Proceedings of the Second International Conference on Electronics and Renewable Systems (ICEARS-2023).
IEEE Xplore Part Number: CFP23AV8-ART; ISBN: 979-8-3503-4664-0

2023 Second International Conference on Electronics and Renewable Systems (ICEARS) | 979-8-3503-4664-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEARS56392.2023.10085043

Voice-based Virtual Assistant with Security

Cintamaria Simon
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, Tamil Nadu, India
cintamariasimon@karunya.edu.in

Dr. M. Rajeswari, M.E, Ph.D
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences
Coimbatore, Tamil Nadu, India
rajeswari@karunya.edu

***Abstract*— With the advancements in speech recognition and AI technology, there is a growing demand for convenient and efficient ways to interact with technology. A Voice-based Virtual Assistant is a technologically advanced solution that uses speech recognition and artificial intelligence to provide users with a convenient and efficient way to interact with devices, access information, and perform tasks. It makes use of the GPT-3 language processing model by OpenAI to respond intelligently. With the help of spoken commands and natural language processing, users can communicate with the virtual assistant using this technology, which is intended to function as a chat interface. However, as with any technology that collects and stores personal information, security is a major concern. The added security feature addresses this concern by incorporating encryption, authentication, and access controls to ensure that personal data remains secure. This not only protects the user's privacy but also helps to prevent unauthorized access to sensitive information, such as bank account numbers, passwords, and other confidential data.**

Keywords—Automatic Speech Recognition, Artificial Intelligence, Natural Language Processing, Virtual Assistant

I. INTRODUCTION

The idea of a smart assistant has grown significantly in popularity during the past ten years. Commercial products like Amazon Alexa, Google Home, and Mycroft may communicate with users using speech synthesis and recognition, provide a number of network-based services, and connect to smart home automation systems to give them a more sophisticated user interface. Such speech-enabled smart assistants are becoming more and more common, largely as a result of the accessibility of several network services and the rising number of new talents or capabilities that can be quickly added to the smart assistants.

Due to its time-consuming nature, accessibility restrictions, lack of multitasking, increased risk of error, and physical strain, voice-enabled gadgets are becoming more and more popular. A virtual assistant is a technology solution that provides users with a convenient and efficient way to interact with devices, access information, and perform tasks. This technology allows users to interact with the assistant through voice commands, text inputs, or touch-based interactions.

Virtual assistants have become increasingly popular in recent years, as users are seeking more convenient and efficient ways to interact with technology. Many other uses for virtual assistants are possible, such as entertainment, information retrieval, home automation, and personal productivity. They can be personalized to each user's unique needs and incorporated with other technologies and services, such as smart home systems and virtual personal assistants, to offer a complete and integrated experience.

Virtual assistants is being integrated into a variety of gadgets, such as smart speakers, smartphones, and home appliances. Additionally, they are employed in a wide range of settings, including hospitals and schools as well as offices and residences. This will change the way we engage with technology and make it easier for us to complete jobs.

The purpose of virtual assistant is to provide a solution that is both accessible and simple to use for people with disabilities while also addressing the growing demand for convenient and secure methods to connect with technology[15]. Many technological solutions currently available are inaccessible to people with disabilities, such as the blind. By adding speech recognition technology and making it simpler for people with impairments to engage with their devices and access information, the virtual assistant tries to alleviate this problem. Privacy issues have become more common as more personal data is stored on devices and in the cloud.

II. RELATED WORKS

Virtual assistants have become increasingly popular in recent years, with many tech giants such as Amazon, Google, and Apple investing heavily in their development. Amazon Alexa, Google Assistant, Apple Siri, Microsoft Cortana, Samsung Bixby, IBM Watson Assistant, are are just a few examples of the many voice-based virtual assistants that have been developed and studied in recent years.

Numerous studies [11] have concentrated on enhancing the natural language processing and generation skills of virtual assistants to make them more human-like in their interactions. Integration with internet of things (IoT) devices has been the subject of numerous studies in an effort to give seamless control over smart home appliances and other linked devices. Virtual assistants for healthcare have been developed to help people manage their drugs, keep track of their symptoms, and communicate with their healthcare providers, among other tasks [7].

