



Voice Assistant and its OS Integration

Saurabh Trivedi¹, Himanshu Shishodia², Raman Prasad Sharma³

^{1,2,3} School of Computing Science and Engineering, Galgotias University, Greater Noida (U.P.)

DOI: <https://doi.org/10.55248/gengpi.5.0124.0344>

ABSTRACT—

Voice Assistant for Accessing Real-World Applications is aimed at developing personal assistants for real-world applications. Utilizing vocals commands to interface with real world virtual assistant like Cortana, Siri, and Alexa. Its purpose is to offer an interface that is easy for users to navigate to perform various tasks by employing well-defined and specific commands. [1][2]Users can engage with the assistant via interaction verbal instructions. Due to the swifter and more effective nature of voice-driven tasks than completing tasks manually. In its role as a personal assistant, Jarvis employs various technologies such as ASR (Automatic Speech Recognition) and TTS (Text to Speech) to help people do things in their daily lives, such as common human conversations, internet searches, video searches, and downloads. Assist users with various activities. Pictures, definitions, opposites, dispatching emails, or overseeing tasks. Whats App messages, creating documents, launching applications present in the system, etc. The advent of voice-based technologies has dramatically changed the way humans interact with computers, resulting in personal assistants designed for real-world applications. Built with an intuitive user interface that explores the concept of commands, Jarvis aims to simplify tasks through voice commands. In this era of technological advancement, the functionality and speed offered by voice-based communication provides a compelling case for integrating such personal assistants into our daily lives. Jarvis uses state-of-the-art technologies, primarily automatic speech recognition (ASR) and transmission to speech (TTS), to provide a seamless voice interface that enables the ASR system to understand spoken language and converts it to simplify the text, while TTS ensures that the assistant responds in natural and understandable language. Jarvis' design is inspired by the success of virtual assistants, which focuses on user-centered principles to create simple and efficient communication. By allowing individuals to engage with technology through voice commands, Jarvis bridges the gap between humans and machines, creating a hands-free and easy-to-use experience Jarvis is more than just a voice-activated device; It is a high-performance personal assistant designed to perform a variety of tasks encountered in daily life. Users can engage in normal human-like conversations, have Jarvis search and play videos, take pictures, and even understand complex language tasks that provide meaning and contrast The aide extends his authority to important communications roles such as Sen

Introduction

The Artificial Intelligence (AI) technologies are one of the rapidly growing technologies. The fact that these technologies are highly growing but some people never tend to use them at all. Since a long time, technologies adoption had been examined, and there are multiple models that describe it. The technologies are mostly new but they use older technologies for finding new applications. The emerging technologies are perceived to be capable of changing the ways of solving problems.

We are constructing a model involving the method of interaction in the project, which we hope to better represent the adoption of AI based technologies. The model we are creating will conduct basic functions on the OS layer with the help of various modules, and it will fulfil the basic daily necessary functions. The model will be providing a good understanding of the intelligent assistant that can interpret the voice commands given by the user. The model can readily understand and respond to the commands given by the user using voice commands.

The voice based technology that works by speech recognition, which takes the voice input and then convert the speech into the text. There are both types of speech recognition system, offline and online. The online recognition system when compared to the offline recognition system tend to process the voice commands fast and give the best response. The Speech Recognition technology also referred to as Automatic Speech Recognition, where it takes the user's voice and automatically converts it to text. The speech recognition works by breaking the sound commands into words by punctuality marked sentences. The recognition models also used the pronunciation to convert the speech into text.

In the present period, there had been very spontaneous improvements in the way that users do interact with the Virtual Assistants. We use the assistants for several works, such as house automation, using google by giving voice commands. This is the new way of interacting with the upcoming technologies. Now days, several handy robots are being made for personal use, they are personalized by giving specific information about the user, so they respond to the commands as per the user's choice.

We will be using the Voice Assistant technology, which would consist of basic functions that include mail exchange, music playing, weather information, Wikipedia search, the initial project aims to build a voice assistant. Voice assistant are not to be confused with the virtual assistants. They both are used for different works. For example, the voice assistants can be used for HR related tasks, voice based search and for teaching kids.

The project will provide a good understanding of an intelligent voice assistant that can understand the commands given by the user. The assistant can respond to the voice commands given by the user with the help of speech recognition system. The assistant takes care of the user's requests and make their work easier.

The voice assistant listens to the user's command and it will announce "done listening" after the end of voice command and will complete the task assigned by the user. We used the gTTS package to make the voice assistant speak. The gTTS analyses the command given by the user via the microphone and converts into text. When the user says, "Mail to random@gmail.com", the assistant responds by asking the message content and after the getting the message response, the assistant sends the mail to the specific person. When the user says "Open YouTube", the assistant will open the YouTube web app on the installed browser on the system.

