WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An Experimental Investigation On Voice Memo Using Python

Atharv Zade[1], Upasana Badgujar[2], Aarya Shinde[3], Hitesh Yadav [4] ,Prof. Madhura Mahindrakar[5] , Prof. Tanuja Bhoir[6]

Students and Lecturer from G. V. Acharya Polytechnic, Shelu, Karjat, Dist. Raigad, Maharashtra
Corresponding Author Email id: prathmeshpatil19@gmail.com

ABSTRACT :

In the age of artificial intelligence, voice assistants are an essential component in the interaction between human and computer. The purpose of this project is to create a voice assistant based on Python that is able to accomplish a range of tasks like responding to queries, launching applications, retrieving weather information, creating reminders, and performing simple commands.

The system employs speech recognition, natural language processing (NLP), and text-to-speech (TTS) technologies to decode and answer user commands. Libraries such as Speech Recognition, pyttsx3, and OpenAI's GPT API support the assistant in comprehending and responding with human-like answers. The assistant also makes use of Google APIs and other web services to retrieve real-time information and present correct results.

The voice assistant is also user-friendly, efficient, and customizable to fit personal or professional needs. Utilizing Python's large libraries and machine learning features, the project intends to enhance day-to-day automation, accessibility, and productivity

The advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have revolutionized human-computer interaction, leading to the development of voice assistants. A voice assistant is an AI-powered software that understands voice commands and performs various tasks, such as retrieving information, controlling devices, managing schedules, and automating workflows. This project aims to develop a voice assistant using Python that can understand, process, and respond to user commands efficiently.

In the current age of speed, automation serves a crucial purpose in enhancing productivity and accessibility. Voice assistants make it easy for users to accomplish tasks without using their hands, thus being ideal in multitasking settings. Voice-based interfaces also grant accessibility to individuals with disabilities, enhancing inclusivity of technology.

This desktop voice assistant project uses Python's artificial intelligence features to enable an interactive experience based on vocal interaction. The project involves designing a voice assistant based on Python that can proficiently perform a variety of tasks including turning alarms on, delivering meteorological reports, sending electronic mails, streaming songs, etc. Designed with utmost ease of use and understandability, the assistant has a simple and smooth interface through which communication can be achieved through verbal language commands. The project leverages AI-powered voice recognition tools to enable accurate understanding and execution of commands by users. To realize the capabilities of the voice assistant, the project uses a set of Python libraries. Speech Recognition is utilized for converting words to text, and Py Audio is used to manage the auditory inputs and outputs. Text-to-speech functionality is managed through gTTS, and playback of audio files is coordinated using the play sound library

1. Introduction :

1.1General

Today, in the era of digitalization, artificial intelligence (AI) has greatly changed the way humans communicate with technology. The biggest achievement in AI, however, is the innovation of voice assistants, which allow people to control devices and apps with their voice. A voice assistant is an AI-based system that has the ability to process voice commands, understand the natural language, and execute tasks like gathering information, scheduling, sending emails, and controlling smart devices. This project is about creating a voice assistant in Python through the use of speech recognition, natural language processing (NLP), and text-to-speech (TTS) technologies. The assistant is able to identify voice commands, process them with AI-based algorithms, and return responses intelligently using synthesized speech. By incorporating a range of Python libraries and APIs, the assistant is capable of doing tasks like web search, retrieving current weather conditions, reminders, launching apps, and even controlling simple system functions.The core aim of this project is to develop an efficient, smart, and easy-to-use voice assistant that makes daily tasks simpler and increases productivity. With additional enhancements, this assistant can be extended to enable smart home automation, customized responses, and multilingual support, making it a handy tool for personal and professional use.

1.2 History of Voice Assistant

With the times of high efficiency demands, the onset of smart home automation and IoT products has kindled a craving for solutions to make daily operations smoother. Serving this purpose, virtual assistants have been introduced as AI-powered facilitators, making it possible for users to choreograph their actions and control intelligent devices using command language. This project presents a desktop voice assistant based on Python, aimed at facilitating user interaction with their computers in a seamless and efficient way. The system is developed with a range of Python libraries, including speech recognition, natural language processing, and text-to-speech modules, to create a virtual assistant that is reactive and friendly. The assistant is voice-controlled, enabling users to achieve functions like setting reminders, web searching, playing audio, and composing emails easily. The software acts as a virtual assistant, making functions like weather checking, reminder setting, and to-do lists creation via voice or text inputs easy. It needs to be activated with phrases before the user gives instructions. This voice assistant is optimized for use by all user segments to help increase productivity by performing mundane tasks and retrieving online data. Traditionally, 'virtual assistants' were web-based service providers, but the current voice assistants differ in that they target three primary functions: text-to-speech conversion, text interpretation for intent, and intent-to-action conversion. Our voice assistant is being continuously developed to increase its functionality. In contrast to human contractor virtual assistants, voice assistants are computerized systems that actively meet our needs, all thanks to the development of AI-powered voice technology.

