Voice Assistant

Sanket Hanagandi- 201IT154
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
sankethanagandi.201it154@nitk.edu.in

Mayur Jinde - 201IT135
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
jindemayur.201it135@nitk.edu.in

Shaulendra Kumar - 201IT159
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
shaulendra.201it159@nitk.edu.in

Kartikeya Ranjan – 201IT127
Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
kartikeyaranjan.201it127@nitk.edu.in

***Abstract*— Today's technical advancements are developing at a breakneck speed. Previously, we only had access to a computer system that allowed us to accomplish a restricted number of tasks. Machine learning, artificial intelligence, deep learning, and other technologies, on the other hand, have enhanced computer systems to the point where we can now complete any task. If individuals are still struggling to communicate via numerous input devices in this day and age, it's not worth it. As a result, we included a voice assistant into the system. A voice assistant's main purpose is to eliminate the use of input devices like a keyboard and mouse. It will also save on hardware space and money.**

I. Introduction

Everything that a human being can perform is being replaced by machines in this technological era. Changes in performance are one of the key factors. We educate our machines to think like humans and do tasks on their own in today's environment. As a result, the notion of a virtual assistant was born.

The voice assistant technology (such as *Siri* or *Alexa*) used today has been under development for over a century. It has come a long way from the first listening and recording devices to cutting-edge technologies found now in every family. Here is a brief history of where speech recognition comes from.

Computers were created with direct human interaction in mind up until the 1950s; they were envisioned as becoming actual social interpreters that could engage in vocal interaction. But, technology was still developing, and machines could not yet interpret the natural language to the point where they could communicate with people. After all, rather than serving as people's discussion companions, computers quickly evolved into communication facilitators.

In the early 2000s, the story of the voice revolution reached a decisive turning point: the *question-answering system*, *Watson* competed with the best champions of the popular television quiz Jeopardy! and defeated them in total points. Thus, becoming the first system capable of processing natural language with the same speed and confidence as a human.

A virtual assistant is a digital assistant that recognizes voice commands and performs related actions as desired by the user using voice recognition capabilities and language processing algorithms. A virtual assistant may filter out ambient noise and offer useful information based on specific commands supplied by the user.

Virtual assistants are entirely software-based, although they are increasingly integrated into a variety of devices, and some, such as Alexa, are created exclusively for a single device.

Because of the rapid advancement of technology, it is now more important than ever to train our machines using machine learning, deep learning, and neural networks. With the aid of Voice Assistant, we can now communicate with our machines. Today, every major corporation employs

Voice Assistant, allowing users to communicate with the machines using their voices. So, with the Voice Assistant, we've progressed to the next level, where we can communicate with our machines.

The voice assistant is regarded as a very powerful tool and is used in various industries. With evolving technology, we witnessed growth in the power and reliability of voice assistants. Voice assistants are the latest technological advance in consumer electronics that are making their way into people’s lives. These devices evidence the impressive development and capability of artificial intelligence and present a tangible contrast to the depictions of this technology in iconic films. Now, they are a key component in smartphones, smart TVs, smart home components, and various other technological products. With time this technology will get better as we see growth in the AI field.

I. Literature Survey

1. Dekate, Abhay & Kulkarni, Chaitanya & Killedar, Rohan. (2016). Study of Voice Controlled Personal Assistant Device. International Journal of Computer Trends and Technology. 42. 42-46. 10.14445/22312803/IJCTT-V42P107.

Summary:

The study on the use of voice-controlled personal assistant devices is presented in the paper. The paper offers a thorough review of the literature on the state of the art in voice assistant technology and research. The performance of "Jarvis," a voice-controlled personal assistant device, in recognizing and carrying out voice commands is then the subject of an empirical investigation by the writers. They assess the device's reliability, response time, and capacity to understand orders in a variety of languages and accents.

