Voice based Virtual Assistant for Desktop

Mukund Rathi[1] Sandesh More[2] Ajay Borde[3] Chetan Kute[4] Dr. A.V.Brahmane[5]
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]Sanjivani College of Engineering, Kopargaon, Ahmednagar, Maharashtra, India

Abstract — Voice control is a major growing feature that change the way people can live. The voice assistant is commonly being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through Speech Recognition Module. in this project we are build voice assistant application that work online as well as offline. We use google recognition and vosk modules that will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.

Keywords: Voice Assistant, AI, ASR, Python, Vosk

I. INTRODUCTION

Voice-based virtual assistants for desktops are becoming increasingly popular in recent years. These assistants, such as Siri for Mac, Cortana for Windows, and Google Assistant for Chromebook, provide users with the ability to control their desktops with voice commands. With a voice-based virtual assistant, users can perform tasks such as opening applications, creating reminders, checking email, and searching the web, all without needing to use a keyboard or mouse. This can be particularly useful for individuals with disabilities, those with mobility issues, or those who simply prefer a hands-free approach. In addition to the convenience factor, voice-based virtual assistants for desktops also incorporate artificial intelligence and natural language processing technologies, allowing them to better understand and respond to user commands over time. As users interact with the system, it can learn their preferences and adapt to their speech patterns, becoming more accurate and effective in performing tasks. Overall, voice-based virtual assistants for desktops offer a powerful and convenient tool for users to control their computers with voice commands, while also incorporating cutting-edge AI and natural language processing technologies.

II. LITERATURE SURVEY

The article "Artificial Intelligence Based Voice Assistant" by Subhash S [1] et al. discusses the development and implementation of a voice-controlled system using natural language processing and machine learning algorithms. The authors address the challenges of background noise and different accents and dialects, and provide experimental evidence of the effectiveness of their system. Overall, the article provides a comprehensive overview of the technologies and techniques involved in developing an artificial intelligence-based voice assistant [1].

The paper "Short Research on Voice Control System Based on Artificial Intelligence Assistant" by Tae-Kook Kim [2] examines the potential of voice control systems based on AI assistants. The author presents a voice control system that uses natural language processing and machine learning to interpret user commands and perform tasks. The system achieved a high level of accuracy and effectiveness in experiments, but further research is needed to address issues around accuracy, privacy, and user experience. The paper provides a valuable contribution to the field, but lacks detail on technical aspects and could benefit from larger sample sizes in future research [2].

The paper "ABYS(Always By Your Side): A Virtual Assistant for Visually Impaired Persons" by M. R. Sultan and M. M. Hoque [3] presents a virtual assistant designed to assist visually impaired individuals in their daily activities. The system utilizes voice commands and natural language processing to interact with users and perform tasks such as reading emails, searching the internet, and managing schedules. The authors conducted user testing with visually impaired individuals, demonstrating the effectiveness and usability of the system. However, the paper lacks detail on technical aspects such as the algorithms and methods used in the system, which could be addressed in future research. Overall, the paper provides a valuable contribution to the field of virtual assistants for individuals with disabilities [3].

However, a study published in the Journal of Medical Systems by authors Hamed Hassanzadeh [4] et al. raised concerns about the accuracy and reliability of voice-based virtual assistants in healthcare settings. The paper, "Voice-Activated Intelligent Agents in Healthcare: An Investigation into the Current Situation and Future Prospects," found that virtual assistants can struggle with understanding medical terminology and accurately interpreting voice commands, potentially leading to errors in patient care [4].

Overall, the literature suggests that voice-based virtual assistants for desktops have the potential to improve productivity and efficiency for users, but also raise concerns about privacy and accuracy. Further research is needed to fully understand the benefits and limitations of these systems and to develop solutions to address potential issues.

III. TECHNOLOGY USED

- Python: - Python is an interpreted high-level general-purpose programming language. The version python 3.6.0 is used in the development of voice assistance project.
- VS Code: It is the most popular IDE for Python It includes great features such as excellent code completion and inspection with advanced debugger and support for web programming and various frameworks.

