VOICE ACTIVATED DESKTOP ASSISTANT USING PYTHON

Dhiraj Pratap Singh, Deepika Sherawat, Sonia

School of Computer Science & Engineering, Galgotias University

Greater Noida-203201

ABSTRACT

The advancement in technology over time has been unmeasurable. From the first digital computer built by Eniac having a clock speed of 100KHz to Summit developed by the US Department of Energy has a performance of 148.6 petaFlops, we have come a long way in technological advancement. In such an era of advancement if people are still struggling to interact with their machine using various input devices then its not worth it. For this reason, many voice assistants were developed and are still being improved for better performance and efficiency.

The main task of a voice assistant is to minimize the use of input devices like keyboard, mouse, touch pens, etc. This will reduce both the hardware cost and space taken by it.

KEYWORDS- Desktop Assistant, Python, Machine Learning, Text to Speech, Speech to Text, Language Processing, Voice Recognition, Artificial Intelligence, Internet Of Things(IoT), Virtual Assistant.

I.   INTRODUCTION

In the 21st century, human interaction is being replaced by automation very quickly. One of the main reasons for this change is performance. There's a drastic change in technology rather than advancement. In today's world, we train our machine's to do their tasks by themselves or to think like humans using technologies like Machine Learning, Neural Networks, etc. Now in the current era, we can talk to our machines with the help of virtual assistants. There are companies like Google, Apple, Microsoft, etc with virtual assistants like Google Now, Siri, Cortana, etc. which helps their users to control their machine by just giving input in the form of voice.

These types of virtual assistants are very useful for old age, blind & physically challenged people, children, etc. by making sure that the interaction with the machine is not a challenge anymore for people. Even blind people who couldn't see the machine can interact with it using their voice only[1].

Some of the basic tasks that are supported by most of the virtual assistants are :

- Checking weather updates
- Sending and checking mails
- Search on Wikipedia
- Make and receive calls
- Stream music
- Open applications
- Text messages etc.

The voice assistant we have developed is a desktop-based built using python modules and libraries. This assistant is just a basic version that could perform all the basic tasks which have been mentioned above but current technology is although good in it is still to be merged with Machine Learning and Internet Of Things(IoT) for better enhancements.

The understanding and executing commands are still to reach a new level like the virtual assistant of the iron man named Jarvis. This is although fictional yet this is what that can be achieved using virtual assistants. All you need to do is give a command to the assistant and the rest will be performed by the assistant.

With the help of voice-activated virtual assistants, there will be no need to write long codes to perform a task, the system will do so for us. The machine will work in three modes- supervised, unsupervised or reinforcement learning depending upon the usage for which the assistant is developed. This is all possible with the help of machine learning.

Now what the IoT does is it will help the assistant to interact with the neighboring smart devices and will act as a single interface that will control everything in the surrounding. With the involvement of IoT, it will be possible to control other smart devices that will in-turn interact among themselves over the internet.

So with a capable virtual assistant, we will be able to control many things around us single-handedly with only one platform.

Figure 1: Timeline Of Main Voice Assistants

II. LITERATURE SURVEY

This field of virtual assistants having speech recognition has seen some major advancements or innovations. This is mainly because of its demand in devices like smartwatches or fitness bands, speakers, bluetooth earphones, mobile phones, laptop or desktop, television, etc. Almost all the digital devices which are coming nowadays are coming with voice assistants which help to control the device with speech recognition only. A new set of techniques is being developed constantly to improve the performance of voice automated search[2].

As the amount of data is increasing exponentially now known as Big Data the best way to improve the results of virtual assistants is to incorporate our assistants with machine learning and train our devices according to their uses. Other major techniques that are equally important are Artificial Intelligence, Internet Of Things, Big Data access and management, etc.

With the use of voice assistants, we can automate the task easily, just give the input to the machine in the speech form and all the tasks will be done by it from converting your speech into text form to taking out keywords from that text and execute the query to give results to the user.

Machine Learning is just a subset of Artificial Intelligence. This has been one of the most helpful advancements in technology. Before AI we were the ones who were upgrading technology to do a task but now the machine is itself able to counter new tasks and solve it without need to involve the humans to evolve it.

This has been helpful in day-to-day lifestyle. From mobile phones to personal desktops to mechanical industries these assistants are in very much demand for automating tasks and increasing efficiency[3].

Figure 2: Voice Controlled Appliances Affecting Our Daily Life

III. SYSTEM ARCHITECTURE


```
import speech_recognition as sr
import os
import sys
import re
import webbrowser
import smtplib
import requests
```

Fig.3: Modules Imported

- **SPEECH RECOGNITION**

The speech recognition module used the program is Google's Speech Recognition API which is imported in python using the command "import speech_recognition as sr".This module is used to recognize the voice which is given as input by the user.

