Grenze International Journal of Engineering and Technology, June Issue

From Voice Commands to Healthcare: A Voice-based Email Assistant for Visually Impaired Users

Bhargavi Peddi Reddy[1] and D. Chandana Sudha[2]
[1-2]Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, Telangana, India
Email: bhargavi@staff.vce.ac.in, chandanasudha4@gmail.com

***Abstract*— This project focuses on developing a voice-based email assistant to help visually impaired users perform email operations without using a keyboard or screen. The system was implemented using Python and Django, and it allows users to compose, send, and read emails using voice commands. Speech recognition was used to convert spoken input into text, and text-to-speech was used to provide audio responses to the user.**
The generative AI module was integrated to structure the spoken input into properly formatted email content. SMTP and IMAP protocols were configured to enable sending and receiving emails securely. The system was tested using real-time voice input, and it successfully performed email communication tasks. This implementation improves accessibility and helps visually impaired users communicate independently using voice interaction.

***Keywords*— Voice-based Email Assistant, Speech Recognition, Text-to-Speech, Generative AI, Accessibility, Visually Impaired Users, Healthcare Communication, Django Framework.**

I. INTRODUCTION

Digital communication technologies have become an essential part of modern daily life, particularly email, which serves as a primary medium for personal, professional, and healthcare-related interactions. However, most existing email platforms are designed with graphical user interfaces that heavily depend on visual interaction. This creates significant accessibility barriers for individuals with visual impairments, limiting their ability to independently access and manage digital communication.

Although assistive technologies such as screen readers provide basic accessibility, they often require users to remember multiple keyboard shortcuts and navigate complex interfaces. This increases cognitive effort and makes tasks such as composing structured emails more difficult.

These challenges become more critical in healthcare scenarios, where timely and accurate communication is necessary for appointment scheduling, receiving medical information, and maintaining communication with healthcare providers.

Recent advancements in speech recognition and artificial intelligence have created new opportunities to improve accessibility.

Voice-based interaction allows users to communicate with systems naturally through speech, reducing reliance on visual interfaces. Generative AI models further enhance this capability by converting spoken input into properly structured and meaningful written communication.

This project implements a voice-based email assistant using speech recognition, generative AI, and text-to-speech technologies.

Grenze ID: 01.GIJET.12.2.316_26

The system was developed using Python and Django framework and tested using real voice input. The implementation allows users to perform email operations such as composing, sending, and reading emails using voice commands.
This system improves accessibility and provides a practical solution for visually impaired users, especially for healthcare-related communication.

II. Problem Statement

Despite the availability of assistive tools, visually impaired individuals continue to face major difficulties when accessing and managing email systems. Most email platforms require visual navigation, keyboard interaction, and manual input, which are not suitable for users who depend entirely on audio-based interaction.
Screen readers provide basic functionality but often lack intelligent assistance for composing structured and meaningful email content.
These limitations are especially problematic in healthcare communication, where patients must interact with hospitals and medical professionals through email for appointments, prescriptions, and important medical updates. Existing systems do not provide sufficient automation, intelligent assistance, or fully voice-based interaction tailored to the needs of visually impaired users.
Therefore, there is a need for a voice-enabled email assistant that allows users to perform all email operations using speech while automatically generating well-structured email content. Such a system can improve accessibility, reduce dependency on external assistance, and enhance communication efficiency, especially in healthcare environments.

III. Background Work

Voice-based interaction systems have emerged as an important solution for improving accessibility for visually impaired users, as they allow interaction without relying on visual elements. Traditional assistive tools such as screen readers, including NVDA and JAWS, convert visual content into audio output. While these tools provide basic accessibility, they often require extensive keyboard navigation and familiarity with commands, which can increase user effort and reduce efficiency during complex tasks like email composition and management.
Commercial voice assistants such as Google Assistant and Amazon Alexa have demonstrated the effectiveness of speech-driven interfaces for performing general tasks. These systems can read messages and respond to simple commands; however, they are limited when handling structured communication such as drafting professional emails. They typically lack the ability to understand detailed user intent and generate properly formatted content suitable for formal or healthcare communication.
Recent developments in generative artificial intelligence have introduced advanced capabilities in natural language understanding and content generation. Large language models can analyze user input, interpret context, and produce coherent and meaningful text output. These capabilities make generative AI highly useful for applications that require automated text
creation, such as email drafting. In accessibility-focused systems, generative AI can assist users by transforming spoken input into clear and professionally structured messages.
Several research studies have explored voice-controlled email systems designed for visually impaired users [2], [3], [4], [6]. These systems demonstrated the feasibility of using speech recognition and text-to-speech technologies to enable email access through voice commands. However, many of these solutions focused mainly on executing basic commands such as reading or sending emails and did not include intelligent content generation features. As a result, users still needed to manually adjust or refine email content.
Speech recognition and text-to-speech technologies play a critical role in voice-based accessibility systems [8], [9]. Modern speech recognition models provide improved accuracy in converting spoken language into text, even in real-world environments.
Similarly, advanced text-to-speech systems generate clear and natural audio responses, improving user understanding and interaction. These technologies reduce user effort and make communication systems more accessible.
Building upon these existing advancements, the proposed system integrates speech recognition, generative AI, and secure email communication into a unified platform. By combining intelligent email generation with voice-based interaction, the system provides a more effective and accessible solution for visually impaired users. This approach improves communication independence and enhances accessibility, particularly in important areas such as healthcare communication.

TABLE I. COMPARISON OF EXISTING SYSTEMS AND THE PROPOSED SYSTEM

Criteria	Existing System	Proposed System
AI-Email Composition	Limited	AI-Assisted
Voice Interaction	Partial	Voice-Controlled
Email Support	Basic SMTP/IMAP	SMTP/IMAP + AI
Audio Feedback	Standard TTS	gTTS
Accessibility	Limited Support	Accessibility-Focused

A. Recent Research Developments

Recent studies have focused on combining artificial intelligence, speech processing, and assistive technologies to improve accessibility for visually impaired individuals. Recent studies published during 2024 and 2025 have demonstrated the effectiveness of large language models in generating structured textual content from natural speech input.

These models have shown significant potential in reducing user effort during communication tasks and improving the quality of generated content [10], [11], [12].

Several accessibility-oriented systems have also explored the integration of speech recognition and text-to-speech technologies for hands-free interaction.Recent investigations reported improvements in speech recognition accuracy, adaptive voice interfaces, and multilingual support, enabling more natural communication experiences for users with disabilities [8], [9].

In healthcare communication environments, voice-enabled assistive systems have been applied to appointment scheduling, medical information access, and patient-provider communication. These developments indicate a growing interest in intelligent voice-driven solutions that promote accessibility and independent digital communication.

Motivated by these advancements, the proposed system integrates speech recognition, generative AI, and secure email communication into a unified framework for visually impaired users.

IV. METHODOLOGY

The proposed system is designed to provide a fully voice-controlled email interaction platform for visually impaired users. The architecture integrates multiple functional components that work together to capture voice input, process commands, generate email content, manage communication, and deliver audio feedback. The design focuses on accessibility, reliability, and efficient performance using open-source technologies and standard communication protocols.

A. Speech Interaction

The system begins by capturing the user's spoken commands through a microphone interface. The speech_recognition library is used to convert the captured voice signals into text format. This conversion allows the system to understand user instructions such as composing, reading, or sending emails. By enabling natural voice input, this module eliminates the need for manual typing or visual interaction, making the system more accessible for visually impaired individuals.

The speech recognition process is designed to handle variations in speech patterns and provides reliable transcription of user commands. The module was tested using a laptop microphone, and it successfully converted voice input into text during real-time testing.

B. Audio Response Generation

To ensure continuous interaction, the system provides audio feedback for every operation performed. The gTTS (Google Text-to-Speech) library is used to convert system-generated text into spoken output. The generated speech is then played using the playsound module.

This feedback mechanism informs users about system status, confirms executed actions, and guides them through various operations. Providing clear audio responses helps users understand system behavior and improves overall usability.

C. Artificial Intelligence Processing

Once the user's spoken input is converted into text, the system uses the google.generativeai module to process and refine the content. The generative AI model analyzes the input and converts it into a structured and meaningful email format.

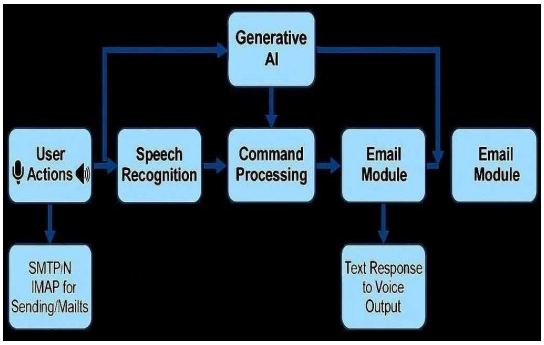

Fig. 1. System Architecture

It improves grammar, organizes the content logically, and ensures that the message is suitable for professional communication. This automated content generation reduces user effort and ensures accurate email composition without requiring manual editing.

D. Email Communication

The system uses Python-based email libraries such as smtplib, imaplib, and email.mime to manage email operations. These libraries enable the system to send new emails, retrieve incoming messages, and manage attachments securely. The system interacts with email servers using standard SMTP and IMAP protocols, ensuring reliable communication. Users can perform all email-related tasks using voice commands, including composing, sending, and reading messages.

E. Web Framework Management

The Django framework serves as the core platform for managing system functionality and user interaction. It handles user authentication, manages sessions, and coordinates communication between different system modules.

Django provides a secure and scalable environment that supports efficient system performance and ensures safe handling of user data.

F. System Utilities

Supporting modules such as os, threading, and time are used to manage system-level operations. These modules assist in file handling, background processing, and task scheduling.They help maintain smooth system performance by ensuring efficient execution of multiple processes simultaneously. These utilities contribute to system stability and improve overall responsiveness.

V. Algorithms and Technical Components

This system does not introduce a completely new machine learning algorithm but integrates existing advanced technologies to build an effective voice-based email assistant. The system combines speech recognition, generative AI, text-to-speech synthesis, and standard email communication protocols to deliver accessible and intelligent functionality. The strength of the system lies in the efficient integration of these technologies to enable seamless voice-driven email interaction.

A. Speech Recognition

The speech recognition module is responsible for converting spoken input into text. The system uses the speech_recognition library, which interacts with advanced speech processing models to perform accurate voice-to-text conversion. These models are trained to recognize different speech patterns, accents, and variations in pronunciation. The module captures audio input through a microphone and processes it to identify spoken words. This approach allows users to interact with the system naturally using voice commands. By using reliable speech recognition services, the system ensures accurate command interpretation and smooth user interaction.

B. Generative AI For Email Composition

The generative AI component plays an important role in converting spoken input into structured email content. The system uses Google's generative AI model to analyze user input and generate properly formatted email

messages. The model understands the meaning and intent of the spoken text and produces clear and meaningful written communication.
This process improves grammar, organizes information, and ensures professional email formatting. By automating the email composition process, the system reduces user effort and enables visually impaired users to communicate effectively without manual typing.

C. Text to Speech Synthesis

The text-to-speech module converts system-generated text into spoken audio output. The gTTS (Google Text-toSpeech) library is used to generate natural and understandable speech. This allows the system to provide audio feedback for commands, notifications, and email content.
The audio output helps users understand system responses and interact confidently without needing visual confirmation. This feature improves accessibility and ensures continuous communication between the user and the system.

D. Email Protocol Libraries

The system uses standard Python libraries such as smtplib and imaplib to handle email communication.
These libraries allow the system to send emails, receive messages, and manage attachments securely. The implementation follows established email communication protocols such as SMTP and IMAP.
These protocols ensure reliable and secure communication between the system and email servers. Error handling mechanisms are included to maintain system stability and ensure proper operation during communication processes.

VI. RESULTS

A. Experimental Setup

The developed voice-based email assistant was assessed through a series of functional evaluations involving five participants. Testing was performed on a Windows-based computing platform using a standard microphone and noise-reduction audio devices to improve speech capture quality.
The application was implemented using Python and the Django framework, with integrated modules for speech recognition, text-to-speech conversion, generative AI processing, and email communication through SMTP and IMAP protocols.
During the evaluation process, participants executed various email-related operations, including command recognition, email drafting, message transmission, inbox access, and attachment management.
The purpose of the assessment was to examine the operational performance, accessibility, and reliability of the proposed system under practical usage conditions.
The experiments were conducted in an indoor environment with moderate ambient noise levels. System performance was
analyzed by monitoring the successful execution of user requests and the overall completion of assigned tasks.
The observations obtained from these evaluations were used to assess the effectiveness of the proposed voice-driven email communication platform.

B. Functional Evaluation Results

The developed system was subjected to functional testing to verify its ability to perform voice-driven email operations. Five participants were asked to execute a set of common tasks, including voice command recognition, email composition, email transmission, email retrieval, and attachment handling. The outcome of each task was recorded based on successful completion without requiring manual intervention.

TABLE II. FUNCTIONAL EVALUATION RESULTS

Task Performed	Number of Trials	Successful Executions	Success Rate (%)
Voice Command Recognition	5	5	100
Email Composition	5	4	80
Email Sending	5	5	100
Email Reading	5	4	80
Attachment Handling	5	4	80

The evaluation results indicate that the proposed system was capable of performing the majority of email-related operations successfully through voice interaction. Voice command recognition and email transmission achieved

the highest performance, while email composition, reading, and attachment processing occasionally required repeated input due to speech interpretation variations.
Overall, the system demonstrated reliable functionality and provided an accessible communication platform for visually impaired users.

C. User Feedback Analysis

A qualitative evaluation was conducted to understand the usability and accessibility of the proposed system. The five participants who tested the application were asked to provide feedback based on ease of interaction, system responsiveness, voice communication quality, and overall user experience.
The collected responses indicated that the voice-based interface simplified email operations and reduced dependence on conventional input methods. Most participants reported that the
audio feedback mechanism improved navigation and made the system easier to operate.
The results suggest that the proposed solution provides an accessible and user-friendly communication platform for visually impaired users.

TABLE III. USER FEEDBACK ANALYSIS

Evaluation Parameter	Rating (out of 5)
Ease of Use	4.4
Voice Interaction Quality	4.2
Accessibility Support	4.6
System Responsiveness	4.1
Overall Satisfaction	4.5

D. Discussion

The testing results demonstrate that the proposed voice-based email assistant can successfully perform essential email operations through speech-driven interaction. The integration of speech recognition, generative AI, and text-to-speech technologies enabled users to compose, send, and retrieve emails with minimal manual effort. Although minor variations in speech recognition were observed in noisy environments, the overall system performance remained stable. The findings indicate that the proposed solution can improve accessibility and support independent communication for visually impaired individuals. Future enhancements may focus on multilingual support, improved noise handling, and integration with additional assistive services.

E. Limitations

Although the proposed system demonstrated satisfactory performance during functional evaluation, certain limitations were identified. The accuracy of speech recognition may vary depending on environmental noise conditions, microphone quality, and user pronunciation.
The current implementation primarily supports English-language interaction and was evaluated using a limited number of participants. Additionally, cloud-based AI services require internet connectivity, which may affect response time in low-network environments. Future improvements can focus on multilingual support, enhanced noise reduction techniques, and large-scale user evaluation.

VII. CONCLUSION

The developed system successfully enables visually impaired users to perform email operations using voice commands. The system was implemented using Python, Django, speech recognition, and generative AI modules. Testing confirmed that the system can compose, send, and read emails without requiring visual interaction. The integration of speech recognition and text-to-speech improved accessibility and made the system easy to use.
The integration of generative AI improves the quality and clarity of email content by converting spoken input into structured and meaningful communication. Text-to-speech functionality ensures continuous audio feedback, allowing users to interact confidently with the system. The use of standard email protocols ensures secure and reliable communication, making the system suitable for real-world applications, including healthcare communication.
The proposed solution enhances accessibility, reduces user effort, and promotes independent digital communication for visually impaired individuals. Future improvements may include multilingual support, improved response speed, and integration with additional healthcare and scheduling systems. This work contributes to the advancement of assistive technology by providing a practical and scalable solution for inclusive communication.

REFERENCES

[1] S. Noel, "Human computer interaction (HCI) based Smart Voice Email (Vmail) Application - Assistant for Visually Impaired Users (VIU)," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 895-900, doi: 10.1109/ICSSIT48917.2020.9214139.

[2] B. V. Patil and K. Sreelakshmi, "Implementation of Voice Based E-Mail System for Visually Challenged," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-9, doi: 10.1109/INCOFT55651.2022.10094386.

[3] N. K. Prakash, T. Monlam, R. Singh, R. Aravindhan, P. V. Reddy and R. Jain, "Voice based E-Mail with Attachment for Blind," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 01-07, doi: 10.1109/ICOEI53556.2022.9776840.

[4] A. Sharma, V. Ahmed, S. Sharma, B. Jana and K. Rani, "An Effective Approach to Speech-based Email Assistance for Visually Impaired People," 2022 8th International Conference on Signal Processing and Communication (ICSC), Noida, India, 2022, pp. 32-35, doi: 10.1109/ICSC56524.2022.10009538.

[5] M. Rajole and R. Phursule, "Voice Based E-Mail System for Visually Impaired Peoples Using Computer Vision Techniques: An Overview," 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCUBEA58933.2023.10391988.

[6] T. S. Roy, N. Namratha and T. Y. J. Naga Malleswari, "Voice E-Mail Synced with Gmail for Visually Impaired," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 802-807, doi: 10.1109/ICAIS56108.2023.10073879.

[7] B. Sharma, B. Roshan, M. Gaur and P. De, "An Email Assistant System based on Voice Command," 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAEECI), Tiruchengode, India, 2023, pp. 1-5, doi: 10.1109/ICAEECI58247.2023.10370972.

[8] S. A, M. B and S. R. R, "Voice-Activated Email System: Enhancing Accessibility with Speech-to-Text Integration," 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICSTSDG61998.2024.11026142.

[9] C. Sasikala, R. M, S. A. S, M. U and H. A. M, "Enhancing Blind Communication With Voice-Assisted Email," 2024 International Conference on Expert Clouds and Applications (ICOECA), Bengaluru, India, 2024, pp. 307-312, doi: 10.1109/ICOECA62351.2024.00062.

[10] K. Vignesh, B. Baranitharan, D. Sushmanth, Y. S. Kumar, A. K. K. Reddy and P. Srinuvasulareddy, "A Voice-based Email System for Visually Impaired Users," 2025 5th International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN), Coimbatore, India, 2025, pp. 25-32, doi: 10.1109/ICECMSN68058.2025.11383340.

[11] S. P. Singh Rathore, S. Yadav, N. K. Sharma, N. Yamsani, B. L. R and S. Muthusundari, "Enhancing Email Interaction for Visually Impaired Users Through a Voice-Controlled AI Interface," 2025 Modern Electronics Devices and Intelligent Communication Systems (MEDCOM), Greater Noida, India, 2025, pp. 1003-1007, doi: 10.1109/MEDCOM67532.2025.11405170.

[12] K. Rani, L. Verma, A. Grover, H. Saini and O. Satapathy, "Enhancing Email Usability with Speech and NLP for Blind and Visually Impaired Users," 2025 Seventh International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2025, pp. 652-656, doi: 10.1109/CCICT65753.2025.00104.