

Collate Virtual Assistance

Pooja prabhakar fulsundar,
MCA Student, TMV,
poojafulsundar1026@gmail.com

Purnima Jangle,
pournimakj78@gmail.com
Assistant Professor, TMV_Kharghar

Abstract

21's Century has been a great evolution of the technologies which actually made a difference in day-to-day life. In late 2011, Apple introduced during lead generation called "Siri" which is a virtual assistance in personal devices, and many more virtual assistances were launched after those couple of years such as Google Assistant by Google, Alex by Amazon. Whereas Google Assistant can be use within their own products & service, and several third-party service, On the other hand Alex by Amazon is more compatible within the household, entertainment, and personal too. So here we are going to elaborate on the prototype which is been used, workings of the Natural Language Processing (NLP), efficiency brought through machine learning and gave abilities such as speed, accuracy, and contextual. Behind the scenes, algorithms also play a vital role to process and excel as per the desired model. We will also glance over how each one of them is more sophisticating within small computing devices up to day-to-day big tasks.

Keywords : Virtual assistant, Echo, Duplex, NLP, speech recognition.

Period of study:

The study was made in summer vacation 2019 – a time when everyone have a free time at home and they usually used their virtual devices this period is flexible to know about the experience which everyone had with their own devices it was super fun to explore more features about the technology.

Introduction:

Virtual Assistant is also known as Intelligent Virtual Assistant (IVA) or Intelligent Personal Assistant (IPA). Basically, it is a software service with some embedded hardware components such as handheld devices or smart gadgets. Combination of hardware and software empowers the device to work as user needs. It has ability to inculcate from every different resource because its core components are based on the Natural

Language Processing, Machine Learning & Artificial Intelligence. The algorithms are designed in such a way, as a user interprets the task viz given through voice command that has been processed according to the command then final result is responded through synthesized voices. In other words, it thoroughly emulates the human interaction and also responds similarly to that particular task which may sound like a customer service or chat bot.

If we look from other side, Virtual Assistant has greatly impacting into day-to-day life as people are more depended over IVA or IPA because it has a quirk to manage all of the sessions which will be going to happen with you, we can have an example like from setting up an alarm to To-Do list, playing music, reading E-Mails, Calling a person, reminder in the calendar, any math

problem and many more. Other than this it can help with home appliances such as light, fan, speakers, home theatre where you can manage this equipment through single command.

After Apple's "Siri" many more Virtual Assistant were developed with time-to-time upgradation. So will have a glance on the chronologically development of the intelligent virtual assistant.

First Virtual Assistant was installed on phone by Apple called as "Siri" during 2011 which was able to make calls, setting up an alarm, checking weather after number of years Amazon at 2014 introduced "Alexa" & "Echo" where they were able to provide real-time information to the user on the go, on the other hand it could also handle home automation services. Later on, Google also announced its product known as "Google Assistant" based on android devices during 2017 where it also has access to Chromecast, Google Home & Nest.

Types of Virtual assistant

Currently when it comes to choosing an Intelligent Personal Assistant there are number of options out in the market. Such as Microsoft's Cortana, Blackberry, Baidu's Duer, Apple's Siri, etc. Well, we have mainly focused here on the devices like Amazon Alexa, Google Assistant and Apple Siri, which are used on a large scale.

Let's have a detail look about these devices each and what uniqueness they possess.

[1] Amazon's Alexa

Amazon is the huge company which provides various services and now also produces smart devices. The list of such smart devices includes *echo earbuds, ring indoor cam, echo dot, echo frames and loops etc.* Echo devices connect to the voice-controlled intelligent virtual assistant service **Alexa**, which will respond when a user says the name "Alexa". Users may change this wake word to "Amazon", "Echo" or "Computer". In 2014 amazon released Amazon Echo, that could take the commands from the user over just voice. Soon after that amazon launched various by-products and updated version of Echo such as *Amazon Echo Dot, Amazon Tap, Amazon Echo Show and Amazon Echo Look.* All these devices have common job to do which was to follow the commands and respond, only there was variation in their size, colour, price and different voice recognition quality. Not just that but there have been different generations of just Amazon Echo device. Amazon Echo is currently available in 8 languages and is planning to add more. [1] Following are :-

- **Echo Dot**

Originally in 2016, Amazon launched Echo Dot 1st Gen, which is a hockey puck -sized version of the Echo designed to be connected to external speakers due to the smaller size of the onboard speakers, or to be used in rooms such as the bedroom as an alternative to the full-sized Echo. Later there is Echo Dot 2nd and 3rd Gen which has lowered price, improved voice recognition and different size and structure. Recently in 2020, the Echo Dot 4th Gen was launched which is 30% smaller in size and has new spherical design.

- **Echo spot**

The device was launched in 2017 and has a 2.5-inch circular screen, and looks like an alarm clock, which works similar to Echo Show.

- **Echo Loop and Echo Frames**

These are the wearables launched in 2019. Echo Loop is a smart ring which activates Alexa with a button. The **Echo Frames** smart glasses, which support prescription lenses. [2]

[2] Google Assistant

After Amazon Echo, there comes another popular virtual assistant which is Google Assistant. Google Assistant is an Artificial-Intelligence powered assistant which primarily available on mobile and smart home devices. Google Assistant is capable of two-way

conversation. Google Assistant is currently available in over 30 countries and more than 90 languages and is used by more than 500 million users monthly. It is primarily used to respond to the natural voice but user can also interact over the input text. Google assistant can be used with third-party developers. Google in 2018 also released their first assistant-powered smart displays. The Assistant is able to search on the Internet, hardware settings, schedule events and alarms on the user's device, and show information from the user's Google account. Google Assistant launched their device using the voice of Kiki Baessell for the American female voice, the same actress for the Google Voice voicemail system since 2010. Also in 2018, Google Duplex was released, an extension of the Google Assistant that allows it to carry out natural conversations by mimicking human voice, in a manner not dissimilar to robocalling. Duplex was created to speak in a more natural voice and language by incorporating speech disfluencies such as filler words like "hmm" and "uh" and using common phrases such as "mhm" and "gotcha", along with more human-like intonation and response latency. [3]

[3] Apple's Siri

Apple's Siri is the oldest and first virtual assistant build for people using iOS devices which was released in 2011 [4]. The assistant uses voice queries, gesture-based control, focus-tracking and a natural-language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services. Returned results are individualized. Apple's products, having been adapted into other hardware devices over the years, including newer iPhone models, as iPad, iPod touch, Mac etc. Siri supports a wide range of user commands, including performing phone actions, checking basic information, handling device settings, searching on the Internet, scheduling events and reminders, navigating areas, finding information about entertainment, and is able to engage with iOS-integrated apps. People gave although mixed reviews, where some really enjoyed this feature, there were others who criticised it as well [5]. Apple's Siri voice assistant supports 21 languages. [6]

[4] Cortana

Cortana is a virtual assistance developed by Microsoft. Cortana currently available in some languages such as English, Portuguese, French, Italian, Japanese, Spanish, Chinese so on. Cortana can set as reminders without any keyboard input it can recognize natural voice. Using information from Bing search engine.

[5] Bixby

Bixby usually used as Samsung's smart assistant. Samsung's android phones come along with their own voice assistant generally its pre-installed in so many Samsung phones. You can use it as a text or voice so it will give you information related to weather, it will remind you about your meetings, so many new articles and so on it will help you out to gather information around you. Bixby learn so many individual voices so it's a personalized answer depending on a particular person who asks.

Use of NLP in Virtual Assistants

[14] Natural Language Processing, a subset of Artificial Intelligence, enables chatbots to understand language as we humans speak it. There are many processes involved in natural language processing but most commonly used is speech recognition. Speech recognition has a couple of applications in sales. This will help out in call centers to transcribe thousands of calls by customers. This is one such technology that is empowered by AI to add convenience to its users. This particular technology has the power to convert voice messages into text. And it also has the ability to recognize a voice of individual person based on their voice command. Hence, this AI-powered speech recognition technology gained mammoth importance among tech giants such as Google,

Facebook, Apple, Amazon, etc. Amazon Echo, Google Home, Microsoft and Siri are some of the apps and devices that flooded the market with speech recognition features.