1.3 Objective of the Study

In 21st century, everything is inclined towards automation, whether it's your house or vehicle. There is an unthinkable change rather a development in technology in the last few years. Believe it or not, nowadays in this era you can talk to your machine. What is talking to a machine? Clearly giving it some input, but what if the input is not in the form of typing, instead it is your very own Voice. What if you are conversing with the machine, giving orders and needing the machine to engage with you like your personal assistant? What if the machine is not only providing you answers by displaying you the optimal results but also by guiding you with a better option. Simple access to machine through voice commands is the revolutionary mode of human system interaction. In order to do this, we have to employ speech to text API for interpretation of input. A number of businesses such as Google, Amazon and Apple are attempting this in universal terms. Is not wonderful how reminders could be sent just by uttering remind me to. And an alarm set by waking me up at. Understanding the importance of this we have decided to create a system which can be placed anywhere in vicinity and you can request it to assist you with anything for you just by talking to it. Along with this, you can also connect two such devices with WiFi and make them communicate with each other in future

2. Review of Literature :

The paper begins with an exposition of voice assistants, tracing their rise to fame in the digital era. It follows the genealogy of the systems from the early IBM Shoebox to modern versions such as Siri, Google Assistant, and Alexa. The paper then goes into the architecture of voice assistant systems, breaking down aspects such as speech recognition, natural language processing, dialogue orchestration, and speech generation. It explains the set of methodologies and computational models that support each element, evaluating their efficacies and limitations. The discussion ends with the introduction of a voice assistant, designed using Python and AI, which can read and carry out user-given commands. Using NLP, the assistant converts voice or text inputs into functional commands, thus automating processes and saving time. In recent years, voice assistant systems based on artificial intelligence have drawn considerable attention because of their potential to transform the interaction between computers and humans. Such systems employ artificial intelligence methods to interpret and interact with user questions and commands through natural language. This literature survey seeks to examine the developments, challenges, and future trends in AI-based voice assistant systems.

Voice assistant has a very long history with multiple waves of significant innovations. Voice assistant for dictation, search, and voice commands has become a typical feature on smartphones and wearable devices. The study is based on an overlooking literature review to share generic knowledge (theory and concepts) regarding voice control, virtual assistants, fields of use and more. When considering a number of now available intelligent programs with natural language processing, there are several examples in everyday life occupying many different roles. The first speech recognition system, called Audrey, was developed by Bell Laboratories in 1952. Audrey was quite primitive and had limited technology, capable of only recognizing ten digits - uttered by specific individuals (Pieraccini, 2012). Approximately 10 years after, IBM created and showcased their Shoebox Machine. The machine identified and reacted to 16 spoken words, all of the ten digits "0" to "9" and also calculation instructions like "plus" or "minus" (IBM, 2018).
Shoebox Machine identified and reacted to 16 spoken words, the ten digits from "0" through "9", only in English by a specific speaker.

Mid 1970's arrived the Hidden Markov Model (HMM) (Rabiner,1989). The HMM significantly changed the process of constructing a viable speech recognition software. Through the aid of HMM speech recognition began employing a statistical approach quantifying the probability of unfamiliar sounds as words. Now, the ability to identify an unlimited number of words became possible because the approach permitted the amount of comprehensible words increase up to a few thousands. These observation choices of distribution in each state of the model enable precise modeling of almost unlimited types of data. The initial mass available voice command system was introduced by Apple Inc. when they launched the virtual assistant Siri in 2011 (Bostic, 2013).The smart bot Siri is available as a standard feature on Apple mobile devices today and is regarded as a fundamental element on these devices

