

Desktop Voice Guide using Python and Artificial Intelligence

Ananya Gupta* and Sagun Gupta

Department of Computer Science (IT), Axis Institute of Technology and Management/AITM, India

Corresponding Author*

Ananya Gupta,

Department of Computer Science (IT), Axis Institute of Technology and Management/AITM,

India

Email: guptaananya468@gmail.com

Copyright: ©2022 Kalhor B. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 20-May-2022, Manuscript No. IJIRSET-22-65227; **Editor assigned:** 27-May-2022, PreQC No. IJIRSET-22-65227(PQ); **Reviewed:** 30-May-2022, QC No. IJIRSET-22-65227(Q); **Revised:** 04-June-2022, Manuscript No. IJIRSET-22-65227(R); **Published:** 10-June-2022, DOI: 10.35248/ijirset.22.3(5).63-69.

Abstract

In today's world the technological advancement is increasing day by day. Earlier only there was a computer system in which we can do only limited tasks. But now machine learning, artificial intelligence, deep learning, and few more technologies have made computer systems so advance that we can perform any type of task. In such era of advancement if people are still struggling to interact using various input devices, then it's not worthwhile. The foremost task of voice assistant is to attenuate the utilization of input devices like keyboard, mouse etc. It will also reduce the hardware space and cost. In recent future all the electronic devices are going to be worked by utilizing the remote helper which is certainly not hard to urge to yet it needs weakness. This structure affirmed the clients to urge to the framework by the voice orders. User can request to the assistant that anything are often done by the framework, for instance Music, Open Specified Application, Open Tabs, Open Websites then so forth. Voice associates are programming specialists which will decipher human discourse and react through orchestrated voices. Clients can pose their associate's inquiries, control home gadgets and media playback by means of voice, and oversee other essential errands, for instance, email, daily agendas, and schedules with verbal orders.

Keywords • Artificial Intelligence • Desktop Assistant • Python, Text to Speech • Virtual Assistant • Voice Recognition • Language Processing

Introduction

Voice assistants are defined as the software agents who interpret or convert human speech and it retorts through coordinated voices, i.e. Siri (Apple Inc.), Alexa (Amazon), Cortana (Microsoft), so forth these are most popular voice assistants. In this era of technology everything that human being can do are being supplanted by machines. One of the main reasons is alteration in performance. In today's world we train our machine to think like humans and do their job by themselves.

Hence, there came a concept of virtual assistant. A virtual assistant is a digital assistant that uses voice recognition features and language processing algorithms to recognize voice commands of user and perform pertinent tasks as requested by the user.

Based on specific commands given by the user a virtual assistant is capable of sifting out the ambient noise and return relevant information. Virtual Assistant are completely software based but nowadays they are integrated in diverse devices and also some of the assistants are designed explicitly for single devices like Alexa.

These sorts of virtual assistants are very useful for old age, blind & physically challenged people, children, etc. by making sure that the dealing 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].

Here are some of the tasks that can be done with the help of voice assistant:

- Reading Newspaper
- Getting updates of mail
- Search on web
- Play a music or video
- Setting a reminder and alarm
- Run any program or application
- Getting weather updates

So, with the help of virtual assistant, we will be capable to control many things around us single handedly on one platform.

Literature Review

This arena of virtual assistants having speech recognition has seen some major advancements or innovations. This is mostly because of its need in devices like smartwatches or fitness bands, speakers, bluetooth earphones, mobile phones, laptop or desktop, television, etc.

Approximately 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 advanced constantly to improve the performance of voice automated search (2).

J. B. Allen et al described about the Language that is the greatest significant means of communication and speech is its major interface. The interface for human to machine, speech signal was transformed into analog and digital wave shape as a machine understood. A technology extremely utilized and has limitless applications. Speech technologies allow machines to react appropriately and consistently to human speeches and offers valuable and appreciated services.

The research provides a summary of the speech identification procedure, its basic model, and its application, techniques and also explains reasonable research of several techniques that are utilized for speech recognition system.

This has been beneficial 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 efficacy.

Bassam A, Raja N. et al, written about report and speech which is most significant. In the communication between human and machine procedure was done through analog signal which is converted by speech signal to digital wave. This technology is vastly utilized, it has limitless uses and permit machines to reply appropriately and regularly to user voices, also offers useful and appreciated facilities. Speech Recognition System (SRS) is rising gradually and has indefinite applications. The research has revealed the summary of the procedure; it is a simple model.

Motivation

The main motive is to improve the efficiency of the personal assistant application by giving the agent the ability to learn. Because the agent probably performs a large number of repetitive activities, previous experiences can be applied to similar future scenarios. We recommend a learn by doing agent that will aid the user in completing tasks. The job at hand is to manage the user's desktop or laptop profile. Every user has a daily routine that requires them to place their desktop or laptop at a distance for a period of time.

Problem Formulation

This part describes the description about the problem formulation. As we know each human have their own features and every developer applies his own method and approaches for development of a product. One assistant can create speech more qualitatively, another can more accurately and without additional explanations and corrections perform tasks, others are able to perform a narrower range of tasks, but most accurately and as the user wants.

Hence, there is no such assistant that can perform all the work and tasks equally. The set of features that an assistant has depends on the area on which developer paid more attention. Since all systems are grounded on machine learning and use for their creation huge amounts of data collected from various sources and then trained on them, an important role is played by the source of this data. Despite the various approaches to learn different algorithms, the principle of developing voice assistant remains the same. The technologies that are used to build a voice assistant that can communicate with the humans are speech recognition, Teach-To-Speech, voice biometrics, dialog manager, natural language understanding and named entity recognition [3].

Proposed Approach

- The proposed system will have the following features:
- The system will keep listening for commands and the time for listening is variable which can be changed according to user need.
- If the system is not able to collect information from the user input it will keep asking again to repeat till the desired no. of times.
- The system can have both male and female voices fitting to user requirements.
- Features reinforced in the current version include playing music, emails, texts, search on Wikipedia or opening system installed applications, opening anything on the web browser, etc.

System Architecture

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

Figure 1. Modules

Speech to text

In this module or layer a speech is to be converted into the text, which can be comprehensible by the system through the installed libraries (Figure 1).

Speech Recognition

The speech recognition module made use of the Google Speech Recognition API, which can be imported into Python with the command "import speech_recognition as sr." This module is used to recognize the voice inquiry that the user has provided as input. This is often a Google-provided and supported API that is available for free of charges. This is a small API that aids in the compression of our application.

Date-time

To sustain date and time formats, the date and time module was imported. The Datetime module comprises classes for manipulating dates and times. These classes offer a spread of capabilities for working with dates, times, and time intervals. In Python, date and datetime are objects. The user may, for instance, want to know the current date and time or schedule a task for a selected time. In short this module supports classes to control date and time and perform operations consistent with it only.

Pytsx3

The pywhatkit module is a Python module. It is in charge of playing everything you want to search for on YouTube. For instance, if a user wishes to play a song from YouTube, they can say "play song ms dhoni" and the song will be played immediately.

Web Browser

This module allows the system to show information from the internet.

It is a Python built-in module that offers every assistance to the user in obtaining information from the internet. For instance, if a user says "open you tube," the question is processed through the web browser module, and you tube is opened.

OS

The OS module in Python offers functions for communicate with the operating system. OS falls under Python's standard utility modules. This module provides a portable way of using OS-dependent functionality.

TTS & STT

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

Wikipedia

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

Smtplib

Python has this module for within the standard library for working with emails & email servers. The SMTPLIB defines an object called as "SMTP client session object" which is used to deliver mails by the user. There are 3 steps - initialize, sendmail(), quit. When the elective parameters which are host and port, are provided connect method is named with these arguments during the primary step which is initialization [4].

Steps of proposed system

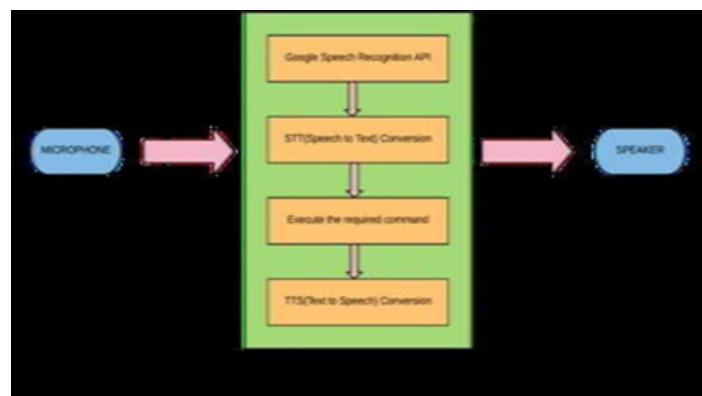


Figure 2. Steps of proposed system

The overall design of our system comprises of the following phases: (Figure 2)

- Receiving input from the user in the form of voice.
- Converting the speech into text to be handled by the assistant.
- The converted text is now processed to get the needed results.
- The text contains one or two keywords that conclude What query is to be executed. If the keyword doesn't match any of the queries in the code then the assistant asks the user to speak again.
- The result which is in the form of text is converted to speech again to provide results to the user.