When the voice assistant detects a pause in the voice command, assistant gets to know that the user has completed their command. The voice assistants are beneficial in a variety of fields, including education, home automation and daily life works. They are also great for the people who do not like to use screen tech products such as phone and laptop, they can interact with the voice assistant with the voice commands only, these people can enjoy these technological advancements without any problem.

Literature Survey

In today's generation voice assistant is playing a crucial role in daily life. This literature survey explores the evolution, and increment in voice assistant technology, in both academic and industry perspectives. The survey aims to provide a comprehensive understanding of voice assistant. Considering some following key facts:

- Over half of individuals (52%) prefer to keep their voice assistant speakers in the living room, while 25% place them in the bedroom, and 22% in the kitchen this is the data provided by google.
- During the 2017 holiday shopping season, one in every four shopper used voice assistant software this why we say voice assistant is playing a crucial role
- According to data provided by google 41% of user that interact with voice assistant feel like they are conversing with a friend.
- According to Geo Marketing 65% of consumers owning Amazon Echo , Google Home, or similar devices can't imagine life without them.
- The voice assistant market was introduced to be worth \$601 million by 2019 provided by Technavio's analysis.

User Interaction and Experience:

In this research we focus on the understanding of the user interactions with the assistant. [3]Famous works that associate with natural language understanding, contextual awareness, and the psychological aspects of human-computer interaction through voice.

Voice Assistant System Components:

- Voice assistant takes action, when a programmed keyword is detected and then responding with a greeting or executing a command.
- The System components include a microphone, voice recognition, Voice Commands, and main virtual assistant program.

Development Details :

The system displays real-time information, including date, time, weather, and news. It provides facility such as playing music and searching topic on internet , enhancing user experience and saving time.This project aims to create an advanced assistant capable of recognizing and responding to user queries and commands.

In conclusion, this project provides user preferences in voice assistant placement, shopping trend, and the increasing reliance on voice assistants. The system boasts efficient interaction through voice commands, offering a comprehensive and user-friendly experience.

Purpose, Scope and Applicability

Purpose

Virtual assistants are designed to facilitate vocal engagement, encompassing tasks such as playing music, creating task lists, setting alarms, streaming podcasts, playing audio books, and providing real-time information on weather, traffic, sports, and news. These digital companions empower users to issue natural language voice commands, seamlessly operating the associated devices and applications.

In the realm of consumer behavior, millennials are demonstrating heightened awareness and an increased level of comfort. In the dynamic digital landscape, where speed, efficiency, and convenience are continually prioritized, there is a discernible trend toward reduced reliance on traditional screen interactions.

Scope

The evolution of voice assistants is steering us towards more personalized interactions. As these AI-driven entities enhance their capacity to differentiate between voices, the user experience becomes increasingly tailored. However, the responsibility for navigating the intricacies of voice technology extends beyond developers. Brands, too, must comprehend the functionalities of each device and ensure seamless integration, aligning with their brand identity.

Maintaining a consistent user experience is crucial as the complexity of voice interactions grows. The absence of a visual interface with voice assistants underscores the challenge users face in visually perceiving or physically interacting with the interface.

Applicability

In practical terms, the widespread adoption of artificial intelligence in everyday life is driving the surge in voice-centric interactions. The proliferation of IoT devices, including smart thermostats and speakers, enhances the utility of voice assistants in the connected lives of users. Smart speakers, in particular, stand out as a primary conduit for voice interactions. Industry experts predict that, within the next five years, nearly every application will integrate some form of voice technology.

The use of virtual assistants extends beyond individual convenience; it also plays a crucial role in enhancing the Internet of Things (IoT). Looking two decades into the future, it is envisioned that major players like Microsoft and their counterparts will offer personal digital assistants providing services typically associated with full-time employees, a luxury once reserved for the elite.

Conclusion

In conclusion, the virtual assistant landscape is rapidly evolving, offering a plethora of capabilities ranging from routine tasks to personalized interactions. The influence of millennials in driving this shift is evident, as is the growing reliance on voice interactions in the broader context of our digital lives. The intricate balance between device integration, user experience, and the absence of a visual interface necessitates a holistic approach from both developers and brands. As voice technology continues to mature, its ubiquity in various applications and the seamless integration into our daily routines are poised to define the future of human-computer interactions.

Methodology

Every project depends on the system design, its performance and its execution. To maintain this project you need to have the system with features having inbuilt or an external microphone, which is used to take voice command.

The below diagram shows the system architecture

- 1) Command Modules i.e. Python script takes the voice as input.
- 2) Then, on the basis of the voice command, the specific task will be performed.
- 3) If the given voice command is not present, then the assistant will give voice feedback saying command cannot be fulfilled.

Hardware Requirements

- Laptop or Desktop
- Internal or External Microphone

Software Requirements

- Python
- OS
- Python pip packages

Modules

1) SPEECH RECOGNITION

Speech Recognition is one of the most important features in various applications such as home automation, voice-based searching. The Speech Recognition is used by Python to recognize and understand the voice command by the user and convert the command into text with the use of the microphone.