Literature Review

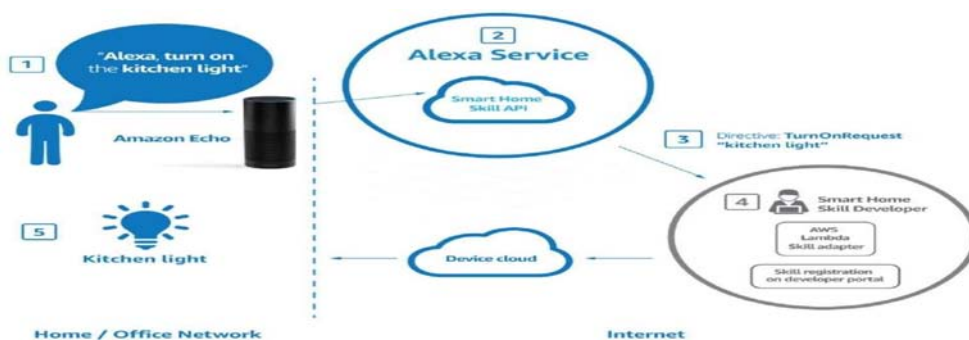
Berdasco, A.; López, G.; Diaz, I.; Quesada, L.; Guerrero, L.A. User Experience Comparison of Intelligent Personal Assistants: Google Assistant, Alexa, Bixby, SiriCortana. *Proceedings 2019, 31, 51.* [8]

In this paper proposed by Berdasco, A. in which they took responses from 92 participants about different virtual assistants like amazon's Alexa, apple Siri, google assistant and Microsoft's Cortana. Here they asked different questions that checked the accuracy and how efficiently these devices really work. They came with the conclusion that Google assistant and amazon's Alexa were excellent in factors like quality and correctness, whereas Siri and Cortana performed worse and produced average results. Although the authors of the page stated to conduct similar studies in different context, to gather more empirical evidence on the use of intelligent personal assistants.

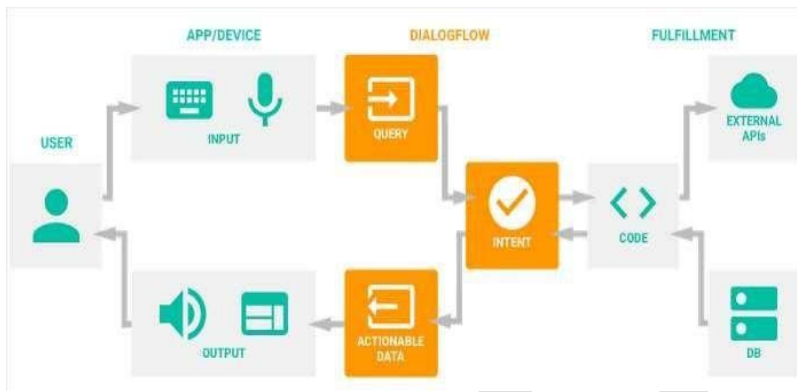
[9] In an article by Bot Penguin, which was to which is the better personal assistant between Alexa, Siri and Google. They considered three different factors and how these three virtual assistants really work in such situation. It is found that users mainly ask many different queries to these devices, where here what really matters is the accuracy and certainty. Google assistant answers 88% of all the questions correctly. And Siri answered 75%, whereas Alexa answered 72.5% of the problems. As you can see Google Assistant came out to be the best by the figures, but it isn't over yet. Another maximum time used command is Play Music, Here Alexa and Siri used their own in-built music applications such as amazon music and apple tunes respectively, although google assistant involves third party like Spotify, Pandora and more. Now coming to the last factor which is shop with assistant, here the competition is more between Alexa vs google. As we already know, Amazon is one of the biggest best retail stores online, which has almost everything you need and from fashion to day-to-day appliances. Since with the help of amazon Alexa, it has a homegrown advantage of letting users find and shop products instantly with just one command. It helps the users get a seamless experience, and all can be done from their comfort zone. Whereas Google assistant is also quite helpful when recommending, Navigating and guiding during online shopping. In the case of Siri, it isn't developed enough to handle orders yet.

