Artificial Intelligence-based Voice Assistant

Subhash S
Department of Telecommunication Engineering
Dayananda Sagar College of Engineering
Bengaluru, India
ssubhash977@gmail.com

Prajwal N Srivatsa
Department of Telecommunication Engineering
Dayananda Sagar College of Engineering
Bengaluru, India
prajwal.nagendra@gmail.com

Siddesh S
Department of Telecommunication Engineering
Dayananda Sagar College of Engineering
Bengaluru, India
sureshchandra.siddesh@gmail.com

Ullas A
Department of Telecommunication Engineering
Dayananda Sagar College of Engineering
Bengaluru, India
ullasullu29@gmail.com

Santhosh B
Department of Telecommunication Engineering
Dayananda Sagar College of Engineering
Bengaluru, India
santhosh-tce@dayanandasagar.edu

***Abstract*—Voice control is a major growing feature that change the way people can live. The voice assistant is commonly being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through GTTS (Google text to speech). GTTS engine will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.**

Keywords— Voice control, AI-based Voice Assistant, GTTS Engine, Playsound, Python.

I. Introduction

In recent times only in the Virtual Assistants we can experience the major changes, the way user interacts and the experience of user. We are already using them for many tasks like switching on/off lights, playing music through streaming apps like Wynk Music, Spotify etc., This is the new method of interacting with the technical devices makes lexical communication as a new ally to this technology.

The concept of virtual assistants in earlier days is to describe the professionals who provide ancillary services on the web. [1] The job of a voice is defined in three stages: Text to speech; Text to Intention; Intention to action; Voice assistant will be fully developed to improve the current range.[6] Voice assistants are not befuddled with the virtual assistants, which are people, who work casually and can therefore handle all kinds of tasks. Voice Assistants anticipate our every need and it takes action, Thanks to AI based Voice Assistants.

AI-based Voice assistants can be useful in many fields such as IT Helpdesk, Home automation, HR related tasks, voice based search etc., and the voice based search is going to be the future for next generation people where users are all most dependent on voice assistants for every needs. In this proposal we have built the AI-based voice assistant which can do all of these tasks without inconvenience.

II. Background

A. History of Voice Assistants

If we come to the history of Voice Assistants,[2] the first voice activated toy to be released is 'Radio rex' in the year 1911. Foundation of smart virtual assistants was laid by IBM Simon in the year 1994 as we know today. Digital speech recognition technology has become an aspect of personal computers in the 90s with Microsoft, Apple, Philips, etc., After various researches 'Siri' was introduced as the first modern digital Voice Assistant as the feature in iPhone 4S in 2011.

Many companies have used oral dialog system to design such system devices like Amazon Alexa, Microsoft Cortana, Apple's Siri, Google Assistant, etc., [3] General dialogue systems have six components which include Voice recognition, voice language apprehension, dialog manager, natural language generation, text to speech convertor, and knowledge base.

B. Future Applications

In future voice assistants can be used for two developments: First quality of dialogue recognition will increase because broadband allows more complex data processing in powerful data centres. Second, from the users perspective, VAs aid for interaction. In the companies, voice assistants can be used to automate repetitive tasks, for example Amazon's Alexa can open video conferencing and book meeting rooms etc.

C. Aim of this Study

The main aim of our project is that we have created a function, Intelligent Personal Assistant which can perform mental tasks like turning on/off smart phone applications with the help of Voice User interface (VUI) which is used to listen and process audio commands.

III. Proposed Design

The project will give a fair knowledge about the intelligent assistant which is capable of understanding the commands given by the user. Our assistant can easily understand the commands given by the user through vocal media and responds as required. Our assistant performs the most frequently asked requests from the user and makes their task easier. Our voice assistant listens to the command given by the user through the microphone. After listening it will say "done listening" and displays what the user said and acts accordingly.

In our project we have installed gTTS engine package to make the voice assistant speak like a normal human being. We have defined a function called 'voice assistant speak', as

978-1-7281-6823-4/20/$31.00 ©2020 IEEE

explained in (1) The gTTS will analyze the command given by the user through microphone and searches in the browser the required response and convert that response into text.

tts = gTTS(text=audio_string, lang='en') (1)

gTTS is basically used to convert the audio string into text. This audio string is nothing but the response which the voice assistant is supposed to give the user. The language of the text is chosen to be English, the code for English is 'en'. We save this entire function into 'tts'. We are saving this text, that is the audio file with the '.mp3' extension. Each audio file is given a random number from 1 to 20000000. The random number can be generated using the command 'random.randint()'. This whole '.mp3' extension file is saved under the name 'audio file'. Finally to save this audio file we have used the command as mentioned in (2).

tts.save(audio file) (2)

This command (2) saves the audio file in the system. (Ex - 'audio24854.mp3').

