

# Vmail: A Voice-Assistive Solution Enabled with a Chatbot

DHANASHREE J K<sup>1</sup>, JESICA M<sup>2</sup>, YAAMINI N<sup>3</sup>, T. SUBASHINI<sup>4</sup>

<sup>1,2,3</sup>*Department of Computer Science and Engineering Velammal Engineering College, Chennai*

<sup>4</sup>*Assistant Professor, Department of Computer, Science and Engineering, Velammal, Engineering, College, Chennai*

**Abstract-** *Because the majority of conventional email systems rely on keyboard interaction and visualization, they are not accessible to blind or visually impaired people on their own. Even with the help of screen readers, writing and navigating emails is still very difficult and takes a long time. This paper describes VMAIL, an email system that enables visually impaired people to use voice commands to accomplish all email functions. VMAIL uses Text-to-Speech to read email content aloud and Speech-to-Text to record user commands. Through OAuth 2.0 authentication, VMAIL gives users safe access to Gmail accounts and the Gmail API. Additionally, VMAIL offers an A.I.-based chatbot to help users with email drafting, formal message rewriting, and content summarization. To avoid mistakes, voice confirmation will be done before any emails are sent. Based on the VMAIL email system's experimental use, we found that it improves digital communication's inclusivity for people who are blind or visually impaired, increases email access, and requires less effort from users when communicating via email.*

**Index Terms-** *Gmail API, Chatbot Support, Accessibility, Assistive Technology, And Speech Interface.*

## I. INTRODUCTION

Although most email services use a keyboard and visual elements for input, electronic mail is a significant communication tool in both professional and academic contexts. Unfortunately, blind users have accessibility issues due to these systems' design. Independent email use is limited by intricate layouts, images, and menu-driven navigation.

Currently, screen readers and keyboard shortcuts are commonly used by blind users to access email services. These tools may provide basic text-to-speech functionality, but before they can accomplish tasks, users must learn a large number of commands

and follow numerous steps. As a result, activities like writing emails, reading emails, and organizing folders quickly become tiresome and lead to errors. Additionally, the solutions that are currently available on the market are either difficult to use or do not support natural voice interaction.

In order to enable blind users to use a full email system through spoken interaction, this paper proposes a voice-enabled e-mail system (VMAIL). In the suggested system, text-to-speech technology will read email content aloud to users, giving them auditory feedback, and speech will be recorded using a speech-to-text recognition process to interpret user commands. Users will be able to securely access Gmail services by using the Gmail API and OAuth2 authentication. Furthermore, an artificial intelligence (AI) bot will assist users with email message composition, formal email message rewording, and email message summarization. The suggested system combines AI support with speech interaction.

## II. PROPOSED METHODOLOGY

### A. Background

In recent years there has been a dramatic increase in the number of voice activated assistive products that have been developed for people with disabilities (e.g. blindness). At present, most assistive technologies available for visually impaired people rely on screen reader technologies that take text written on the computer and present it via audible speech. In order to access the Internet email through a screen reader, a user is often required to use keyboard shortcuts and also has to navigate through complex user interfaces, resulting in a significant increase in a user's cognitive load compared to if they had been able to communicate with the computer via voice first.

With advances in speech-to-text (STT), text-to-speech (TTS), and conversational artificial intelligence (AI), the quality of personal computer (PC) communication is rapidly improving. Voice first technology allows a user to perform actions using voice commands rather than relying on visual (graphic) information. Existing examples of applications that are voice enabled offer limited features; therefore, they may not provide all the functions needed to create a fully functioning email workflow. A unified system that incorporates speechbased interaction and provides secure access to email and creates content using conversational AI will help to address many of the current limitations of existing voice enabled email applications.

#### System Design and Data Flow

The VMAIL Framework consists of a Voice Commands enabled Client/Server Architecture. The users speak voice commands through the Web Browser, which serves as the client side of the VMAIL framework; at the same time, the computer is connected to the server and uses Voice Commands/Voice Instructions to perform actions. The Voice Instruction is converted to Text through Speech-To-Text technology, then VMAIL analyzes the Text to determine what Action to perform for the user (i.e., Show Inbox, Compose E-mail, or Use ChatBot). The Node.js\_Server, utilizing Express.js, performs Authentication, Routing and communicates with External Services. The use of OAuth 2.0 Authentication (Authentication Protocol) and the Gmail APIs provide access to Gmail services for Users in a Secure manner. Through the VMAIL framework, mail that has been sent or deleted by the User will be accessed in real-time by the user's Browser as a Voice Recordings through the Voice to Text component of the VMAIL Framework will read the email content back to the user.

#### B. Voice-Controlled Email Tasks

The main function of the VMAIL framework will support (1) Voice Activated Login (2) Navigation through their Email (3) Read Email, and (4) Compose Email. The User will speak their Email Address and Password as a Voice Command for Login. The VMAIL framework will securely validate the user's Email address and Password. After having been

authenticated, the user can navigate through a voice guided menu of (Gathering) Items and command by saying what he/she wants to access (e.g., Inbox, Compose, Sent, Trash, Chat, etc.).

Composing an email involves providing the bot with three elements: (1) recipient/or email address, (2) subject/title, and (3) the actual content of the email through voice command. Once the email is created, the system reads the entire email (recipient, subject and content) back to the user to confirm what was created before it will send it. The performance of the overall system is assessed based on how accurate the function is, the accuracy of the response given to the user, and the ease of use for the user, rather than on quantitative method to evaluate prediction, due to the fact that the evaluation is being performed on an application basis.

#### C. AI-Assisted Content Processing

An A.I. powered bot has been integrated into the email communication system to improve the quality of email communication. The bot will help users perform tasks related to composing an email, changing the tone of the email from informal to formal, and summarizing emails using the voice input from the user and the audio response produced via TTS technologies.

By enabling the use of artificial intelligence to make structured email messages more efficient for creation by users while improving the clarity of recipients' communication with others, this approach provides a solution that entails the combination of three technologies: a speech recognition system, Google's Gmail Application Programming Interface (API), and other forms of conversational AI in order to create an end-to-end, voice-enabled email system. The end-to-end, voice-enabled email system will be fully evaluated for its overall functionality and user experience and subsequent enhancements will include both non-English language testing and creation of a user-specific interface (i.e., Interface Design).

### III. LITERATURE REVIEW

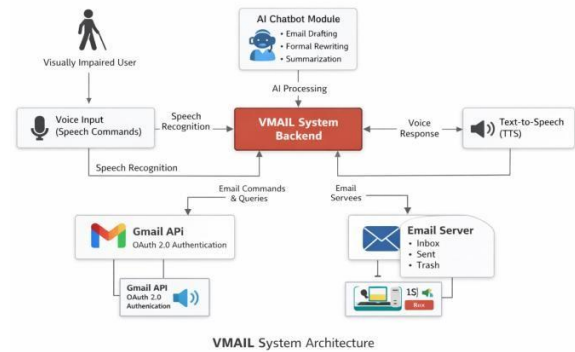
To allow visually impaired individuals access to e-mail, innovative ways of developing an email system that allows visually impaired individuals use e-mail through a method of providing another means of accessing e-mail via the development of a voice interface to be able to conduct standard email functions through voice command (i.e., in lieu of using visual interfaces) has been performed thus allowing for the reduction and/or eventual elimination of reliance upon visual interfaces.

Unfortunately, due to the limitations of features and lack of intelligent assistant features present in voice-based email systems would result in the limited ability to effectively perform the full range of functions for the majority of users. A newly developed voice e-mail system was implemented for visually impaired individuals in 2022 which allows people to send and/or receive e-mail through a voice command. While this system has provided an ability to send and/or receive email through voice command, it has no AI capability of drafting e-mails; there is no integration with Gmail for real time e-mail access; and there is no ability to confirm e-mail created or sent prior to sending it. As a result, users continue to make unintentional mistakes when composing and/or sending their e-mails.

In 2021, a voice e-mail system that used SMTP was developed allowing people to use their voice to send e-mails. This system allowed users to perform basic e-mail functions using traditional SMTP protocols, however it did not have the capability of integrating with Gmail via the use of the Gmail API. In addition, the system did not include any type of natural language processing (NLP) capability to assist users in drafting and/or understanding information contained in an e-mail message.

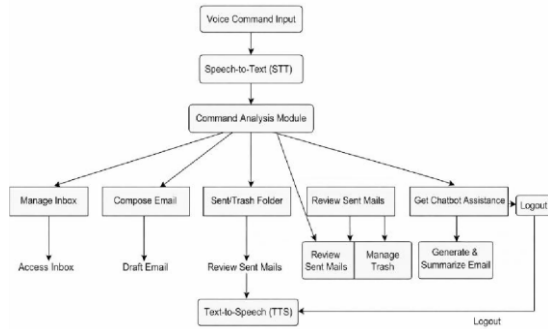
The advancements of Speech Technology have led to the Voice Activated E-mail System being developed in 2024. Voice Activated E-mail System utilizes Speech-to-Text technology in providing users hands free ability to interact with e-mail systems. This new e-mail system supports handsfree operation of e-mail, but does not include support for automatic intelligent drafting; enhancing the quality of e-mail content; or conducting a voice verification prior to sending out an e-mail.

Architecture Diagram



In order for blind people using the VMAIL system, a voice-enabled e-mail program, to retrieve their e-mails by way of voice, a structural architecture must be developed to support the delivery of voice interaction with the e-mail program and to support security and efficiency. The architecture must provide alternative ways for the user to perform all functions in relation to their e-mail through the use of voice commands, and as a direct result of the development of this architecture; the user will rely less on the display of any electronic device to access and fulfill their responsibilities in relation to their e-mail, thereby virtually eliminating the user's dependency on any visual display.

The first component of the architecture will be the ability of the user to input their voice into a web-based client application. The voice will be the only method for communicating with the e-mail program; all commands that enable the user to log, navigate, retrieve, compose, or delete e-mail will come from the voice input. The voice input will constitute the user's command to log in, navigate, retrieve, create, or delete email messages, etc. The user will say a command out loud, and it will be turned into text by a machine almost immediately following their spoken words.



Once the user speaks an audio input to the system, STT technology will convert their voice to text to allow the system to properly interpret what the user is trying to do. There will be additional steps involved in this conversion process that are intended to improve the accuracy of the recognition process by performing basic audio processing on the user's input including removing background noise and cutting off the user's voice altogether, and producing high-quality STT output from which to generate.

Once the audio input is converted into text using STT technology, it will then be transferred to the command interpretation system that will help interpret the command. This command interpretation system will then be used to determine the user request to either check an email in their inbox, create a new email, reply to an existing email, or seek assistance from the chatbot. The user's intent is determined by command analysis so that appropriate actions can be performed to fulfill the user's intent.

In the case of email actions, a secure connection to the user's Gmail account is established via the Gmail API with OAuth 2.0 authentication. This allows access to the user's inbox, sent mail, deleted mail (trash), and to send email messages. For requests for user assistance around content, the command analysis module relays these types of requests to the AI Chatbot module, which assists the user with composing email messages, revising email messages into a more formal tone, and summarizing email messages.

Once the user has requested assistance from the TTS process, the user will receive an audio confirmation of their actions prior to the user finalising any actions using their voice. Finally, the system will close out the interaction with the user by providing an audio response confirming that the user's task has been completed.

#### IV. RESULT

VMAIL has been tested to evaluate how completely auditory-based access is to sending and using eMail for users without vision. The evaluation was based on functional correctness and accuracy of response as well as effectiveness of auditory functionality across the various functional units of VMAIL.

To date, testing has been limited to a combination of scripted evaluations and functional verification using pre-defined test cases that define the specific features of VMAIL including voice login, navigation menus, reading inboxes, composing email, requesting help from the chatbot, and sending email. All responses from VMAIL have been verified using completion examples of completed task (i.e., each task) as well as varying spoken forms of feedback for all task examples.

Preliminary evaluation results indicate that VMAIL integrates speech-to-text efficiently, accurately captures user commands, and provides adequate audio feedback to users. VMAIL allows users to view messages in their inboxes, messages and messages in the trash using Gmail (Google web-based eMail) and delivers and retrieves all securely through Google's secure eMailbox architecture.

VMAIL also allows users to progress from simple-to-complex AI technology through its chatbot (i.e., allows users to create drafts, alter email message content, learn other features of VMAIL).

## V. FUTURE WORK

The VMAIL team will continue improving the complete voice-based email solution for visually impaired users with speech-to-speech interactions facilitated by AI support. With the goal of being able to have low-latency, real-time voice interaction and high accuracy in speech recognition, future enhancements are planned.

By making VMAIL support multiple languages, we would be able to achieve a broader reach with our solution, allowing users to interact with the system in their native or regional language. In conjunction with supporting multiple languages, we could develop a mobile application or dedicated webbased dashboard to create a more flexible and user-friendly interface.

Using advances in natural language processing (NLP) technology, we could help provide intelligent suggestions for responding to email messages and analyze emotionally charged messages in determining how to structure the response. Continued enhancement would allow for personalization of the responses based on improved AI capabilities that consider the user's preferences and communication channels when responding.

Security-related enhancements could include investigating the implementation of voice-based authentication and other biometric verification methods to provide safe access for the user, while avoiding substantial complexity in the solution. The enhancements that could occur to VMAIL including adding more powerful command processing offline during outage times, making it more fault tolerant during network down times, and enhancing system reliability through redundant elements of its architecture.

By performing extensive user testing, (based on pragmatic measurements such as time spent completing a task, error rates, and customer satisfaction) the value VMAIL is delivering to visually impaired users can be further advanced.

## VI. CONCLUSION

This project addresses how to allow individuals with vision loss to independently communicate via email

through a voice-based platform that requires only the use of voice command instead of visually and by using a computer keyboard. This type of task is significantly time-consuming for those with vision loss. To facilitate this accessibility challenge, VMAIL provides a complete voice-assisted email experience that combines Speech-to-Text, Text-to-Speech, Gmail API access, and an AI-based chatbot.

The pipeline that provides a way to interact with their email on VMAIL is built upon the OAuth 2.0 security model, which uses the actual Gmail services to allow users to perform some basic tasks with email (i.e., reading email in inbox, writing emails, deleting or managing sent or deleted emails) and receiving audio responses from the VMAIL service. The VMAIL email system includes an AI-based chatbot that helps users create formal and well-written emails by drafting emails, rewriting them in a formal writing style, and summarizing their drafts. This adds efficiency by decreasing the amount of effort it takes for the user to write email and helps improve the quality of email communications. Voice confirmation before sending email has also been found to increase reliability and usable features of the email system.

As a result, through pairing both a speech-based user interface with AI support from the chatbot, visually impaired individuals will be provided with an effective and workable means of being able to access email independently. This further confirms that use of voice-controlled interfaces may reduce the need for a user to utilize a screen reader and eliminate difficult-to-follow navigation paths which would provide a more user-friendly and effective overall experience.

## REFERENCES

- [1] G. Kalyani and B. S. Murthy, "Voice Based Email for Visually Impaired Person," *International Journal of Engineering & Science Research*, vol. 15, no. 2s, pp. 397–401, Apr.–June 2025, PDF available: [Download PDF – Voice Based Email for Visually Impaired Person](#)
- [2] S. Prakash, *Voice Based Email System for the Visually Challenged*, ResearchSquare, 2024,

PDF available: Download PDF – Voice Based Email System for the Visually Challenged

- [3] Md. Jalal U. Chowdhury and A. Hussan, “A review-based study on different Text-to-Speech technologies,” arXiv, Dec. 2023, PDF available:  
<https://arxiv.org/pdf/2312.11563.pdf>
- [4] L. Clark et al., “The State of Speech in HCI: Trends, Themes and Challenges,” arXiv, Oct. 2018, PDF available:  
<https://arxiv.org/pdf/1810.06828.pdf>
- [5] “User Experience and Usability of Voice User Interfaces: A Systematic Literature Review,” ResearchGate, 2023, PDF available: Download PDF – User Experience and Usability of Voice User Interfaces
- [6] M. D. Mulfari et al., “A Voice User Interface on the Edge for People with Speech Impairments,” Electronics, vol. 13, no. 7, 2024, PDF available: Article – A Voice User Interface on the Edge for People with Speech Impairments
- [7] P. Surekha, “Voice-based e-mail for Information Systems: An aid to accessibility,” E3S Web Conf. 2023, PDF available: Voice-based e-mail for Information Systems (abstract/full text)
- [8] “Voice Based E-Mail Generating System Using AI,” 2025, PDF available: Voice Based E-Mail Generating System Using AI (IJARSCT)
- [9] Speech Recognition & Synthesis, Wikipedia (overview of speech recognition and TTS technologies), 2025, HTML available:  
[https://en.wikipedia.org/wiki/Speech\\_recognition\\_%26\\_synthesis](https://en.wikipedia.org/wiki/Speech_recognition_%26_synthesis)
- [10] Whisper (speech recognition system), Wikipedia — details on OpenAI’s Whisper model used for transcription in voice systems, 2025, HTML available:  
[https://en.wikipedia.org/wiki/Whisper\\_\(speech\\_recognition\\_system\)](https://en.wikipedia.org/wiki/Whisper_(speech_recognition_system))