AI BASED DESKTOP VOICE ASSISTANT

Shweta Lilhare[1], , Abhishek Rout [2], Sahil Thombare [3], Aditya Shinde [4]
[1](Professor, Department of Computer Engineering., Savitribai Phule Pune University.)
[2](Student , Department of Computer Engineering., Savitribai Phule Pune University.)
[3](Student , Department of Computer Engineering., Savitribai Phule Pune University.)
[4](Student , Department of Computer Engineering., Savitribai Phule Pune University.)
[1]shweta1641991@gmail.com, [2]abhirout14@gmail.com., [3]sahilthombare20001@gmail.com .,
[4]adishinde611@gmail.com.

Abstract:
Nowadays voice control is a significant feature that impacts the life of people. We have many voice assistants out there which are mainly used in smartphones, computers or IoT devices but in the context of computers, none of them really provide system control and a good user interface. Voice assistants are AI based systems that can understand and respond and react to it using built in voices. This assistant provides a smart and effective solution to multitasking since wegive input in speech format and the assistant would simply evaluate our query and give theoutput in speech. This not only becomes feasible for anyone without detailed knowledge of technology but also for someone looking to use it like a personalized one-to-one assistant providingthe luxury of controlling the system using only voice. It is an endeavor of narrowing the visible gap between users who are trying to achieve sometask on their computers but are often hindered bysome help regarding search or any small tasks. Voice control is an attempt at balancing user heed at the main task while speeding up the actual work .Moreover, also for the users working with multiple devices or dealing with multiple tasks can make use of it to speeden up their work be they of any age group. Speech as a technology is quite underrated and can be thefuture and it is an effort to make the best possibleuse of it.

Keywords— Voice Control, AI, Voice Assistant, System Control, Text to Speech, Mass Audience

INTRODUCTION

Nowadays the use of technology has increased to avast extent and we spend a good amount of time with electronic devices and thus it is important to achieve productivity while performing any task and that's where a voice assistant is very helpful. A voice assistant is kind of like a smart computer that directlyspeaks the required query. Thus the program that can understand and talk to you when you speak to it. It uses special technology to figure out what you want and then does things or gives you information using only your voice. This means you can control things and get help without touching anything, just by talkingto the assistant. Various assistants out there don't provide system control and are limited to tasks like searching and it is where our assistant shows distinctiveness compared to other assistants. A good voice assistant should be able to interpret user commands and understand the exact requirement ofthe user. While the basic need of our project was to create a unique assistant that is different from all of the existing voice assistants that can serve a mass audience of every age group. A voice assistant ensuresthat there is no such difficulty while multitasking and enables efficient work flow or learning as the userneed not to manually type the query or command and can make use of a wake word to directly speak the required query. Thus the tasks from day to day life canbe executed more efficiently.

BACKGROUND

History of Voice Assistants

The concept of AI voice assistance is rooted in the broader field of artificial intelligence, which seeks to create machines , AI's ability to imitate human intelligence and perform tasks that typically necessitate human cognitive abilities is what is being referred to such as natural language understanding and speech recognition. Voice assistants aim to enable human-interaction through spoken language, making tech accessible and use-friendly.

History:

1960-1970-EarlyText-Based AI Systems:

Voice assistance can be traced back to early computers that could understand and respond to inputs .Systems like ELIZA, developed in the first attempts of natural language process. While not voice based ,they laid the ground for development.1980s-1990s Speech Recognition Advances: During this period , significant progress was made in speech. IBMs Voice Type was introduced, it is possible for users to communicate with computers through spoken language, albeit to a limited extent.

2000s - Emergence of Voice Assistants: The 2000smarked the emergence of voice assistants with more widespread adoption. In 2001,Microsoftintroduced Clippy, a virtual assistant. However ,it wasn't until 2011 when Apple introduced Siri that voice assistants gained significant attention. Siriwas integrated into the iPhone 4S and could perform tasks based on voice commands. 2010s- Proliferation and Advancements: The 2010s saw a proliferation of voice assistants, with Amazon's Alexa, Google Assistant, and Microsoft's Cortana entering the market. These assistants were integrated into smart speakers, smartphones, and other devices, making them more accessible to the general public. Advancements in NLP and machine learning greatly improved their conversational capabilities.

2020s - Continued Growth: In the 2020s, AI voice assistants continued to evolve and expand their capabilities. They became integral parts of smart homes, helping control lights, thermostats, and appliances. The integration of AI voice assistants into various industries, from healthcare tocustomer service, became increasingly common.

