AI-Based Voice E-Mail System For Visually Impaired

Ms S.Revathi[1], M.Nathiya[2], V.Preethi Vihashini[3], G.Veeramani[4]
[1]Assistant Professor, Dept of Computer Science and Engineering
[2, 3, 4]Dept of Computer Science and Engineering
[1, 2, 3, 4] Perundurai,Erode,TamilNadu,India.

Abstract- *The email system to assist visually impaired individuals. The system converts text emails into synthesized speech, enabling users to listen to messages. It also utilizes speech recognition for dictating replies. Additionally, it can analyze emails as they are composed and offer real-time suggestions to improve clarity and readability. The system's recommendations become more accurate over time through machine learning algorithms that learn from user interactions. By leveraging natural language processing and adaptive technology, this email system aims to make electronic communication more accessible for people with visual disabilities.*

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

I. INTRODUCTION

In today's digital world, email has become a vital communication tool. However, because email interfaces rely heavily on visual elements like graphics, colors, and visual cues, they can be difficult for people with visual impairments to use. Natural language processing (NLP), a type of artificial intelligence, aims to help computers understand and generate human language. By leveraging NLP techniques like text-to-speech and speech recognition, we can convert text-based emails into voice formats. This improves email accessibility and comprehension for the visually impaired.

A proposed system would integrate text-to-speech, speech recognition, and NLP to offer a complete email communication solution for blind users. 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. To ensure a smooth, simple user experience for visually challenged users, the system will incorporate large letters, high contrast colors and intuitive interfaces. When users write emails, it will offer recommendations and corrections, relying on machine learning to analyze interactions and continuously improve the relevancy and accuracy of suggestions.

1.1 VOICE E-MAIL

Voice mail revolutionized communication by transcending the limitations of conventional textual messages. It enables people to convey ideas, emotions, and information easily through the natural power of their voice. Voice mail improves productivity and facilitates more personal, dynamic interactions, adding a human touch to digital communication.

1.2 VOICE SYSTEM

Voice systems have revolutionized the way we interact with technology by to the enabling advanced voice recognition and synthesis. These technologies allow voice systems to understand nuances in human speech, rather than just recognizing set commands. As a result, voice systems have become dynamic, conversational interfaces on platforms ranging from customer service IVRs to virtual assistants. Looking ahead, voice systems will likely become an even more powerful conduit between humans and AI as we enter an era defined by seamless integration of intelligent machines into our daily lives. With hands-free operation and natural language capabilities, voice systems provide a convenient way to communicate and control our increasingly connected world.

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.

II. LITERTURE REVIEW

2.1.IRISMATH:AWEB-BASED COMPUTER ALGEBRA SYSTEM THAT IS BLIND AND FRIENDLY

Mateo N. Salvador, DaniloPilacuan, and Ana M. Zambrano. 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. In light of this, this paper describes the creation of IrisMath, is 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 the web application that was created with a layered architecture offer 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 shown to be a valuable tool for engineering students to are blind or visually impaired.[1]

2.2. A VISUALLY IMPAIRED PEOPLE'S ASSISTANCE TRACKING AND OBJECT RECOGNITION SYSTEM BASED ON CNN

Modern technology has demonstrated its presence in every field, and inventive gadgets help people in the 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 while 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 with 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 was 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]

2.3. ACTOR PERSPECTIVE ACTIVITIES FOUND IN EMAILS USING SPEECH ACT DETECTION

WalidGaaloul, NassimLaga, NourAssy, OumaimaAlaoui Ismaili, Marwa Elleuch 2022 The goal of process mining is to identify various viewpoints on business process (BP) from events 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. Nonetheless, 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 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 is based on voice act recognition from email text. We validate our method using an openly available email dataset. As a first step towards assuring reproducibility in the examined region, our data are made publicly available.[3]

2.4 STEREOPILOT: A WEARABLE TARGET LOCATION SYSTEM USING SPATIALLY-AUGMENTED AUDIO FOR BLIND AND VISUALLY IMPAIRED

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 the four sight volunteers participated in trials where three haptic and aural presentation modalities compared with SAR. The Fitts law test experimental findings shown 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 factors 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 half when

Compared the 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.[4]

2.5. STATES-BASED ASSISTIVE SYSTEM HISTOGRAM FOR NEUROLOGICAL DISORDERS RELATING TO SPEECH IMPAIRMENT

S. Malini, S. Chandrakala, and S. Vishnika Veni 2021 Due to the compromised characteristics of dysarthria speech, including 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 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 embedding. 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 is 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]

III. EXISTING SYSTEM

In many aspects, neural end-to-end text-to-speech is better than traditional statistical techniques. Still, there is the exposures bias problem, which results from the autoregressive models mismatch between the training and also inference process. 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 knowledges to distillation network for the Tacotron2 TTS model together with a unique decoding knowledge transfer approach.