IBM Watson is used to create a chatbot or virtual assistant that can answer consumers' questions and give them pertinent information. With the help of Watson's natural language processing (NLP) abilities, users may ask questions and receive precise answers[11].

One of the main concern with a virtual assistant is security. Some assistants might make use of facial recognition to boost security and provide authentication. However, it also poses privacy issues because it might be able to gather and exploit private biometric information without the user's knowledge or approval[14].

Many of the system's limitations [12] [13], was that the study was initially limited to archival analysis. Second, a lot of primary data was gathered from blind persons regarding the difficulties and barriers they encounter when utilizing the frameworks of educational services now in place. Another

drawback where that most of the systems operated on mouse clicks, thus making it not fully voice activated [1].

III. LITERATURE SURVEY

Technology Used	Advantages	Drawbacks
[1]Speech Recognition, Python Backend, WolframAlpha, gTTS, Pyaudio, Smtplib	● Listen to the users voice only and will not be activated from environment noise. ● Add more features in the program without disturbing the functionalities.	● Doesn't support email features. ● Supports only English language.
[2]Speech Recognition, PYTTSX3	● Entire system is based on verbal input. ● Voice instructions can only be accessed by permitted individuals, according to the framework	● Current systems understanding and reliability need to be greatly enhanced. ● Doesn't support multi-lingual application. ● Capability is limited to just working online.
[3]Speech Recognition, Speaker Identification, AIML, gTTS	● Capable of recognizing the user's voice its gesture. ● Challenges in speech recognition like environment factors were removed.	● Designed for Windows OS only.
[4]Automatic Speech Recognition, Google Dialogflow	● Supports various applications. ● Object detection feature.	● No security feature.
[5]Speech Recognition, Python Backend, Text-to-Speech, Content Extraction	● It will prompt the user to repeat the process if the system is unable to extract information.	● Voice assistant does not work online. ● System does not provide voice-based mail delivery.
[6]Speech Recognition, Python Backend, Text-To-Speech(TTS)	● Provides assistance to disabled personalities over the website. ● Start up the voice assistant by the proposed wakeup word.	● Not be capable of choosing the correct meaning as intended by human beings as they lack emotions and senses. ● Issue of privacy arises from anyone can access a voice activated device. ● Limited to high-speed networks. ● No read-aloud feature.
[7]AT&T Speak4it R voice search, Voice Recognizer	● Capability to work with and without internet connectivity. ● It directly can be opened by pressing power button. ● Language barrier independent.	● Doesn't provide a full PC functionality.
[8]Speech recognition, TTS, STT	● Entire system works on the verbal voice input.	● Doesn't include modern technologies. ● Less accuracy.
[9]Speech recognition systems, Speech recognizer and synthesizer, Natural language processing	● Users can get inside/outside of anything just by speaking out, proper voice commands and speedy computation in a secure manner.	● It produces plenty of data, encryption of them can increase processing load and time.
[10]Speech Recognition, AKIRA Module, Google Text to Speech	● It can be easily deployed and used by any operating system. ● Voice-based authentication password.	● It does not support social gaming. ● The accuracy and security features need to be improvised.

IV. METHODOLOGY

Here is an high-level overview for creating a voice-based virtual assistant with added security feature.

1)Voice Data Collection: The process starts with gathering and storing a large number of voice samples from authorized users in a database to train the model.

2)Voice Biometric Model Training: With the help of the voice data gathered, machine learning algorithms are trained to develop a model that can identify every authorized user based on their particular voice traits.

3)Integration of Voice Biometric model into Virtual Assistant: The speech biometric model records and analyses a user's voice when they talk to the virtual assistant to determine whether they are an authorized user.

4)Voice Interaction Design: The virtual assistant is designed to interact with users via voice commands. It can use speech recognition technology to convert voice commands into text and natural language processing (NLP) to understand the user's intent.

5)User Interface Design: The verbal commands of users can be used to communicate with the virtual assistant. It can grasp the user's intent using NLP and speech recognition technology to translate voice commands into text.

6)Email System Integration: The virtual assistant is integrated with an email system to provide a fully accessible email experience for blind users. The text-to-speech (TTS) and speech-to-text (STT) technologies can be used by the virtual assistant to communicate with the user and deliver audio feedback.

V. PROPOSED SYSTEM

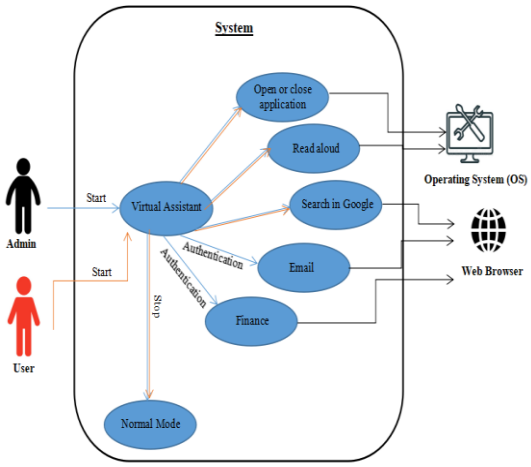

Fig. 1. UML diagram

As shown in the Fig. 1, the user will speak the commands they want to run rather than typing them in. The assistant utilises a wakeup phrase to start the process of full voice activation. The software then creates the command that will be performed by translating the speech input into text. Supported actions include playing music, sending emails or SMS, browsing Wikipedia, using system-installed programmes, viewing any website through a web browser, and more. In secure apps like email, finance, and other areas, the user must authenticate verbally. If only the voice matches the voice saved in the system, the system permits the user to proceed. It provides simple accessibility through integration with many desktop functionalities.

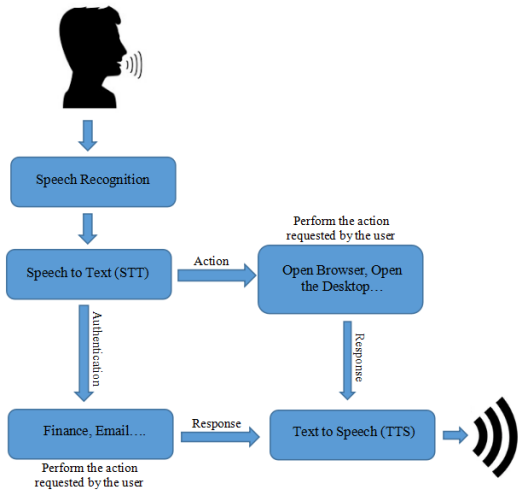

Fig. 2. High-level architecture diagram

It makes use of the following concepts to achieve these results as given in Fig. 2:

- Speech recognition: To effectively understand the user's speech and respond appropriately, virtual assistants use advanced algorithms and machine learning techniques for speech recognition. It includes receiving audio input from the user's device, pre-processing the audio signal to remove background noise and improve the signal, converting the audio signal into a set of features that represent the characteristics of the speech signal, creating a statistical model of the sound units, such as phonemes or words, based on the extracted features, predicting the likelihood of a sequence of words given the preceding words in a sentence and finally to use the models to decode the speech input and generate a text output. Using SpeechRecognition API in python, speech-to-text conversion is achieved in the system.
- Voice biometrics: The system would use voice biometrics for authentication, allowing users to access their virtual assistant using their voice. GMM (Gaussian Mixture Model), which is used to train on extracted MFCC characteristics from audio files, is utilized to achieve voice biometrics authentication. When performing speech recognition tasks, features called Mel-frequency cepstral coefficients (MFCC) are taken from speech signals. The user's newly derived MFCC vector is compared to previously saved pre-trained GMM models to establish authentication.This would increase security by guaranteeing that only the authorized user may access the virtual assistant and any services that go along with it.
- Natural language processing (NLP): NLP techniques would be used by the system to decipher the user's intent and offer relevant responses. Generative Pretrained Transformer 3 (GPT-3) is the language processing model used in the system that can produce text responses that are human-like to a variety of stimuli. The size of the GPT-3 is what distinguishes it from earlier variants. GPT-3 is 17 times larger than GPT-2 and has 175 billion parameters. GPT-3 has been trained on an exceptionally huge dataset, which aids in executing a variety of functionalities such as generating answers to queries and tasks it has never seen before

and being built to produce natural language text without any additional training.

- Text-to-speech synthesis: The system would produce human-like speech via text-to-speech synthesis, making it simpler for the user to comprehend and communicate with the virtual assistant.

VI. System Architecture

A voice based virtual assistant can be designed and implemented with the following architecture as given in the Fig. 3:

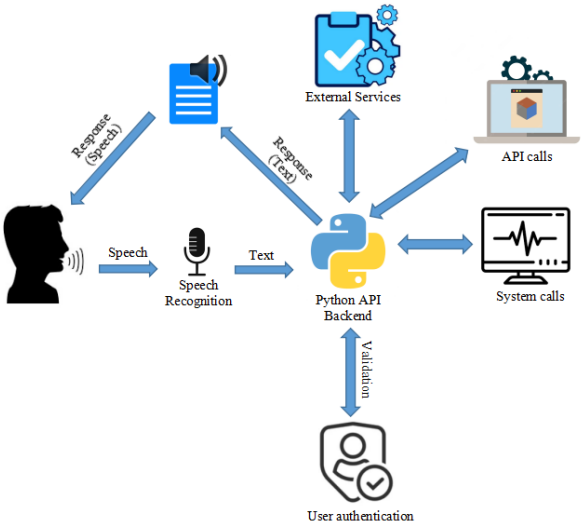

Fig. 3. Proposed System Architecture

- Speech-to-Text (STT) Component: This part is in charge of translating voice commands from the user into text using tools like Google Speech API.
- Intent and Dialogue Management Component: This element controls the conversation's flow and links the user's intention to the proper course of action. The component can conduct the mapping using rule-based or machine learning-based techniques.
- Speaker Recognition Component: The phases of voice enrolment, voice authentication, speaker verification, and speaker identification can be used to achieve speaker recognition in a virtual assistant. This can be done using libraries like CMUSphinx, Kaldi.
- Email Management Component: The user's email accounts will be managed by this component, which will also carry out tasks like email creation, reading, and replying.
- Action and API Integration Component: This component interacts with the APIs of many services, such as reminders, weather, and others, to carry out the necessary action.
- Text-to-Speech (TTS) Component: The virtual assistant's written response is translated into speech by this component. This can be done using TTS technologies like Google Text-to-Speech API or festival TTS.

VII. Experimental Prototype And Architecture

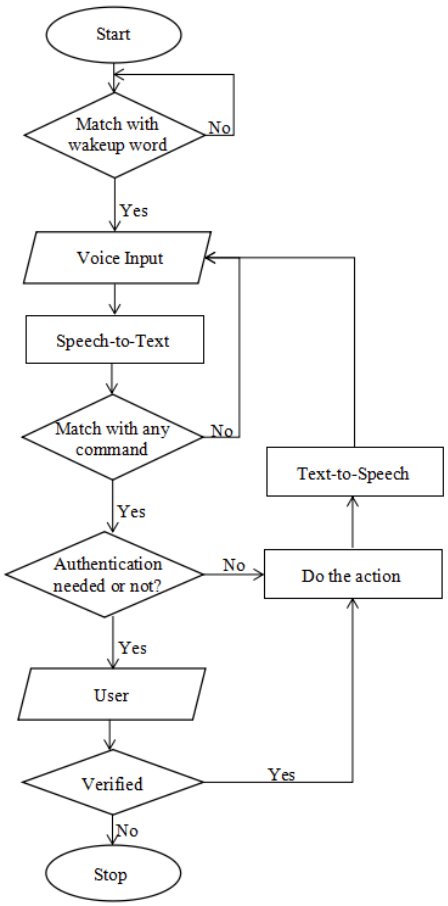

Fig. 4. Flowchart

The prototype of this assistant would use voice biometrics to identify users and NLP technologies to comprehend user inquiries. As a result, the virtual assistant would be able to offer tailored suggestions and responses depending on the user's preferences and prior experiences with the assistant. The control flow of the system is as given in the Fig. 4.

For each of the functionalities in the system, following modules can be used to achieve those:

1. SpeechRecognition: It is the ability of a machine to convert speech to text.
2. gTTS: It is a Python module that communicates with Google Translate's text-to-speech API for text-to-speech conversion.
3. OpenAI ChatGPT: It is used to compute expert-level answers using AI technology. It act as a question-and-answer prompt.
4. NeMo Speaker Recognition API: It is used to authenticate the user by matching the input voice with the stored voice.
5. smtplib: It is a Python module for sending emails using the Simple Mail Transfer Protocol (SMTP).
6. pyaudio: It is used to play and record audio on a system.
7. tkinter: This Python-built package is used to create graphical user interfaces.

8. webbrowser: This module is a convenient web browser controller. It offers a user interface that enables users to view documents hosted on the Web.
9. datetime: This is an inbuilt module in python and it works on date and time

When evaluating a system, there are several metrics that can be used to assess its performance. Some of the common evaluation metrics includes:

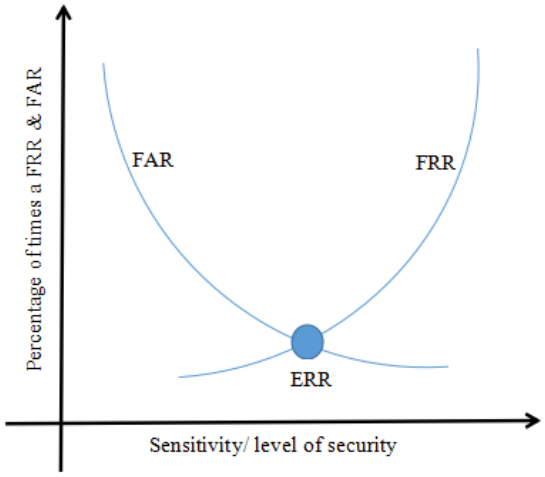

Fig. 5. FAR and FRR equilibrium

A single point of intersection exists between the rates of false rejections and false acceptances. At this time, the percentage indicators coincide, indicating that the rates are equal. The equal error rate is a moniker for such a point (EER). The threshold values of the rates are predetermined by the algorithm.

False Acceptance Rate (FAR): It measures the percentage of times the system mistakenly accepts an unauthorized voice for a real one.

False Rejection Rate (FRR): It measures the percentage of times the system mistakenly accepts an authorized voice for a fake one.

Equal Error Rate (EER): It represents the point at which the FAR and FRR are equal as given Fig. 5. It is frequently used as a common benchmark to assess how well voice authentication systems function.

		Actual Values: Positive	Actual Values: Negative
Predicted Values	Positive	True Positive (TP): number of instances that were correctly classified as positive	False Positive (FP): number of instances that were incorrectly classified as positive
Predicted Values	Negative	False Negative (FN): number of instances that were incorrectly classified as negative	True Negative (TN): number of instances that were correctly classified as negative

Fig. 6. Confusion matrix

Accuracy: It measures the overall accuracy of the system in correctly identifying authorized and unauthorized voices.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Precision: It measures the proportion of true positive voice identifications among all the voice identifications made by the system.

$$Precision = \frac{TP}{TP + FP}$$

Recall: It measures the proportion of true positive voice identifications among all the authorized voices in the system.

$$Recall = \frac{TP}{TP + FN}$$

F1 Score: It is a combination of precision and recall.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

VIII. Conclusion And Future Scope

In conclusion, blind people could significantly benefited from a voice-based virtual assistant. Such a system would give individuals a safe and convenient means to carry out daily activities, connect with others, and get access to crucial data. While the voice-based interface and email system would allow for simple and convenient use, the incorporation of security mechanisms would assure the privacy and protection of important information. By granting blind people greater independence and control over their daily routines, this technology has the potential to significantly improve their quality of life.

Virtual assistants have a lot of room for growth and improvement in the future, and this is an exciting field for innovation and advancement. It can incorporate more sophisticated security methods, including multi-factor authentication, to further protect sensitive data. By making the system compatible with a variety of platforms and gadgets so that blind people may use it on their smartphones, computers, and other gadgets. With additional features like the capacity to make purchases and handle financial transactions, scheduling and reminders, and connection with other assistive technologies can be more efficient.

Overall, the integration of voice biometrics with virtual assistants is a significant step forward in the development of intelligent and personalized technology. These virtual assistants have the potential to significantly improve the user experience.

Refernces

[1] Ujjwal Gupta, Utkarsh Jindal, Apurv Goel, Vaishali Malik,"Desktop Voice Assistant",International Journal for Research in Applied Science & Engineering Technology (IJRASET),ISSN: 2321-9653
[2] 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), London, UK, 2020, pp. 593-596
[3] Kolte, Vrushali S. et al. "Voice-Based Intelligent Virtual Assistant for Windows using Speech Recognition and Speaker Identification Technology." International Journal of Scientific Research in Science and Technology 5 (2020): 98-103.
[4] Patil, Dr & Shewale, Atharva & Bhushan, Ekta & Fernandes, Alister & Khartadkar, Rucha. (2021). A Voice Based Assistant Using Google Dialogflow and Machine Learning. International Journal of Scientific Research in Science and Technology. 06-17. 10.32628/IJSRST218311.
[5] Chinchane, Ayush. "SARA: A Voice Assistant Using Python." International Journal for Research in Applied Science and Engineering Technology 10.6 3567–3582. Web.
[6] Khushboo Sharma, Disha Bahal, Aman Sharma, Ankita Garg, Neeta Verma, "ARA- A Voice Assistant for Disabled Personalities",

International Journal of Engineering Applied Sciences and Technology, 2022, Vol. 7, Issue 1, ISSN No. 2455-2143, Pages 106-109

[7] Kshama V. Kulhalli ,Kotrappa Sirbi, Abhijit J. Patankar, "Personal Assistant with Voice Recognition Intelligence", International Journal of Engineering Research and Technology. ISSN 0974-3154 Volume 10, Number 1 (2017)

[8] Pritam Kakde, Piyush Shah, Anmol Meshram, Yogesh Umap, Yogendra Misal, Akhil Anjikar, "JARVIS - The Virtual Assistant", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022

[9] Saunshimath, Nirmala & Thakur, Sakshi & Singh, Shrooti & Singh, Ranavijay & .K, Kumuda. (2022). SAHARA -A Smart Personal Assistant. 10.13140/RG.2.2.33808.48642.

[10] Bhonsle, V.S., Thota, S., Thota, S. (2022). AKIRA—A Voice Based Virtual Assistant with Authentication. In: Gu, J., Dey, R., Adhikary, N. (eds) Communication and Control for Robotic Systems. Smart Innovation, Systems and Technologies, vol 229. Springer, Singapore.

[11] T. J. Swamy, M. Nandini, N. B, V. Karthika K, V. L. Anvitha and C. Sunitha, "Voice and Gesture based Virtual Desktop Assistant for Physically Challenged People," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 222- 226.

[12] Tariq S Almurayziq, Gharbi Khamis Alshammari, Abdullah Alshammari , Mohammad Alsaffar Saud Aljaloud, "Evaluating AI Techniques for Blind Students Using Voice-Activated Personal Assistants", IJCSNS International Journal of Computer Science and Network Security, VOL.22 No.1, January 2022

[13] Akshat Jain, Apoorv Garg,"Voice Based Emailing System For Visually Impaired People in Django",International Research Journal of Modernization in Engineering Technology and Science,Volume:04/ Issue:05/May-2022

[14] Belekar, Aishwarya & Sunka, Shivani & Bhawar, Neha & Bagade, Sudhir. (2020). Voice based E-mail for the Visually Impaired. International Journal of Computer Applications. 175.

[15] Pranjal Gajendragadkar, Seema Rajput, Harjeet Kaur, "Artificial intelligence based virtual voice assistance ", Application of Communication Computational Intelligence and Learning.