IV. METHODOLOGY

A. Automatic Speech Recognition (ASR)

Applications for voice assistants use the Automatic Speech Recognition (ASR) system to function. ASR systems capture speech, decode it into phonemes, and then convert the

phonemes into text. A phoneme serves as the fundamental unit of measurement for human speech recognition, not words or syllables. The method of word decoding produces worse results than phoneme recognition because it prefers to evaluate words as independent units without considering context constraints. In a nutshell, the process starts with the device gathering audio with the microphone.

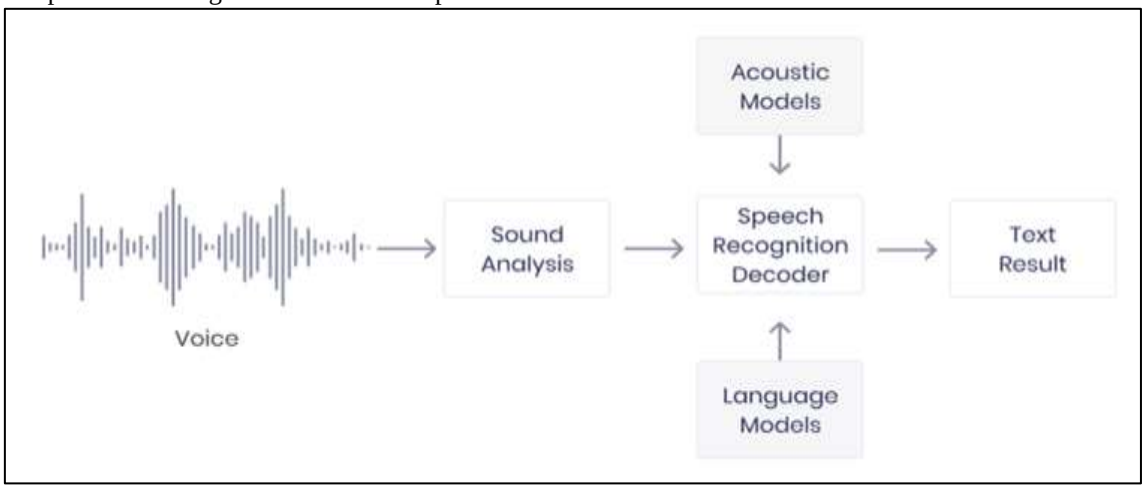

Fig. 1: ASR Architecture

Recorded speech waveforms get straight to acoustic analysis, which is performed on three different levels:

- Acoustic modelling, which displays the phonemes that were pronounced and the words they make up.
- Pronunciation modelling, which examines how phonemes are spoken and determines whether an accent or other vocal apparatus quirks contribute to the phonetic variability of speech.
- Language modelling, whose goal is to identify contextual probabilities based on the phonemes that were recorded.

The advances made possible by machine learning are what caused the low mistake rate to become so low. The data from the speech waveforms is then sent to the decoder, where it finally becomes text for use in commands or dictation.

V. System Architecture

The system architecture of a proposed system for the Voice Based Virtual Assistant on basis of the voice it performs the task. The system design includes the Recognition of voice of the person if voice not clear it will throw error that unable to recognize. It will take command in form of voice and perform accordingly.