The plan is to the 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 offers a sufficient neural TTS training platform, reducing the discrepancy between training and the run-time inferences while the student model.

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

3.1 EXISING DISADVANTAGES

The distilled model may not be able to capture all the nuances of the original teacher models, leading to a loss of information and potentially suboptimal performance.

i. Knowledge distillation methods may not generalize well to new domains or tasks, as the distilled knowledge is often specific to the new settings.
ii. Multi-teacher knowledge distillation networks may require significant computational resources and time to train and deploy, which can be a significant drawback in practical applications.

iii. Distillation-based approaches are often difficult to interpret and understand, as the resulting models may not be easily explainable or transparent in their decision-making processes.

IV. 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 want to send or read the most recently email. In the event that the user decides to view the most recented 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 the 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 speak 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 voicebased and accessible to people with disabilities including vision impairments.

4.1 ADVANTAGES

i. It provides an alternative means of accessing email for individuals with visual impairments or other disabilities, allowing them to read and send emails using speech recognition and text-to-speech technology.
ii. This improves accessibility and inclusivity in the digital world. The system can also benefit individuals who have busy schedules and need to multitask. For example, someone who is driving can use the system to check and respond to emails hands-free, increasing their productivity and efficiency.
iii. The use of speech recognition and text-to-speech technology in the system allows for a more natural and intuitive interaction with the email system..
iv. The system can be easily adapted and customized to meet the specific needs of different users, such as changing the language of the text-to-speech output or adjusting the sensitivity of the speech recognition.

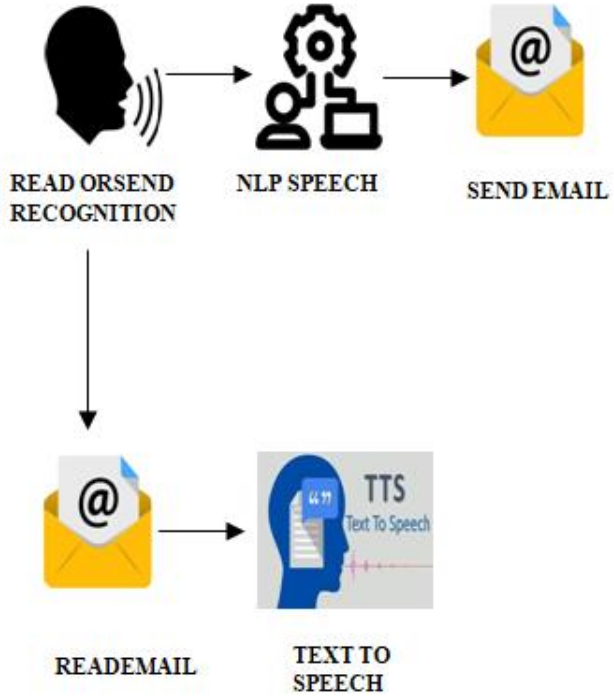


Figure 1. Proposed System Block Diagram

V. MODULE DESCRIPTION

- Audio input
- Preprocessing
- Acoustic modeling
- Language modelling
- Postprocessing

5.1 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 ordered to be processed further. Pre-processing the digital audio data to get rid of any undesired artifacts or background noise is the next step.

The audio stream is usually divided into smaller components, such 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.

5.2 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 audio input system's whole pipeline 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.

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 is 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.

5.3 ACOUSTIC MODELING

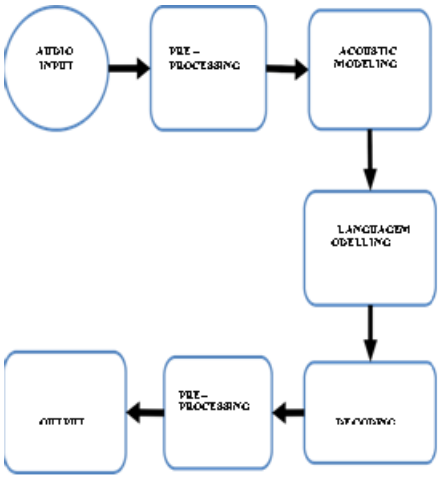
An essential part of the speech of recognition systems is the acoustic modelling module, which associates acoustic characteristics of speech signals with phonetic or linguistic units. This module's several sub-modules cooperate to translate audio data that has already been processed into text. Feature 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 is that capture the spectral properties of the 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.