Aim of this Study

The goal of our project is to develop an AI- powered Desktop Voice Assistant capable ofperforming multiple tasks.. based on the input from the user , these tasks can vary from different categories offering a wide area of use cases , thus making maximum use of the technology and reducing the effort by giving thecommands/prompts to the computer in the vocal (voice) format.

LITERATURE REVIEW

[1]Artificial Intelligence - based Voice Assistant written by authors Subhash S, Prajwal N Srivatsa,Siddesh S, Ullas A and Santhosh B published in the year 2020.It discusses the growing significance of voice control and the integration of AI-based voice assistants in various devices such as smartphones and laptops. It outlines the process by which these voice assistants convert spoken commands into text, utilize the Google Text-to-Speech (GTTS) engine to convert text into audio, and play the audio using Python'sPlaysound package. The introduction highlights the transformative impact of voice assistants on user interactions with technology. It distinguishes voice assistants from virtual assistants, emphasizing the roleof AI-based voice assistants in anticipating and fulfilling user needs. The primary aim of their project was to create an advanced intelligence,it is capable of performing a wide range of tasks as a personal assistant. based on voice commands, enhancing user convenience through Voice User Interface (VUI).The paper outlines the design of the AI-based voice assistant, emphasizing its ability to understand user commands, perform various tasks, and utilize the GTTS engine to convert text to speech. Tasks Performed by the Voice Assistant: The paper lists several tasks that the voice assistant can perform, including playing music, searching the web, providing weather updates, capturing screenshots, and more. The methodology section explains the process behind automatic speech recognition (ASR), which is fundamental regarding the operation of voice assistants that are powered by AI.. It details the steps involved in ASR, from recording and acoustic analysis to decoding and text processing.

[2]Artificial Intelligence based Vision and Voice Assistant written by authors R S Sai Dinesh, R Surendran, D Kathirvelan and V Logesh published in the year 2022. It focuses on the domains of Artificial Intelligence (AI), Natural Language Processing (NLP), and Computer Vision It addresses the need for input devices in computer systems and explores the potential of AI and machine learning to perform tasks at human-like levels of competence and even beyond. However, a significant limitation in AI is the absence of consciousness, which hinders machines from perceiving their environment as humans do. The core idea is to enable users to directly communicate with computer systems using Natural Language Processing. This interaction would involve programming the system based on its specific requirements. Traditionally, users have had to resort to machine-level programming codes to operate computer operating systems, a process often associated with complexity. The proposedapproach seeks to simplify this interaction through the use of Python programming. In summary, the literature review on this proposed work highlights the integration of AI, NLP, and Computer Visionto create a system that enables users to engage with computers. using natural language, reducing the reliance on conventional input devices and potentially enhancing the user experience with technology.

[3]AI-Based Desktop Voice Assistant written by authors Shubham Thorbole, Anuradha Pandit,Gayatri Raut and Tejas Sirsat published in the year 2022. It discusses the integration of Artificial Intelligence (AI) into desktop voice assistants, with a particular focus on a voice assistant named Jarvis AI. The literature review section of the paper references prior research on voice control frameworks, the importance of voice assistants in various devices, and the analysis of speech recognition systems. It mentions the utilization of technologies like Google Assistant and database query language transformations. The methodology section outlines the key components of Jarvis AI, including Python for programming, theQue.py framework for natural language question transformation, Pyttsx3 for text-to-speech conversion, and the use of NLP and voice recognition techniques for command interpretation. SQLite is employed for efficient data storage, and various Python libraries and modules are

leveraged for specific functionalities. The system architecture is presented, highlighting the role of modules like GTTS for text-to-speech conversion, Datetime for displaying date and time, Wikipedia for knowledge retrieval, Web Browser for web browsing, and others for tasks such asemail sending, making HTTP requests, and handling audio.

[4] AI-based Desktop Voice Assistant written by authors Pankaj Kunekar, Ajinkya Deshmukh, Sachin Gajalwad, Aniket Bichare, Kiran Gunjal and Shubham Hingade published in the year 2023.It highlights the rapid advancement of Artificial Intelligence (AI) lately , primarily focusing on the application of Natural Language Processing (NLP) within the domain. The paper outlines the fundamental premise of desktop voice assistants, emphasizing their ability to recognize and respond to human voices through integrated voice systems. However, it also alludes to the existence of certain challenges and limitations associated with voice assistants, which the paper intends to explore in-depth. In summary, the literature review provides an overview of the evolving landscape of AI, particularly in the context of NLP and voice assistants. Ithighlights the broad adoption of these technologies, both in homes and educational institutions, and allows us to explore the deeper aspects of the challenges and constraints faced by the voice assistants nowadays in practice.