Proposed Work (Working of Assistants)

We have been always wondering about the way technologies evolving around us, so we are trying get a picture of working virtual assistance in the most of the devices we use in day-to-day life. [10] So "Alexa & Echo" is a well-known service by Amazon Inc. As Alexa has been made using natural language processing where this process initializes the conversion of speech into sound or word. After ages, cloud computing has given more computing power to process huge process in order to work more efficiently as Amazon needs to record the speech interpreted by the user or client which directly send over the servers of the Amazon. These speeches are firstly clean but there must be some kind disturbing or ambient noise such as people gossiping, machinery noise & many more. So, to filter such disguise noise, noise cancellation has been performed to understand what user is trying to interpret.



If we trying get a closer look, Communication is been always depended on the signal, domain like sound as signal processing will only help to improvise the audio which has been recorded at user- end. Amazon has term called as Wake-Word-Detection here it sets a flag sounds more like a Boolean method as device hears the word “Alexa”, the device has to get activate or else device will sleep automatically. As Alex hears a command these audios are converted into text then it will analyse on the basis of the frequency and pitch the way user has pronounced the sentence. Hereafter the deep learning is applied to compare audio and transcripts which are been generated. After collecting the speech words spoken by the user, those words are been broken down to understand the need of the command and provide result in response. Amazon has huge database of the words as it is necessary to update the words so user will not encounter by any while searching. Now the word is been compare to the most synonym or relevant to interpreted word by user. Later on, it processes to identify words for example “series” or “movies”, so relevant app has to be open like “Netflix” or “Amazon Prime”. This is how servers will be responding to these keywords. If we look into the Google’s Google Assistant, it awakes when we call “Hey Google” and continue our desired question. Later on, those recorded speech is converted using speech-to-text viz also called as TTS in generally. Here we have an entry point into interaction which is built for assistant that is requested by user called as “Action”, whereas intent is known as what user wants to do for example make a call



Then we have fulfilment response by Google Assistant. But actually, working is depended upon the Action SDK (Software Development Kit) which has ability to fulfil request. Between Intents and Fulfilment, there is a training phrases as here NLP, Machine Learning & Knowledge Graph is performed. These processes are connected to backend of the fulfilment which uses technologies like python java, PHP, etc. where it is been hosted.[13]

Since 2011, Siri was featured in the Apple iPhone 4S. Later on, many upgradations were made to make as fluent and efficient as possible. It works kind of Hands-Free Access in other words we can say that there is no need to press any button to activate the Siri as we only have to call the invocation word called as “Hey Siri”. As it detects the word “Hey Siri” it will automatically concatenate rest of the command query which is been uttered from user. The invocation detector uses Deep Neural Network (DNN) to convert the acoustic pattern of the speech at each instant into a probability distribution over speech sounds. As we know, Siri learn our voice pitch and intensity while setting up profile so if the score is higher means when it matches with the saved value of the voice it. If we take a closer look, it contains five vectors, a corresponding speech vector is extracted for each test utterance. Its cosine score is computed against each of the speech vectors already in the profile, if the average of these scores surpasses a specific threshold (λ), the device processes the user’s command. The most predominant part of any speaker recognition system is the speaker transform. The speaker transform was further improved by using explicit enrolment data, enhancing the front-end speech vector, and switching to a non-linear discriminative technique in the form of deep neural networks (DNNs). A DNN from 16,000 users each speaking ~150 utterances were trained. After the DNN is trained (using the speech vector as an input), the SoftMax layer is removed. The output of the linear activation layer is then used as a speaker vector.

Privacy

Earlier people were only able to interact with computer using devices like mouse, keyboard and touch screen. But after some years technology of the voice or speech recognition has brought huge difference as

we are able to control using voice control technology. Most of the people are engaged to get help from smart devices or else we can say many of the people are depended for any future task by using smart devices. Today's Virtual Assistant are more advanced because of natural language processing as it connected to the internet means it is accessible globally from anywhere. As these gadgets are interconnected to each other so there are some significant numbers of security and privacy that has be addressed. Issue is generally related to the authorization, speech recognition, profiling & the technology integrated with the cloud or third party companies. [16]

Aspects of Virtual Assistants

Wake-up Words

To activate any virtual assistant, there is a wake-up word for example "Hey Alexa" or "Ok Google" through this method devices are activated. Before they are activated, it must be authenticated so they are using wake-up word to authenticate user which can help anyone to attack on it. That's why there is weak authentication which can be use as proxy for more elaboration.

Always On, Always Listening

Speech command interpreter constantly listens to the user as it waits for the wake-up word. This may lead to a privacy issue because it is default on by itself as it waits for wake-up word but if some spoke a word which sounds same like a wake-up word then it might upload the further conversation which is been spoken after that as it can also contains sensitive information. Recently, due to this feature, private conversation of a couple was accidentally recorded and sent to a random contact with the echo.

Multi-user Environment

Without any proper authorized functional can prevent users from people who and why the resources are given to access it. In multi-user environment, anyone can put into recording mode and give out instruction to it. So even after main primary user set the permission level for secondary user, security is still improper because any one the family member can modify the network connection, sound, any many more without actually user's permission.

Attacks

Many of the attackers study the architecture of the working system and try to identify the flaws in the system where it will be easy to exploit the weak authentication and other third party integrated technology can also be a major flaw. While device is connected to the cloud services, many attacks can be exploit architecture through a hidden traffic during on-going interconnection as this can be a possibility of attacking. At point of the architecture virtual assistant need to exchange the information to get back response as a service through cloud so no matter how much it is encrypted there will be a privacy vulnerability because can give protection up to some extent. Voice squatting allows an attacker to use a malicious skill with longest matching skill name, similar phonemes, or paraphrased name to hijack the voice command of another skill. The feasibility of this type of attack is high, particularly in SPA, such as Alexa that allows multiple skills with the same invocation name. This attack can be used to damage the reputation of a legitimate skill as any poor service of the malicious skill will be blamed on it. Equally, in voice masquerading attack, a malicious skill pretends to invoke another skill or fake termination. Then, the skill keeps recording the user's utterances. This attack could be used to snoop on the conversations of the user.

Methodology

we selected 100 users there are 45 males, and 55 females and from them there are 30 frequent users of virtual assistant and 70 users are non-frequent users. Among them all are native English speakers. Usually, we focused on best virtual assistant. Among them 100 users are regularly using four virtual devices for build their own experiences and majority of them are not regularly active users. (fig 1&2).

our purpose of survey is involved only one stage. Were users provided multiple choice of questions and multiple choice of choice devices which they supposed to ask to all of these multiple virtual devices. The basic questions is basically based on Email, messaging, Weather, Call log, Reminder for meeting schedules,

music, Sports, Social sites, Personality Calendar & Personality. The evaluation between different virtual assistant is find the best virtual device which used to access each personal assistant are mentioned in Table no.1

Survey on different virtual assistant which are: -

Google assistant, Siri, Alexa, Cortana

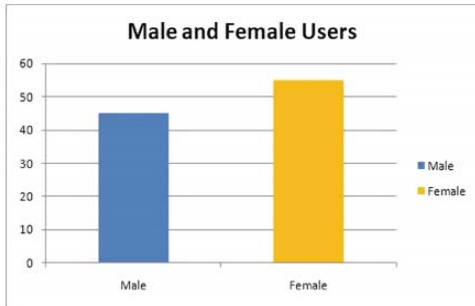


Fig. 1. Numbers of male and female.

