International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

VOICE ASSISTANT

[1]Zaid Ahmad Khan , [2] Dr. Umesh Dwivedi , [3] Shreya Mishra , [4]Tushar Bandhiya, [5] Sourabh Chaurasia , [6] Shubhi Kesarwani , [7] Shoaib Jafri

[1]Student, [2]Professor, [3]Student, [4]Student, [5]Student, [6] Student , [7]Student,

Computer Science and Engineering,

Babu Banarasi Das Northern India Institute Of Technology, Lucknow, Uttar Pradesh, India

Abstract :- **Voice assistants are virtual companions powered by artificial intelligence that facilitate hands-free and intuitive interactions using natural language. They rely on advanced speech recognition and natural language processing technologies to comprehend and respond to user commands and inquiries. These versatile assistants offer a broad spectrum of functionalities, including information retrieval, task execution, and smart device control, making them an integral part of daily life. By leveraging extensive language models and personalized algorithms, they generate precise and contextually appropriate responses. Privacy concerns are effectively tackled through encryption and privacy controls, enabling voice assistants to evolve further and become increasingly personalized and context-aware. Their seamless integration into technology enhances productivity and convenience, revolutionizing our interactions with devices and the accessibility of information.**

Index Terms - **Artificial Intelligence, Desktop Assistant, Python, Text to Speech, Virtual Assistant, Voice Recognition**

1. INTRODUCTION :

In this era of rapidly advancing technology, machines are replacing various human tasks, primarily due to their improved performance capabilities. A significant development in this regard is the training of machines to simulate human thinking and independently execute tasks, leading to the emergence of virtual assistants.

Virtual assistants are digital counterparts that utilize voice recognition and language processing algorithms to understand user commands and carry out relevant tasks. By effectively filtering ambient noise and providing pertinent information based on user instructions, virtual assistants have become an integral part of various devices. They are predominantly software-based, although they are now integrated into diverse hardware platforms, including devices like Alexa designed specifically for voice assistance. With the advent of technologies like machine learning, deep learning, and neural networks, there is a tremendous opportunity to enhance machine capabilities.

Voice assistants have become commonplace, enabling users to interact with machines through spoken commands. This evolution has been embraced by major companies, incorporating voice assistants to facilitate user-machine interactions. The integration of voice assistants represents a significant advancement, allowing us to communicate with machines and embrace a new level of technological sophistication.