Siri is an intelligent personal assistant that utilizes natural language processing in responding to queries and delegating requests to web services, which will then be executed on behalf of the user. Likewise, chatbot HAL was developed by Zabaware Inc to act as a computer user's virtual assistant. The bot also employs natural language processing techniques to interact with the user and note what the user is uttering in a bid to sort out the information provided to it. IBM has put much resource into this area and developed Watson, a program designed to play on the Jeopardy! TV show. This system shows the existing ability of intelligent systems with natural language recognition since it managed to defeat the two top human performers of the show. Contrary to these roles is Kari, who is a chatbot playing the role of a virtual girlfriend. This system interacts with the user and by applying similar techniques natural language recognition attempts to elicit social dialogue with the user. The program tries to provide companionship on an individual level and to simulate human interaction as accurately as possible through the use of algorithms that work to enable the program to learn from its data.

3.Methodology :

1. Voice Input & Speech Recognition

•\tThe user's voice is recorded by the system through a microphone.
•\tThe recorded audio is processed through SpeechRecognition (Python library) to translate speech into text.
•\tIf the speech is not clear or not recognized, the system asks the user to repeat the command.
Tools Used:
✓SpeechRecognition – Translates speech into text.
✓pyaudio – Records audio from the microphone

2. Natural Language Processing (NLP) & Intent Analysis

• The converted text is analyzed to identify the intent behind the user's command.
• Natural Language Processing (NLP) techniques are applied to extract relevant keywords.
• Predefined command sets are compared, and if needed, external APIs (such as OpenAI's GPT, Google Search API, or Wolfram Alpha) are used for intelligent responses.
Tools Used:
✓ NLTK / spaCy – For NLP and text processing.
✓ GPT API – For advanced conversational responses.
✓ Google Search API – For retrieving online search results.
.

3. Task Execution

• Once the intent is determined, the system translates the command to a pre-defined function.
• The actions of opening an application, retrieving weather updates, setting an alarm, sending an email, or online searching are performed accordingly.
• The assistant can also conduct basic system automation operations like opening files, volume control, or shutting down the system.
Examples of Executions:
✓ If the action is "Open Notepad", the assistant invokes the application using the os module.
✓ If the action is "Tell me the weather", it retrieves live information by invoking a Weather API.
✓ If the action is "What is the capital of France?", it searches Google or Wikipedia and responds.
Tools Used:
✓ OS & Subprocess – For opening applications.
✓ Requests – For handling web APIs.
✓ Weather API / Google API – To fetch real-time information

4. Response Generation & Voice Output

• The assistant creates a text response after running the command.
• The response is rendered as speech output by a Text-to-Speech (TTS) engine.
• The response is heard by the user as audio output (speakers).
Tools Used:
✓ pyttsx3 – Renders text to speech.
✓gTTS (Google Text-to-Speech) – Alternative to voice synthesis.

5. Continuous Learning & Improvement (Optional)

•\tThe assistant can discover user preferences with time through machine learning algorithms.
•\tThe assistant records user commands and enhances responses through past interactions.
•\tMultilingual support can be integrated for enhanced accessibility.
Advanced Additions:
✓ Machine Learning (ML) – Enhances voice understanding with time.
✓ Database Integration – Storing user preferences and logs.

✓ Voice Authentication – Incorporating security through identifying the user's voice.

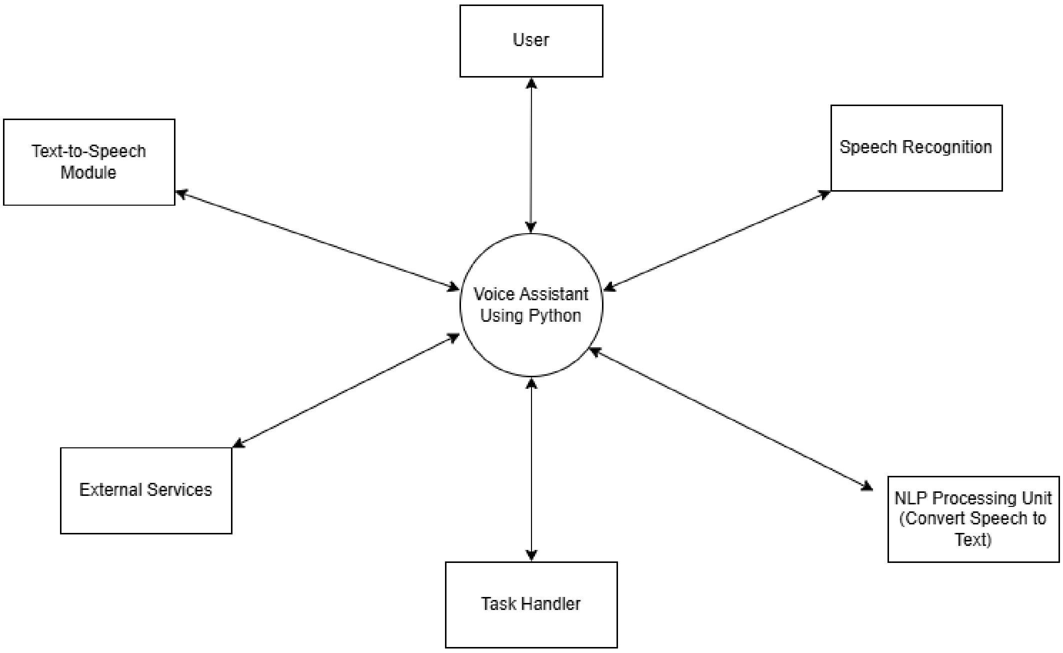
4.Result and Discussions :

