assistants are hands free technology and doesn't require user to physically interact with the interface it can be used as a guide to the visually impaired people. It can fully utilize the most useful features of a phone to the visually impaired person such as tracking GPS location, automatically reading out messages along phone number or name of the sender, the time and date, reading out system information like battery level and much more.
- ❖ **Remove Language Barrier:** With Virtual Assistants companies can overcome language barriers and can easily do business with a user base that speaks a different language. Virtual Assistants like Alexa and Google Assistant already provide support for multiple languages. Also language translation technologies are usually free of cost hence can be easily integrated with any service. In 2019 Google launched the interpreter mode which enable two users to communicate in a different language with real-time translation. As of now this feature supports about 27 languages with plan to include more languages in the future.
- ❖ **Predictive Analysis:** Virtual Assistants have the ability to analyse user interactions and patterns and suggest users related things according to the history of interactions to provide a more personalized experience. Organizations that provide services through voice assistants can use the assistant to analyse historic user data and suggest changes accordingly for increasing profits with the use of this data.
- ❖ **Support for E-Commerce:** In a large organization providing customer support to a large customer base can become very expensive and also unfeasible sometimes. Using Virtual Assistants you can automate tasks such as taking user queries and customer feedback. As interactions are more conversational than graphical it provides user a more engaging user experience. Also unlike humans a Virtual Assistant would never take a break hence can be available to users throughout the day.

7. Limitations and Security Concerns

Virtual Assistant Technology is becoming more and more sophisticated every day. But this doesn't mean that they will replace humans altogether. As number of Virtual Assistant users increase every day there are some concerns being raised about the Security and Privacy of the users. Here are some of the major issues related to Virtual Assistants.

- ❖ **Privacy concern:** When a user gives a command to the Virtual Assistant. The voice command is recorded and sent to the server for further processing. So ideally every Virtual Assistant has a wake word which when spoken activates the Assistant. For Google Assistant its "Ok, Google", For Alexa it is simply "Alexa". Its only when the wake word is spoken the Assistant starts the conversation and the recording. A common misconception is that the Virtual Assistant is always recording the user's speech throughout the day. But Actually a Virtual Assistant is always listening for the wake word that means it can only start recording the speech after it hears the wake word. It does not record anything spoken prior to the wake word. Also every Virtual

Assistant has an exit intent which is activated if there is a long period of inactivity during an assistant conversation the Assistant will exit the conversation rather than keep listening.

- ❖ **Security Concern:** Virtual Assistants like Google Assistant only store user data with the permission of the user. These audio files given the permission by user are sent to the cloud and used by Google to improve performance of Google Assistant. The voice commands provided by users are available to the providers of the Virtual Assistant in an unencrypted format. Hence there is a concern that those commands could be shared with any unauthorized third party by the provider. Those voice commands may contain some sensitive information about the user like Biometric Identity, personal details like passport number or phone number, location, medical history, search patterns. As per the New York Times newspaper [6] research shows that some voice commands can be directly embedded into music, advertisements, or other common sounds which are inaudible to the human ear. These embedded commands have the ability to wake the Virtual Assistant and could perform certain malicious actions like sending messages, opening malicious websites, even transfer money without the user even noticing.
- ❖ **Impersonation:** Most of the Intelligent Virtual Assistants provide a voice training function which can distinguish between voices and only start listening when the wake word is received by the correct user. But sometimes it can be difficult to distinguish between similar sounding voices and the voice training feature might fail and could give access to an unauthorised person. If an unauthorised person gets access to someone's personal Virtual Assistant they could fool the system by impersonating the user and gain access to personal information like bank details.

Although Virtual Assistant Technology might seem futuristic like any piece of technology it is vulnerable to attacks. From attacks that impersonate users to attacks meant to bypass device authentication in the wrong hands these can be very malicious.

8. Conclusion

With latest revolution in Voice User Interface (VUI) and Natural Language Understanding it is safe to say that Personal Assistant technology are definitely here to stay. Voice Assistance puts all kinds of information at the users fingertips which can be accessed easily from anywhere and anytime without the need to interact with a graphical user interface. It lets users get the most out of their devices very efficiently by just talking to a device. Users just need to trigger the wake command and access the information they need. As Voice Assistants are available to user 24/7 hours so service providers like e-commerce websites can provide customer support throughout the day any time. Other tasks such as setting reminders, creating a to-do list, scheduling events can be easily automated.

In the last decade Voice Assistant technology has grown exponentially and has integrated into day to day life of millions of users. Many large organizations have already started exploring the benefits of Voice Assistant Technology and many small businesses have also started to join the trend. Chances are that the products and services you provide will

definitely benefit from an integrated voice assistant. This is the best time to join into the action as Virtual Assistant technology is becoming more advanced every day with new developments being released regularly.

7. References

- [1] George Terzopoulos and Maya Satratzemi. Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education*, 2020, Vol. 19, No. 3, 473–490.
- [2] Matthew B Hoy. Alexa, Siri, Cortana and More: An Introduction to Voice Assistants. January 2018. *Medical Reference Services Quarterly*, 37(1):81-88, DOI: [10.1080/02763869.2018.1404391](https://doi.org/10.1080/02763869.2018.1404391) .
- [3] ERIC HAL SCHWARTZ “The Decade of Voice Assistant Revolution” December 31, 2019.
- [4] Diana Ramos “How Artificial Intelligence Assistants Are Changing Our Lives Every Day”, Apr 16, 2018.
- [5] Bert Kinsella “Juniper Estimates 3.25 Billion Voice Assistant Are In Use Today, Google Has About 30% of Them”, February 14, 2019.
- [6] Craig S. Smith “Alexa and Siri Can Hear This Hidden Command. You Can’t”, May 10, 2018.