2. LITERATURE SURVEY :

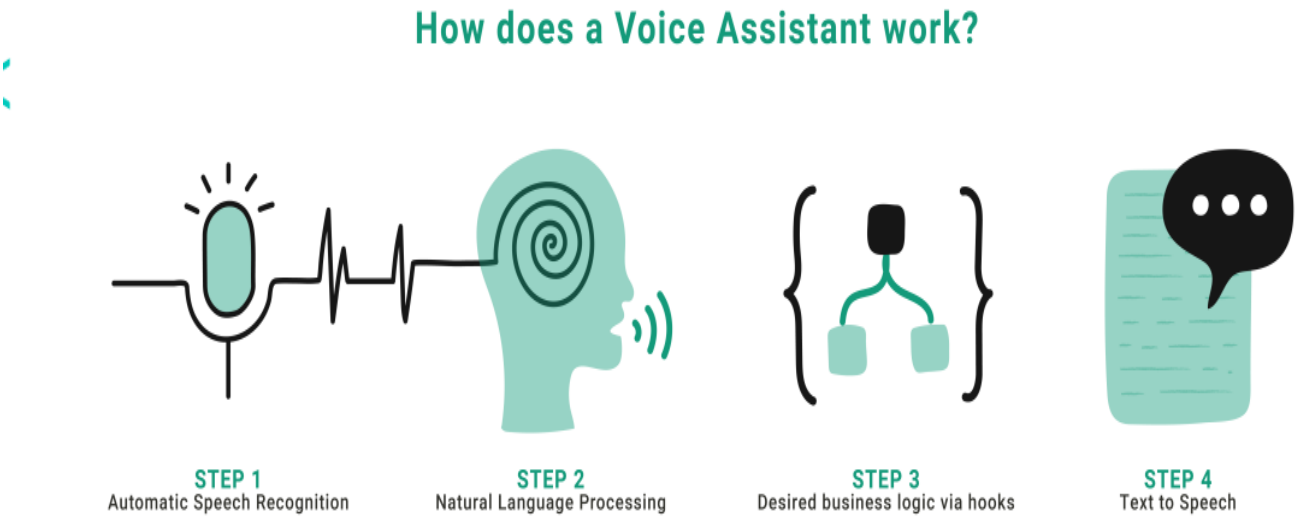
Several researchers have explored different aspects of speech and language processing, highlighting the significance of these technologies in human-machine communication. Bassam A. and Raja N. discussed the conversion of analog signals to digital waves for effective communication between humans and machines, emphasizing the widespread utilization and limitless applications of this technology. B. S. Atal and L. R. Rabiner focused on speech analysis, particularly in combination with pitch analysis, presenting a pattern recognition technique for classifying speech signals as voiced, unvoiced, or silence. V. Radha and C. Vimala highlighted the popularity of autonomous speech recognition and the use of Mel Frequency Cepstrum Coefficients (MFCC) for feature extraction in speech recognition systems. T. Schultz and A. Waiel addressed the challenges of porting large vocabulary continuous speech recognition systems to new destination languages, emphasizing the need for effective acoustic models. J. B. Allen emphasized the importance of speech as a means of communication and its significant role as an interface between humans and machines. Mugdha Bapat, Pushpak Bhattacharyya, and others described a morphological analyzer for Indian languages, specifically focusing on the Marathi language and its inflectional patterns. G. Muhammad, M. N. Huda, and colleagues presented an Automatic Speech Recognition (ASR) model for Bangla digits, highlighting the challenges of dialectical variance and gender-based trials. Finally, Sean R. Eddy and co-authors discussed the use of Hidden Markov Models (HMMs) for sequence analysis and the improvement of homolog recognition in protein reverse folding using HMM techniques. Collectively, these research studies shed light on the advancements and challenges in speech and language processing, showcasing their diverse applications and potential for future development.

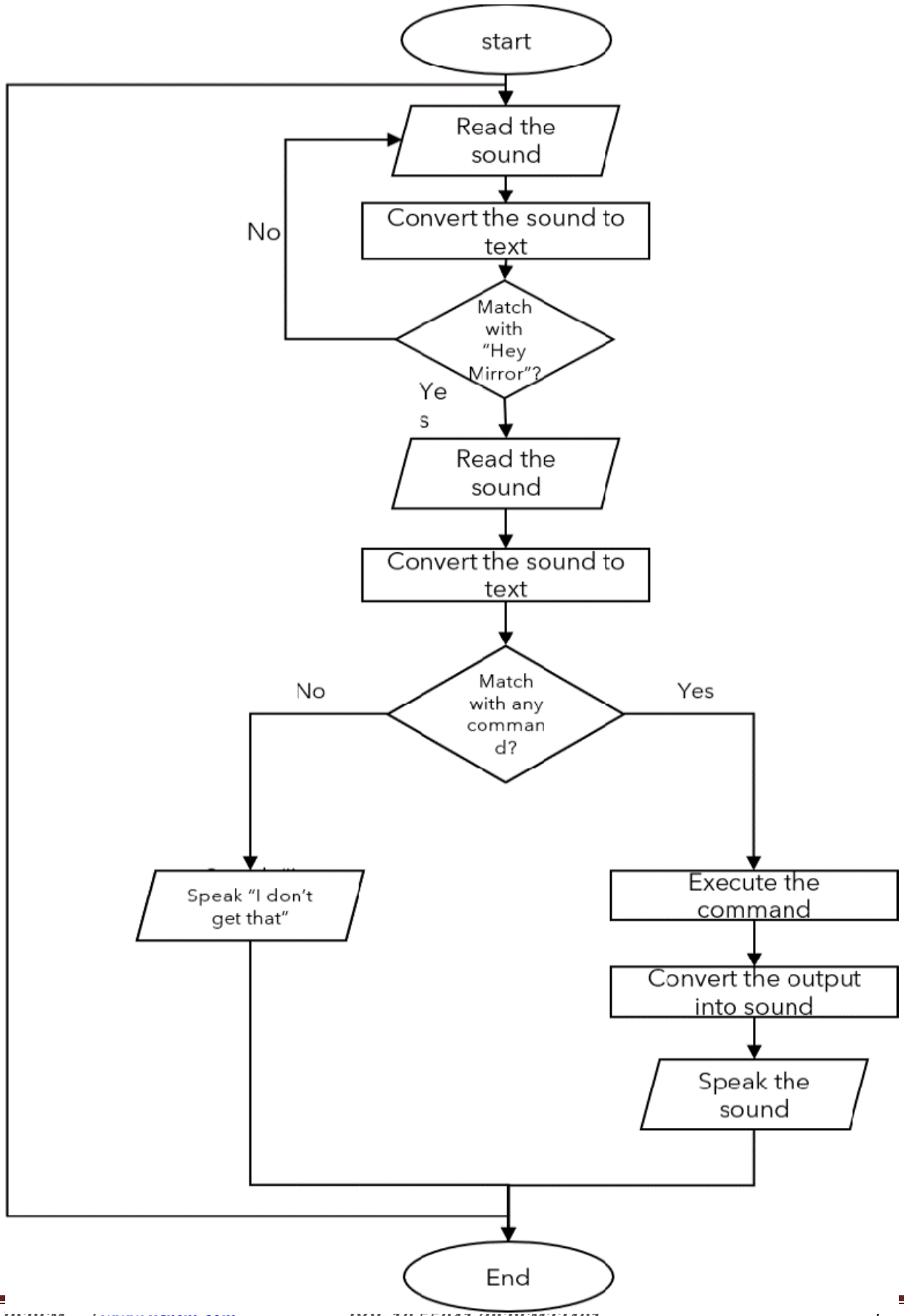
3. PROBLEM FORMULATION :