The new voice assistant system as proposed will address some of the problems of current system as well as bring in new features for improved quality and usage. So, let's give a brief overview of the new revised version of the voice assistant. Rather using pattern recognition technique which we have used in all prior models, we utilize Natural Language Processing (NLP) techniques to identify the text which is context based instead of the conventional pattern based. This Runs in online and offline mode. System application operates in offline mode, while web based operations operate in online mode. Data is Stored within Application itself, not in cloud which saves Time and Space Complexity. It even decreases the economic cost because of lowering high bundles of data usage.

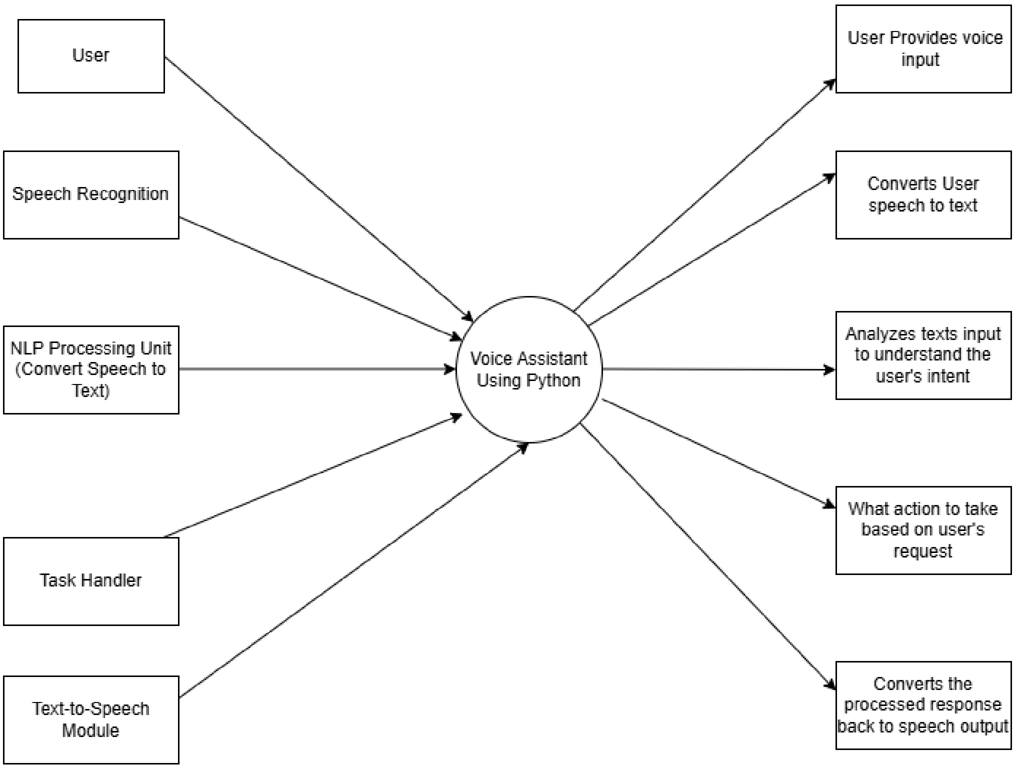
The Voice Assistant does simple operations like controlling computer tasks and operations, requesting temperature, humidity, date, time, and year. Adding, reading and deleting notes by voice commands and playing YouTube videos on demand. The above operations can be done using some methodologies in which every technique has its own functionality and various operations to be executed. Every technique has various process logic to be executed.

Diagrams:

DFD Level 0:


User
Text-to-Speech Module
Speech Recognition
Voice Assistant Using Python
External Services
NLP Processing Unit (Convert Speech to Text)
Task Handler

DFD Level 1:


User
Speech Recognition
NLP Processing Unit (Convert Speech to Text)
Task Handler
Text-to-Speech Module
Voice Assistant Using Python
User Provides voice input
Converts User speech to text
Analyzes texts input to understand the user's intent
What action to take based on user's request
Converts the processed response back to speech output

5. Conclusion :

In summary, the desktop voice assistant project represents a landmark innovation for consumers, providing a handsfree means to get information and control computer functionality. Beyond the current limitations of vocabulary range and the accuracy of speech recognition, the advancements in natural language processing and AI portend a new era for voice assistant potential. Essentially, the desktop voice assistant project contains huge potential to make user interaction better and daily activities easier, hence constituting a huge resource for both commercial businesses and individual consumers. Summarily, the desktop voice assistant project is more than just a technological tool but a transformative gadget that recasts user interfaces to digital platforms. It serves as a testament to human innovation, using the power of AI to streamline complicated tasks and improve productivity

Voice Controlled Personal Assistant System will implement the Natural language processing and can be combined with Machine learning approaches to obtain a smart assistant able to take action on different applications and will provide comfort to human life. The system will comprise the following phases: Voice as data collection form; Voice analysis and text conversion; Data processing and storage; producing speech out of the processed text output. This application will also simplify life for the physically disabled and all general users who are interested in voice recognition. In academic sense, creating awareness on systems like this for students can help students achieve deeper knowledge of subjects such as Artificial Intelligence, Neural Networks, Natural Language Processing, Machine Learning and Human Computer Interaction and also enhancing user experience in app development. The formulated solution is able to process voice commands offline allowing users to cut down on the cost of data bundles.

This also serves to make it quicker compared to other programs such as Apple's Siri, Google assistant, etc. Additionally, the solution is able to perform a wide range of tasks without any difficulty such as giving the date and time, playing videos/music, dialing phone numbers, locating weather, temperature, googling facts etc. This paper may also serve as a prototype to most sophisticated programs.

6. REFERENCES :

1. Speech and Language Processing by Daniel Jurafsky& James H. Martin – Covers NLP and speech recognition basics.
2. Python Machine Learning by Sebastian Raschka – Discusses machine learning principles that can be applied to enhance voice assistants.
3. A Hands-On Introduction to Learning with Scikit-Learn, Keras, and TensorFlow by Aurélien Géron – Good for applying AI-powered voice assistants.
4. A Review on Virtual Voice Assistants and Their Technologies(IEEE) – Discusses voice assistant architectures.
5. Speech Recognition: A Review(IJCA) – Explores speech recognition techniques used in modern voice assistants.
6. Artificial Intelligence-Based Smart Assistants (Springer) – Covers AI advancements in virtual assistants.
7. Allen, A. A., Shane, H. C., & Schlosser, R. W. TM (2018). The echo as a speaker- independent speech recognition device to support children with autism: An ex ploratory Study. Advances in Neurodevelopmental Disorders, 2(1), 6974. doi:10. 1007/s41252-017-0041- 5.
8. Angelini, L., Caon, M., Carrino, S., Bergeron, L., Nyffeler, N., Jean-Mairet, M., &Mugellini,E. (2013). Designing a desirable smart bracelet for older adults. In Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing adjunct publicationUbiComp '13 Adjunct. In UbiComp '13 Adjunct (pp. 425–434). New York, New York, USA: ACM Press. doi:10.1145/2494091.2495974.
9. Azvine, B., Djian, D., Tsui, K. C., &Wobcke, W. (2000). In B. Azvine, D. D. Nauck, & N. Azarmi (Eds.), The intelligent assistant: An overview (pp. 215–238)). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/10720181_9.
10. Bellegarda, J. R. (2014). Spoken language understanding for natural interaction: The Siri experience. In J. Mariani, S. Rosset, M. Garnier Rizet, & L. Devillers (Eds.), Natural interaction with robots, knowbots and smartphones (pp. 3–14). New York, NY: Springer New York. doi:10.1007/978- 1- 4614-8280- 2_1.