Process of proposed system

- The proposed system will have the following functionality:
- The system will keep listening for commands and the Time for listening is not constant which can be changed according to user requirements.
- If the system is not able to collect information from the user input it will keep asking again and again to repeat till the desired no. of times.

- The system can have both male and female voices according to user need.
- Features offered in the current version include playing music, emails, texts, search on Wikipedia, or opening system installed applications, opening anything on the web browser, etc (Figure 3) [5].

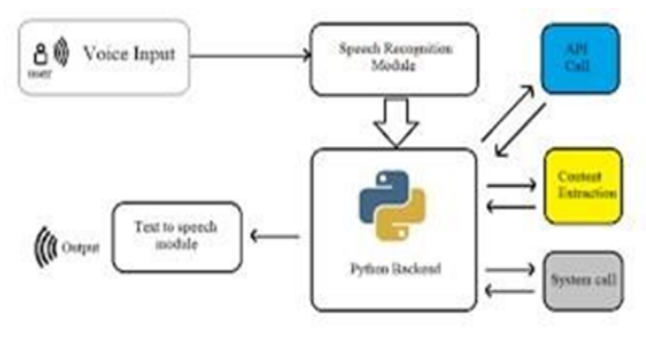


Figure 3. Process of Proposed System

Future Scope

Virtual assistants are now available and are quick and responsive, but there is still an extended way to go. The current system understanding and reliability got to be greatly enhanced. In critical situations, the helpers available now are still unreliable. Virtual assistants will be combined with Artificial Intelligence, such as Machine Learning, Neural Networks, and IoT, in the future of these assistants. We will be able to reach new heights by integrating these technology. What virtual assistants can accomplish is way beyond what we have accomplished so far. Although Jarvis, a voice-activated virtual assistant created by Iron Man, is fictional, it has set new expectations for what we will achieve with voice-activated virtual assistants (Figure 4).

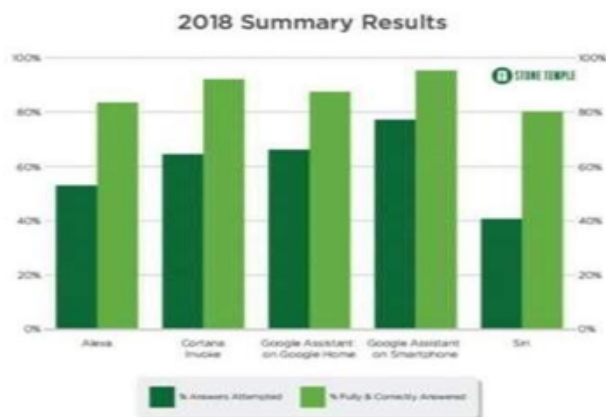


Figure 4. 2018 Summary Results

Conclusion

In this paper we have deliberated 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 existing system is limited to working online only. The upcoming informs 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 practise of the assistant will get offline also for features that don't require an internet connection. Desktop Assistant has various functions as like mobile phone like managing various applications on the voice commands. It helps to access the system hands-free and to get rid of typing chaos. This technology progression is taking the world to the next level. If we compare, two decades ago our words may very carried as we could have imagined.

References

1. Radha and Vimala C. "A review on speech recognition challenges and approaches" doaj org 2.1(2012):1-7.
2. Shaughnessy D O. "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis" proc. THE IEEE 91.9(2003):1272-1305.
3. Nguyen P., et al "Speech recognition with flat direct models" J Sel Top Signal Process 4.6(2010):994-1006.
4. Canbek N G, and Mehmet E M. "On the track of artificial intelligence: Learning with intelligent personal assistants" J Hum Sci 13.1(2016):592-601.
5. Nguyen P., et al. "Speech recognition with flat direct models" J Sel Top Signal Process 4.6(2010):994-1006.