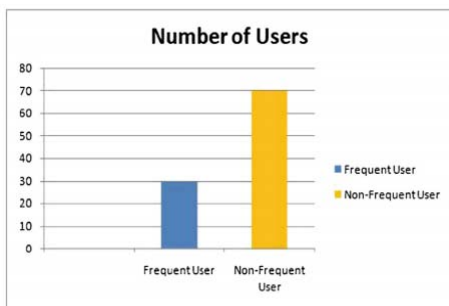


Fig. 2. Numbers of frequent and none frequent users.

Table no.1

Virtual Assistant	Device
Alexa	Echo Dot
Siri	iPhone 6s
Google Assistant	All android phones
Cortana	Windows 10 Laptops
Bixby	Samsung smartphones

Statement Of Problems

Probability of messages leaked by virtual network to another. interception of information in this case sometimes attackers attack on sensitive data by using sniffing and eavesdropping skills. In Alexa or apple phone have Siri like this device take voice commands and nlp convert that voice commands into text format but the problem in devices which is adding too many skills so there is no such way to memorize the list of voice commands can or can't give the proper ai assistant. As per tasks AI perform so many tasks so every time user expect device can able and to understand what they ask to tell it. Another problem with voice assistant is that they are not suitable for large, complex and multi tasks. Like if someone want to make a reservation which would require going back to another screen that would be very challenging in this sort of works. But in case of minimal tasks like playing music, open browser or setting reminder in this kind of tasks the voice assistant work very neatly. Problems with integrate too many commands into smart speakers like digital devices stored our sensitive data like location, conversations, images and so on they used it for machine learning for improvement of themselves. There are so many examples of leaking a data by mistakenly. For example, Alexa device or amazon assistant accidentally remain unplug if conversation between

two people were recorded by them and it will be mistakenly sent to the random person from their contact list so there is no particular secrecy maintained by devices. Or they are maintaining our collective privacy.

Result & Discussion

After coming across many virtual [10] assistants, it still looks that Alexa and Google Assistant are used widely by audience. It is still difficult to say which is preferable as there are many changes and updated versions of virtual assistants present. But considering the factors such as accuracy, security, ease-to-use, price and compatibility these three virtual assistants are best in their own way, like Amazon Alexa is Best for device compatibility, Google Assistant is best for responding and last but not least Apple's Siri is most popular mobile assistant. [15]

Conclusion

In this research paper, we have come across various points related to different virtual assistants. The main objective of this paper is to study the technologies behind these systems. Even though using NLP, speech recognition and market intelligence etc. When natural processing language is used with artificial intelligence, it enriches the working of virtual assistant. As we conclude that with growing technologies and innovation, there are many updated versions of such devices coming along. Still nothing has yet defeated Amazon Alexa and Google Assistant.

References

1. https://en.wikipedia.org/wiki/Google_Assistant
2. <https://summalinguae.com/language-technology/language-support-voice-assistants-compared/>
3. Hey Siri: An On-device DNN-powered Voice Trigger for Apple's Personal Assistant - Apple Machine Learning Research
4. Tilak, G. Usage of Visual Communication Design on Consumer Behaviour.
5. Natural Language Processing ft. Siri | by Komal Saini | MyTake | Medium
6. <https://www.foholicresearch.com/blog/Impact-of-artificial-intelligence-in-speech-recognition-technology/>
7. [1903.05593] Smart Home Personal Assistants: A Security and Privacy Review (arxiv.org)
8. <https://www.reviews.com/home/smart-home/best-voice-assistant/>
9. Tilak, G., 2021. Robotics and Artificial Intelligence: Impact of Robotics at Intelligent Homes. International Journal of Applied Engineering Research, 6(1), pp.11-30.
10. https://en.wikipedia.org/wiki/Virtual_assistant#:~:text=The%20first%20modern%20digital%20virtual,4S%20on%204%20October%202011