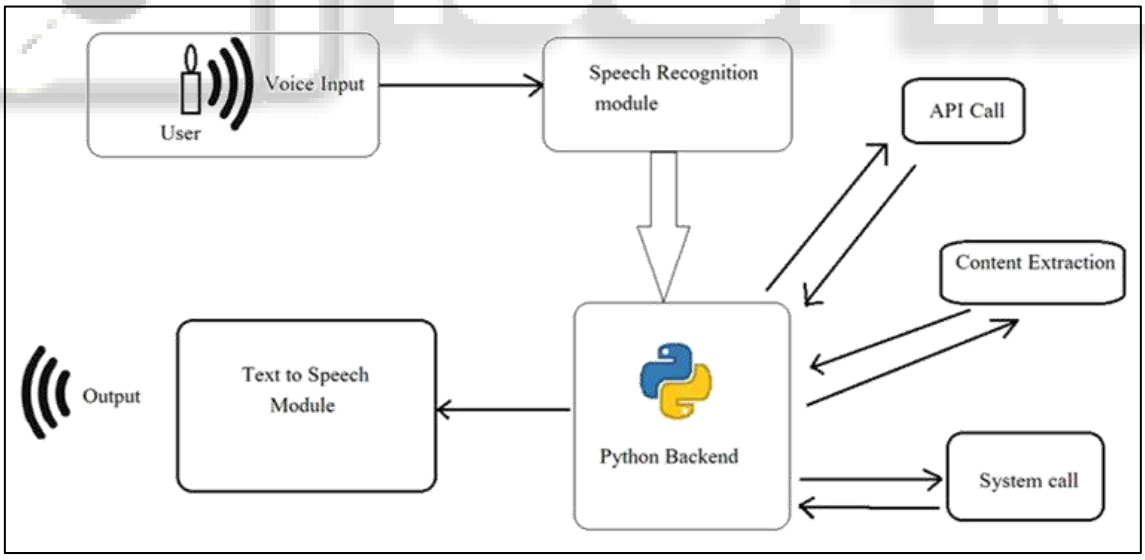

Fig. 2: System Architecture

System architecture consist of Five different model which helps voice assistant better

A. Voice input

In this module we are taking user voice input from user with accessing users system mic and some programming line to access this system.

B. Speech recognition module:

This is the main module in our system here we are going to recognize the user voice and process that voice convert it into the text format so it will interact with further modules. Here we use Vosk model which is based on ASR technology which helps our assistant working as Online as well as offline.

C. Python Backend

Python backend consist of program that communicate with different modules which required to execute different task of assistant.

D. API Calls, Content Extraction, System call

These are the modules help to execute task like searching Wikipedia, playing video on YouTube, open application on the system etc.

E. Text to Speech Module

In this module assistant generate output along with text format which will convert into speech form so the user get an appropriate communication with voice assistant.

VI. RESULTS

The required packages of Python programming language has been installed and the code was implemented using VS Code and the python code we have developed runs in both Python 2.7 and Python 3.x, and below are the few outputs which we have received in our AI-based voice assistant.

A. Google Search Output

As shown in below fig 3. When we ask the voice assistant to search 'python', it receives the request and performs the action by searching google.