This is a free API that is provided and supported by Google. This is a very light API that helps in reducing the size of our application[4].

- **TTS & STT**

The voice which is given as input is first converted to text using the speech recognition module. The text is then processed to give the result of the query given by the user. The final step is the conversion of the result of the processed query to speech which is the final output. The most time consuming among the two is STT because the system first has to listen to the user and different users have different, some are easy to understand while some are not easily audible. This is the step upon which our total execution time depends. Once the speech is converted to text executing commands and giving the results back to the user is not a time-consuming step[5].

- **IMPORTED MODULES**

A. PYTTSX3

The pyttsx3 is an offline module that is used for text to speech conversion in Python and it is supported by both Python 2 & 3. The run and wait functionality is also in this module only. It

determines how much time the system will wait for another input or in other words the time interval between inputs.

This is a free module available in the python community which can be installed using the pip command just like other modules.

B. DATETIME

The DateTime module is imported to support the functionality of the date and time. For example, the user wants to know the current date and time or the user wants to schedule a task at a certain time. In short this module supports classes to manipulate date and time and perform operations according to it only. This is an essential module, especially in tasks where we want to keep a track of time. This module is very small in size and helps to control the size of our program. If the modules are too large or heavy then the system will lag and give slow responses.

C. WEBBROWSER

This module allows the system to display web-based information to users. For example, the user wants to open any website and he gives input as "Open Google". The input is processed using the web browser module and the user gets a browser with google opened in it. The browser which will be used is the default set web browser.

D. WIKIPEDIA

Wikipedia is a library in python which it possible for the virtual assistant to process the queries regarding Wikipedia and display the results to users. This is an online library and needs an internet connection to fetch the results.

The no. of lines that the user wants to get as a result can be set manually.

E. OS MODULE

OS Module provides an operating system dependent functionalities. If we want to perform operations on files like reading, writing, or manipulate paths, all these types of functionalities are available in an OS module. All the operations available raise an error "OSError" in case of any error like invalid names, paths, or arguments which may be incorrect or correct but just no accepted by the operating system.

F. SMTPLIB

Python has this module for in the standard library for working with emails & email servers. The SMTPLIB defines an object known as "SMTP client session object" which is used to send mails by the user. There are 3 steps involved - initialize, sendmail(), quit. When the optional parameters which are host and port, are provided connect method is called with these arguments during the first step which is initialization[6].

● DESIGN

The overall design of our system consists of the following phases:

(a) Taking input from the user in the form of voice.

(b) Converting the speech into text to be processed by the assistant.

(c) The converted text is now processed to get the required results.

(d) The text contains one or two keywords that determine what query is to be executed. If the keyword doesn't match any of the querries in the code then the assistant asks the user to speak again.

(e) The result which is in the form of text is converted to speech again to give results to the user.
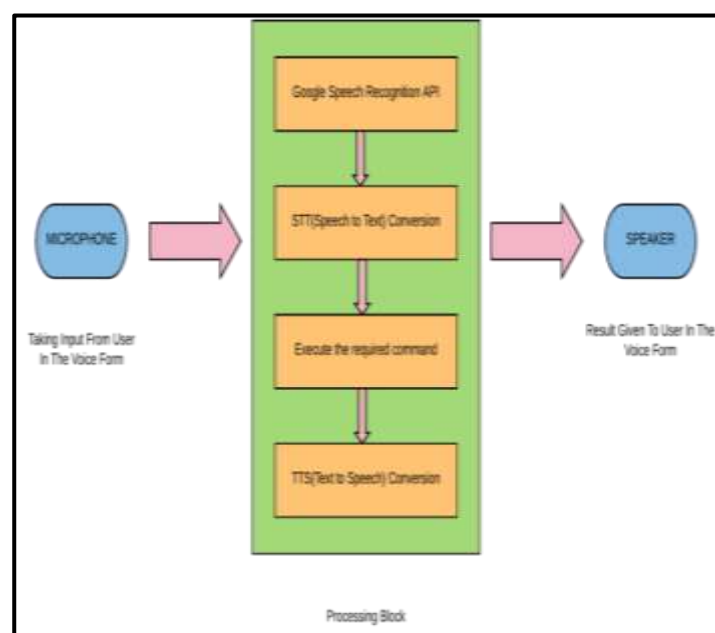

Figure 3: Processing Block Of STT To TTS

● PROPOSED SYSTEM

The proposed system will have the following functionality:

(a) The system will keep listening for commands and the time for listening is variable which can be changed according to user requirements.

(b) If the system is not able to gather information from the user input it will keep asking again to repeat till the desired no. of times.

(c) The system can have both male and female voices according to user requirements.

(d) Features supported in the current version include playing music, emails, texts, search on Wikipedia, or opening system installed applications, opening anything on the web browser, etc.