This section provides an overview of the problem formulation in the development of voice assistants. Each developer applies their own methods and approaches, resulting in assistants with varying characteristics. Some assistants excel in speech synthesis, while others are more accurate in task execution. The set of characteristics an assistant possesses depends on the developer's focus area. Machine learning and data collection play a crucial role in building voice assistants, and different algorithms may be employed. Common technologies used in voice assistant development include speech recognition, Teach-To-Speech, voice biometrics, dialog management, natural language understanding, and named entity recognition.

The proposed system aims to have several functionalities. It will actively listen for commands, with the listening duration customizable to meet user requirements. If the system fails to gather information from the user's input, it will prompt for repetition until a desired number of attempts is reached. Users can choose between male and female voices for the assistant. The current version supports features such as playing music, sending emails and texts, searching on Wikipedia, opening applications, and browsing the web.

The system implementation involves Python, machine learning, and AI. It aims to assist users in their tasks through voice commands. The process includes two phases: converting the user's audio input into English phrases using Speech Recognition API, and searching for and executing the desired task on a Linux server using the HTTP protocol.




start
Read the sound
Convert the sound to text
No
Match with "Hey Mirror"?
Ye
s
Read the sound
Convert the sound to text
Match with any comman d?
No
Yes
Speak "I don't get that"
Execute the command
Convert the output into sound
Speak the sound
End