IV. Tasks Performed by the Voice Assistant

- Can remember any person name till the usage session.
- Voice assistant name can also be changed unlike in other voice assistants.
- Play/download a song or video from YouTube. When user asked 'can you play/download me a song', 'play movie' the assistant open YouTube and plays the required content for the user or download the requested video/song.
- Searches anything from google and tells the required content. If asked 'google search' the assistant searches the content asked from the google and opens the required content in browser.
- Opens the maps and tells the exact location the user asked for. When asked for 'find location' or 'google maps' the assistant ask for the location the user wants and opens the google maps and highlight the location user asked for.
- Tells the accurate weather of the location the user asks for. When asked for 'current weather in' the assistant tells the exact weather of the desired location of both maximum and minimum in degree Celsius.
- Takes a screenshot of the display. When asked for "capture", "capture my screen", "my screen", "screenshot", "take screenshot", the assistant captures the display the user is using and stores it in the path specified.
- Gives the live news around the world. When asked 'news for today', 'tell me the news', 'what's the news', 'what is the news', 'news', the assistant reads the first 5 updated news headlines from the website.
- Can able to tell whether the password has been hacked or not.
- If any person is in danger, our voice assistant can able to send the user's location to the police or to the relatives by giving command as "I'm in danger".
- It sends a mail to the username specified by the user. When told "send mail", the assistant asks to whom the mail has to be sent and it will send a mail according to that.
- It can translate the words the user speaks into any language and displays the words of that language which is specified by the user.
- Can able to shut down or restart the system by just user command etc.

These are some features we have added to our AI-based voice assistant as of now and we are working on many more features to embed into this assistant.

V. Methodology

Voice assistants are all written in programming languages, which listens the verbal commands and respond according to the user's requests. In this project we have used Python Programming language to build the AI-based Voice assistant. A user can say, "Play me a Song" or "Open facebook.com", the voice assistant will respond with the results by playing that particular song or by opening Facebook website. The Voice assistant waits for a pause to know that users have finished their request, then the voice assistant sends users request to its database to search for the request.

- The request asked by the user gets split into separate commands, so that our voice assistant can able to understand.
- Once within the commands list, our request is searched and compared with the other requests.
- The commands list then sends these commands back to the Voice assistant.
- Once the voice assistant receives those commands, then it knows what to do next.
- The voice assistant would even ask a question if the request is not clear enough to process it, in other words, to make sure it understands what we would like to receive.
- If it thinks, it understands enough to process it, the voice assistant will perform the task which the user has asked for.

A. Working of ASR

As shown in Fig 1. Automatic Speech Recognition which is termed as ASR is the main principle behind the working of AI-based Voice Assistant.[4] ASR systems, at first it records the speech, then the wavefile has been created by the device which consists of the words it hears, later the wavefile will be cleaned so that the background noise would get deleted and the volume will be normalized, then it will break down into elements and it will be analysed in sequences, then the ASR software examines these sequences and it implements statistical probability to find out the entire words and then it will get processed into text content.[5] The better method to recognise elements is Element Recognition as it provides better results than the method of word decoding.

Fig. 1. Process of ASR

It does not matter what kind of speech recognition software we may use, because all the work happens in its ASR. During a nutshell, at first the method starts with the device gathering audio with the source, where source is microphone, then the Recorded speech waveforms will be sent to acoustic analysis, which will be performed on three different levels, as shown in Fig 2.,

B. *Acoustic Analysis*

- **Acoustic Modelling:** In this process, it represents that the elements were pronounced or not and what are the words which can complete these elements.
- **Pronunciation Modelling:** That analyses the way, where how these elements are pronounced, it will check whether there is any accent or other peculiarities.
- **Language Modelling:** This is often aimed toward finding contextual probabilities counting on what elements were captured.

All the data which were recorded get processed by Artificial Intelligence without any human interaction, then the speech waveforms data is transmitted to the decoder, where it finally transforms into text for further use like command.

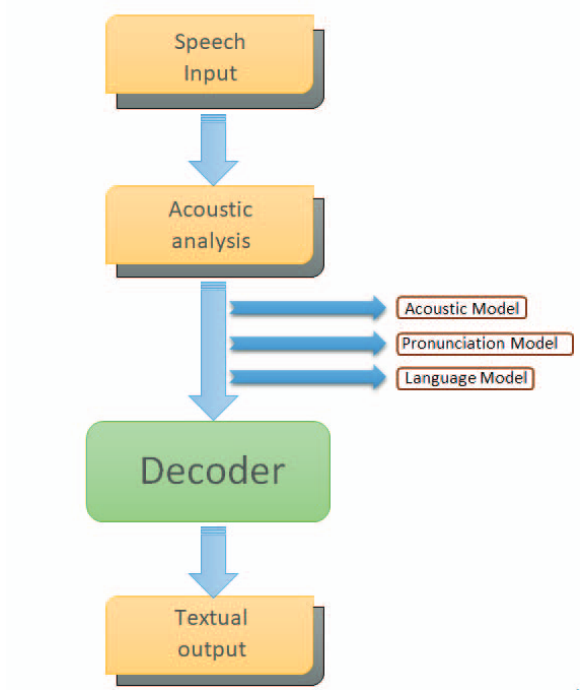

Fig. 2. Acoustic Analysis

VI. How our Voice Assistant is Useful

In this section, we explain a set of situational events and we can tell how our voice assistant is useful. For e.g., if we want to go to a particular location instead of opening google maps and typing the destination takes a lot of time instead of that with our VA just by command to find the particular location it opens the map and highlight the particular location. In the modern VA's, playing a song on YouTube is just a feature, but they can't download it. For e.g., if you want to download the song "give me some sunshine", in other voice assistants we can't download it but in our VA, the song is searched in the YouTube database and the Video ID is noted, with this the video is downloaded.