PROPOSED DESIGN

This project aims to provide users with some what comprehensive understanding of an intelligent assistant that can effectively comprehend their commands. Our assistant is equipped to understandvocal commands and respond accordingly. It listens touser commands through the microphone.

The microphone will be used to capture voice input from the end user or the user of the application.The Speech Recognition Module is used for converting the vocal information from the user to the textual format.The extracted data is then acted upon by the application which then works for fulfillment. This is done from multiple sources based on the type of the request. If the prompt is for API calls for the real time information then respective APIs get called for theinformation. The content extraction from the modules present in the python e.g. Wikipedia module is used

in the process of searching for any particular information. If the prompt is for API calls for the real timeinformation then respective APIs get called for the information. The content extraction from themodules present in the python e.g. Wikipedia module is used in the process of searching for anyparticular information. If the request is concerned with the System tasks then the System Calls are performed.

After any of these data sources the data is taken in the textual format which is then given to the “TextTo Speech " module which converts this data from the textual format to the vocal form. This data is then delivered to the end user in the voice format form the output device of the system Speakers.

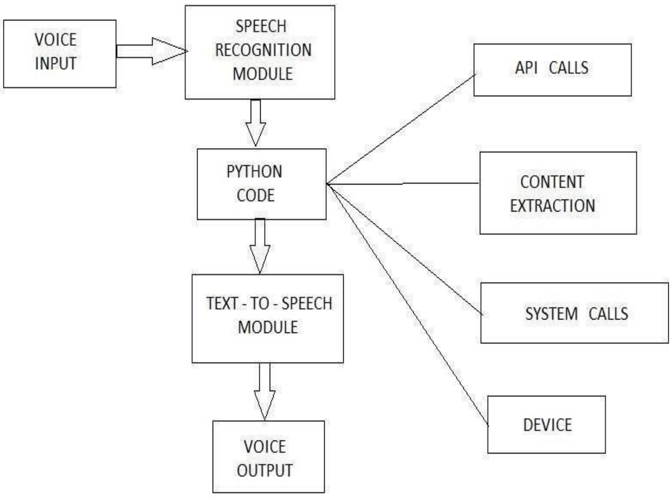

Fig 1.1 : Block Diagram

Tasks Performed By The Voice Assistant

1)Can remember the name of the end user or the user of the application.

2)Playing songs from 'Youtube' or 'Device' as requested by the user.

3)The assistant has the ability to perform "Google Search" for the information the user wants.

4)The assistant opens up Google Maps and pinpoints the exact location specified by the user. Ifthe user asks to "where is " or "locate on map " theassistant prompts them to specify which location they want to view and then displays that location on the map.

5)Tells the accurate weather of the location the user asks for. When asked for 'current weather in’ the assistant tells the exact weather of the desired location.

6)If you require assistance capturing your screen,simply request the assistant to "take screenshot". The assistant will promptly take a screenshot of your current display and save it to the designated path for your convenience.

7)Stay up-to-date with the latest news from allcorners of the globe. Simply ask the assistant for news, and it will provide you with the 'top most recent headlines' thus saving time and delivering the important content.

8)Up-to-date time and date information which keeps the user updated.

9)System Control enables the assistant to perform various system related tasks like emptying the recycle bin, shutting down the computer, restarting the computer.

10)Launching various installed apps from the'Desktop' as well as Opening any application on the 'Browser'

11)Sleep mode allows the assistant to remain active only if the user is requesting; if not then it goes into sleep mode until it is woken up back by the user.
12)Note taking is also one of the features allowing the assistant to take the notes from the user in the voice format and storing it in the form of the txt file in the computer.
13) Touchless brightness and volume control through hand gestures.
14)Enhanced security by identifying the authenticate user through face recognition