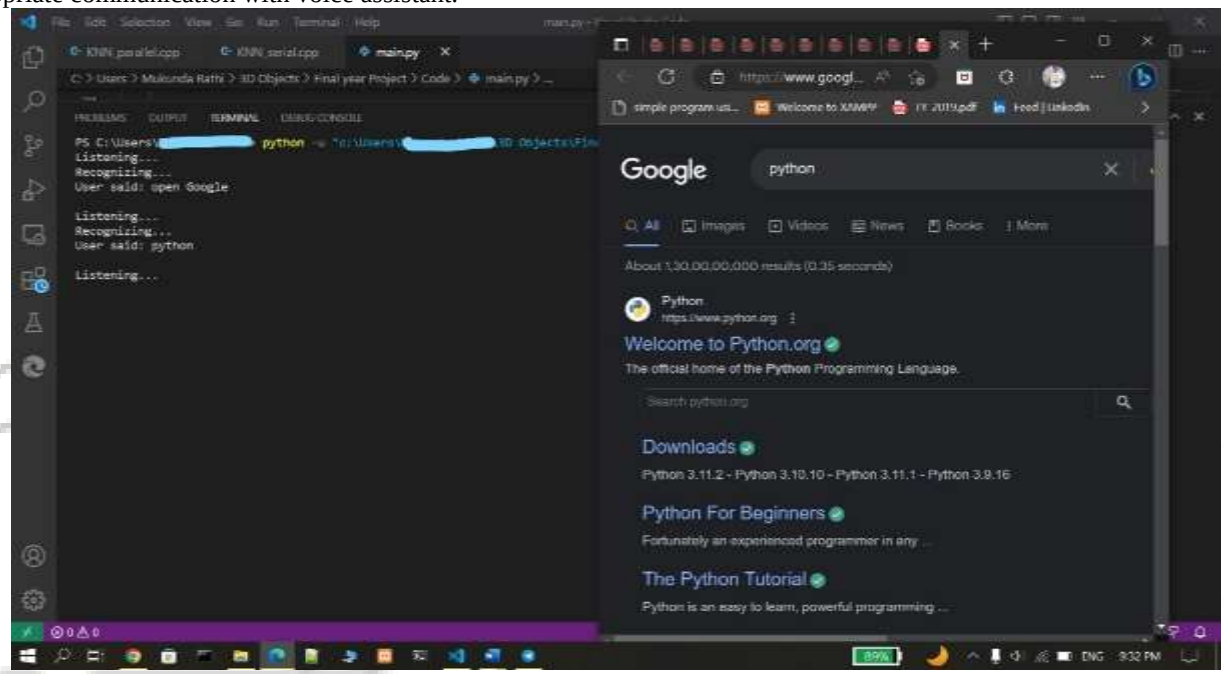

Fig. 3: Output screen of performing Google Search

B. Application opening

As shown in below fig 4.When we ask the voice assistant to open notepad, it receives the request and performs the action and opening the notepad application.

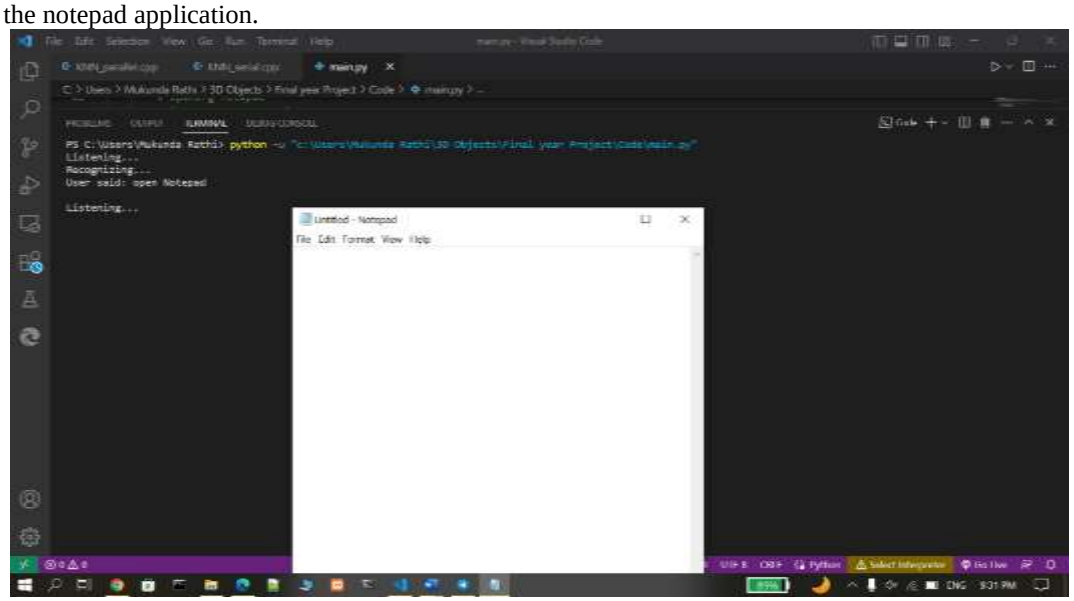

Fig. 4: Application opening

C. Note down the points

As shown in below fig 5.When we ask the voice assistant to open notepad and write the point like 'hello', it receives the request and performs the action and opening the notepad application and note down it.