According to the study, Jarvis can recognize voice instructions with a high degree of accuracy in a variety of languages and dialects and with a reaction time of under a second. The authors draw the conclusion that voice-controlled personal assistant devices could offer a comfortable and effective user interface for a variety of applications, including home automation, healthcare, and entertainment.

Overall, the paper provides insights into the technology and performance of voice-controlled personal assistant devices.

Drawbacks:

•The study was conducted on a small sample size, which may limit the generalizability of the results.

•The authors did not perform a comparative analysis of different voice assistant devices, which could have provided more insights into their strengths and weaknesses.

•The study did not consider the impact of user demographics such as age and gender, which could influence user satisfaction and adoption of voice assistant devices.

•The authors did not investigate the ethical and legal implications of using voice assistant devices, such as data privacy, security, and ownership.

2. Gupta, Siddhi & Joshi, Devanshi & Kolpate, Krutika & Sharma, Dhruv & Parmar, Mrs. (2022). Personal Voice Assistant. International Journal for Research in Applied Science and Engineering Technology. 10. 2127-2131. 10.22214/ijraset.2022.42730.

Summary:

The concept and execution of a voice-controlled personal assistant system are presented in the paper. The system can do a variety of things, like playing music, checking the weather, creating reminders, and sending messages. First the paper gives a general description of the idea of personal assistant systems and their rising popularity. The paper continues by outlining the voice assistant system's core technology, which includes speech recognition, natural language processing, and text-to-speech conversion. The voice assistant system, which was developed using the Python programming language and a number of open-source libraries, is also covered in the article along with its design and implementation. The system architecture and the many parts necessary for it to operate are described by the authors. The paper also provides an analysis of user input and the system's performance evaluation, which includes accuracy testing. The outcomes show a high level of speech recognition accuracy and user satisfaction with the system's functionality.

The overall discussion of the design and execution of a voice-controlled personal assistant system in the paper is instructive and thorough. The study shows how voice assistant technology has the potential to offer a hands-free and natural user interface for carrying out a variety of tasks, increasing user productivity and convenience.

Drawbacks:

•The planned voice assistant system is not compared to other voice assistant systems that are already in use, which would have shed more light on its advantages and disadvantages.

•The study may not be comprehensive enough to provide reliable and valid results due to limitations in methodology, sample size, or data collection.

•The paper does not provide a detailed technical analysis of the performance of the voice assistant system, such as response time and system resource utilization, which could have provided more information on its accuracy and efficiency. The study may not provide sufficient evidence to support the claims and conclusions made by the authors.

•Similar to the previous paper, the paper does not investigate the impact of user demographics such as age and gender on the user satisfaction with the voice assistant system.

3. S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas and B. Santhosh, "Artificial Intelligence-based Voice Assistant," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.

Summary:

The design and implementation of a voice assistant

system based on artificial intelligence (AI) technology are presented in the paper. The system is made to carry out a number of functions, including playing music, setting reminders, answering queries, and managing smart home gadgets, among others. The paper starts off by giving a general introduction to the idea of voice assistants and their rising popularity. They continue by outlining the AI-based voice assistant system's core technology, which includes speech recognition, natural language processing, and machine learning algorithms. The voice assistant system, which was developed using the Python programming language and a number of open-source libraries, is also covered in the article along with its design and implementation. The system architecture and the many parts necessary for it to operate are described by the authors. The paper also provides an analysis of user input and the system's performance evaluation, which includes accuracy testing. The outcomes show a high level of speech recognition accuracy and user satisfaction with the system's functionality.

Overall, the paper offers a comprehensive and exhaustive explanation of the development of an AI-based voice assistant system overall. The study highlights how AI technology has the ability to improve voice assistant functionality and offer users with more tailored and effective user experiences.

Drawbacks:

•The study does not provide a comparative analysis of the proposed AI-based voice assistant system with other existing voice assistant systems, be it AI-based or non-AI-based, which could have provided more insights into its strengths and weaknesses.