METHODOLOGY

Voice assistants have gained immense popularity in the technology world for their ability to respond to voice commands. This assistant is created through programming, and our AI-based Voice Assistant is built using the versatile Python ProgrammingLanguage.. A user can command "Open specific website on web browser", the intelligent voice assistant will efficiently navigate and access the appropriate folder or website to display the desired results, saving you time and effort.. The Voice assistant takes input through a mic which is in format of voice then speech recognition recognizes voice and task is further processed. Using natural language processing (NLP), virtual assistants can match the user's voice or text input with executable commands, allowing them to operate machines such as laptops or computers based on the user's commands. Voice assistant applications operate using Speech Recognition (SR) systems. These systems recordspeech and then analyze it into phonemes, which are subsequently converted into text. A phoneme, rather than words or syllables, is the fundamental component used

Speech recognition: The virtual assistant can start by listening for voice input from the user through a microphone or other input device. Python libraries like Speech recognition involves converting the user's spoken words into text. After this, the AI-powered assistant utilizes natural language processing (NLP)techniques to comprehend the user's intent and extract vital information from the input.

Task Execution : After analyzing the user's input and

updates, or calendar events. Interacting with APIs and external services to fetch data or perform specific actions.

Text-to-Speech (TTS): After processing the user's request and generatinga response , the virtual assistant can use a Python library like pyttsx3 to convert text responses into speech. This allows the assistant to communicate its responses audibly to the user.

Face Recognition: We employ face recognition technology alongside the Haar cascade classifier for enhanced security measures. The Haar cascade classifier is utilized due to its effectiveness in detecting frontal faces in images or video streams, making it an ideal choice for real-time applications. This technology offers benefits such as rapid processing speed and low computational requirements, ensuring efficient face detection within our assistant. Additionally, by leveraging computer vision algorithms and machine learning techniques, we achieve accurate face recognition using libraries like OpenCV. OpenCV, an open-source computer vision library, provides a wide range of functionalities for image processing, feature extraction, and face detection. Its extensive documentation and active community support make it a preferred choice for implementing face recognition systems. By combining the robustness of face recognition with the efficiency of the Haar cascade classifier, our system ensures reliable user authentication and heightened security levels.

Gesture Recognition: Gesture recognition plays a crucial role in providing intuitive interaction modalities alongside voice commands. We utilize computer vision techniques and machine learning models to detect and interpret hand gestures in real-time. OpenCV, a versatile computer vision library, serves as the foundation for gesture recognition within our system. OpenCV offers a plethora of tools and algorithms for image processing, motion tracking, and gesture recognition, making it well-suited for our application. By leveraging OpenCV's capabilities, we can track hand movements accurately and identify predefined gestures associated with volume control and brightness adjustment. The integration of gesture recognition enhances user accessibility and experience by providing alternative control methods beyond voice commands. This technology empowers users to interact seamlessly with our assistant, catering to diverse preferences and facilitating intuitive control of desktop functions.

ARCHITECTURE

Voice assistants is a software where it takes input from mic to increase the interaction of the user to electronic devices such as computers. After taking input from mic the useful information is extracted from the text which is given by the user through mic and converted by speech recognition to text. Then a query is searched to match the executable command in our code if matched then code is executed to desire output. If the query is not matched then the code will give a prompt to the user please try again in speech format to convert text to speech it uses pyttsx library

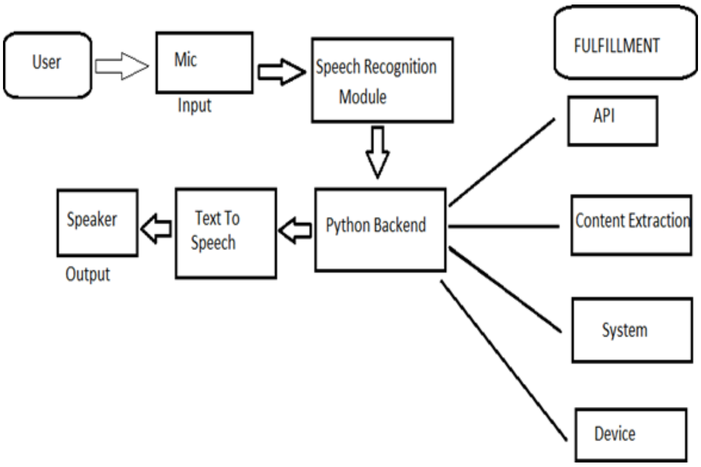

Fig 1.2 Architecture

Speech Recognition:

This is the initial step where the system listens to the user's spoken commands. It involves audio input processing, such as capturing audio through a microphone. The captured audio is then converted into text using speech recognition.Voice Assistant: After speech recognition, the user's spoken command in text format is passed to the voice assistant. It is the core component of the system responsible for understanding and acting upon user commands. Python Backend: The voice assistant relies on a Python backend, which includes code and logic to interpret and respond to user commands. The backend contains functions and scripts for various tasks and interactions.