2) PYTTSX3

pyttsx3 library in python is used to convert text to speech. This library also works finely in offline mode with Python. The application invokes the pyttsx3.init() factory function to get a reference to the pyttsx3 engine instance. The library provides both male and female voice.

3)OS

OS module of Python provides the seamless facility of establishing the engagement between the user and the operating system. The module offers various useful function that are used to perform OS based tasks and get information related to the OS.

4)SMTPLIB

Simple Mail Transfer Protocol(SMTP) is a protocol, which sending e-mails between the mail servers. Python provides the SMTPLib module, which defines a SMTP client session object that is used to send email to any machine over the Internet. If you are not running the SMTP server on the local machine, SMTPLib client can be used to communicate with a remote SMTP server.

5)WEB BROWSER

Web Browser is a Python module, which is used to seamlessly control the web browser. The module provides a high level interface that displays web based documents to the users. It takes the URL as the argument with optional parameters such as -n, which opens the URL in the new browser window and -t, which opens URL in new browser tab.

6)WIKIPEDIA

Wikipedia is a module provided by Python, this module allows to get and parse the information from Wikipedia. This module is known as a little scrapper that can scrap only a limited amount of data.

7)DATE TIME

date time module is a module that can be imported into python projects to work with the date as well as time. It supplies classes to work with date and time, so on manipulating them, the objects of the class are manipulated and not the string or the timestamps.

Conclusion

The objective of this Project Voice Official assistant, with a focus on interacting with applications in the real world , primarily revolves around creating an assistant designed to recognize and respond to various queries and commands derived from user input, the system in question is constructed using Python programming. Diverse libraries were utilized, encompassing Speech Recognition, Text-to-Speech conversion, and Short Mail Transferring Protocols (SMTP). It provides information regarding the weather, News, it can play music, it can search for topics on Wikipedia, Display the current date and time. User can collect information through this application. It minimizes both human resources and time expenditure.

There's no requirement to pose inquiries in an overly stringent or narrowly defined manner. Users should be familiar with the fundamental principles of the English language. The objective is to offer individuals a fast and straightforward means to obtain answers to their inquiries. It can make the human daily activities go in a simpler, faster and less time consumption way. It is very Useful and efficient upon making in doing the given tasks.

Working on this Project has helped me to understand what is my area of interest. I have gained an immense amount of technical knowledge from this experience and I plan to continue it for my future career. I believe that my commitment as a Software Developer won't solely enhance my career path however additionally I have to learn new technologies to improve my ability to create changes on my career path.

To extend, while working on the Following project I have learned python and its different modules that could be used for integration of different feature in it for machine learning. I used python and a few major libraries for project development and while working with it, I found it easily learnable, usable, and there is a bright scope for me to choose to Machine learning as a future path. Also, I

set myself some achievement list before the beginning of my project for my improvement and I am pleased to inform that my achievement list was achieved precisely.

Conversely, I've acquired the ability to adapt and collaborate within both development teams and professional settings. As a developer, I gain so much confidence dramatically due to my internship. I am prepared to collaborate within a professional setting, enthusiastic about contributing to the advancement and prosperity of the organization through the application of my acquired skills.

Acknowledgment

The authors are thankful to Dr. T. Poongodi Department of Computing Science And Engineering, Galgotias University for his valuable suggestions and guidance while preparing this paper.

References

Yuqi Haung, "Research on the Development of Voice Assistants in the Era of Artificial Intelligence", SHS Web of Conferences 155, January 2023.

Sakshi R Jain, and Prof Feon Jason “Personal Desktop Voice Assistant”, International Journal of Advanced Research in computer and communication Engineering, Vol. 12, March 2023

Debajyoti Pal, Chonlameth Arpikanondt, Suree Funilkul, Vijayakumar Varadarajan, “User Exrience with Smart Voice Assistants: The Accent Pirective”, 2019 10th International Conference On Computing, Communication and Networking Technologies, July 2019

K. R. Srinath, “Python- The Fastest Growing Programming Language”, International Research Journal of Engineering and Technology, Vol. 04, Dec 2017

John Goerzen, “Simple Message transport Protocol”, Foundations of Python Networking Programming (pp. 197-210), January 2004

Xiaojie Chen, “Introduction And Analysis Of python Software”, Frontiers in Computing and Intelligent Systems 5(2):41-43, September 2023.

Neha Jain, and Somya Rastogi, “Sppech Recognition Systems- A Comprehensive Study Of Concepts And Mechanism”, Acta Informatica Malaysia 3(1):01-03.

Sneha Tamboli, Pratiksha Raut, Lavkush Sategaonkar, Anjali Atram, Shubham Kawane, Prof. V. K. Barbudhe “A Review Paper On Text-To-Speech Converter”, Vol 3, no 5, pp 3807-3810, May 2022.