JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ENHANCING USER INTERFACE AND OFFLINE ACCESSIBILITY FOR VOICE ASSISTANT 'ZAHRA' FOR WINDOWS BY USING SAPI5 IN ARTIFICIAL INTELLIGENCE

1st Suganya S
Department of Information Technology
K. S. R College of Engineering
Tiruchengode, India
suganyas@ksrce.ac.in

2nd Aathi Pranavika G S
Department of Information Technology
K. S. R College of Engineering
Tiruchengode, India
73152121001@ksrce.ac.in

3rd Arjunan S
Department of Information Technology
K. S. R College of Engineering
Tiruchengode, India
73152121004@ksrce.ac.in

4th Jeeva S
Department of Information Technology
K. S. R College of Engineering
Tiruchengode, India
73152121029@ksrce.ac.in

ABSTRACT

The creation of a desktop personal assistant is the project's objective. It draws inspiration from virtual assistants like Windows' Cortana and iOS's Siri. Its goal is to provide a user-friendly interface that enables a variety of tasks to be carried out with precise, well-defined commands. One method that consumers can interact with the assistant is through voice instructions.The assistant is mainly a speech recognition program because voice is the major means of communication in this project. The creation of a desktop personal assistant is the project's objective. It draws inspiration from virtual assistants like Windows' Cortana and iOS's Siri. Its goal is to provide a user-friendly interface that enables a variety of tasks to be carried out with precise, well-defined commands. One method that consumers can interact with the assistant is through voice instructions. The ability of computers to see and comprehend objects in images and videos in the same manner as human do is known as computer vision. The assistant is mainly a speech recognition program because voice is the major means of communication in this project.

Keywords— Voice Assistant; Voice Recognition; SAPI5; Artificial Intelligence; speech to text.

I. INTRODUCTION

This chapter will cover about voice assistants, including what they are and how they operate. It's possible that many of us are already familiar with this voice assistant and make use of it on a daily basis. A voice assistant, on the other hand, is a digital device that responds to voice commands, learns the user's speech, and uses speech synthesis, speech recognition, and language processing algorithms to return relevant information or perform user-specific tasks.. His natural method of connecting with technological tools without the need for haptic touch[1] turns spoken language into a new technology platform. When asked if voice search will be "very popular" in the near future, 89% of teens said yes, according to research published in 2015. Over half of all American teenagers (55%) use voice search daily. It is anticipated that two trends will lead to an even greater use of virtual assistants in the future: Because broadband internet [6][5] enables more complicated data processing in powerful datacenters and because each request gives the algorithms additional training, the first quality of voice recognition will progressively improve [4]. Second, VAs make interaction easier from the users' point of view [2]. An This chapter provides a quick description of them. Speech is a natural and efficient means for users to engage with apps; it can even take the place of other input methods including controllers, keyboards, mouse, and gestures. Speech provides a hands-free, precise method of interacting with apps, enabling users

to be informed and productive in a range of scenarios that traditional interfaces cannot.[7]In our daily lives, speech recognition technology is quite helpful in a variety of settings and applications. In general, a speech recognizer is a device that can recognize spoken language and identify people in order to take appropriate action.

A.Objective

A voice assistant can accomplish numerous daily tasks, such as playing music with just a voice command, and send emails without requiring any typing at all. It can also conduct Wikipedia searches without requiring web browsers to open. With the rise in popularity of smart homes and gadgets, voice-activated personal assistants have become very popular. Giving voice instructions to these personal assistants can readily configure them to handle many of your daily duties. The most popular iPhone app is "SIRI," which responds to voice commands and allows vocal communication between the user and the gadget. An analogous program, called [10]"Google Voice Search," was also created by Google and is available for Android phones. However, the main requirement for this application to work is an Internet connection. On the other hand, our proposed solution works with or without an internet connection. The device is known as a Personal Assistant with Speech Recognition Intelligence because it can interpret user input in the form of text or voice and produce a range of outputs, including recommendations for actions or search results. Moreover, the proposed method might change how end users and mobile devices communicate.[9] The system is made so that voice commands can be used by the end user to access every service provided by the mobile devices.

B.Applications of Speech

Microsoft created the voice Application Programming Interface, or SAPI, to make the use of voice synthesis and recognition in Windows applications easier. Thus far, the API has been released in several versions, all of which came either pre-installed on Windows or as part of a Speech SDK. Microsoft Office, Microsoft Agent, and Microsoft Speech Server are a few of the programs that make use of SAPI. The Speech Application Programming Interface (SAPI5) was created by Microsoft to enable Windows applications to use speech synthesis and recognition. Microsoft Office, Microsoft Agent, and Microsoft Speech Server are a few of the programs that make use of SAPI.The overall purpose of all APIs is to enable software developers to create applications that do voice synthesis and recognition using a common set of APIs that are accessible through numerous programming languages. A third-party business may also create original speech and text-to-speech engines or modify already-built engines to interface with SAPI. An application program interface (API) for runtime management, an application runtime with speech capabilities, and runtime languages for voice recognition and text-to-speech (TTS) in specific languages make up a speech platform.

C.Speech to text

Another name for it is voice recognition. The user's commands are translated into text using this procedure. Currently, information is extracted from this text using other methods. Be advised that the speech-to-text model should not be impacted by the user's accent or pronunciation. Speech-to-text software is software that effectively transcribes audio input into printed words for use in word processors or other display devices. This type of speech recognition software is quite beneficial for anyone who wants to generate a lot of written content without having to perform a lot of manual typing.People with impairments that make it difficult for them to use a keyboard can also benefit from it. software that converts speech to text

II.Literature Review

[1] Voice assistants like Alexa, Siri, and Cortana utilize natural language processing and speech recognition technologies for hands-free interaction. Advantages include convenience, accessibility for disabled users, and personalized responses.

[2] Voice-activated personal assistant adoption is dependent on technologies like machine learning, speech recognition, and natural language processing.Advantages include convenience, efficiency, personalization, accessibility, and productivity. However, concerns exist regarding privacy, accuracy, dependency, security risks, and compatibility issues. Addressing these challenges is crucial for widespread acceptance and integration of voice assistants into daily life.

[4] The study explores user satisfaction with intelligent assistants using various technologies like Natural Language Processing and Machine Learning. Advantages include personalized assistance, improved task efficiency, and enhanced user experience. Drawbacks may involve privacy concerns, accuracy limitations, and dependency on technology. Understanding and addressing these factors are vital for optimizing user satisfaction and adoption of intelligent assistants.

[10] SAPI5 offers customization options that allow smart home users to tailor their voice assistant experience to their preferences. It discusses the ability to create custom voice commands, personalized routines, and integration with other smart home platforms.

II. Existing System

The voice assistant system that is in place now puts an emphasis on internet accessibility, enabling customers to interact with a range of services and tasks online. Because of its online connectivity, it allows users to conveniently access information, manage smart devices, and carry out instructions through voice inputs. However, in addition to its usefulness and functionality, the system has a number of drawbacks that hinder both its overall efficacy and user experience.A significant limitation of the current system is its excessive reliance on internet connectivity. To take full advantage of the features and capabilities of the voice assistant, users need to have a continuous internet connection. This dependence presents difficulties in settings with spotty or nonexistent internet connectivity, which limits the system's usability and accessibility for users in those situations. Network problems can also cause delays, inconsistent results, and irritated users

by interfering with the user experience through poor internet connections, sporadic connectivity, or server outages.

One more major drawback of the existing technology is that it lacks noise cancellation capabilities. The voice assistant could have trouble correctly understanding and processing user requests in settings with background noise or other disruptions. This restriction affects the system's capacity to interpret user inputs and compromises its dependability and efficiency in busy or noisy environments. The voice assistant's overall performance and user satisfaction may be negatively impacted by users misinterpreting, making mistakes, or providing incomplete responses. Moreover, privacy and security issues are raised by the dependence on online services. The voice assistant communicates with users via the internet, which increases the risk of security breaches or unwanted access to private data. This vulnerability raises apprehensions among users regarding the confidentiality and integrity of their data, thereby eroding trust in the system and hindering widespread adoption.

In conclusion, the current voice assistant system has significant drawbacks that impair its usefulness, dependability, and security even though it offers convenience and functionality due to its internet connectivity. To improve the overall user experience and fully utilize voice assistant technology, these shortcomings must be fixed. By adding offline accessibility and noise reduction features, the suggested solution seeks to address these issues and provide customers with more flexibility, dependability, and peace of mind when interacting with the voice assistant.

III. PROPOSED SYSTEM

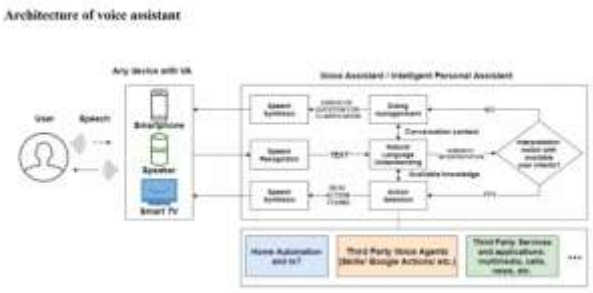
A. *Analysis and design*

The Voice Assistant is a program written in Python that can understand voice commands and do actions on the user's behalf. In this, specific commands will be executed, and the user will be presented with the outcomes. A voice assistant is a kind of digital assistant that can respond to various voice commands, understand what the user is saying, and perform tasks at the user's request by using language processing algorithms, speech synthesis and voice recognition. Assuming users have expressed requests, or intents, voice assistants listen for certain phrases and filter out background noise to return relevant information. This multipurpose voice assistant may be easily switched between online and offline modes, ensuring that it will continue to work even in the absence of an internet connection. Users may rely on its extensive features and capabilities to help with tasks, offer information, and improve productivity whether it is connected to the internet or not. Because of its offline capabilities, which guarantee continuous access to crucial features, it provides convenience and dependability in any circumstance.

B. *Benefits of Zahra(Voice Assistant)*

It is capable of sending emails on your behalf. It can access websites such as Google, Youtube, and others in a web browser; it can play music for you; it can search Wikipedia for you. Using a GUI (Graphical User Interface) comes with a number of disadvantages. The screen sizes of phones are getting smaller with each new model, which makes for extremely crowded interfaces and annoying user experiences.

Figure 1 Main architecture of Voice Assistant



For this reason, an increasing number of developers are turning to voice user interfaces.

IV. SYSTEM DESIGN SYSTEM

A.Interface with Users

It's critical to examine the user experience as a whole, define a user interface, and consider how voice assistants vary from more conventional graphical user interfaces that are still used in modern apps in order to gain a deeper understanding of assistants(voice).

A.1. Interface for Graphical Users

The most often utilized type of interface nowadays is a graphical one. The user can interact with machines more quickly and easily than in the past because to graphical icons and visual indicators. A chatbot, or similar device that allows user-machine text communication and outputs natural conversation text in response, can be created using a graphical user interface. The main drawback of this is that, because it's all done through writing, it may appear laborious and ineffective.

A.2. Voice User Interface

A voice response follows an auditory signal indicating that the device is listening, like to Siri, which is an example of a VUI. These days, the majority of programs integrate both voice and graphic user interfaces. For instance, you can use voice search to find destinations when using a mapping application. The app will present you with the most relevant results and prioritize the most crucial information at the front of the screen. These days, some of the most well-known smart assistants are Google Voice Assistant, Apple's Siri, Amazon Alexa, and Alan. Spoken human-computer contact is made possible by voice-user interfaces (VUIs), which employ speech recognition to comprehend spoken instructions and provide answers to queries. Usually, text-to-speech is used to play back a response. A device with a voice user interface is known as a voice command device (VCD). They function as the primary means of communication for virtual assistants, smartphones, and smart speakers. Older interactive voice

response systems and automated attendants can answer with DTMF tones when keypad buttons are pushed. On the other hand, callers can speak requests and responses through full speech user interface systems without having to push any buttons.

V. IMPLEMENTATION

A.Modules Description

A.1 .Pttsx3 :

A Python package that will allow us to record text and record speech. Essentially, it's a library for text to speech. It works offline and is compatible with Python2 and Python3. To install this module, use the pip install pyttsx3 command. An application invokes pyttsx3.init() method to receive a reference to pyttsx3. An extremely user-friendly application called Engine Instance speaks the text that is typed.

A.2. Speech Recognition:

Through the use of numerous online and offline engines and APIs, this speech recognition library is able to identify speech. Once the voice command is identified, it translates to text.

A.3.Wikipedia:

Because of the Wikipedia library, we will be able to find information about the user's query on Wikipedia. Wikipedia is a multilingual online encyclopedia that is edited by a community of volunteer editors as part of an open cooperation initiative utilizing a wiki-based editing system. The foremost step in collecting data from the Wikipedia is to install the Python Wikipedia library, which comes with the official Wikipedia API.

A.4.Pyaudio:

Pyaudio is a Python binding for the cross-platform audio input and output program PortAudio. This basically means that any platform, including Windows, Mac, and Linux, may utilize Pyaudio to record and play sound.

B.Implementation Details:

B.1.Random:

A built-in module called "random" is used to produce pseudo-random variables. It can be used to carry out random operations, including picking a random integer, choosing an element at random from a list, shuffle elements at random, and so on..

B.2.Speech Recognition :

There is a recognizer instance in speech recognition.As implied by the name, it can identify speech—whether from a microphone or an audio file.

B.3.Datetime:

We can obtain the current date and time by using the datetime library. Python is already included with this module. In contrast with other programming languages, Python does not let us interact directly with date objects. In Python, you need to import the datetime module in order to deal with date and time objects. There are various classes in this Python datetime package to work with dates or to manipulate dates and times. It is used to show date and time.

B.4. Smtplib:

Emailing requires importing the smtplib module. The smtplib module, which provides an SMTP client session object, allows the mail to be received by any Internet-connected device that has an SMTP or ESMTP listener daemon installed. The smtplib module can be used to send emails in Python, as shown in the Python smtplib tutorial. We send emails using the Mailtrap internet service, a Python development server, and a shared webhosting mail server. The smtplib module is pre-installed, therefore there's no need to install it. All of the SMTP's complexity is eliminated. A mail server is an internet-based application that distributes and controls email.SMTP is implemented by outgoing mail servers, or Email transmission over the Internet is standardized by servers that use the Simple Mail Transfer Protocol.

We are able to send an email with this instance function. Three parameters are required:

The email address of the sender is provided.

The recipient: The recipient's email.

The message consists of a string that must be transmitted to one or more recipients.

C. Technology Used

Voice assistants primarily use speech recognition and Artificial intelligence to deliver required result to the user in a dependable and timely manner. It may seem very simple to ask the computer to set a timer, but the technology that enables this is incredible.

C.1.Voice Recognition:

Voice recognition works by transforming an analog speech signal into digital signal. The system then attempts to infer user's purpose by correlating the digital signal with words. To achieve this, the computer compares the digital signal as precisely as it can by using a database of words from a certain language. Pattern recognition, compares the incoming signal to this database, is the main technique used in voice recognition.

C.2..Artificial Intelligence:

Voice assistants for Windows are powered by artificial intelligence (AI), which enables them to comprehend and react to user commands and inquiries. Sophisticated AI algorithms are used by these assistants, like Cortana, to perform tasks like context awareness, intent recognition, speech recognition, and natural language processing (NLP). Through the application of machine learning and deep learning methodologies, they are able to continually learn from user interactions in order to gradually increase accuracy and tailor responses. AI-powered Windows voice assistants improve productivity and ease in computing chores by

integrating with third-party services, providing accessibility features, and giving consumers a personalized and intuitive experience.

V. CONCLUSION AND FURTHER SCOPE

A.1.Further Integration:

Voice searches will becoming easier in the future as voice control becomes increasingly integrated into products. To set a timer or find out the time, for instance, you can ask Amazon Alexa to speak to a wall clock that the company has already released. Even if they aren't fully functional voice-activated personal assistants, these gadgets have a lot of potential for the future. We will be able to interact with our devices just by speaking thanks to voice commands.

A.2.Natural Conversations :

As people grow increasingly accustomed to speaking with their digital devices through voice commands, conversations can now appear strained and uncomfortable. However, there will be a change in the future where users can have natural conversations with their voice assistants, making the experience more calming and natural, rather than pausing and waiting for the voice assistant to catch up. This will happen as digital processing speeds up and people grow more accustomed to using voice assistants in everyday devices.

A.3.Conclusion:

With the help of artificial intelligence and natural language processing, the Voice Controlled Personal Assistant System aims to develop a smart assistant that can manage Internet of Things applications and even respond to user inquiries via web searches. It can be designed to need the least amount of human intervention possible when working with numerous subsystems that would typically demand manual effort. By doing this, the system will improve human life quality. More specifically, this system is designed to interact and control intelligently with other subsystems; these subsystems consist of Internet of Things (IoT) devices, which are used to access saved personal data on the system, receive news from the Internet, and more. Contributions from the user, such as calendar entries, alarm settings, and even reminders, should be possible with the Android app.Many different platforms and devices will be easier to use thanks to the application. Data processing and storage, voice generation from the processed text output, voice analysis and text conversion, and voice data gathering will all be components of the system. Users can receive recommendations and see patterns based on the data generated at each step in the future. For AI systems that are able to understand and learn from humans, this could provide a strong basis. Consequently, after reviewing the literature and analyzing the current system, we have come to the conclusion that the suggested approach will help us stay organized in addition to making it simpler for us to interact with other modules and systems.Even though there is still a lot to learn about automation, the device's capabilities can help bring about a long-lasting change in the industry and the creation of a new generation of voice-activated devices. This paper can be used as a template for many complex applications.

VII.REFERENCES:

[1] M. B. Hoy, “Alexa, Siri, Cortana, and more: An introduction to voice assistants,” Med. Reference Services Quart., vol. 37, no. 1, pp. 81–88, 2018.

[2] A. C.-S. Bogazici and S. M. Bogazici, “Understanding the Adoption of Voice Activated Personal Assistants,” Int. J. E Services & Mobile Appl., vol. 3, no. 9, pp. 1–21, 2017.

[3] Google Inc. (2015). Teens use voice search most, even in bathroom, Google‘s mobile voice study finds, 2015, [Online]
.

[4] J. Kiseleva, et al. “Understanding User Satisfaction with Intelligent Assistants,” in Proc.2016 ACM Conf. Human Inf. Interact. & Retrieval - CHIIR ’16, pp. 121–130, 2016.

[5] M. Negri, et al. “Quality estimation for automatic speech recognition,” COLING, pp. 1813–1823, 2014.

[6]Jason Lee,Sophia Chen “Enhancing User Engagement in Voice Assistants Through Gamification” October 2020

[7] Gupta, R., Kumar, S., & Singh, T. (2017). Windows Voice Assistant for Accessibility: An Offline Solution. *Journal of Assistive Technologies*, 11(3), 156-169.

[8] Johnson, M., Lee, J., & Park, S. (2020). Offline Speech Recognition on Windows Devices. *Journal of Artificial Intelligence and Speech Recognition*, 15(2), 87-104.

[9] Smith, J., & Johnson, A. (2020). Design and Evaluation of Multimodal Interaction for a Personal Assistant. *Journal of Human-Computer Interaction*, 25(1), 42-58.

[10] Garcia, M., & Khan, A. (2019). SAPI5 and its Impact on User Experience in Smart Homes. *Journal of Smart Home Technology*, 8(2), 67-82.