If we want to get any information, we have to open google and search for it. But in VA by just giving it in the form of command for e.g., "what are super computers?" it collects the best information available on Wikipedia and gives us and if we want to know about what is happening in the world by giving the command "what are the top five news of the day?" it grabs the information and tells us. From this we can gain knowledge. Concerning the aspect of security, if a girl says "I'm in danger" the Voice Assistant extracts the current location of the girl and sends it to their parents/relatives and the nearby police station.

VII. Results

The required packages of Python programming language has been installed and the code was implemented using PyCharm Integrated development environment (IDE) and the python code we have developed runs in both Python 2.7 and Python 3.x, and below are the few outputs which we have received in our AI-based voice assistant.

A. *Google Search Output*

As shown in below Fig 3. When we ask the voice assistant to search 'MS DHONI', it receives the request and performs the action by searching google.

Fig. 3. Output screen of performing Google Search

B. *Playing song/video on YouTube*

As shown in below Fig 4. When we ask the VA 'Play me a song', it responds by saying 'which song you want me to play' and we ask the VA a particular song/video then the VA will perform the task by playing song/video on YouTube.

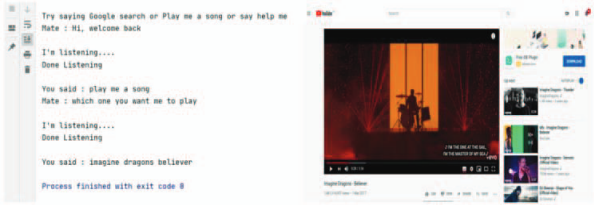

Fig. 4. Output screen of Playing song on YouTube

C. Searches any location

As shown in below Fig 5. When we ask the voice assistant to search any location, it receives the request and it searches that particular location using Google maps.

Fig. 5. Output screen of Searching Location

D. Current News

As shown below in Fig 6. When we ask current news to the voice assistant, it receives the request and responds back by giving top 5 news for that day along with website name.

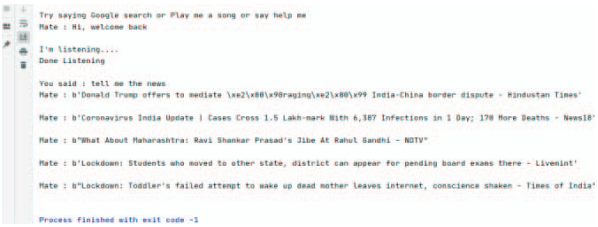

Fig. 6. Output screen of Showing Current News

VIII. Conclusion

In our project we have implemented many things compared to other assistants. Now a days it is very useful in human life because it is a hands-free application. It is a very simple application. As well as it is used in a business field also for example in laboratory, the person wears gloves and body suits for their safety purpose so it is difficult to type, through voice assistant they can get any information so that their work becomes easy.

Voice assistants are useful in many fields such as education, daily life application, home appliances etc. and voice assistant is also useful for the illiterate people they can get any information just by saying to the assistant, luxury is available for people, thanks to AI based voice assistants.

Voice assistant is developing more and more in daily life. Many companies of voice assistant trying to improve interaction and more features to the next level and many of the youth started using voice assistant in daily life and from many sources the result showing very good feedback. Compared to last 2 years voice assistants have been developed more and more.

References

[1] Saadman Shahid Chowdury, Atiar Talukdar, Ashik Mahmud, Tanzilur Rahman, "Domain specific Intelligent personal assistant with bilingual voice command processing," IEEE 2018.

[2] Polyakov EV, Mazhanov MS, AY Voskov, LS Kachalova MV, Polyakov SV "Investigation and development of the intelligent voice assistant for the IOT using machine learning," Moscow workshop on electronic technologies, 2018.

[3] Khawir Mahmood, Tausfer Rana, Abdur Rehman Raza, "Singular adaptive multi role intelligent personal assistant (SAM-IPA) for human computer interaction," International conference on open source system and technologies, 2018.

[4] Veton Kepuska and Gamal Bohota, "Next generation of virtual assistant (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," IEEE conference, 2018.

[5] Piyush Vashishta, Juginder Pal Singh, Pranav Jain, Jitendra Kumar, "Raspberry PI based voice-operated personal assistant," International Conference on Electronics And Communication and Aerospace Technology, ICECA, 2019.

[6] Laura BURbach, Patrick Halbach, Nils Plettenberg, Johannes Nakyama, Matrina Ziefle, Andre Calero Valdez, "Ok google, Hey Siri, Alexa. Acceptance relevant of virtual voice assistants," International communication conference, IEEE 2019.