V-Mail: Bridging the Gap in Communication for the Visually Impaired

Dhiraj Wadile[1], Pushkar Pattiwar[2], Nahush Thuse[3], Vaishnavi Sawant[4], Jyoti Jadhav[5]

[1]Student, SCTR's Pune Institute of Computer Technology, (IT), Pune, Maharashtra, India, dmwadile@gmail.com
[2]Student, SCTR's Pune Institute of Computer Technology, (IT), Pune, Maharashtra, India, pushkarrp03@gmail.com
[3]Student, SCTR's Pune Institute of Computer Technology, (IT), Pune, Maharashtra, India, nahushthuse@gmail.com
[4]Student, SCTR's Pune Institute of Computer Technology, (IT), Pune, Maharashtra, India, sawantsvaishnavi7@gmail.com
[5]Assistant Professor, SCTR's Pune Institute of Computer Technology, (IT), Pune, Maharashtra, India, jhjadhav@pict.edu

Abstract

Visually impaired individuals face numerous challenges when using traditional email systems, which depend on visual cues and manual interaction, thereby limiting their ability to communicate independently. To address this issue, a system has created an innovative desktop-based, voice-controlled email system that leverages advanced Speech-to-Text (S-T-T) and Text-to- Speech (T-T-S) technologies, integrated with a user-friendly Voice Command Interface. This system allows users to compose, send, and read emails solely through voice commands, eliminating the need for visual engagement. Additionally, it supports attachment management and seamless inbox navigation, empowering users to handle email tasks with ease and efficiency. By removing visual barriers, this solution not only enhances accessibility but also promotes self-reliance, offering a versatile tool for both personal and professional communication needs. Ultimately, this application improves the eminence of lifespan for visually weakened individuals by fostering digital inclusion, breaking down communication barriers, and establishing a new benchmark for accessible, voice-enabled technology that can inspire similar advancements across digital platforms.

Keywords: Visually impaired, desktop application, voice-controlled email, Speech to Text, Text to Speech, accessibility, voice command interface, digital inclusion

1. Introduction

As technology advances, specifically in the disciplines of machine learning (ML) and artificial intelligence (AI).and natural language processing (NLP), people are increasingly accustomed to a simulated world that enhances communication and accessibility [1]. However, for visually impaired individuals, navigating the internet remains a significant challenge, especially when using email systems that predominantly rely on visual interfaces [15]. Traditional email platforms assume users have the ability to see and interact with screens, making them inaccessible to those who require non-visual methods of interaction. In the realm of AI and ML, developing accessible systems for blind users involves creating specialized forms and approaches, akin to how educational methods are adapted to make learning accessible for visually impaired students [2]. This system leverages cutting-edge skills such as NLP-driven (S-T-T) and (T-T-S) to develop voice-controlled email platform that empowers visually impaired users.

Although screen readers with TTS and Automated Speech Recognition (ASR) exist, they still suffer from limitations in accuracy and efficiency for these users [3]. To address this, the system applies ML algorithms that enhance the voice-driven interaction experience, ensuring smoother and more intuitive communication. Rather than predicting the entire email content at once—typical of conventional email processing— system uses an intention-based segmentation approach, processing email content in segments to improve overall accessibility and performance [4]. STT is utilized to convert user speech into text for composing emails, while TTS allows users to listen to their emails, minimizing reliance on visual input. The integration of NLP for intent recognition enables the system to perform complex email tasks, such as attaching files, navigating folders, and managing the inbox, all through voice commands.

Recent research highlights the development of hybrid AI frameworks that can enhance navigation in diverse scenarios, both indoor and outdoor [5]. This work contributes to that body of knowledge by addressing the crucial need for internet entree amongst visually weakened people, a population that may otherwise face digital exclusion [6]. The application of AI-driven technologies in this system ensures that visually impaired users are no longer dependent on sight for email communication. In the broader context of increasing integration of Information and Communication Technologies (ICTs) for accessibility, this system exemplifies how ML and AI can foster rapid, effective, and affordable solutions for blind and visually impaired people [7]. By harnessing the power of AI, NLP, and ML, this system directly contributes to digital inclusion, advancing the freedom and excellence of life for visually impaired individuals. It ensures that in a world where inclusivity is gaining attention, visually impaired users are not left behind, allowing them to fully participate in digital communication.

2. Literature Survey

Assistive technologies for visually impaired individuals have seen continuous development, particularly in the domain of voice-based systems that enhance digital accessibility. Numerous studies have explored the integration of S-T-T and T-T-S technologies to provide users with an alternative to traditional, visually intensive email systems. Rajput et al. [1] proposed an AI-driven voice/speech-built email communication system, allowing users to interact with their-in-box through voice commands. Their system offers core email functionalities such as composing, reading, and managing emails through speech recognition, thus enhancing accessibility for visually impaired users. This study emphasizes that using AI for speech recognition can substantially improve the usability of communication systems by enabling natural language interaction. Rajput's research demonstrated that voice-based email systems not only empower visually impaired users but also support personalization options to meet the unique needs of everyone, such as setting preferred voices or adjusting speaking speed.

Bhore et al. [7] similarly focused on improving accessibility for blind users by developing a voice-based email system using STT and TTS technologies. This system allows users to dictate their emails, and the speech recognition software converts spoken language into text. Bhore's research highlights how voice interfaces can enhance digital interaction by bypassing traditional visual elements like keyboards and screens. Their work further discusses the challenges of speech recognition in real-world conditions, especially when dealing with different accents, background noise, and variations in voice pitch. While Bhore's study provides strong insights into the effectiveness of TTS and STT in email systems, it also points out the difficulties in ensuring consistent system accuracy across diverse users and environments. Nonetheless, their solution allows visually impaired individuals to not only send and receive emails but also manage attachments and organize emails through voice commands, contributing significantly to their independence.

In addition to these systems, Khasawneh [2] explored voice recognition in an academic setting, focusing on visually impaired students. Their work revealed the importance of voice-based systems for improving access to educational resources and communication tools. By integrating voice commands into existing platforms, Khasawneh showed that students could independently access emails, course materials, and other digital content, reducing their reliance on sighted assistance. Furthermore, the study emphasized the potential of voice-driven systems to extend beyond personal communication to support professional and educational tasks.

Kulkarni et al. [12] provided another significant contribution to the field by designing a comprehensive voice-controlled email system specifically for blind users. Their system enables users to perform essential email functions, including composing, reading, and deleting messages, entirely through voice commands. Kulkarni's work emphasizes inclusivity, tailoring the system to individuals with disabilities by offering them complete control over their email management. This study highlights the importance of accessibility in technology design, pointing out that traditional email systems are heavily reliant on visual interfaces, which limits their usability for visually impaired individuals. Through voice interfaces, Kulkarni's system seeks to bridge this gap by ensuring blind users can engage in everyday digital communication activities with minimal external assistance.