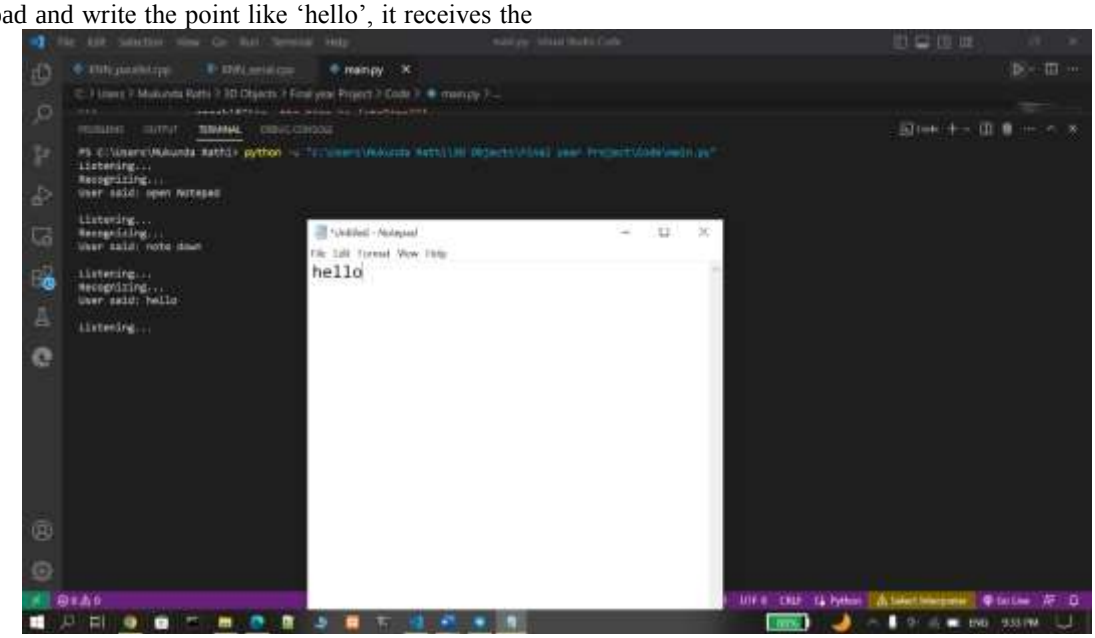

Fig. 5: Note down the points

VII. Conclusion

In our project we've got implemented several things compared to other assistants. currently a days it's very useful in human life as a result of it's a hands-free application. it's a very easy application. moreover because it is employed in a very business field also for example in laboratory, the person wears gloves and body suits for his or her safety purpose therefore it's difficult to type, through voice assistant they can get any information so their work becomes simple. Voice assistants are helpful in several fields like education, daily life application, home appliances etc. and voice assistant is additionally helpful for the illiterate people they can get any information just by locution to the assistant, luxury is obtainable for people, due to AI primarily based voice assistants. Voice assistant is developing a lot of and a lot of in existence. several companies of voice assistant trying to improve interaction and a lot of options to the next level of the youth started using voice assistant in daily life and from many sources the result showing very good feedback. Compared to last a pair of years voice assistants are developed more and more.

References

[1] 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), 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.

[2] G. Iannizzotto, L. L. Bello, A. Nucita and G. M. Grasso, "A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments," 2018 11th International Conference on Human System Interaction (HSI), 2018, pp. 50-56, doi: 10.1109/HSI.2018.8431232.

[3] T. -K. Kim, "Short Research on Voice Control System Based on Artificial Intelligence Assistant," 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1-2, doi: 10.1109/ICEIC49074.2020.9051160.

[4] C. Yan, G. Zhang, X. Ji, T. Zhang, T. Zhang and W. Xu, "The Feasibility of Injecting Inaudible Voice Commands to Voice Assistants," in IEEE Transactions on Dependable and Secure Computing, vol. 18, no. 3, pp. 1108-1124, 1 May-June 2021, doi: 10.1109/TDSC.2019.2906165.

[5] M. R. Sultan and M. M. Hoque, "ABYS(Always By Your Side): A Virtual Assistant for Visually Impaired Persons," 2019 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038603.

[6] P. Wang, "Research and Design of Smart Home Speech Recognition System Based on Deep Learning," 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), 2020, pp. 218-221, doi: 10.1109/CVIDL51233.2020.00-98.

[7] S. Ganvir and N. Lal, "Automatic Speaker Recognition using Transfer Learning Approach of Deep Learning Models," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 595-601, doi: 10.1109/ICICT50816.2021.9358539.

