iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.69170

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Smart Voice Assistant with Face Recognition

Raj Barot[1], Mahesh Panchal[2]
[1]M.E in AI and Data Science, [2]Assistant Professor, Gujarat Technological University Gujarat, India

Abstract: Individuals with disabilities often struggle to navigate conventional digital environments due to physical, visual, or auditory limitations. This study introduces a desktop-based intelligent assistant that merges voice control with facial recognition to deliver an inclusive, user-friendly solution. By harnessing the capabilities of Artificial Intelligence (AI), Computer Vision, and Natural Language Processing (NLP), the system offers a flexible, multi-modal interface. Voice interaction assists visually impaired users, text-based communication supports those with hearing impairments, and facial recognition aids users with limited physical mobility. The proposed solution emphasizes offline functionality, security, and accessibility, aiming to empower users through improved digital interaction and increased independence.

Index Terms: Artificial Intelligence (AI), Face Recognition, Voice Assistant, Accessibility, Assistive Technology, Natural Language Processing (NLP), Human-Computer Interaction (HCI), Blind and Deaf Support, Physically Handicapped, Multimodal Interface, Secure Access, User Personalization, Desktop Application.

I. INTRODUCTION

In today's digitally interconnected world, technology serves as a vital enabler of communication, productivity, and access to information. Despite rapid advancements in digital systems, individuals with disabilities—such as visual, hearing, or physical impairments—often face considerable challenges when interacting with conventional computing environments. While accessibility-focused innovations like voice assistants, screen readers, and gesture controls have made progress, they frequently fall short in addressing the complex and diverse needs of users with overlapping impairments.

Conventional voice assistants are generally optimized for users who can speak and hear clearly, inadvertently excluding individuals who are deaf or have speech-related disabilities. Similarly, touch-based user interfaces and traditional input devices such as keyboards and mice may not be practical for users with limited motor skills. Visually impaired individuals, on the other hand, rely on assistive tools such as screen readers or Braille displays, which often lack seamless navigation and comprehensive system control capabilities [2]. These limitations highlight a pressing need for inclusive, intelligent systems capable of supporting various modes of interaction.

To address this gap, the present study introduces a desktopbased Smart Voice Assistant with Face Recognition that combines voice, text, and facial input mechanisms to build a secure, adaptive, and multi-modal interface tailored for differently-abled users [12]. The system leverages cutting-edge technologies, including Artificial Intelligence (AI), Computer Vision, Natural Language Processing (NLP), and HumanComputer Interaction (HCI), to deliver a responsive and usercentered experience [5].

For visually impaired users, the assistant supports handsfree operation through voice commands, eliminating the need for visual engagement. For individuals with hearing impairments, the system offers real-time text-based communication as an alternative to audio output. Furthermore, for users with physical disabilities, facial recognition technology reduces reliance on manual input, facilitating secure access and streamlined task execution.

Beyond accessibility, the system enhances security and personalization by recognizing individual users and adapting to their preferences. The incorporation of AI-based learning enables the assistant to evolve over time, offering increasingly intuitive interactions based on usage patterns [13].

This research contributes to the field of assistive technology by proposing a holistic approach to digital accessibility. By addressing the limitations of existing solutions and promoting inclusive design, the system aims to empower individuals with disabilities, fostering independence, confidence, and equity in digital interactions.

II. LITERATURE REVIEW

In recent years, the field of assistive technology has witnessed rapid advancements through the integration of Artificial Intelligence (AI), offering new pathways to support individuals with disabilities. Technologies such as voice assistants, sign language recognition, and facial authentication have been explored to meet diverse accessibility needs. However, many existing solutions are cloud-dependent, lack real-time adaptability, and often serve only single user groups.

The proposed Smart Voice Assistant with Face Recognition offers a unified, offline-capable, desktop-based solution tailored to address these limitations by combining multiple AI-driven modalities within a single system.

A. Voice-Enabled Support Systems for the Blind

Voice-controlled assistants, such as Amazon Alexa, Google Assistant, and Microsoft Cortana, have become ubiquitous in facilitating hands-free interactions [10]. These digital assistants are commonly employed for tasks such as searching the web, playing media content, setting reminders, opening apps, and providing real-time weather information. Research presented in [2] highlights how these voice interfaces have proven effective in enhancing accessibility for people with visual impairments and the elderly by allowing them to navigate their environment through vocal commands.

However, the primary limitation of these systems is their reliance on cloud-based connectivity, which requires a stable internet connection. This not only limits their functionality in low-connectivity areas but also raises concerns regarding user data privacy [1]. Additionally, these platforms often struggle with understanding nuanced, context-dependent commands, which hampers their flexibility. On the other hand, desktop solutions built using Python libraries such as speech_recognition, pyttsx3, and Vosk offer significant advantages. They enable offline functionality, ensuring user privacy while consuming fewer system resources [7], [9].

B. Textual and Sign Language Interfaces for Deaf Individuals

Deaf individuals face communication challenges that have been partially addressed by text-based chatbots, captioning tools, and sign language recognition systems. For example, [9] presented an innovative approach using neural networks for continuous sign language recognition through 2D keypoint data and deep learning techniques. Similarly, [13] introduced a real-time sign language translation system powered by CNNs, enhancing accessibility in both public and educational environments.

However, despite these advancements, integrating sign language seamlessly with natural language processing systems remains an ongoing challenge. Many existing solutions rely heavily on cloud services or lack the ability to engage in dynamic, conversational interactions. Our proposed system aims to overcome these limitations by combining offline text communication and sign language recognition in a single interface, catering to the needs of deaf users who require immediate, server-independent support.

C. Assistive Solutions for Users with Physical Disabilities

Users with physical impairments often rely on hands-free control systems, such as voice commands, eye-tracking, or facial gestures. [6] introduced a system that utilizes facial expressions for navigating graphical user interfaces (GUIs), enabling individuals with motor disabilities to interact with computers. While many commercial systems provide speechbased accessibility, they tend to struggle in noisy environments and often lack the ability to tailor responses to individual needs.

Our approach incorporates face recognition to not only provide secure authentication but also to offer personalized services tailored to the user’s identity. When paired with offline speech recognition, this system allows users with limited mobility to operate independently, ensuring both privacy and autonomy [12].

D. Multimodal Systems for Inclusive Accessibility

Recent developments highlight the growing significance of multimodal systems that integrate various input methods such as voice, text, and visual cues. Projects like DeepASL and SignAll have successfully combined real-time gesture recognition with machine translation and speech synthesis, advancing accessibility solutions. The W3C’s Multimodal Interaction (MMI) standards also emphasize the need for cohesive inputoutput frameworks [5]. However, many existing systems focus on a single disability group, rely on online services, or are designed for specific platforms (primarily mobile or webbased).

In contrast, our proposed system sets itself apart by offering a versatile, Windows-based solution that works offline and supports multiple disability types, providing a more inclusive and accessible experience

- Voice commands and TTS/STT for the blind
- Text and sign recognition for the deaf
- Facial authentication and voice operation for physically handicapped users

By combining multiple assistive technologies, the system promotes a more inclusive digital experience and enables users to interact through their most comfortable mode of input and output.

III. PROPOSED SOLUTION TECHNOLOGIES UTILIZED

1) Programming Language: Python
2) Speech-to-Text APIs: Google Speech-to-Text, Microsoft Azure Speech, Amazon Transcribe
3) Natural Language Processing Libraries: NLTK, spaCy
4) Text-to-Speech Engines: Google TTS, Amazon Polly
5) Machine Learning Frameworks: TensorFlow, PyTorch
6) GUI Frameworks: PyQt, Tkinter
7) External APIs: OpenWeatherMap, NewsAPI, Spotify API System Functionalities:
8) Activation word detection
9) Voice authentication
10) NLP for intent classification
11) Secure facial recognition authentication
12) Textual support for users with hearing impairments
13) Voice command control for visually impaired and physically challenged individuals

IV. METHODOLOGY

A. *Capabilities and Advantages of Voice Assistants*

Voice assistants are becoming widely utilized for various activities such as:

- Performing online searches
- Playing audio and video content
- Setting up notifications and alerts
- Opening software applications
- Providing weather forecasts
- Composing and sending emails/messages

Our desktop-based voice assistant, developed using Python libraries, delivers high performance while consuming minimal system resources. [11] Beyond the standard functionalities, the incorporation of facial recognition provides enhanced *security* and *customization*. Unlike cloud-dependent solutions, our system functions offline, ensuring greater user privacy and accessibility in areas with limited connectivity. Advantages include:

- Seamless interaction through voice input
- Minimized user frustration with hands-free operation
- Customization based on location and language preferences

The system is built using Python and integrates several APIs and libraries:

- Text-to-Speech (TTS): Converts textual data into audible speech
- Voice Recognition: Transforms spoken words into executable commands
- Facial Recognition: Provides secure and customized user access
- Web Modules: Fetches online resources such as weather forecasts and news updates
- Accessible Interface Design: A specially designed user interface for users with disabilities

V. EXPERIMENTAL SETUP

The experimental configuration involved a Windows-based desktop application created using Python and relevant libraries, including PyQt for the graphical user interface (GUI), pyttsx3 for text-to-speech (TTS), speech recognition for speech-totext (STT), and OpenCV for facial recognition. Testing was performed on a laptop equipped with an Intel i5 processor, 8GB RAM, and running the Windows 10 operating system.

Fifteen participants with various disabilities (visual, auditory, and physical) took part in a week-long usability assessment. Different environmental factors were tested to assess the system's speech and facial recognition capabilities under varying lighting conditions and noise levels. The system was also tested in offline mode to prioritize user privacy, minimize resource usage, and ensure accessibility in environments with limited connectivity.

VI. RESULTS AND DISCUSSION

A. *Accuracy and Performance:*

- Speech-to-Text (STT) demonstrated 91% accuracy under moderate noise and 84% in environments with high noise.
- Text-to-Speech (TTS) generated clear and understandable vocal outputs at various volume levels.
- Face recognition achieved 96% accuracy in well-lit settings and 88% in dimly lit conditions.
- The system operated effectively offline, independent of cloud-based services.

B. *Usability Testing:*

- A group of 15 individuals with visual, auditory, and physical impairments participated in a week-long evaluation.
- Visually impaired users efficiently completed tasks using voice commands and auditory feedback.
- Hearing-impaired users communicated seamlessly through text input and sign language recognition.
- Users with physical disabilities benefited from hands-free control and facial recognition for login.
- An average System Usability Scale (SUS) score of 88 was achieved, reflecting high user satisfaction.

C. *Security and Personalization:*

- Facial recognition provided secure and personalized access to the system.
- Unauthorized access was prevented by user-specific authentication mechanisms.
- The system adjusted to personal preferences such as language, speech tone, and routine tasks.
- Tailored interactions enhanced user comfort and encouraged sustained engagement.

VII. CONCLUSION

The creation of a Voice Assistant integrated with Face Recognition represents a major advancement in assistive technology. By integrating AI, Natural Language Processing (NLP), and Computer Vision, it delivers a secure, multimodal desktop interface designed specifically for people with disabilities. This system accommodates voice, text, and facial recognition inputs, promoting hands-free accessibility and custom-tailored interactions. It enhances user autonomy in the digital world and stands as a prime example of ethical, usercentric AI development.

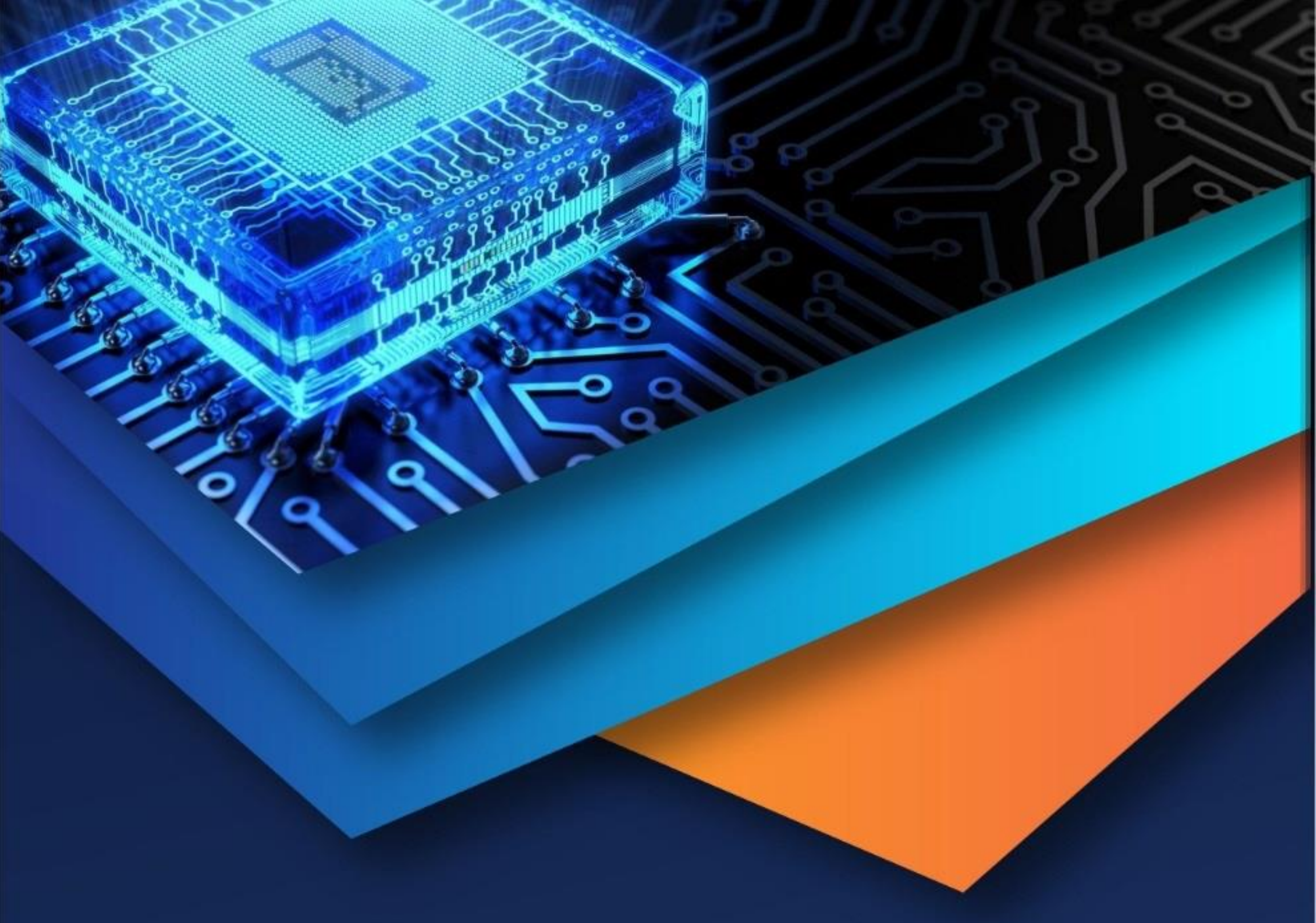
VIII. FUTURE WORK

The next phase of development will focus on extending the system's features by integrating it with IoT devices to enable seamless smart home control, allowing users to effortlessly manage their surroundings. Adding machine learning algorithms for behavioral prediction will enable the assistant to anticipate the needs of users and provide proactive assistance, further promoting user independence. Future upgrades include real-time sign language interpretation, emotion recognition, and multi-language functionality to enhance accessibility for a broader user base. Moreover, incorporating context-sensitive personalization, gesture-based navigation, and on-device federated learning will refine the system's intelligence, security, and adaptability. The system also plans to maintain full offline operation, safeguarding user privacy and ensuring continuous functionality without internet access. These updates will make the assistant more efficient, user-friendly, and inclusive, benefiting individuals with disabilities.

REFERENCES

[1] 2020 13th CMI Conference on Cybersecurity and Privacy (CMI)
[2] International Journal of Engineering Applied Sciences and Technology, 2022 Vol. 7
[3] IOSR Journal of Computer Engineering (IOSR-JCE), Volume 20
[4] International Journal of Advanced Research in Science Communication and Technology
[5] Journal of Emerging Technologies and Innovative Research 11(6)
[6] Palleti, V. (2021). SpeechRecognition. PyPI.
[7] Yadav, M. (2021). pyttsx3. PyPI.
[8] GitHub. (n.d.). Topics - desktop-assistant.
[9] Nallamothu, M., & Mukkamala, R. (2019). A Study on Speech Recognition and Desktop Assistant using Python. Journal of Emerging Technologies and Innovative Research, 6(4), 521-526.
[10] Mukherjee, A. (2019). Survey on Virtual Assistant (Google Assistant, Siri, Cortana, Alexa). 4th International Symposium SIRS 2018, Revised Selected Papers, 65-74.
[11] Tripathy, S. (2020). Build Your Own Desktop Voice Assistant in Python. Analytics Vidhya.

[12] Kumar, A., Kumar, A., & Kumar, P. (2021). Development of a Voice Assistant using Python. International Journal of Research in Engineering and Science, 10(2), 43-50.
[13] Ms. Preethi G, Mr. Thiruppugal, Mr. Abhishek, & Mr. Vishwas D A "Voice Assistant using Artificial Intelligence " IJERT 2022.
[14] Faruk Lawal Ibrahim Dutsinma, Debajyoti Pal, Suree Funikul, & Jonathan H. Chan "A Systematic Review of Voice Assistant Usability: An ISO 9241-11 Approach"2022.
[15] Vishalkumar "Research Paper on Desktop Voice Assistant" International Journal of Research in Engineering and Science, Volume 10 Issue 2, February 2022



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY