AI-based Voice E-Mail System for Visually Impaired

N.P.Shangara Narayanee[1], M.Nathiya[2], V.Preethi Vihashini[3], G.Veeramani[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College (Autonomous) Thudupathi, Erode, Tamilnadu, India

[2,3,4] Final Year, Department of Computer Science and Engineering, Erode Sengunthar Engineering College (Autonomous) Thudupathi, Erode, Tamilnadu, India

E-mail: [3]phashini73@gmail.com

Abstract

Electronic communication tools, such as email are visual in nature, and accessing as well as using them might be difficult for visually impaired people. This research present an email communication system designed to beeasy to use and accessible, aiding individuals in overcoming obstacles associated with visual impairment The suggested method transforms the text-based email communications into speech using natural language processing algorithms, in order to makes it simpler for those with visual impairments to understand and reply to the emails received. Additionally, the system incorporates the speech-to-text technology, enabling users to dictate and send emails using their voice. To assist users, write emails that are more accurate and readable, the system may also be set up to offer recommendations and edits to them while they are writing them. Over time, the system can offer more precise and pertinent recommendations by using machine learning algorithms to learn from user interactions. The proposed study involves an in-depth review of existing methods designed to assist the visually impaired. Additionally, the study introduces a suggested method aimed at enhancing email communication for individuals with visual impairments.

Keywords: Voice E-mail, Voice System, AI, Visually Impaired AI Tool, E-mail Impaired System, Voice Recognition.

DOI: https://doi.org/10.36548/jucct.2024.1.002
Received: 24.12.2023, received in revised form: 21.01.2024, accepted: 08.02.2024, published: 19.02.2024

1. Introduction

In the current digital era, email communication has become a vital form of communication. However, because email interfaces are visual, visually impaired people may find email to be extremely challenging to use. These interfaces can be challenging for people with visual impairments to navigate and utilize since they frequently rely largely on graphics, colors, and visual cues. The goal of the artificial intelligence (AI) in natural language processing (NLP) is to empower computers to comprehend, interpret, and produce human language. Text-based email communications may be turned into voice using NLP techniques, which will increase their accessibility and comprehension for people are visual impairments. The suggested method combines text-to-speech, speech recognition, and natural language processing to provide a full email communication system for the blind persons.

The system will be made using large letters, high contrast color schemes, and user-friendly interfaces to ensure a smooth and simple user experience for visually challenged users. Additionally, when users are writing emails, the system will be built to offer recommendations and corrections. Machine learning algorithms will be used to learn from user interactions and enhance the relevancy and accuracy of recommendations over time.

1.1 Voice E-Mail

Voice email is one of the revolutionary inventions to come around in the fast paced digital society we live in today. Imagine being able to transcend the constraints of conventional textual communication and effortlessly communicate your ideas, feelings, and the messages through voice emails. The power of the voice email improves productivity and allows for more natural and productive communication while also giving communications a more personal touch. Voice email is a dynamic tool that bridges the gap between written and spoken communication in this day of connectivity and time constraints. It is altering the way we interact and communicate in the digital sphere.

1.2 Voice System

Voice systems represent a major development in the field of modern technology, revolutionizing the way we interact with and manage the digital surroundings. Advanced voice recognition and synthesis technologies have allowed voice systems to go beyond their conventional functions and become dynamic interfaces that can adapt to the subtleties of

human speech. These technologies have revolutionized user experiences on a variety of platforms, from interactive voice response systems that expedite customer contacts to virtual assistants who comprehend and carry out commands. Voice systems become an increasingly potent conduit as we go through a time marked by the merging of artificial intelligence and smooth human-machine contact. They provide a hands-free and natural way to communicate with an ever more linked environment.

1.3 E-Mail System

The email system is a vital component of today's communication environment, enabling quick and easy information sharing between people all over the world. Email systems have developed from basic text-based communications to multimedia rich platforms that provide the smooth transfer of documents, photos, and data. They are now an essential part of both personal and professional correspondence.Email systems, which provide instantaneous communication, organizing, and archival capabilities, have grown to be indispensable in both personal and professional contexts. Email's widespread use has revolutionized how we communicate, work together, and do business by creating a digital world where concepts and data move at previously unheard-of speeds and with unparalleled accessibility. The email system is still a crucial tool that evolves with technology to meet the ever-changing demands of a connected and dynamic society. While the email system offers several advantages, making it a widely used and essential communication tool in both personal and professional environments, there are many challenges endured by individuals who are visually impaired. Most email systems rely on visual interfaces and image-based content, posing difficulties for visually impaired users. Additionally, many voice-based emails developed are still subpar, as they are unable to meet the requirements of visually impaired people.

This study provides an in-depth review of various supporting methods developed to assist the visually impaired and suggests a voice email system enabled with speech-to-text conversion and text-to-speech conversion. This system aims to offer hands-free interaction for visually impaired individuals, as well as for older people who may face difficulties in reading and replying to emails.

Objectives:

- To provide an in-depth review about the methods developed for assisting the visually impaired.
- To suggest a voice based email system that helps the visually impaired and elderly.

2. Related Study

In 2023, higher education presents difficulties for those are visually challenged, especially in technical engineering degrees. Their entrance, retention, and graduation from higher education institutions are significantly hampered by the lack of specialized tools and resources that permit effective growth in academic activities., This research describes the creation of IrisMath, a blind-friendly Computer Algebra System (CAS). People with visual impairments may now do mathematical procedures routinely employed in engineering thanks to this device.

Iris Math is a web application that was created with a layered architecture and offers modularity, drawing inspiration from Jupyter Notebooks. It provides a range of output formats, such as audio, JSON, CMathML, and Latex. After a thorough evaluation of its functional, non-functional, and usability criteria, our CAS has proven to be a valuable tool for engineering students who are blind or visually impaired.[1]

Modern technology has demonstrated its presence in every field, and inventive gadgets help people in their daily lives. For VIPs, a clever and intuitive system is created in this work to support their movement and guarantee their protection. The suggested method uses an automated voice to deliver real-time navigation. VIPs can perceive and imagine their surroundings even when they aren't able to see anything in their immediate surroundings.

In addition, an online application is created to guarantee their security. With this program, the user can choose to compromise privacy by revealing his or her location to family members on-demand.VIPs' family members would be able to follow their whereabouts (receive location and photos) when they were at home, thanks to this app. As a result, the gadget ensures VIP protection and lets them see their surroundings.

A device with this level of comprehensiveness is lacking in the body of current literature. Because Mobile Net architecture has a low computational complexity and can operate on low-power end devices, it is used by the application. Six pilot tests have been conducted to evaluate the accuracy of the proposed system, with good results. With an accuracy of 83.3%, a deep Convolution Neural Network (CNN) model is used for item identification and recognition, and the dataset has over 1000 categories.[2]

The goal of processes mining is to identify various viewpoints on business process (BP) from event logs produced by BP management systems. Nonetheless, BP can be carried out fully or in part outside of these systems. Emails are a popular substitute method for working together on BP activities. A number of attempts have been launched recently to expand the use of BP mining to include email logs. Nevertheless, the majority of them have primarily ignored other equally significant information in favor of learning about the BP activity perspective. One of the key pieces of information that can guarantee a greater understanding of people acting in order to carry out BP operations is the actor's perspective. Extra information on the specific role that each participant played in carrying out BP operations may be obtained by mining such a perspective from email records. Such information includes requests, notifications, planning, and activity execution observation in addition to activity execution itself.

This study first formalizes what we could learn about actor viewpoints from emails. Next, it presents a method for obtaining such information based on voice act recognition from email text. We validate our method using an openly available email dataset. As a first step towards ensuring reproducibility in the examined region, our data are made publicly available [4].

For those with poor vision, object identification and spatial cognition are significantly affected by visual loss. Making up for this with other sense modalities, like touch or hearing, is difficult. In order to aid BVI's spatial cognition, this research presents a wearable target locating system.

By donning an RGB-D headmounted camera, the environment's three-dimensional spatial data is calculated, then transformed into navigational cues. By utilizing Spatial Audio Rendering (SAR) technology is possible to convey navigation signals in a 3D sound format

that can be recognized by human sound localization instincts, allowing for the distinction of sound orientation. Three BVI and four sight volunteers participated in trials where three haptic and aural presentation modalities were compared with SAR.

The Fitts law test experimental findings show that, in comparison to standard speech instructional feedback, SAR reduces positioning error by 40% and boosts Information Transfer Rate (ITR) for spatial navigation by a factor of three. Furthermore, compared to other signification techniques like voice, SAR has a smaller learning impact. In trials including desktop manipulation, Stereo Pilot achieved accurate desktop object localization while cutting the target grabbing task completion time in half when compared to voice instruction techniques. In conclusion, Stereo Pilot offers a cutting-edge wearable target localization system that quickly and easily communicates environmental data to BVI people in the real world.[3]

Due to the compromised characteristics of dysarthria speech, including a breathy voice, strained speech, distorted vowels, and consonants, using assistive speech technology might be difficult. For degraded voice recognition, it is crucial to learn compact and discriminative embedding for dysarthria speech utterances.

We provide a Histogram of State (HoS) based method for learning word lattice-based compacts and also discriminative embedding using the Deep Neural Network-Hidden Markov Model (DNN-HMM).

A dysarthria speaking utterance is represented by the best state sequence selected from the word lattice. We next utilize a discriminative model-based classifiers to identify these embeddings. Three datasets are used to assess the effectiveness of the suggested method: a 50-word dataset from the TORGO database, 100-common word datasets from the UA-SPEECH databases, and 15 acoustically related words. For all three datasets, the suggested HoS-based strategy outperforms the conventional Hidden Markov Models and DNNHMM-based techniques by a large margin. The suggested HoS-based embedding' compactness and discriminative power result in the highest accuracy possible for impaired speech recognition.[5]

3. Existing System

In many aspects, neural end-to-end text-to-speech is better than traditional statistical techniques. Still, there is the exposure bias problem, which results from the autoregressive model mismatch between the training and inference processes. When test data is outside of the domain, it frequently results in a decline in performance. In order to tackle this issue, we suggest a multi teacher knowledge distillation network for the Tacotron2 TTS model, together with a unique decoding knowledge transfer approach.

The plan is to pre-train two Tacotron2 TTS instructor models in planned sampling and teacher forcing modes, then feed the pre-trained information to a student model that it can decode naturally. We demonstrate that the MT-KD networks offer a sufficient neural TTS training platform, reducing the discrepancy between training and the run-time inferences in the student model.

Simultaneously, they learn to mimic the actions of their two teachers. The MT-KD system consistently outperforms to the competitive baselines in terms of naturalness, robustness, and expressiveness for both in domain and out of domain test datasets, according to experiments conducted on both Chinese and English data.

The existing voice based email system highly relies on the Speech To Text (STT) for email composing and also converts Text To Speech (TTS) for reading emails [6],an approach based on speech act identification from textual content of emails for identifying knowledge linked to the actors' views [9]. An email system that would enable even untrained, visually impaired users to use communication services, and includes the integration of speech recognition technology and natural language processing to allow the system to effectively interpret and respond to user requests [12-15]. Though the existing system offers a voice based email system they are still subpar, as they are unable to meet the requirements of visually impaired people. So the proposed method is devised with the aim to address the challenges of the existing by integrating speech-to-text technology, natural language processing, and adaptive machine learning algorithms, to improve the text-to the speech conversion comprehension for the incoming messages and also enable effortless speech-to-text capabilities for outgoing communication.

4. Proposed System

Email system that is voice-based and uses text-to-speech and speech recognition to let users read and send emails. A number of libraries, including "smtplib," "speech_recognition," "pyttsx3," "email," and "imaplib," are used by the Python software. Upon initializing, the software asks the user if they would like to send or read the most recent email. In the event that the user decides to view the most recent email, the application asks for the user's email address and password before logging in using "imaplib." After that, the application looks through the inbox for the most recent email and uses the email library to extract the email's text, sender, and topic.

Lastly, the application reads the user's email's body, sender, and subject using text-to-speech technology. Should the user decide to send an email, the application will ask them to provide the recipient's name, the email's title, and its text. The "smtplib" library is then used by the application to send the email. The 'speech_recognition' library is used by the application for speech recognition, while the pyttsx3 library is used for text-to-voice conversion. The "imaplib" library is used to retrieve the most recent email from the user's account and log in, while the email library is used to handle email messages. For emailing, the "smtplib" package is utilized. This proposal aims to show how text-to-speech and speech recognition technologies may be applied to provide an email system that is voice-based and accessible to people with disabilities, including vision impairments.

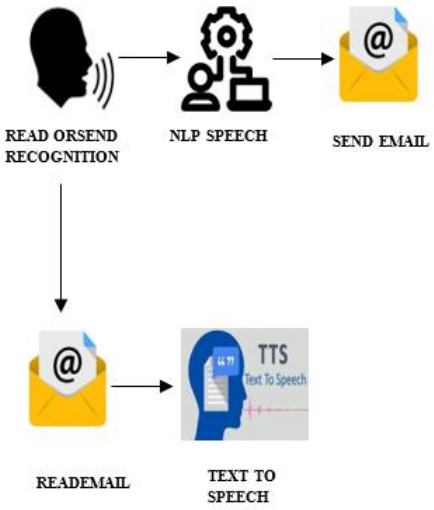

Figure 1. Proposed System Block Diagram

4.1 Audio Input

The audio input module records user voice input and transforms it into a digital format that the system's other components may use. Usually, this module requires using a microphone or another type of audio recording equipment. The audio input module's initial action is to record user spoken input.

Using a microphone or another audio recording device will help achieve this. Usually, the audio is recorded as an analog signal that needs to be converted to a digital format in order to be processed further. Pre-processing the digital audio data to get rid of any undesired artifacts or background noise is the next step. Filtering and noise reduction are two methods of digital signal processing that may be used to accomplish this. The audio stream is usually divided into smaller components, such as phonemes or words, after pre-processing. Since most speech recognition systems work under the premise that speech may be divided into smaller pieces that can be modelled individually, the subsequent module in the speech-to-recognition pipeline, which usually includes acoustic modelling, receives the segmented audio. This module serves as the foundation for additional speech input processing and analysis by using statistical models to match the audio signal to a collection of potential speech sounds or words.

4.2 Preprocessing

In an audio input system, the pre-processing module is in charge of preparing the unprocessed audio signal for additional processing by cleaning it up. Usually, the whole pipeline of the audio input system's begins with this module. Typically, the pre-processing module is made up of a number of smaller modules that cooperate to enhance the audio signal's quality.

There are possible sub modules here.

1. Noise Reduction: The audio signal's undesired background noise is eliminated by this module. Several methods, including spectral subtraction, Wiener filtering, and adaptive filtering, are employed for noise reduction.

2. Filtering: Any high- or low-frequency noise that could be present in the audio signal is eliminated by this module. High-pass, low-pass, and band pass filtering are common filtering methods.

3. Normalization: This module is in charge of modifying the audio signal's volume levels to guarantee that they remain constant between recordings. LUFS normalization, RMS normalization, and peak normalization are examples of normalizing approaches.

4. Resampling: The audio signal's sampling rate is adjusted to a standardized rate by this module. When working with audio files that have varying sample rates, this is frequently required.

In order to guarantee that the audio data is of the highest caliber and appropriate for additional processing, the pre-processing module is essential. The accuracy and efficiency of downstream modules, such language and audio modelling, might suffer from improper pre-processing.

4.3 Acoustic Modeling

An essential part of the speech recognition systems is the acoustic modelling module, which associates the acoustic characteristics of speech signals with phonetic or linguistic units. This module's several sub-modules cooperate to translate audio data that has already been processed into text. Feature extraction is the initial sub-module of the acoustic modelling module, and it gathers pertinent characteristics from the audio data that has already been pre-processed.

Mel Frequency Cepstral Coefficients (MFCCs) are often utilized features that capture the spectral properties of speech signals and the Linear Predictive Coding (LPC) coefficients that describe the resonances of the vocal tracts. Acoustic modelling, the second sub-module, links the characteristics that have been retrieved with the appropriate phonetic or language units. Typically, machine learning methods such as Deep Neural Networks (DNNs) or Hidden Markov Model (HMMS) are used for acoustic modelling. Speech signals' temporal relationships are captured by statistical models called HMMs, whereas DNNs learn by utilizing many layers of artificial neurons. N-grams, or word sequences of n words that occur together in a corpus of text, are the standard basis for language models. By choosing the most

likely word order depending on the situation, the language model is utilized to clarify the output of the acoustic model.

4.4 Language Modelling

An essential stage in turning voice into text is language modelling. It assists in forecasting the most likely word order that a user may have said. The acoustic modelling module produces text output that may contain mistakes or inaccuracies. The language modelling module examines this text output and creates a more accurate representation of the spoken words. Language modelling is done using a variety of methods, such as neural modelling and statistical language modelling.

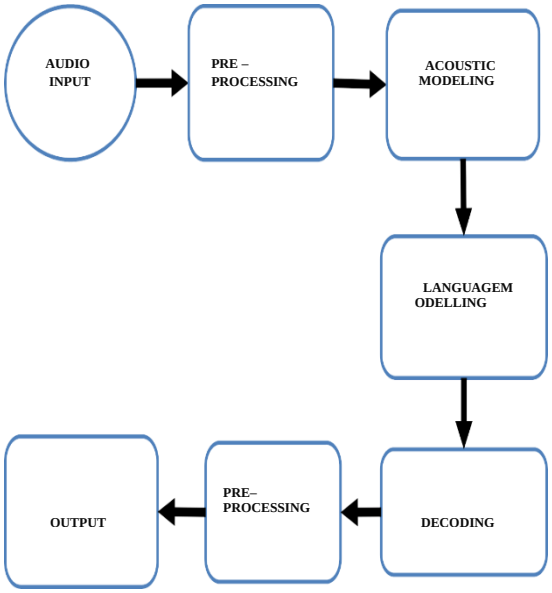

Figure 2. Work Flow of Proposed System

An analysis of the frequency of word sequences in a given text corpus is part of statistical language modelling. This technique employs probability theory to provide a probability score to each conceivable sequence of words that the user may have uttered. The word sequence that has the highest probability score is regarded as the most likely one. A kind of statistical language modelling known as "n-gram language modelling" assigns probability ratings to sets of n words rather than single words.

4.5 Post Processing

The post-processing module is also responsible for adding punctuation to the text output. This can assist to make the writing more legible and simpler to grasp. Punctuation marks like commas, periods, and question marks can be introduced based on the contexts of the text. One of the key functions of the post processing module is formatting the text output. The text output generated by the voice recognition technology may not always be formatted appropriately. The post processing modules in the final stage of the speech recognition pipeline is responsible for improving the output provided by the language modelling modules. There may be instances of inaccuracies in the output produced by the language modelling module, such as misspelled words, improper word order, or improper punctuation. The purpose of the post processing module is to fix these mistakes and improve.

5. Conclusion

In summary, the email communication system that has been suggested here offers a thorough and creative way to overcome the obstacles that visually impaired people encounter when attempting to use electronic communication. Through the smooth integration of speech-to-text technology, natural language processing, and adaptive machine learning algorithms, the system not only improves text-to speech conversion comprehension for incoming messages but also enables effortless speech-to-text capabilities for outgoing communication. Accuracy and customisation are enhanced by the user-configurable parameters and the real-time recommendations and adjustments. Its user friendly interface and security and privacy concerns make it a desirable option**.**

6. Future Work

The future work of this email communication system's development may concentrate on the making of voice based email system that includes the design and the deployment of the model for practical use. Additionally, the system is planned to be designed with more compatibility, new gadgets, and technologies in order to have a seamless integration with speech recognition software and other growing email platforms. Expanding the system's language support, adding user feedback methods, and improving the machine learning algorithms would all help make it more adaptable and efficient for a wider range of user demographics. Furthermore, investigating possible partnerships with accessibility

organizations and makers of assistive technology may yield insightful information for future improvements. To maintain the integrity of sensitive user data, more investigation into cutting-edge security protocols and privacy-preserving technology is necessary.

References

[1] Ana M, Zambrano, Danilo Pilacuan, Mateosalvador, Felipegrijalva, "Irismath: A Web-Based Computer Algebra System That is Blind And Friently", IEEE, vol. 11, July 2023.

[2] Ashiq, Fahad, Muhammad Asif, Maaz Bin Ahmad, Sadia Zafar, Khalid Masood, Toqeer Mahmood, Muhammad Tariq Mahmood, and Ik Hyun Lee. "CNN-based object recognition and tracking system to assist visually impaired people." IEEE Access 10 (2022): 14819-14834.

[3] Hu, Xuhui, Aiguo Song, Zhikai Wei, and Hong Zeng. "StereoPilot: A wearable target location system for blind and visually impaired using spatial audio rendering." IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022): 1621-1630.

[4] Kumar, Sunny, R. Yogitha, and R. Aishwarya. "Voice Email Based on SMTP For Physically Handicapped." In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1323-1326. IEEE, 2021.

[5] Chandrakala, S., S. Malini, and S. Vishnika Veni. "Histogram of states based assistive system for speech impairment due to neurological disorders." IEEE Transactions on Neural Systems and Rehabilitation Engineering 29 (2021): 2425-2434.

[6] Noel, Sherly. "Human computer interaction (HCI) based smart voice email (Vmail) application-assistant for visually impaired users (VIU)." In 2020 third international conference on smart systems and inventive technology (ICSSIT), pp. 895-900. IEEE, 2020.

[7] Sonbhadra, Sanjay Kumar, Sonali Agarwal, Mohammad Syafrullah, and Krisna Adiyarta. "Email classification via intention-based segmentation." In 2020 7th

International Conference on Electrical Engineering, Computer Sciences and Informatics (EECSI), pp. 38-44. IEEE, 2020.

[8] Saha, Sagor, Farhan Hossain Shakal, Ahmed Mortuza Saleque, and Jerin Jahan Trisha. "Vision Maker: An Audio Visual And Navigation Aid For Visually Impaired Person." In 2020 International Conference on Intelligent Engineering and Management (ICIEM), pp. 266-271. IEEE, 2020.

[9] Elleuch, Marwa, Oumaima Alaoui Ismaili, Nassim Laga, Nour Assy, and Walid Gaaloul. "Discovery of activities' actor perspective from emails based on speech acts detection." In 2020 2nd International Conference on Process Mining (ICPM), pp. 73-80. IEEE, 2020.

[10] Zhang, Xiangliang, Chaoxi Lu, Jibin Yin, Hailang Xie, and Tao Liu. "The Study of Two Novel Speech-Based Selection Techniques in Voice-User Interfaces." IEEE Access 8 (2020): 217024-217032.

[11] Sripriya, N., S. Poornima, S. Mohanavalli, R. Pooja Bhaiya, and V. Nikita. "Speech-based virtual travel assistant for visually impaired." In 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), pp. 1-7. IEEE, 2020.

[12] Kulkarni, Omkar, Akshay Alhat, Namdeo Tejankar, and Madhuri Patil. "Voice based e-mail system for blind people." Open access international journal of science and engineering 4, no. 01 (2019).

[13] Mamatha, A., Veerabhadra Jade, J. Saravana, A. Purshotham, and A. V. Suhas. "Voice Based E-mail System for Visually Impaired." International Journal of Research in Engineering, Science and Management 3, no. 8 (2020): 51-54.

[14] K. Venkadesh, P. Santhosh Kumar, A. Sivanesh Kumar, "Voice Based E-Mail System For Visionless People And Object Detection Using Optimization Technique", IJSTR, vol. 9, pp 457-461,February 2020.

[15] PanikalaHemanth Kumar, "Voice based email for Visually Challenged people", IJERT, vol. 17,pp 108-112 July 2020.