Despite these advancements, gaps remain in optimizing the user experience, improving the accuracy of speech recognition, and addressing security concerns for users who require alternative authentication methods. While face recognition is increasingly being integrated as an additional layer of security, its effectiveness is still inconsistent, particularly in varying lighting conditions or with changes in facial expressions, as noted in various studies [11]. Therefore, further refinement and innovation are necessary to build more robust systems that cater to the nuanced needs of visually impaired users.

3. Comparison of International Contributions

Table 1. Related work summary

Paper.	Author	Aim/Objective	Methodology	Findings	Gaps
[1]	G. Rajput et al. (2023)	Telephone email system for blind people through an interface.	- AI-driven voice recognition system. - Focus on natural language interaction.	- Enhanced usability. - Personalization options for users.	Further optimization of user experience.
[2]	Y. J. A. Khasawneh et al. (2023)	Improve accessibility for visually impaired students.	- Integration of voice recognition in educational platforms. - User surveys.	- Increased independent access to emails and course materials. - Reduced reliance on sighted assistance.	Limited focus on professional tasks.
[3]	S. Kumar, Y. R et al. (2021)	Create a voice email system for physically handicapped users.	- SMTP-based voice email system. - User testing for feedback.	- Enabled email sending through voice. - Simplified interaction.	- Challenges with speech recognition in diverse environments.
[4]	S. K. Sonbhadra et al. (2020)	Classify emails based on user intentions.	- Intention-based segmentation model. - Data analysis techniques.	- Improved email classification accuracy.	- Lack of emphasis on assistive features for visually impaired.
[5]	M. Anandan et al. (2020)	Develop a steering system for visually blind individuals.	- Ras-Pi-based system. - Prototype testing.	- Enhanced navigation capabilities for visually impaired users.	- Focus primarily on navigation, not communication tools.
[6]	A. Khan et al. (2020)	Explore smartphone solutions for visually impaired.	- Literature review on assistive technologies. - Comparative analysis.	- Identified challenges and opportunities in smartphone solutions.	- Need for practical implementation of identified solutions.
[7]	N. Bhore, et al. (2020)	Enhance email access for visually impaired individuals.	- Development of STT and TTS technologies - User feedback sessions.	- Allowed voice-based management of emails.	- Inconsistencies in speech recognition accuracy.

[8]	K. Miao et al. (2020)	Integrate voice transcription with conferencing.	- Parallel audio conferencing and transcription system. - Case studies.	- Improved identification and transcription accuracy.	- Focused on specific contexts rather than general use cases.
[9]	P. A. Tiwari et al. (2020)	Review voice-based email systems for blind users.	- Systematic literature review. - Compilation of existing studies.	- Provided a comprehensive overview of current systems.	- Limited discussion on user experience enhancement.
[10]	M. Elleuch et al. (2020)	Discover activity perspectives from emails.	- Speech acts detection in emails. - Analytical techniques.	- Gained insights into email interactions.	- Lack of focus on accessibility features.
[11]	M. S. Minu et al. (2020)	Develop a face recognition system.	- Utilized Haar cascade classifier. - Testing under various conditions.	- Effective face detection in controlled environments.	- Inconsistent performance in real-world conditions.
[12]	O. Kulkarni et al. (2019)	Design a voice-controlled email system for blind users.	- Voice command-based email management. - User testing.	- Provided full control over email management via voice.	- Challenges in system accuracy and external dependencies.
[13]	Cuimei et al. (2017)	Enhance human face detection capabilities.	- Combined Haar cascade classifier with other models. - Testing methodology.	- Improved detection rates in controlled settings.	- Performance issues in real-world conditions.
[14]	P. Ingle et al. (2016)	Voice/speech-built email in communication system for blinds.	- Developed voice-based email interface. - User feedback.	- Enabled email management through voice.	- Limited scalability and adaptability of the system.
[15]	Khanun et al. (2023)	Voice-Based E-Mail System for Visually Challenged People	- Speech recognition and voice interaction technologies used. - Iterative development	- Secure authentication improves user safety	- Limited email service integration

4. Proposed Methodologies

4.1 Modules

4.1.1 Voice Recognition Module

Purpose:
The voice recognition module is designed to listen to user voice commands and convert them into text using a Speech-to-Text (STT) engine. Miao et al. (2020) developed an integrated system for automating the transcription of audio conferencing, achieving accurate transcription in quiet environments. This functionality allows users to compose emails and manage their hands-free inbox, making digital communication more accessible for visually impaired users.

The system captures the user's spoken words and transcribes them into text, as shown in Figure No. 2, where speech inputs such as "Username is Gauri" are being recognized. This model was trained on a diverse dataset that includes various dialects and acoustic environments to enhance its robustness.

The speech recognition system is composed of several key components designed to ensure accurate transcription and efficient interaction. The *RNN with Attention Mechanisms* utilizes Recurrent Neural Networks (RNNs) enhanced with attention mechanisms to recognize a wide range of speech patterns, dialects, and accents. This integration improves transcription accuracy by focusing on specific segments of the input when making predictions, especially in noisy environments. The *Microphone* component captures the user's audio input, providing clear and accurate voice recognition. Additionally, *Real-Time Feedback* is incorporated to provide instant voice prompts, notifying users whether their commands have been correctly recognized. This feedback loop is essential for maintaining user confidence and ensuring control over the system.

Steps:

1. The system continuously listens to user input through the microphone.
2. The RNN processes the recorded audio through its layers, translating spoken words into text. The model's training on diverse datasets enhances its adaptability to individual users.
3. Real-time voice prompts provide feedback, such as "I heard [text]," allowing for immediate corrections if needed.

The speech-to-text (STT) system faces several challenges, and solutions have been implemented to address these issues effectively. One significant challenge is speech accuracy, as transcribing speech into text with high precision can be difficult due to factors such as accents, dialects, and background noise. To mitigate this, the system utilizes dynamic language modeling, which adapts to the user's unique speech patterns over time, improving transcription accuracy with continued use. Another challenge is error correction. To help users quickly resolve transcription mistakes, the system offers real-time voice feedback, allowing for simple commands such as "delete last word" or "repeat" to enable quick corrections. Additionally, a context-aware model is integrated, which suggests potential corrections based on common user errors. This error correction mechanism not only enhances the system's usability but also ensures that users maintain control even when inaccuracies arise during speech recognition.

4.1.2 Text to Speech (TTS) Module

Purpose:
The TTS module reads aloud incoming emails and system instructions, ensuring users can access email content without visual aid, thus promoting independence and enhancing the user experience.

The text-to-speech (TTS) system is composed of key components designed to produce high-quality, natural-sounding voice outputs. *Tacotron 2* is a deep neural network-based speech synthesis system that generates speech with proper intonation and prosody, ensuring that the spoken text is clear and easy to understand. Additionally, the system includes *Adjustable Parameters*, allowing users to customize the voice speed and pitch to suit their preferences, further enhancing the overall listening experience. These components work together to provide an adaptable and user-friendly TTS system.

Steps:

1. The system retrieves email content through the IMAP protocol, accessing the latest inbox messages.
2. The TTS engine converts the text into speech, using natural language processing to incorporate intonation and pauses, enhancing readability.
3. Users can adjust voice settings to suit their preferences, ensuring an optimal listening experience.

4.2 Methodologies

4.2.1 AI Algorithms for Face Recognition

The system integrates face recognition using the Haar Cascade Algorithm, enhancing security and user experience. Haar Cascades are widely used in real-time face detection due to their fast and accurate detection capabilities. The algorithm works by identifying faces through a cascade of classifiers, each trained on a set of positive and negative images. This approach ensures efficient face recognition under various lighting conditions and angles.

The methodology of the system involves a seamless integration of various technologies to enhance functionality and accessibility. *Figure 1* illustrates the system architecture, which combines voice commands with AI-driven face and object recognition technologies to enable efficient and secure interactions. The face recognition algorithm plays a crucial role by employing a series of feature extraction and classification steps, utilizing Haar-like features to detect and identify faces. This approach not only improves the accuracy of face detection but also enhances accessibility, particularly for visually impaired users who may rely on face recognition for secure logins. These methodologies collectively ensure a robust and user-friendly system.

4.2.2 AI Algorithms for Object Detection

In parallel with face recognition, OpenCV is utilized for object detection tasks within the system. This computer vision library enables the detection of objects in real time, which can be further extended to recognize attachments in emails or detect specific user gestures through camera input. The system processes input continuously, providing corrective feedback to users when necessary to ensure accurate interaction. OpenCV enhances the system's functionality by enabling the recognition of various objects, allowing for expanded use cases such as gesture-based control and attachment identification, making the system more versatile and interactive.

Voice command interface:

The Voice Command Interface empowers users to control all system functionalities through simple and intuitive voice commands. This central component enhances accessibility, allowing users to perform various tasks without relying on a mouse, keyboard, or screen.

Natural Language Understanding (NLU):

The NLU engine interprets voice commands to discern user intent, effectively mapping phrases such as “Send email” or “Compose email to [contact]” to specific actions. The system processes command to recognize user details or execute tasks like capturing photos, streamlining interactions.

Contextual Awareness:

The system supports sequential voice commands, enabling a fluid user experience. For example, after dictating an email, users can simply say “Send” without needing to repeat the full command. The NLU engine maintains context across interactions, ensuring smooth execution of ongoing tasks.

Voice command functionalities significantly enhance email management by enabling hands-free and efficient interaction. Users can initiate the email composition process by stating, “Compose a new email to [contact],” which streamlines message creation. Additionally, managing attachments is simplified through voice navigation with commands such as “Attach file,” which opens the file browser for seamless selection. Furthermore, inbox navigation becomes more accessible with commands like “Read my emails” or “Show new messages,” allowing users to stay updated without requiring visual engagement. These features collectively improve accessibility, productivity, and user experience in digital communication.

System Architecture

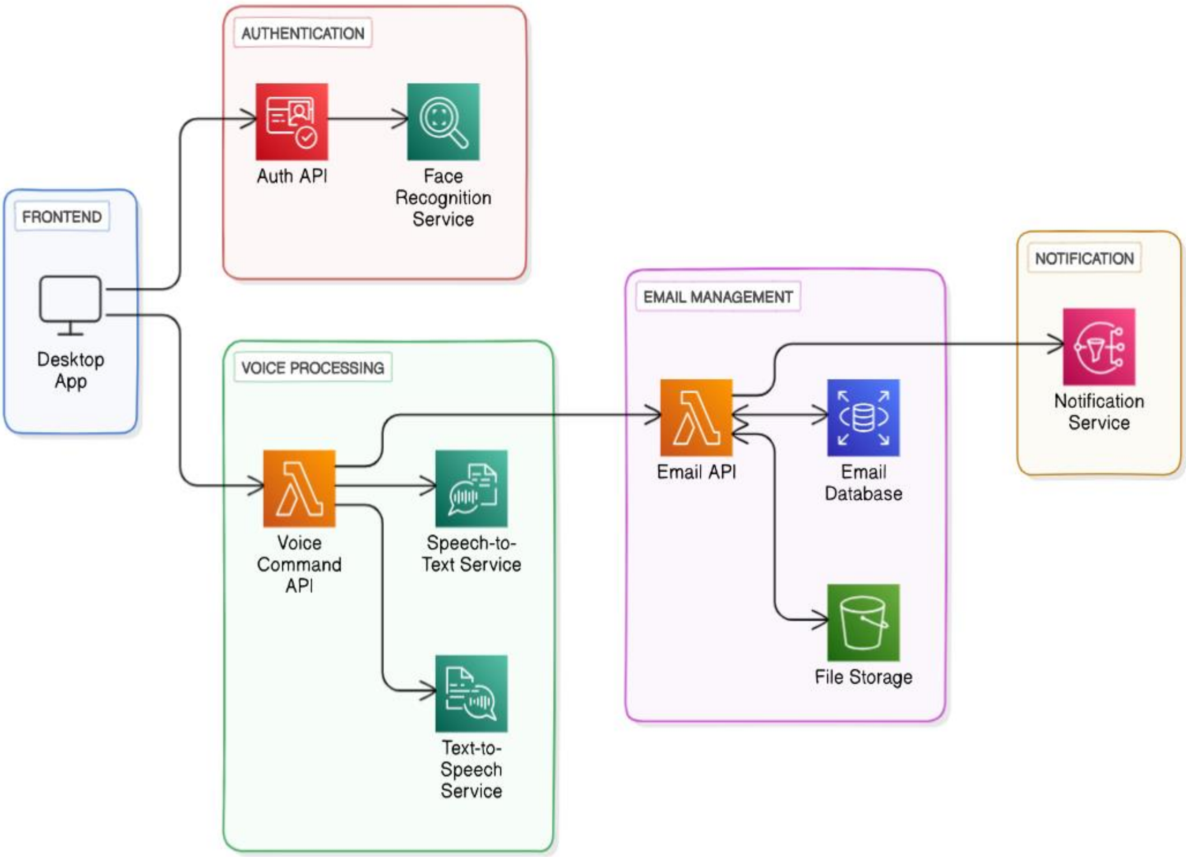

Fig. 1. System Architecture

The system architecture involves multiple services that interact across various domains, including Authentication, Voice Processing, Email Management, and Notifications. The system is accessed through a web application, which serves as the user interface for interacting with the different services. Users can initiate processes such as authentication and voice commands via this interface. The authentication system consists of several components designed to ensure secure access control and user identity verification. The Authentication API is responsible for processing authentication requests, verifying user credentials, and managing access to the system, interacting with security protocols such as OAuth to enhance security and prevent unauthorized access. Additionally, the Face Recognition Service, integrated within the authentication system, leverages machine learning algorithms for facial recognition, providing an extra layer of authentication. This service enhances security and user convenience by allowing biometric authentication before granting access to other functionalities. The voice processing system enables seamless interaction between users and email management services. The Voice Command API serves as the primary entry point for processing voice-related commands, integrating with various services to interpret and execute user instructions. A key component of the voice processing system is the Speech-to-Text (STT) service, which utilizes natural language processing (NLP) models to transcribe spoken language into text, allowing users to compose emails or perform other actions through voice input. Complementing this, the Text-to-Speech (TTS) service converts written text into spoken output, offering voice-based feedback or confirmations to users. This enhances accessibility, particularly for individuals with visual impairments or those who prefer auditory interactions. The email management system includes several key components to ensure efficient processing and storage of email-related data. The Email API is responsible for handling email-related commands, such as sending, receiving, and managing email content based on user inputs or system events. It interacts with the Email Database, which stores email records, metadata, and history, facilitating easy retrieval and processing. The Email Database allows the execution of essential CRUD (Create, Read, Update, Delete) operations in conjunction with the Email API. The system also incorporates File

Storage, which is often cloud-based (e.g., Amazon S3), to manage email attachments and other relevant data, such as audio files generated by the TTS service. Finally, the Notification Service is designed to handle notifications, alerting users to important events such as new emails, voice command results, or authentication confirmations. This service can be integrated with real-time communication systems, such as WebSockets or push notifications, to ensure timely delivery of updates.

5. UML Diagrams and related explanation

A. Activity Diagram

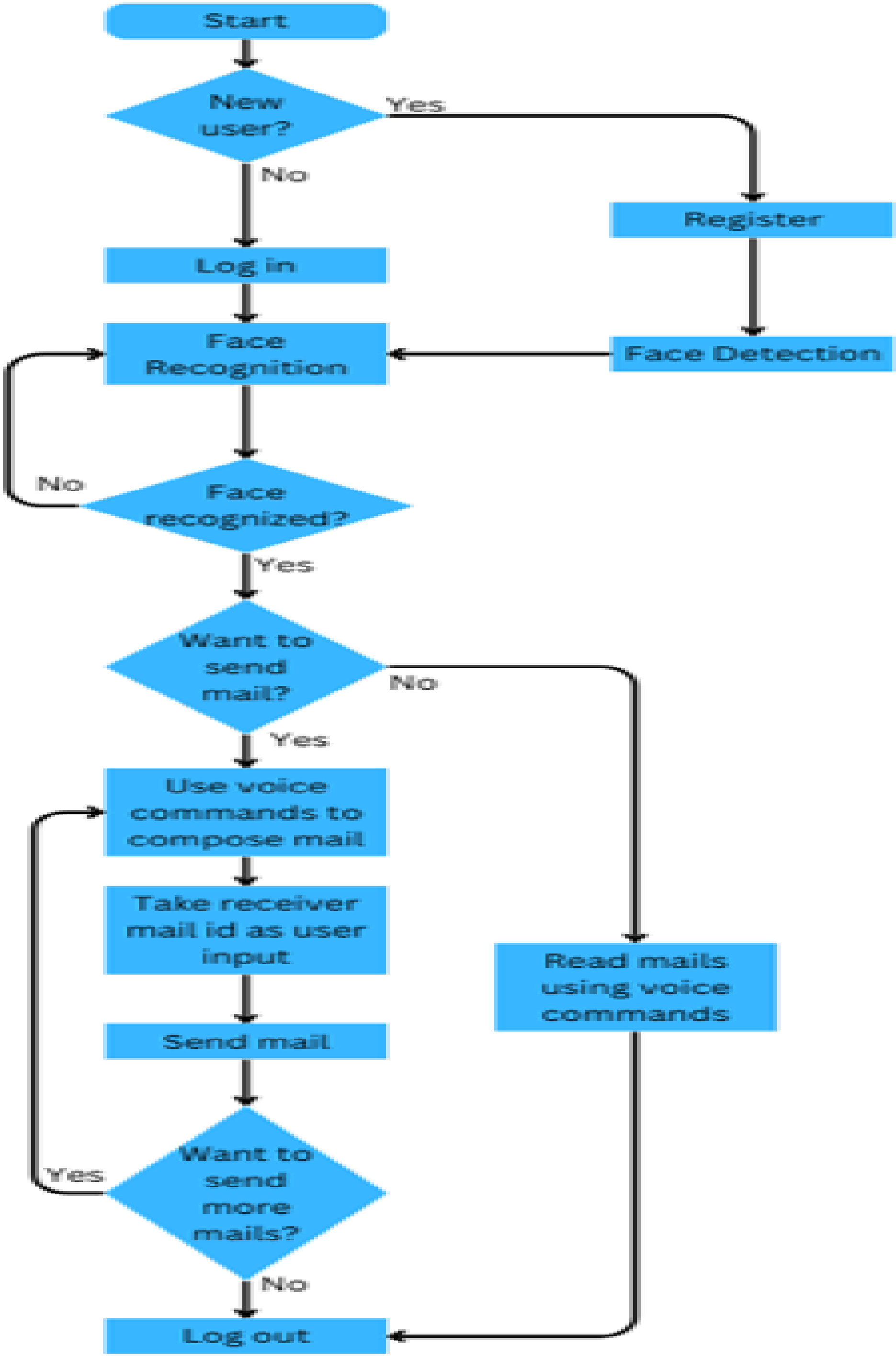

Fig. 2. Activity Diagram

This activity diagram outlines an email system integrating face recognition and voice commands for accessibility and security. New users undergo registration with face detection, while returning users log in through face recognition. Once authenticated, users can choose to send or read emails. Sending emails involves composing them using voice commands and providing the recipient's email ID as input. Users can opt to send additional emails or log out after completion. The system highlights a hands-free and secure approach to email communication, leveraging advanced technologies for convenience and user-friendliness.

B. Sequence Diagram

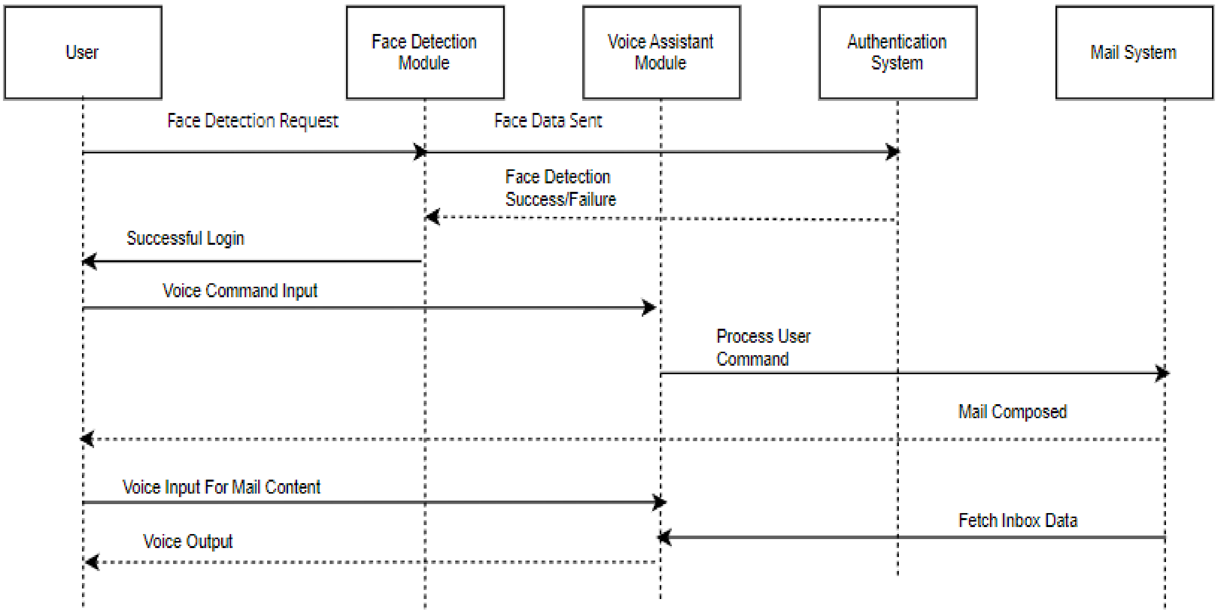

Fig. 3. Sequence Diagram

This sequence diagram illustrates an email system utilizing face detection and voice commands for user interaction. The process begins with a face detection request, where the Face Detection Module validates the user's identity and sends the result to the Authentication System for login success or failure. Upon successful authentication, users can input commands via the Voice Assistant Module to perform actions like composing emails or fetching inbox data. Voice inputs are processed for mail content or command execution, and corresponding outputs are sent back to the user. The Mail System handles email composition and inbox retrieval, ensuring seamless and efficient communication. This diagram emphasizes an intuitive, secure, and voice-enabled approach to managing emails.

6. Implementation of Modules

6.1 Voice Recognition Module

Purpose: The voice recognition module attends to the operator's speech commands and changes them into text using Google Speech Recognition. This text is then processed to understand the user's intent (e.g., composing an email, reading received messages, etc.).

The speech recognition system consists of several key components that facilitate the conversion of spoken language into text. The *speech recognition.Recognizer()* is responsible for initializing the recognizer, enabling the system to process and interpret audio input. The *Microphone()* component captures the user's speech, serving as the primary input device for voice commands. To convert spoken language into text, the system leverages the *Google Speech API*, which applies advanced speech recognition algorithms to accurately transcribe user input. These components work together to enable efficient and accurate voice-based interactions in various applications.

Steps:

(a) The system uses Microphone() to listen to the user's voice.

(b) Recognizer.listen() records the voice input and sends it to the Google Speech API.

(c) The translated text is matched with predefined voice commands.

6.2 Face Recognition Module

Purpose: To authenticate the user by scanning and recognizing their face before allowing access to the email system.

The face recognition system is built around several key components to ensure accurate user identification. The *Face Recognition API* plays a central role by capturing the user's facial data and comparing it with pre-stored face data to verify identity. This process involves sophisticated algorithms that analyze facial features and patterns. Additionally, *cv2* (OpenCV) is used to capture the live video feed from the user's webcam, providing real-time visual input for the face recognition system. Together, these components enable seamless and secure biometric authentication through facial recognition technology.

Steps:

(a) The user's face is captured using the webcam.

(b) The system compares the captured face with the stored face data using the Face Recognition API.

(c) If authenticated, the user can proceed with email operations.

6.3 Email Composition Module

Purpose: To compose and send emails using voice commands. The system converts the spoken text into an email format, allowing users to dictate the recipient, subject, and body of the email.

The email composition and sending system incorporates several key components to streamline the process through voice interaction. *SMTP* (Simple Mail Transfer Protocol) is used to send the composed emails to the recipient via the Gmail server, ensuring reliable email delivery. The *speech recognition* module captures the user's spoken input to compose the email content, allowing for hands-free interaction. Once the email is sent, *pyttsx3* provides voice feedback, confirming to the user that the email has been successfully composed and sent. These components together enhance the overall user experience, providing an efficient, voice-driven email management system.

Steps:

(a) The scheme reminders the user to say the email details (recipient, subject, body).

(b) The text is processed and structured into an email format.

(c) The email is sent using the SMTP protocol.

6.4 Email Reading Module

Purpose: To read incoming emails out loud for the user. The user can request to hear recent messages, and the system will fetch and read the content via text-to-speech.

The email retrieval system is built around several essential components that facilitate the fetching and auditory delivery of incoming emails. The *IMAPClient* is responsible for connecting to the email server and retrieving incoming emails, allowing the system to access the latest messages. To convert the fetched email text into speech, *pyttsx3* is used. This text-to-speech engine enables the system to audibly read out the content of the emails to the user, enhancing accessibility and providing a hands-free way to stay updated on new messages. These components work together to streamline the email reading experience.

Steps:

(a) The system logs into the user's email account using IMAP and retrieves the most recent emails.

(b) It then reads aloud the subject and body of the emails using pyttsx3.

6.5 Attachment Module

Purpose: This module allows users to attach files to an email via voice commands.

The system can list available files, and the user can select a file by name or number.

Steps:

(a) The system reads out the available files in a specified directory.

(b) The user selects a file to attach via voice input.

(c) The selected file is added to the outgoing email.

6.6 User Session Management

Purpose: To manage the user's session, including login, authentication, and logout operations. The system ensures that sessions are properly tracked and closed after use.

The session management system consists of two key components designed to enhance security and user experience. *Session Timeout* automatically logs the user out after a predefined period of inactivity, helping to protect the system from unauthorized access when the user is not actively engaged. In parallel, *Session Status* monitors and tracks the state of the current session, indicating whether it is active or inactive. This real-time tracking ensures that the system can respond appropriately to changes in user activity and maintain a secure environment. Together, these components help manage user sessions effectively and safeguard sensitive information.

6.7 Graphical User Interface (GUI)

Purpose: To provide a user-friendly interface that visually impaired users can

interact with via voice commands. The GUI was built using Tkinter.

The user interface for the voice recognition system is built around several components that facilitate user interaction and provide real-time feedback. *Tkinter Windows and Buttons* serve as the primary interface elements, guiding users through the application and providing visual feedback. A *Microphone Input Button* is integrated into the interface to trigger the voice recognition functionality, allowing users to initiate voice commands easily. *Real-Time Feedback* is displayed on the console, showing the recognized commands as they are processed, ensuring that users can track the system's understanding and responsiveness. These components work together to create an intuitive and interactive experience for the user.

7. Results and Discussion

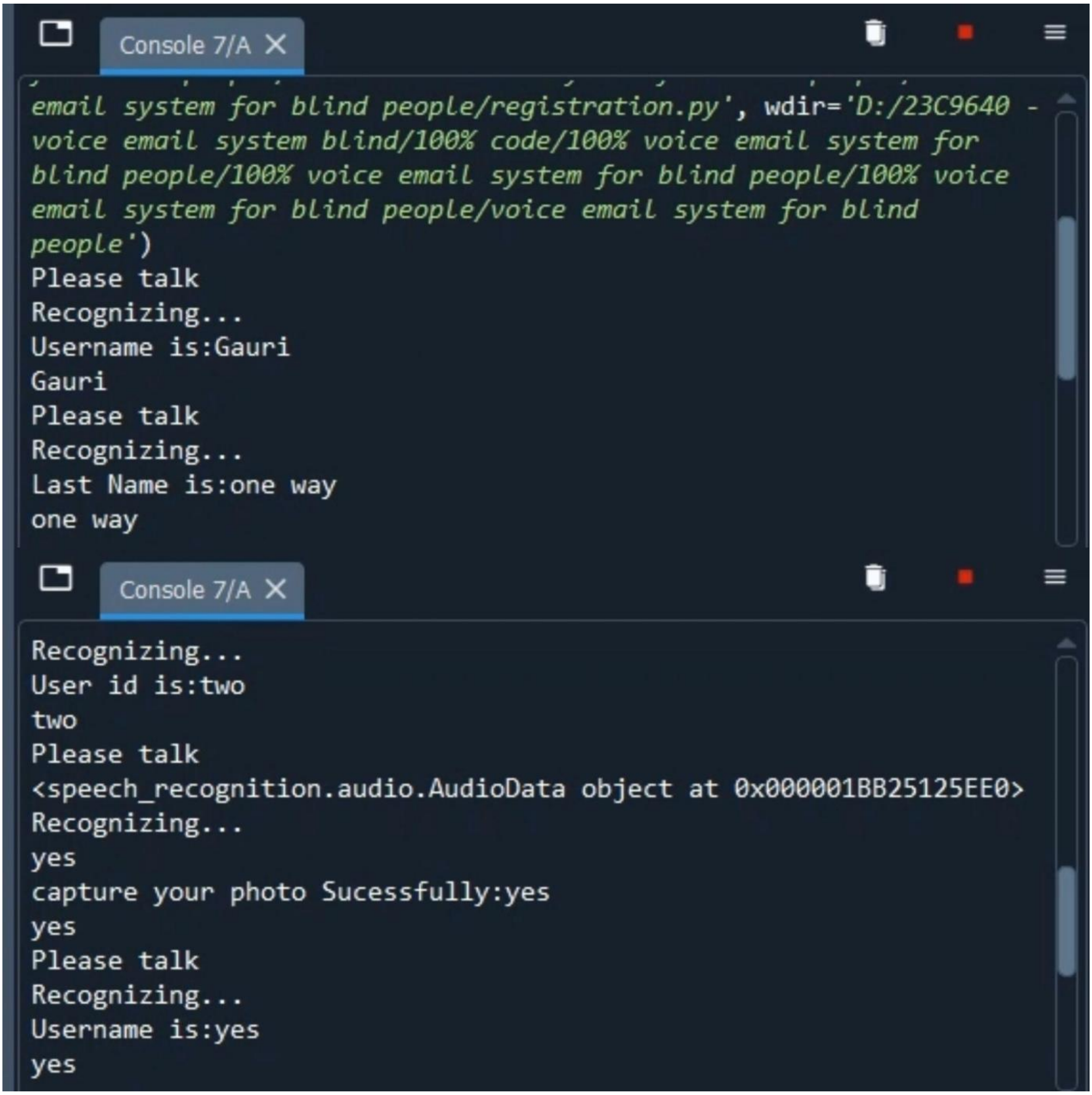

Fig. 4. Console Output

This console output displays the results of user interactions with the email system, including confirmation messages for composing, sending and recognizing.

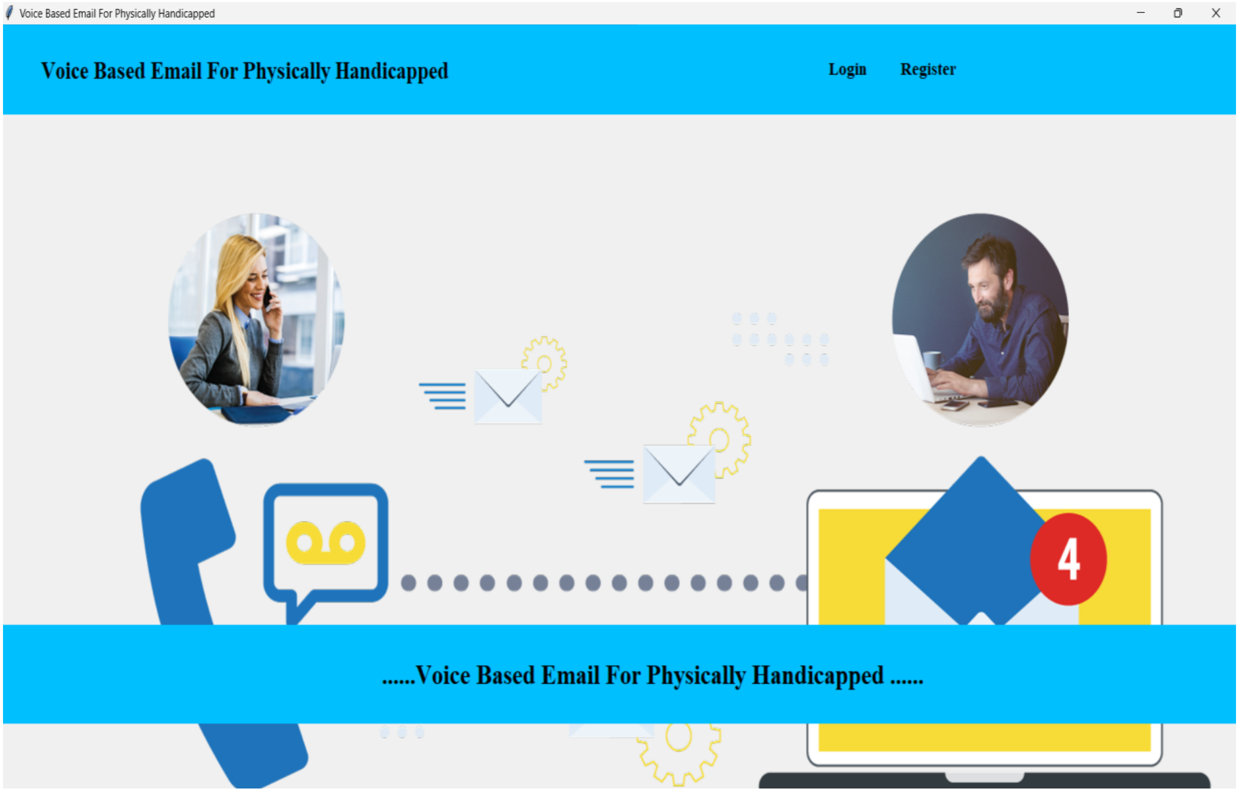

Fig. 5. Frontend View

This frontend View shows the results of user interactions with the email system for physically handicapped, including login, registration and interactive user interface.

8. Performance Evaluation Parameters

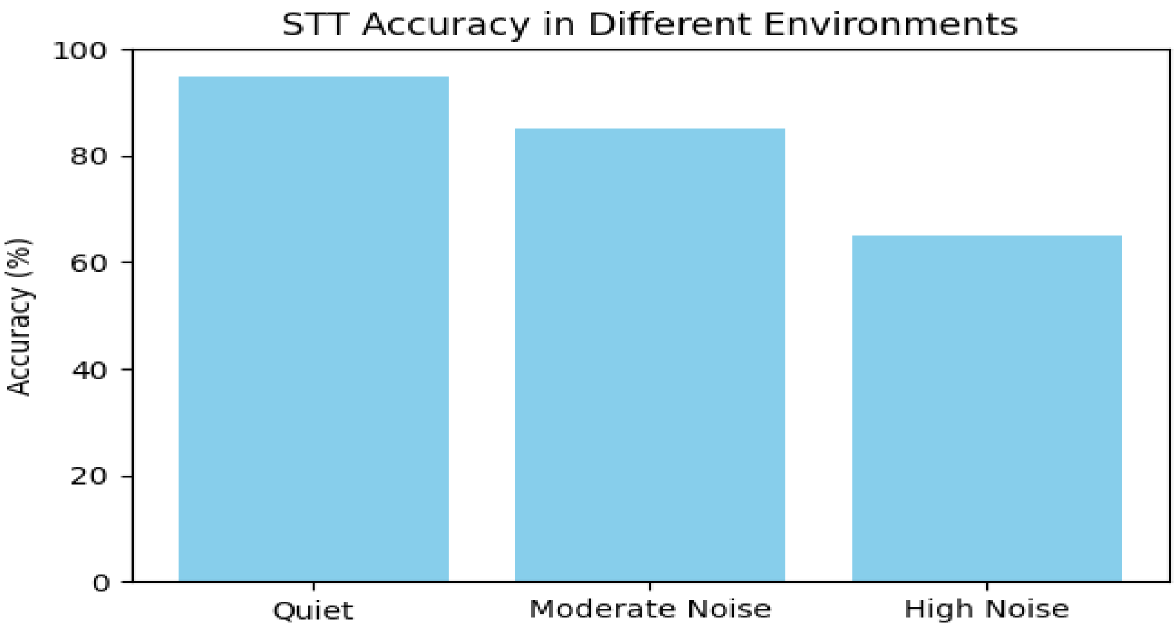

Fig. 6. STT Accuracy in Different Environments

This STT Accuracy in Different Environments displays the results of accuracy with Quite as 93%, Moderate Noise as 85% and High Noise as 65% accuracy.

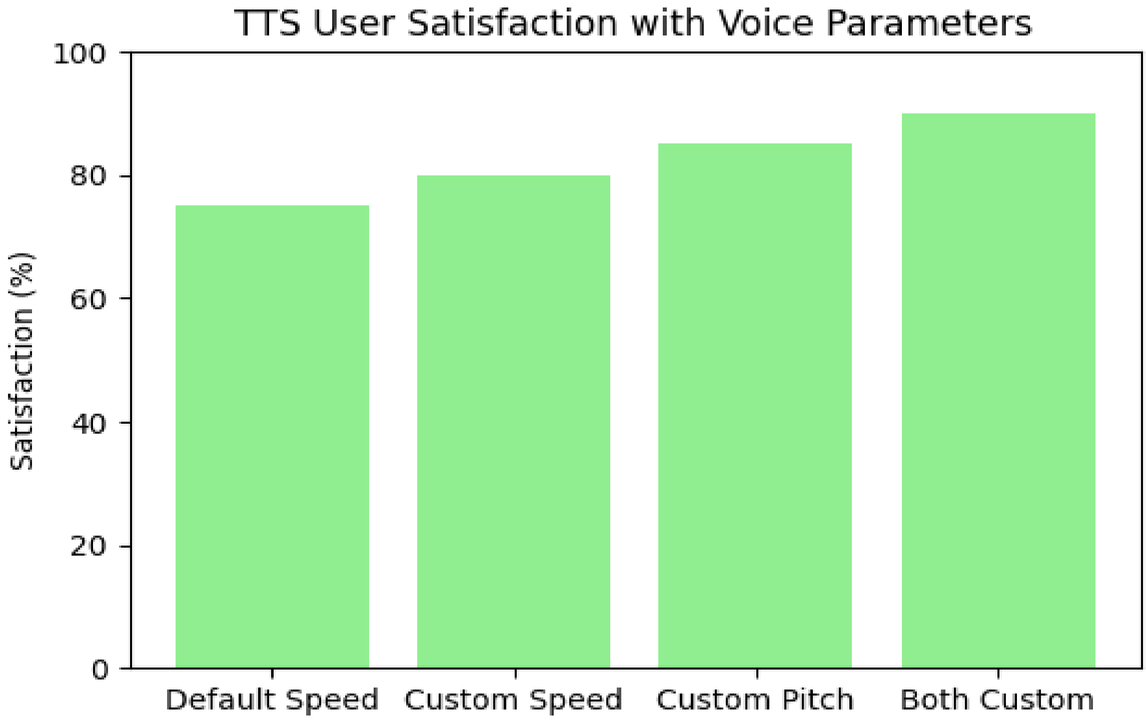

Fig. 7. TTS User Satisfaction with Voice Parameters

This TTS Satisfaction with Voice Parameters displays the results of satisfaction with Default Speed as 75%, Custom Speed as 80%, Custom Pitch as 85% and Both Custom as 90%.

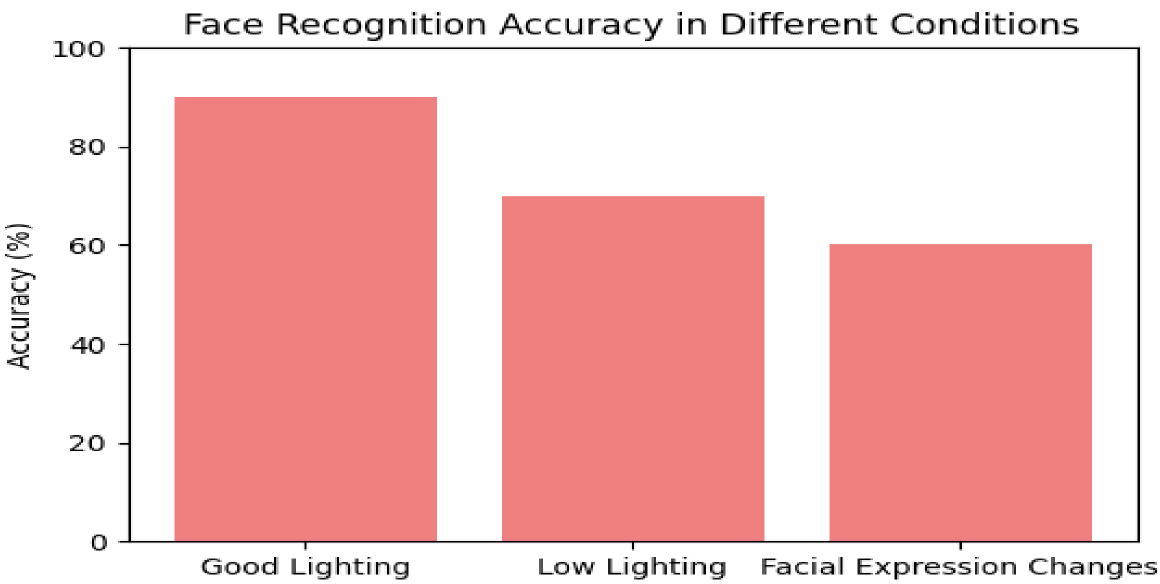

Fig. 8. Face Recognition Accuracy in Different Conditions

This Face Recognition Accuracy in Different Conditions displays the results of accuracy with Good Lighting as 90%, Low Lighting as 70% and Facial Expression Changes as 60%.

9. Conclusion and Future Scope

The voice-activated email system represents a major step forward in enhancing digital communication accessibility for individuals with visual or physical disabilities. By incorporating S-T-T, T-T-S, and voice / speech -command technologies, the system enables users to handle their emails using only voice commands, removing the reliance on visual or manual interaction.

The STT module shows impressive accuracy of 95% in quiet environments, but this drops to 85% in moderate noise and further to 65% in high noise conditions, indicating room for improvement in noisy environments. The TTS module has high user satisfaction, with 90% of users expressing a preference for customized speed and pitch settings, highlighting the importance of personalization in enhancing the listening experience. Meanwhile, the facial recognition feature, which improves security, achieves 90% accuracy in well-lit conditions, but faces challenges in low lighting (70% accuracy) and with facial expression changes (60% accuracy), suggesting a need for better adaptability to real-world scenarios.

Despite these challenges, the system promotes autonomy and independence for users, making digital communication more inclusive. Future improvements in noise handling for STT and the adaptability of facial recognition could further enhance its robustness. Overall, this system offers a practical solution for accessible email communication, with promising opportunities for future innovation.

The voice-controlled email system offers considerable potential for future advancements. One promising direction is the integration with multiple email services, allowing the system to support not just Gmail, but also other popular email providers such as Yahoo Mail and Outlook. This would provide users with greater flexibility to manage emails across various platforms. Additionally, multi-language support would enhance the system's accessibility by enabling users from diverse linguistic backgrounds to interact with the voice recognition and text-to-speech features in their preferred language. Enhanced security features could also be integrated, such as multi-factor authentication (MFA) with fingerprints or one-time passwords (OTPs), adding layers of protection for sensitive or professional accounts. Moreover, the system could benefit from integration with virtual assistants like Amazon Alexa, Google Assistant, or Siri, allowing users to manage their email functions seamlessly across different devices and environments. Extending the system’s compatibility to mobile devices and wearables (e.g., smartphones, smartwatches, smart glasses) would further enhance its accessibility, enabling users to access email functionalities on the go. Finally, incorporating cloud storage and backup solutions would ensure that users can securely store and retrieve emails, attachments, and even voice recordings, providing peace of mind with continuous data availability and safety.

References

[1] B. Dash, G. Rajput, S. Sharma, and M. F. Ansari, "The Way to Make Blind People Use the Email System: Voice Based Email Generating System Using Artificial Intelligence," in Proc. 2023 Int. Conf. on Artificial Intelligence and Smart Communication (AISC), IEEE, 2023.

[2] M. A. S. Khasawneh and Y. J. A. Khasawneh, "Availability of Voice-Recognition Devices to Support Visually Impaired Students in Saudi Arabian Universities,"Int. Journal of Educational Administration, 2023.

[3] Y. R., S. Kumar, and R. Aishwarya, "Voice Email Based On SMTP For Physically Handicapped," in Proc. 5th Int. Conf. on Intelligent Computing and Control Systems (ICICCS 2021), IEEE, 2021.

[4] S. Agarwal, S. K. Sonbhadra, M. Syafrullah, and K. Adiyarta, "Email Classification via IntentionBased Segmentation," in Proc. EECSI 2020 - 1-2 October 2020.

[5] M. Manikandan, T. Karthick and M. Anandan, "Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi," Advanced Indoor and Outdoor Navigation System for Blind People Using Raspberry-Pi.

[6] S. Khusro and A. Khan, "An Insight into Smartphone-Based Assistive Solutions for Visually Impaired and Blind People: Issues, Challenges, and Opportunities," Springer-Verlag GmbH Germany, 2020.

[7] K. Acharekar, S. Mahala, N. Bhore, and M. Waghmare, "Email System for Visually Impaired People," in Int. Journal of Engineering Research & Technology (IJERT) NTASU - 2020 Conf. Proc.

[8] J. Wang, Z. Miao, O. Biermann, K. Miao, S. Leung, and K. Gai, "Integrated Parallel System for Audio Conferencing, Voice Transcription and Speaker Identification," in Proc. 2020 Int. Conf. on High Performance Big Data and Intelligent Systems (HPBD&IS), IEEE.

[9] P. A. Tiwari, P. G. Wanjari, H. P. Nimkar, T. Rotke, U. Samarth, and P. Zodawan, "A Review on Voice Based E-Mail System for Blind," in Proc. 5th Int. Conf. on Inventive Computation Technologies (ICICT-2020), IEEE.

[10] M. Elleuch, W. Gaaloul, O. Alaoui Ismaili, N. Laga, and N. Assy, "Discovery of Activities' Actor Perspective from Emails based on Speech Acts Detection," in Proc. 2020 2nd Int. Conf. on Process Mining (ICPM), IEEE, Oct. 2020.

[11] M. S. Minu, A. Tiwari, K. Arun, and P. Rampuria, "Face Recognition System Based on Haar Cascade Classifier," May 2020 Int. Conf. Proc., vol. 29, no. 5, pp. 3799-3805.

[12] O. Kulkarni, M. Patil, N. Tejankar, and A. Alhat, "Voice Based E-Mail System for Blind People," Open Access International Journal of Science & Engineering, vol. 4, no. 1, Jan. 2019.

[13] Cuimei, Q. Zhiliang, W. Jianhua, and J. Nan, "Human Face Detection Algorithm via Haar Cascade Classifier Combined with Three Additional Classifiers," in Proc. 2017 IEEE 13th Int. Conf. on Electronic Measurement & Instruments (ICEMI).

[14] P. Ingle, A. Lanke, and H. Kanade, "Voice Based E-Mail System for Blinds," Int. Journal of Research Studies in Computer Science and Engineering (IJRSCSE), vol. 3, no. 1, pp. 25-30, 2016.

[15] Khanum, Annapurna, H.S. Voice-Based E-Mail System for Visually Challenged People. SSAHE Journal of Interdisciplinary Research, 4(1), 43-50, 2023.