(e)The system will keep listening for commands and the time for listening is variable which can be changed according to user requirements.

(f) If the system is not able to gather information from the user input it will keep asking again to repeat till the desired no. of times.

(g) The system can have both male and female voices according to user requirements[7].

FUTURE SCOPE

The virtual assistants which are currently available are fast and responsive but we still have to go a long way. The understanding and reliability of the current systems need to be improved a lot. The assistants available nowadays are still not reliable in critical scenarios. The future of these assistants will have the virtual assistants incorporated with Artificial Intelligence which includes Machine Learning, Neural Networks, etc. and IoT. With the incorporation of these technologies, we will be able to achieve new heights. What the virtual assistants can achieve is much beyond what we have achieved till now. Most of us have seen Jarvis, that is a virtual assistant developed by iron man which is although fictional but this has set new standards of what we can achieve using voice-activated virtual assistants[8].
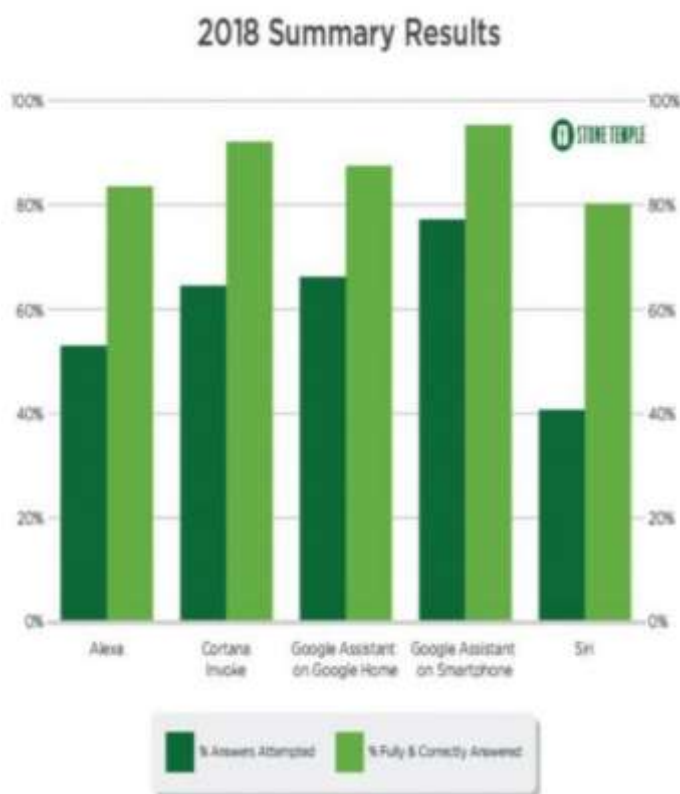

Figure 4: Accuracy Of Results Over Time

IV. CONCLUSION

In this paper we have discussed a Voice Activated Personal Assistant developed using python. This assistant currently works online and performs basic tasks like weather updates, stream music, search Wikipedia, open desktop applications, etc. The functionality of the current system is limited to working online only. The upcoming updates of this assistant will have machine learning incorporated in the system which will result in better suggestions with IoT to control the nearby devices similar to what Amazon's Alexa does.

The usage of the assistant will get offline also for features that don't require an internet connection.

V. REFERENCES

[1] D O'SHAUGHNESSY, SENIOR MEMBER, IEEE, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis" proceedings of THE IEEE, VOL. 91, NO. 9, SEPTEMBER 2003

[2] Kei Hashimoto, Junichi Yamagishi, William Byrne, Simon King, Keiichi Tokuda, "An analysis of machine translation and speech synthesis in speech-to-speech translation system" proceedings of 5108978-1-4577-0539- 7/11/$26.00 ©2011 IEEE.

[3] Nil Goksel-Canbek Mehmet Emin Mutlu, "On the track of Artificial Intelligence: Learning with Intelligent Personal Assistant" International Journal of Human Sciences.

[4] H. Phatnani, Mr. J. Patra and Ankit Sharma' "CHATBOT ASSISTING: SIRI" Proceedings of BITCON-2015 Innovations For National Development National Conference on Research and Development in Computer Science and Applications, E-ISSN2249–8974.

[5] Sutar Shekhar, P. Sameer, Kamad Neha, Prof. Devkate Laxman, " An Intelligent Voice Assistant Using Android Platform", March 2015, IJARCSMS, ISSN: 232 7782

[6] VINAY SAGAR, KUSUMA SM, "Home Automation Using Internet of Things", June-2015, IRJET, e-ISSN: 2395 -0056.

[7] "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010.

[8] Rishabh Shah, Siddhant Lahoti, Prof. Lavanya. K, "An Intelligent Chatbot using Natural Language Processing". International Journal of Engineering Research, Vol. 6, pp. 281-286, 1 May 2017.