5.4 LANGUAGE MODELLING

An essential stage in turning voice to 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 have mistaken 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.



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

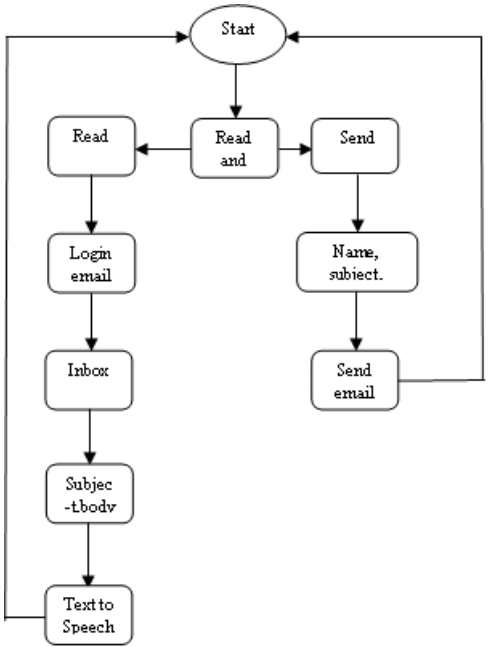
have uttered. The word sequence that has the highest probability score is regarded as the most likely one. A kind of statistical language modelling known as "n-gram language modelling" assigns probability ratings to sets of n words rather than single words.

5.5 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 is the final stage of the speech recognition pipeline and 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.

VI. ARCHITECTURE DIAGRAM



VII. 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 peoples 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 the speech conversion comprehension for the 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**.**

VIII. FUTURE WORK

The future for this email communication system's development may concentrate on making it more compatible with new gadgets and technologies and guaranteeing that it integrates seamlessly 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] Fahad Ashiq, Muhammad Asif, Maaz Bin Ahamad sadia Zafar, Khalid Masood, "CNN-Based Object Recognition And Tracking System to Assist Visually Impaired People", IEEE, vol. 10, February 2022.
[3] Xuhui Hu, ZhikaiWei, " Steropilot: A Wearable Target Location System For Blind and Visually ImpairedSpatial Audio Rendering", IEEE, vol. 30, March 2022.
[4] Sunny Kumar, Yogitha, Aishwarya, "Voice Email based on SMTP for Physically Handicapped", IEEE, vol. 20, June 2022.

[5] Chandrakala S, Malini S, VishnikaVeni S, "Histogram of states based Assistive System for Speech Impairment Due to Neurological Disorder", IEEE, vol. 29, May 2021.
[6] Sherly Noel, "Humen Computer Intraction based smart voice Email(Vmail) Application-Assistant for visually Impaired Users ", IEEE, vol. 12, April 2020.
[7] Sanjay Kumar, Sonabhadar, Somali. Mohammad Syaruallah, Krishna Adiyarta, "Email classification via Intention- Based Segmentation", IEEE, vol. 16, October 2020.
[8] Sagor Saha, FarhanHossain Shakal, AhmedMortuza Saleque, "Vision Maker: An Audio Visual Navigation Aid for Visually Impaired Person", IEEE, vol. 13, January 2020.
[9] Xiangliang Zhang ChaoxiLu, Jib in Yin, Tao Liu, "Discovery of Activities Actor Perspective from Emails based on Speech Acts Detection", IEEE, vol. 15, June 2020.
[10] ChaoxiLu, Jib in Yin, Tao Liu, "The study of Two Novel Speech- based selection Techniques in Voice user Interfaces", IEEE, vol. 8, December 2020.
[11] Sripriya N, Poornima S, Mohanavalli S, PoojaBhaiya R, Nikita V, "Speech based virtual travel assistant for visually impaired", IEEE, vol. 5, November 2020.
[12] Omkar Kulkarni, AkshayAlhat, NamdeoTejankar, MadhuriPatil, “Voice Based email system for Blind People”, IJSTR, vol. 4, January 2020.
[13] A. Mamatha, Veerabhadra jade, J. Saravana, A. Purushotham, A. V. Suhas, “Voice based email system for Visually Impaired”, IJRESM, vol. 3, August 2020.
[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, February 2020.
[15] PanikalaHemanth Kumar, “Voice based email for Visually Challenged people”, IJERT, vol. 17, July 2020.