4. METHODOLOGY

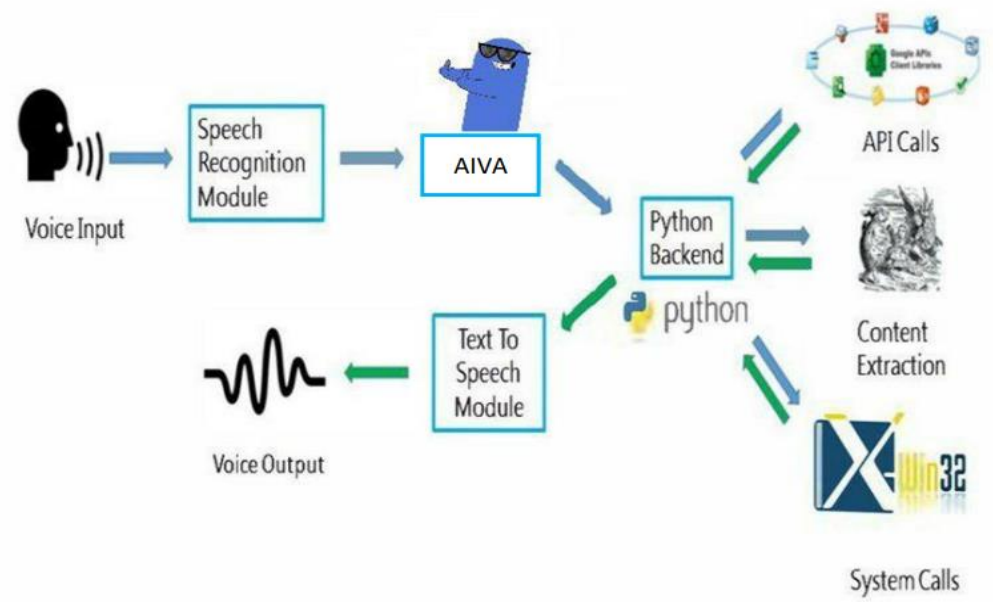


Figure : Detailed Workflow

A. Speech Recognition:

The system utilizes Google's online speech recognition system to convert speech input into text. The user's speech input is temporarily stored in the system and then sent to the Google Cloud for speech recognition. The resulting text is received and processed by the central processor.

B. Python Backend:

The Python backend module receives the output from the speech recognition module and identifies whether it corresponds to an API Call, Context Extraction, or System Call. The output is then sent back to the Python backend to provide the required response to the user.

C. API Calls:

API stands for Application Programming Interface, which enables communication between two applications. The system utilizes API calls to interact with external services or systems, sending requests and receiving responses to fulfill specific functions or retrieve information.

D. Context Extraction:

Context extraction involves automatically extracting structured information from unstructured or semi-structured machine-readable documents. In this system, context extraction is typically performed using natural language processing (NLP) techniques. It aims to process human language texts to extract meaningful information, often applied in activities like automatic annotation and content extraction from multimedia documents.

E. System Calls:

System calls are programmatic requests made by a computer program to the kernel of the operating system it runs on. These calls enable the program to access various services provided by the operating system, such as hardware-related functions, process management, and communication with kernel services. System calls act as an interface between a process and the operating system.

F. Text-To-Speech:

Text-to-Speech (TTS) refers to the capability of computers to convert written text into spoken words. A TTS engine takes written text and converts it into a phonemic representation, which is then transformed into waveforms that can be output as audible sound. Different TTS engines support various languages, dialects, and specialized vocabularies, and they are often provided by third-party publishers.

5. Conclusion :

This paper presents the design and implementation of a Digital Assistance system. The project utilizes open source software modules and is supported by the PyCharm community, allowing for future updates and improvements. The system is designed with a modular architecture, enabling easy integration of additional

features while maintaining the integrity of the current functionalities. The Digital Assistance not only responds to human commands but also provides relevant responses based on user queries or spoken words, including the execution of tasks and operations. It greets the user in a personalized manner, ensuring a comfortable and engaging interaction with the voice assistant. Additionally, the application aims to streamline user tasks and eliminate unnecessary manual work. The entire system operates on verbal input, prioritizing speech recognition over textual input.

REFERENCES :

[1] G. Bohouta, V. Z. Këpuska, "Comparing Speech Recognition Systems (Microsoft API Google API And CMU Sphinx)", Int. Journal of Engineering Research and Application 2017, 2017.

[2] Artificial intelligence (AI), sometimes called machine intelligence. https://en.wikipedia.org/wiki/Artificial_intelligence

[3] B. Marr, The Amazing Ways Google Uses Deep Learning AI.

[4] Cortana Intelligence, Google Assistant, Apple Siri.

[5] Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human– human online conversations and human–chatbot conversations. Computers in Human Behavior, 49, pp.245-250.

[6] K. Noda, H. Arie, Y. Suga, T. Ogata, Multimodal integration learning of robot behavior using deep neural networks, Elsevier: Robotics and Autonomous Systems, 2014.

[7] "CMUSphnix Basic concepts of speech - Speech Recognition process". http://cmusphinx.sourceforge.netlwiki/tutorialconcepts

[8] Huang, J., Zhou, M. and Yang, D., 2007, January. Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI(Vol. 7, pp. 423-428).

[9] Thakur, N., Hiwrale, A., Selote, S., Shinde, A. and Mahakalkar, N., Artificially Intelligent Chatbot.

[10] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp. 79-84).

[11] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology