

Desktop Voice Assistant With Speech Recognition Intelligence (DVAbot)

Aditya Tyagi, Student, ABES Engineering College, adityatyagi765@gmail.com Harshi Singhal, Student, ABES Engineering College, harshisinghal765@gmail.com Mradul Jain, Associate Professor, ABES Engineering College, mradul.jain@abes.ac.in

ABSTRACT- In the advanced current days of reckless aheading technology we can perform tasks that we not once imagined that we could prepare that earlier but to complete these daydreams there is a requirement for a way by which we are able to automate our everyday tasks easily. Thus we developed applications like Voice Assistant that have the capability to communicate with us just by of human interaction. Various applications like Amazon Alexa, Microsoft Cortana are also used as an voice assistant .The voice application called "SIRI" help users to interact with other user's mobile with vocal sound and also answers back to the voice instructions given by the user. We are going to develop a desktop voice assistant which will automate the tasks that are performed on desktop. It can modify the mode of communications between user and machine. This application is made in such a way that all of the facilities provided by the mobile phone can be retrieved by the user by giving the voice commands.

Keywords : Automated Speech Recognition, Desktop Voice Assistant, Natural Language Generation, Natural Language Processing, Natural Language Understanding, Speech Recognition, Speech Synthesis, Text-to-Speech, Voice Assistant.

I. INTRODUCTION

A Voice Assistant has been established in the form of application which runs on desktop that helps users to perform multiple tasks such as opening checking the time, websites, searching different information on the internet and opening programs in computer using voice commands. The methodology that was followed during the development of this Voice Assistant was Object Oriented Programming, and the programming languages used was Python.

1.1. OBJECTIVE

The objectives of our project are as follows:

- Open any website in the browser by giving verbal command to the voice assistant.
- Gives information about Corona; total cases, deaths and recovered cases in particular country through speech.
- Can perform calculations and tells the result through speech.
- Answer queries through speech.
- Tell the current time and date to the user through speech.
- Tell the weather of a city to the user through speech.
- Send email to the desired person.
- Open various programs and applications available within the desktop.
- Open Youtube with desired videos.

• Open wikipedia with information about particular topic.

1.2. PROBLEM IDENTIFICATION AND DEFINITION

Everyone want to have the extravagance to have an assistant who always give attention for your calls, look after each and every need, and takes necessary actions when needed? This opulence is now obtainable because of voice assistants. They can carry out a variety of functions after listening to your voice command. They can also play music, send e-mails answer questions, and can also perform various tasks for you.

[1] PERSONAL ASSISTANT WITH VOICE RECOGNITION INTELLIGENCE

Now-a-days the technology is only well-known for the experience of user, because it is very easy to use such applications and services from anyplace regardless of our location. Android, Windows, Apple, etc. are some well-known and widely used Mobile Operating Systems. All these Structures provide many services and applications to their users. The application allows the user to connect a call or send an SMS to other user. Contacts applications are used to store the contact details. There are many similar types of app in the world via Play Store, Apple Store, etc. All these features generate some different kinds of functionalities to be implemented in the cell phones.

Speech recognition systems has a very long past with most important inventions. Recently, designing of large



terminology speech recognition which are able to work effectively on devices, precisely and with less efforts. It can gather if we use CTC-based LSTM acoustic model which is able to forecast context-independent mobiles and use a mixture of SVD-based quantization for attain the compression ratio of 1:10.The Search components do search tasks and speech recognition. In addition to Search, we integrate a query analyzing module between Search and ASR for various numbers of reasons. Variety of methods are used for improving the performance of voice services.

STEPS TO START OFFLINE VOICE RECOGNIZER

1. Firstly default Android Voice Recognizer should be set to Google.

2. Uninstall the updates for the Google Voice Search applications.

- 3. Reboot your device.
- 4. Update Google Voice Search from the Play Store.
- 5. Reboot.
- 6. Install English US language files.
- 7. Reboot the device.
- 8. Use Utter with connection.
- 9. Switch to aero plane mode.

It was made for the help of Blind persons who can able to work using their Voice. It has the skill of identifying the voice without internet. It also has various features of mobile like handling many applications and network connection just by commands given by voice. It also has features like Voice Pattern Detection, Keyword Learning etc. That is very helpful for end user to use the mobile devices. It is not language dependent, it actively responds back to user's voice. It is faster than other Online Voice Search applications.

[2] GROWTH OF A VOICE-CONTROLLED ASSISTANT FOR ELDERS AND DISABLE:

- Works similarly as a desktop application to support visibility, since elders find it hard to access the phone because the size of the screens is small.
- Have a simple interface that will permit users to interact simply with the device.
- Speech production is the non-natural way to create sound like human being. Speech synthesizer is a system made for the purpose to produce sound like human.
- This project was accepted to supply a means by which the old and disabled in our society can use phones and desktop easily. This was skilful with the growth of a speech recognition application that supports the user to do simple tasks on their system for example: examination of time, opening programs, searching any information over internet and even starting different files and application on their system, and this all can done through voice commands.

[3] THE STRUCTURE OF GENERAL DIALOGUE SYSTEM



Fig 1: General Dialogue System

ASR MODEL

The speech recognition sample will work in real-time by using microphone as input device and with the ASR model in Cloud Servers to recognizing the noises that a user produce unwantedly into a microphone and then modify it to text; then it sends that text to the applications which is in Cloud Servers to examine the text and provides the result.



NATURAL LANGUAGE PROCESSING(NLP):

The most common technology that is used behind voice assistants is Natural LanguageProcessing [5]. NLP uses artificial intelligence that means intelligence used by machines rather than humans and for that it requires a large data set. It has four components that are Automatic Speech Recognition, Natural Language Understanding, Natural Language Generation, Text-to-Speech.

AUTOMATIC SPEECH RECOGNITION (ASR):

ASR model is used to convert all the speech signals into respective words or string of words. Day by day the vocabulary is becoming larger and larger [6]. Single word can be recognised easily rather than a continuous speech. The accent of the person can affect the error rate of the speech [5]. It is evaluated that Japanese and Spanish accented Englisg has 3 to 4 times more error rate than normally spoken English [7].

NATURAL LANGUAGE UNDERSTANDING (NLU):

After conversion of speech to text using ASR, the most important step is to understand the meaning of the text



closer to that of user [8]. The challenges with NLU are ambiguity and variability [8].

NATURAL LANGUAGE GENERATION (NLG):

It is process of converting the process into words. It is generation of a language that has some sense and similar to humans [9]. The NLG process is further categorized into two phases that is what to say and how to say [10].

TEXT-TO-SPEECH(TTS):

Text-to-Speech is also known as speech synthesis. It is the last step in the creation of voice assistants. This process involves two steps that are text synthesis and waveform synthesis. In text synthesis words are translated into speech and the next step that is waveform synthesis involves creation of desired sentence based on previous speech samples [5].

Our Desktop Voice Assistant (DVAbot) is distributed into four layers that are as follows:

- 1. Speech to text.
- 2. Text Analysing.
- 3. Interpret Commands.
- 4. Text to speech.

FIRST LAYER (SPEECH TO TEXT):

It is a piece of software that will convert audio to text.

- For this we used Speech_Recognition.
- It doesn't understand just anything you might say.

SECOND LAYER (TEXT ANALYSING):

- Converted text is just letters for computers.
- A piece of software that converts text to something that is understandable for computer.
- Computer understands commands, so it converts this text to computer command.
- Computer command consists of functions and the parameters of these functions.

THIRD LAYER (INTERPRET COMMANDS):

- In this level it isn't doing much.
- The mapped computer commands go to server through internet.
- Simultaneously, the speech is evaluated locally.

FORTH LAYER (TEXT TO SPEECH):

- It is done at desktop end.
- The desktop voice assistant converts to text to speech and delivers to the user.

III. RESULTS

GREETINGS

Initially it will ask you your name and then desktop voice assistant greet you with "hello, your name." And then greet you according to the current time like "Good Morning sir", "Good Afternoon sir" and "Good Evening sir" and then asks "What can I do for you?"

c:\users\u
DVAbot: What's your name, Human?
djetov
Stop.
You : Aditya
DVAbot: Hello, Aditya
DVAbot: Good Morning sir. I am DVAbot.
DVAbot: What can i do for you?
Speak
Stop.

Fig 3: Greeting to User according to time

OPEN ANY APPLICATION IN DESKTOP

You can open any application or program just by saying "Open application name" like "Open Microsoft Excel" like this you can ask DVAbot to open any application for you and then DVAbot replies by "Opening Microsoft Excel".

PERFORM CALCULATIONS

You can calculate various mathematical expressions by giving voice command to DVAbot.

For example: "Calculate 92 * 402" like this you can ask DVAbot to calculate mathematical expressions for you and then assistant replies by "The Answer is 36984".



Fig4: Calculation of Mathematical Expressions

SEARCH WEBPAGE ON GOOGLE

You can search any webpage or information just by saying "Search name".

For example: "Search Narendra Modi" like this you can ask DVAbot to open the webpage with this information for you.

GIVES UPDATED INFORMATION ABOUT CORONA

DVAbot can provide us all the stats regarding corona at real time. It can tell us the total number of cases, deaths and recovered cases till now and can also tell the number of cases and deaths in every country.

For Example: When you say "Corona Information" it replies with "What do you want to know about Corona?" then you can further ask "How many total cases in world?" or "How many total deaths in world?" or "How many total recovered cases in world?" or "How many total cases in India?" etc then it replies with the actual result.

"Engl "Search name".





Fig 5: Corona Information

ANSWER QUERIES

DVAbot can answer any query that is asked by the user by searching it on Google and then reading out some part of it.

For example: "What is Photosynthesis" like this you can ask DVAbot to answer this query and it replies with the answer of this query like "".



Fig 6: Answering Queries using Web Scraping

TELLS CURRENT DATE AND TIME

DVAbot can tell the current date and current time to the user.

For example: "Tell me the date today?" like this you can ask DVAbot to tell the current date and then DVAbot replies by "2020-06-04" and when you ask "What is the time now?" then it replies with the current time like "11:59 am".

DVAbot: What can i do for you?		
Speak		
Stop.		
You : tell me the date today		
DVAbot: 2020-06-04		
DVAbot: What can i do for you?		
Speak		
Stop.		
You : what is time now		
DVAbot: 11:59 am		

Fig 7: Current Time and Date

TELLS THE WEATHER OF THE CITY

When you speak "Weather Information" then it replies with "Tell city name which you want data about or exit" then after telling the city name; it further asks "What data do you want, average temperature, minimum temperature, maximum temperature or description" then it replies with the desired information.

VAbot: What can i do for you?
peak
top.
ou : weather information
WAbot: Tell city name which you want data about, or exit
peak
top.
ou : New York
VAbot: what data you want, average temperature, minimum temperature, maximum temperature or description.
peak
top.
ou : description
Whot: overcast clouds
WAbot: Tell city name which you want data about. or exit
peak
top.
ou : London
VAbot: what data you want, average temperature, minimum temperature, maximum temperature or description.
peak
top.
ou : minimum temperature
VAbot: minimum Temperature is=10.56degree celsius
WAbot: Tell city name which you want data about. or exit
peak
top.
ou : London
WAbot: what data you want, average temperature, minimum temperature, maximum temperature or description.
peak
top.
ou : maximum temperature
VAbot: Maximum Temperature is12.00degree celsius
WAbot: Tell city name which you want data about. or exit
peak
top.



SEND EMAIL

DVAbot can send mail for you. When the user say "I want to send email" then DVAbot asks for the password like "Tell your email password" after providing the password DVAbot asks "To whom you want to send mail, please type email id for security reasons" then the user have to write the mail id further the assistant asks "Send email to this mail yes/no" after getting the response as "Yes" it asks for the subject followed by the body of mail after that DVAbot replies with "Email sent".



Fig 9: Sending Email



email is sent through Python program of Desktop voice assistant

Fig 10: Email Received



LEAVE

When you want to terminate the program say something like "Bye", "Exit" or "Sleep" then it replies with "Okay Bye, your name. Have a nice day!



Fig 11: Leave

IV. RESULT ANALYSIS

- DVAbot can process voice command in less than 5 seconds.
- Private information like email password is kept private.

V. CONCLUSION

In order to achieve smart assistant we need to create Voice Assistants using Natural Language Processing along with Artificial Intelligence.

This program is designed to control the system like opening all desktop applications, to get information of all almost everything available over Internet using Web Scraping, performing mathematical calculations, sending mails, getting weather information as well as all stats about corona cases of all the affected countries, Our DVAbot can be used by the deaf and dumb people and people with hectic schedule.

FUTURE DIRECTIONS

- Will update the voice for more clear, human voice n Engi and language translation.
- Will work on other languages as well apart from English.
- Will introduce an user interface, with a "more natural and expressive" voice.

REFERENCES

- Voice Assistant Application for a college Website, Deny Nancy, Anushria Sai, M.Ganga, R.S. Abisree, Sumithra Praveen.
- [2] *Study of Voice Controlled Personal Assistant Device*, Abhay Dekate, Chaitanya Kulkarni, Rohan Killedar
- [3] Personal Assistant with Voice Recognition Intelligence, Dr. Kshama V. KulhalliVice-Principal & HOD ITD.Y. Patil College of Engineering and Technology
- [4] Intelligent Android Voice Assistant- A Future Requisite, Ms.Ayushi, M. Krina, Ms.Yesha, CSE

Department, Madhuben and Bhanubhai Patel Women's Institute of Engineering

- [5] Jurafsky, D., & Martin, J. H. (2009). Chapter 13: Syntactic Parsing. Speech and Language Processing: An Introduction to Natural Language Processing. *Computational Linguistics, and Speech Recognition* (2nd ed.). Pearson–Prentice Hall.
- [6] Sen, S., Dutta, A., & Dey, N. (2019). Speech Processing and Recognition System. In Audio Processing and Speech Recognition (pp. 13-43). Springer, Singapore.
- [7] Tomokiyo, L. M. (2001). Recognizing non-native speech: characterizing and adapting to non-native usage in LVCSR. *Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA*.
- [8] Khashabi, D. (2019). Reasoning-Driven Question-Answering for Natural Language Understanding. arXiv preprint arXiv:1908.04926.
- [9] Santhanam, S., & Shaikh, S. (2019). A Survey of Natural Language Generation Techniques with a Focus on Dialogue Systems-Past, Present and Future Directions. arXiv preprint arXiv:1906.00500.
- [10] Daniel Jurafsky and James H. Martin Speech and language processing, 3rd edition draft, 2018.