•The paper does not provide a detailed technical analysis of the performance of the AI-based voice assistant system, such as response time and system resource utilization, which could have provided more information on its accuracy and efficiency.

•Similar to the other papers, the paper does not investigate the impact of user demographics such as age and gender on the user satisfaction with the voice assistant system.

4. B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212,1976.

Summary:

The most significant declarations and speeches have been covered in writing by Bassam A., Raja N., and others. Humans and machines communicated via an analogue signal, which was then converted to a digital wave by speech. This technology is extensively utilized, has many uses, and enables computers to respond to human voice commands with accuracy and consistency while also offering useful and appreciated features. Speech Recognition System (SRS) is gaining popularity and has several uses. The research has provided a summary of the strategy; it is a fundamental model [1]. According to B. S. Atal and L. R. Rabiner et al., speech analysis and pitch analysis are frequently used in conjunction with one another. The study created a pattern recognition method for determining whether a certain slice of a speech signal should be categorised as voiced speech, unvoiced speech, or silence based on signal dimensions.

Drawbacks:

•The paper was published when speech recognition technology was still in its early stages, and the dataset used in the study may be limited in size and variety, which could limit the generalizability of the proposed approach.

•The paper's approach assumes that speech can be modeled as a stationary process, which may not be accurate for all types of speech signals, leading to possible errors in classification.

•The paper did not compare the proposed approach's performance with other state-of-the-art approaches at that time.

• The paper does not provide implementation details of the proposed approach, making it difficult to reproduce the results or to apply the approach in practice.

II. PROBLEM STATEMENT

Designing an interactive voice assistant using Python libraries to ease the work through voice commands of the user.

Objectives

- To make a user friendly, interactive voice assistant.
- To browse on various websites via voice commands.
- To play local music/video via default players.
- Tosend emails via voice commands.

III. METHODOLOGY

The methodology can be divided into three steps namely:

1) Speech to Text
2) Text Analyzing
3) Interpret the Command

- Speech to Text:

  A piece of code converts audio to text. To do this the python library used is *Speech Recognition*. It's a Text to speech library developed by Microsoft sapi5 engine, it's used here because sapi5 is made for windows. Sometimes it does not understand what you say but there's an exception handling to deal with this issue.

- Text Analyzing:

  As the speech is converted to text. The text arguments are passed through the program to interpret the command.

Then the program checks whether the passed arguments are valid or not. This is done by the test analyzing library - *pyttsx3 · PyPI* . If it's valid, it executesthe command, else throws an exception.

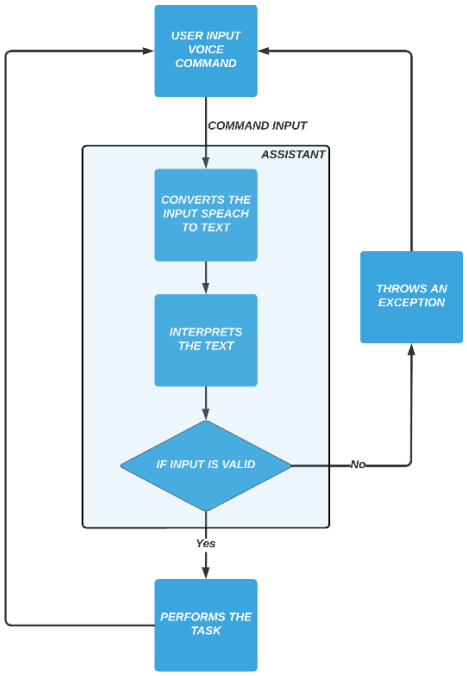
- Interpret Commands:

  If the command is to access a website, then the program contacts the server to open the desired website. The python library which is used to do this is *Web Browser* . It's a Web browser API used for opening web

browsers.

If the command is to access a local file/folder, then it opens it as per the path provided. If the command is to play music/video, it plays through the default music/video player.

Flowchart:



Flowchart of the Algorithm

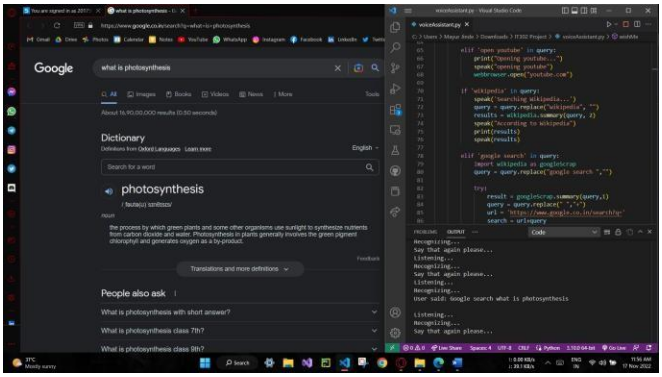
IV. Results and Analysis

The present assistant will have the following functionality:

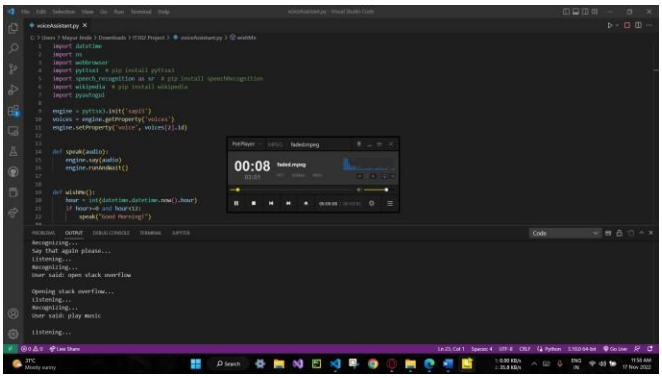
1) The system will continue to listen for orders, and the length of time it spends listening is adjustable to meet the needsof the user.
2) If the system is unable to extract information from the user's input, it will prompt the user to repeat the process until the desired number of times has been reached.
3) Depending on the needs of the user, the system can feature both male and female voices.
4) Playing music, sending emails, searching Wikipedia, and launching system-installed apps, as well as any web browser, are all supported in thepresent edition.

Using voice command to perform different tasks:

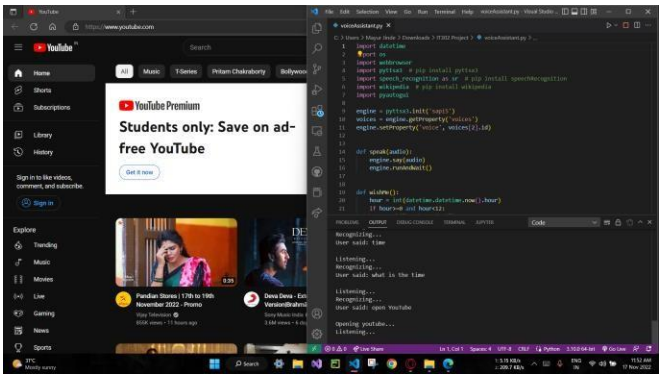
Google Search:



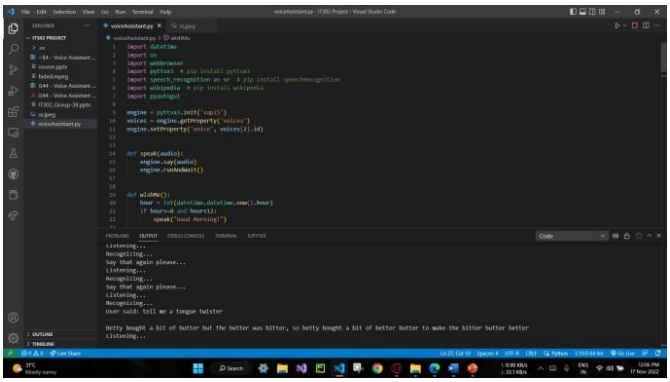
Playing Music:



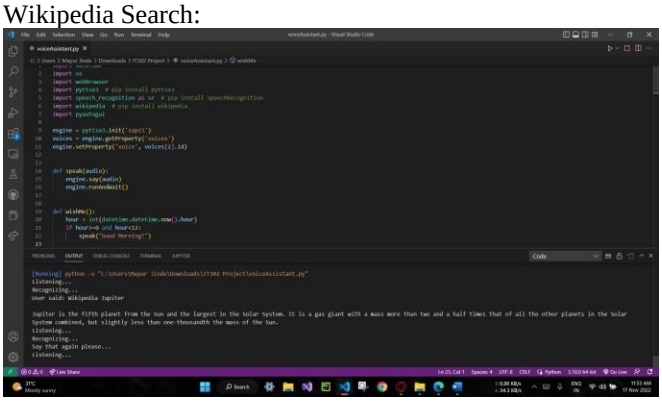
Opening YouTube:



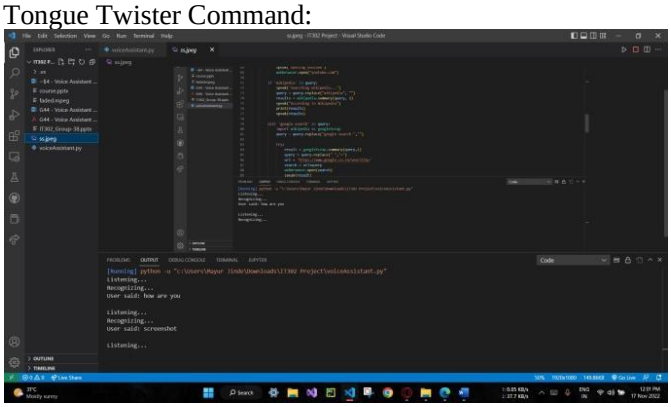
Screenshot command:



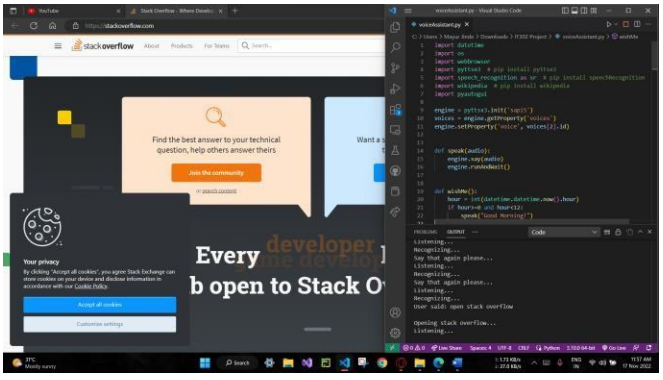
Wikipedia Search:



Tongue Twister Command:



Opening StackOverflow:



V. CONCLUSION

We covered a Voice Assistant written in Python in this article. This assistant now operates as an application that does basic activities like browsing the web, playing music, Wikipedia searches, displaying time, opening MS Office apps and the opening of desktop apps. The present system's capability is limited to solely dealing with applications. Machine learning will be added into the system in future versions of this assistant.

IMPLEMENTED/BASE PAPER

Nivedita Singh, Dr. Diwakar Yagyasen, Mr. Surya Vikram, Gaurav Kumar, Harshit Agrawal. Department of CSE, Babu Banarasi Das National Institute of Technology and Management, Lucknow, India, July2021.

REFERENCES

[1] M. Bapat, H. Gune, and P. Bhattacharyya, “A paradigm-based finite state morphological analyzer for marathi,” in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34,2010.

[2] B. S. Atal and L. R. Rabiner, “A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212,1976.

Click here for the video demonstration of our project