API Call:

In many cases, the voice assistant may need to fetch data or perform actions that require external resources. To do this, it makes API (Application Programming Interface) calls to external services or systems. TheAPI call can be to services like weather data, news, Wolfram Alpha, etc. The voice assistant sends requeststo the API, which returns data or performs the requested action.

Content Extraction:

Once the voice assistant receives data from an API or external source, it often needs to extract relevant information from the data. Content extraction involves parsing and processing the data to extract specific pieces of information that are needed to respond to theuser's query. For example, if the assistant fetchesweather data, it may extract the temperature andweather condition from the API response.

MATH

Haar-like Features:

Formula: f = ∑white pixel value - black pixel value
2.Integral Image:
Calculation: ii(x, y) = Σ(Sum of y)Σ(Sum of x)I(i, j)

3.AdaBoost Algorithm:
N • Weighted error. E = Σ(From 1 to N) Wi*ERRi
4.Cascade Classifier.
Thresholding: Decision based on classifier output exceeding a threshold.
5.Training:
Minimize error using positive and negative samples.
6.Thresholding:
Classifier output comparison to threshold determines region acceptance .

To calculate the distance between the thumb and finger in the provided functions, Euclidean distance is commonly used.

Given the coordinates of two points (x1,y1) and (x2,y2) the Euclidean distance between them is calculated using the following formula:

distance = $\sqrt{[(x2-x1)^2+(y2-y1)^2]}$

In the functions provided, the distance between the thumb and finger is computed using the np.linalg.norm() function, which calculates the Euclidean norm (or distance) of a vector.

length = np.linalg.norm(np.array(p2) - np.array(p1))

Here:
p1' and 'p2" represent the coordinates of the thumb and finger points, respectively. np.array(p2)-np.array(pi) computes the vector from 'pi to p2.
np.linalg.norm()computes the Euclidean norm of this vector, giving the distance between p1' and 'p2.
This distance is then used for further processing, such as mapping to brightness or volume levels.

CHALLENGES AND GAPS

Challenges

1.The assistant sometimes has problems misunderstanding the prompt which may lead to the inaccurate gathering of data or generating wrong results.
2.The voice assistant has to ensure that it has the necessary permissions to perform various system related tasks without compromising the computers security.
3.In a noisy environment it becomes very challenging for the assistant to correctly understand the prompt.
Since voice assistants have access to a lot of users data it raises a concern of security , thus violating the user's privacy.
4.The voice based assistant usually runs on battery powered devices , and since they perform multiple tasks , the power consumption issues arises as these applications require continuous access of the systems hardware like mic and speakers.

Gaps

1. Various system related tasks may related tasks may require time to execute.

2.Getting real time information form the APIs may also require time to load the data.
3.Slowdown of the system while performingcomplex tasks.
4. Accessibility issues as only the person who is able to interact in English will be able to use the application.
5.The assistant may be not useful for the people who are disabled with hearing impairments.

CONCLUSION

Thus in this project we will be implementing a variety of features as discussed and the uniquenessof our project will be the use of System APIs for smooth functioning of the assistant .

Nowadays due to the ever increasing use of electronics the hands free operation and execution of tasks will enable a good and efficient use of devices .

In the field of Desktop computing, it is essential that any user gets the required information withoutthe hassle of traditional input methods of manuallysearching.

Our project will provide the distinct feature of executing any task related to the system and on topof that also give the user the freedom to increase the speed of his task completion by giving therequired information in a matter of seconds.

While we felt that there is lack of a completely flawless assistant in terms of laptop which can alsoperform system related tasks and basiccomputations, our project will the solution to narrow this gap and provide users with an efficientassistant that would allow users with disabilities or empower any user to use this to increase their productivity.

REFERENCES

[1] P. N. ,. S. S. U. A. S. B. Subhash S, Artificial intelligence - based Voice Assistant, 2000.

[2] R. S. S. D. R. S. and D. K. , “Artificial Intelligence based Vision and Voice Assistant,” 2022.

[3] S. T. A. P. and G. R. , “AI-Based DESKTOP VOICE ASSISTANT,” 2023.

[4] P. K. A. D. and S. G. , “AI - based Desktop Voice Assistant,” 2023.