JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enriching Email System for Blind

1st Akshada Sunil Bangar, 2nd Swapnali Sahebrao Pokharkar, 3rd Sayali Laxman Padir, 4th Prof.M.G.Sinalkar

1st Student of JCOE, 2nd Student of JCOE, 3rd Student of JCOE, 4th Prof. of Computer Engineering JCOE

Computer Engineering Department,
Jaihind College of Engineering, Kuran, Pune, India

Abstract- The Web has developed into the central nervous system of the modern world. But without the web, there is no way to get anything finished in modern life. Email, often known as electronic mail, is quickly transforming into one of the most essential components of modern life. The use of email is mostly a private activity. Visually impaired individuals are unable to utilize the majority of the standard postal services that we rely on in our everyday lives. Those who struggle with their vision are unable to use the majority of the mail services that we employ in our regular lives. Color correction and voice assistants are two examples of the kinds of assistive technology that may be made available to persons who struggle with their vision so that they can make better use of the systems available to them. Nevertheless, since these technologies were unable to deliver the appropriate response as a typical system would, they are not nearly as helpful to the persons in question. Because of this, those individuals might benefit from the use of modern technologies since they are unable to provide the appropriate response in accordance with the functionality. Therefore to improve accessibility and engagement of blind individuals in monitoring and managing their mail this research defines a voice-activated email system. The proposed approach utilizes registration and login, composing email by voice, encrypting the email, getting voice notification for the email, listen to email and attachment through OCR. The approach has been tested for its performance and has attained superior results.

Keywords: Accessible Infrastructure, Optical Character Recognition, Voice to Text System

I. Introduction

The World wide web is often regarded as the most significant technological advancement in recent history. The information revolution we are currently living in is one that not only often experiences innovation but also adapts to that development in the appropriate manner. Electronic equipment have low upfront costs and are simple for anybody to learn how to operate. The ability to utilize these solutions is restricted for those who do not possess ideal physical characteristics. Individuals who have trouble seeing well or those who are blind or visually impaired might benefit from specialized programs such as voice-based email providers. The most effective method of interaction so far has been via the use of emails. Email is unquestionably the most popular mode of communication, and it is utilized in many different fields, including teaching and learning. There are numerous other channels of communication accessible. There are some people on the earth who are not familiar with email service providers. A person has to have good eyesight in order to understand what is being presented on the screen. Those who have visual impairments should not attempt to navigate the internet since they will not benefit from doing so if they are unable to view the display. The most effective method of interaction so far has been via the use of emails. Email is unquestionably the most popular mode of communication, and it is utilized in many different fields, including teaching and learning. There are numerous other channels of communication accessible. There are some people on the earth who are not familiar with email service providers. A person has to have good eyesight in order to understand what is being presented on the screen. Those who have visual impairments should not attempt to navigate the internet since they will not benefit from doing so if they are unable to view the display. A person who has trouble seeing may retrieve their email in a timely and effective manner by making use of the voicemail service. As a result, the individual with visual impairment does not require any assistance from third-party programs, which are risky in any case. This implementation is a desktop program written in the Python programming language that is designed for persons who are visually impaired. It uses speech-to-text as well as text-to-speech components to ensure that each user will have the ability to operate their personal profiles using just voice commands. The user would constantly be prompted by this system to carry out the required tasks according to the instructions that relate to those functions. The primary reason why we are working on this initiative is due to the fact that the usage of a keyboard will be significantly reduced, mouse actions will be performed only seldom, and the majority of the user interface will be controlled only by voice input.

II.Related Work

Peng Cheng [1] is an attempt to provide answers to the problems raised above. The authors provide a vocabulary to organize ongoing research activiti s as well as provide a summary of the current state of the art with regard to the protection and confidentiality of personalized voice recognition. The nomenclature and general description of the issue of personalized voice-activated confidentiality and safety are both expanded and improved as a result of this study. The problems with confidentiality and safety that might be caused by the usage of the auditory channel are the primary subject of this study. Since voice inflection assistance is a decentralized computing platform, it is susceptible to the same kind of risks as other computer systems. It is possible to compromise a voice inflections assistant, and once compromised, the device might be utilized, for instance, as a component in a botnet. The software application of personalized voice assistants may be compromised, and

sensitive information about individual customers can be taken. Such time-honored security concerns unavoidably need to be handled inside personalized voice companions as well as the accompanying service architectures.

Blind amputees may benefit from an instinctive contextual awareness help approach that was suggested by Xuhui Hu [2] and is predicated on spatial audio representation. The technique of spatial audio representation virtualizes the expression of realworld information about the surrounding environment. The consumer is able to gain genuine geographical information interpretation to the interacting objects by making full utilization of the human trait on sound navigation. As compared to voicemail transcription, spatial acoustic reconstruction greatly enhances the information transmission rate, cuts the project duration by half compared to voicemail transcription, and restores the organic gripping route. All of these benefits come at the expense of voicemail notifications.

Ryo Iijima [3] suggested a brand innovative voice activated assault that was completely imperceptible and was given the term Auditory Hotspot Attack. Significant user research and tests that could be replicated were used to assess the practicability of the idea. The authors proved that assaults may operate over very large distances when directed noises are generated from adjustable amplifiers and aren't noticed by a close individual. Moreover, these assaults are forgiving despite background disturbances. While the Acoustic Hotspot Attack is just a proofof-concept, viable solutions to make the assaults ineffectual have been offered. The suggested attack involves ultrasonic self-demodulation, being a quantitative phenomenon. From the vantage point of auditory stimuli, the researchers feel that somehow this approach offers fresh light upon data protection research focusing on portable and Edge computing. Section 2 of this research article presents an analysis of the relevant literature; Section 3 explains the research approach; Section 4 discusses the experimental assessments; and Section 5 closes with suggestions for further study in the future.

Rumana Islam [4] introduces field research on speech impairment identification techniques that have previously been published. It has been demonstrated by the research that speech disability identification is an extremely tough task to do due to the fact that the audio signal is highly complex to evaluate. The vocal signals are quite different for each of the many types of disabilities. The literature has included various algorithms published in literature. Nevertheless, neither of these techniques is suited for identifying any particular kind of vocal impairment. As a result, it is of the utmost importance to concentrate on a specific impairment while developing the algorithm. The most significant disadvantage is that the investigators need to concentrate their efforts not only on the spoken as well as the voiceless aspects of the data. This is because there is an indication of pathological identification in the uninflected portion of the audio recordings as well.

June-Woo Kim [5] concentrated on enhancing the effectiveness of conventional ASR systems, which are not very good at identifying voices that are outside the norm, such as those of the geriatric. In order to accomplish this goal, the authors presented verbal combined data by way of an unsupervised phonetics grouping technique and suggested an age-to-age speech transcription by making use of the lexicon data with the goal of improving the language processing effectiveness of the aged. In this regard, the technique that has been provided is an alternate adaption methodology that may be implemented prior to any Speech recognition system that is either commercially available or open source. Researchers 3 established the efficacy of the suggested A2AVT technique and successfully integrated linguistically related information by demonstrating an increase in the reliability of natural language processing obtained from a commercialized Speech recognition system.

III.PROBLEM STATEMENT

To enhance the process of email handling and image attachment analysis for blind users through the use of optical character recognition and text to speech implementations along with instance character encryption for blind persons.

IV. OBJECTIVE

- The system will not let the user make use of the keyboard instead will work on speech In today's age much of the communication takes place through the internet.
- The system will recognize the speech and convert that into text hence user-friendly for them. It will be connected to the internet via LAN or wi-fi hotspot or stand-alone internet connection so that the respective email can be sent to the receiver.
- This application recognizes the user's voice and performs comparisons with pre-sample voice stored in the database and executes the voice command.
- The proposed work aims to develop a mechanism that converts Speech To Text (STT) for email composing and also converts Text To Speech (TTS) for reading emails.

V. SYSTEM ARCHITECTURE

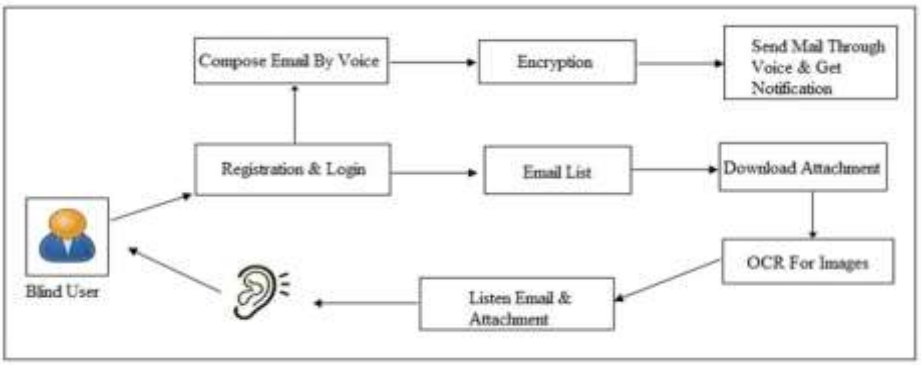

FIG 1.1 SYSTEM ARCHITECTURE

Step 1: Face detection: All users who wish to use this software have to sign up with their faces in this initial stage of the slated system. The first step in the process of capturing faces is importing the cV2 object by installing the opencv package in our Python IDE Spyder. The openCV library helps handle the localhost laptop's camera and external devices. Frames start being captured as soon as the camera is discovered. Using HAAR cascade classifier patterns stored in the XML file, each collected frame is examined to identify the faces using this XML file. The face positions are automatically taken from the cascade classifier output list once it has returned the number of faces. The facial image portion of the captured frame is cropped based on specific facial criteria. Following cropping, the openCV object is used to resize the facial portion to a dimension of 150 x 150. When repeatedly pressing the "SPACE BAR" key on the keyboard, the cropped image with the RGB color channel is likely to be stored in a folder under the respective user name with an increasing number of the count of integers starting with 1. One of the best facial photographs for each of these individuals is chosen and downloaded to be stored in the facial image dataset along with the user's name in jpg format. This procedure also involves tagging the user's name, email address, password, and security key in the database.
Step 2: Composing Emails: The email for blind systems is now operational, and the open CV is going to open the camera and start collecting frames. Every single frame is being HAARcascaded to identify the face. After this it is resized once more on a temporary path. From this point, the images are read using an open CV object before being matched one by one with the database-stored facial images. In order to recognize the face features and, as a result, return the name of the user for whom the face has been identified, the face recognition library has been installed as well as imported. Once a blind user's face has been recognized, his name is used to retrieve his email ID and the secret key, and a window appears requesting the user to choose between two options: compose or read emails with voice guidance from another programmer. The design model opens the compose email window when the user chooses the compose email option in response to the voice command. The user is prompted to say their email ID and stop at the end of an interactive continuous thread as soon as the compose email window opens. The other thread will then question the blind user about whether what he said was correct or not after receiving the blind user's saved email ID. The other interactive thread handles this procedure of rating the speech train. If the blind user's response is correct, the current thread will be stopped and the next one will start. However, if the blind user answers "not correct," the process of asking for the email ID will continue indefinitely until the blind user responds as "correct". After the email ID has been successfully captured using interactive threads, additional threads that function similarly attempt to obtain the email's subject. The identical procedure will be repeated in order to collect the email body by two more threads. Once all the information needed to create an email, such as the email ID, subject and body has been collected, the proposed model will send all these details, along with the sender's username and security key to the function for sending emails using the Gmail host. This entire process allows the blind person to send emails entirely through interactive voice means rather than using the GUI of the software.
Step 3: Reading Inbox – As the user says read email, the reading email window is being opened and asks the user to say read email followed by stop. As this command is identified, the Gmail inbox is accessed by the username and the secret key of the respective account. For the purpose of accessing the Gmail inbox emails, a meme object and a multi-part object are being used by importing the email library of Python. The obtained email data, like from email ID, subject and email body are retrieved in a double-dimensional list to play them in voice for the user to hear. After this, a thread is being asked to say the user the respective email ID number to hear its body content. On saying the email ID number, the respective email ID contents are being played in voice to repeat the process again. Some of the libraries are being used in our methodology by the Python programming language, like "speech_recognition" to recognise the voice, gtts (Google text to speech), "pyglet" to handle the multi-media, and finally the 'email' library to access the respective email account.

VI. ADVANTAGES AND DISADVANTAGES

A. Advantages
Ease of use to blind users for email.
Help develop independent nature in blind users.
Improves the quality of life for blind individuals.

B. Disadvantages
The approach will be demonstrated in a minimal number of emails. Works on specific image formats for OCR.

VII.RESULT

The designed system is developed on a Windows PC that runs Python and uses the Spyder as an IDE. The system's legitimacy is determined using a variety of assessments, some 5 of which are listed below. Every time our system performs complete voice interaction, the user is supposed to verify the system's legitimacy. One of the best human evaluation methods for the system's perfection is the MRR (Mean Reciprocal Ratio) is used for this purpose.
A score between 1 and 6 is given to the output performance of the email system for blind users in MRR, depending on how well it works. For instance, if the system performed effectively, a user would grant a rank of 1, and that rank would also be indicated. If the rank is 2, that rank is also indicated. Finally, the mean rank for the collection of experiments will be calculated using the MRR equation, which is illustrated by equations 1 and 2.

$S = \sum 1 /(Ranki) \; n \; i=1$ _____(1)
MRR=S/N _____(2)

Where n – Number of sample trails The result for MRR is listed in Table 1 below. MRR is calculated using different trails, each of which consists of five test.

Trail No	No of Experiments	MRR
1	5	0.97
2	5	0.9
3	5	0.95
4	5	0.98
5	5	1

TABLE 1 : Recorded MRR

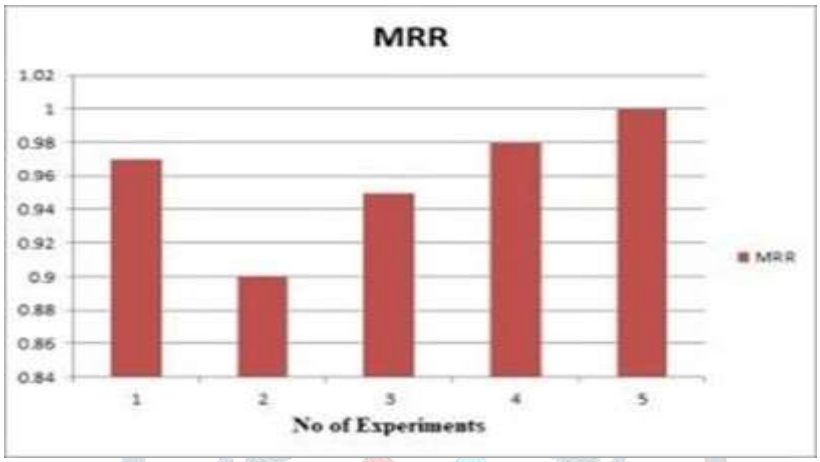

Figure 2: MRR Comparison for different types of images

According to the above plot in Figure 2, our proposed email system for blind people produces an average MRR of 0.96, which we can justify as one of the best email systems for blind people.

VIII.CONCLUSION

The presented approach to enable an effective and useful mechanism for the purpose of achieving a Voice based Email system for blind people has been elaborated in this research article. The defined approach initiates with the blind user interacting with the system using the voice based navigational system. The proposed approach interacts with the blind person that can utilize this system to perform the various tasks for sending, receiving or managing emails. The blind user first utilizes the system to perform the registration into the mail services system. This registration is performed through the user providing the various details that are used to create the account for the user. The user can now utilize these login credentials to login into the system. Once the user logs into the system the email list is read out to the user. The user here can perform a wide variety of actions, the user can then compose the new email to a particular sender using only voice. The email will be encrypted to secure the user from any peeking individuals and other bad actors and after the confirmation from the user the mail can be sent through the voice and a notification for the same will be received. The user can also use the system to red out the mails and also the attachments. The attachments could also have some text in an image which will be extracted through an effective Optical Character Recognition that will be listened along with the contents of the email by the blind user. The approach has been tested rigorously for its performance which has been regarded as highly helpful and assistive.

IX.FUTURE SCOPE

For the future enhancement this model can be enhanced to work as readymade API or library, that can be utilized by the other developers across the globe to develop this kind of System.

X.REFERENCES

[1] P. Cheng and U. Roedig, "Personal Voice Assistant Security and Privacy—A Survey," in Proceedings of the IEEE, vol. 110, no. 4, pp. 476-507, April 2022, doi: 10.1109/JPROC.2022.3153167.
[2] X. Hu, A. Song, H. Zeng and D. Chen, "Intuitive Environmental Perception Assistance for Blind Amputees Using Spatial Audio Rendering," in IEEE Transactions on Medical Robotics and Bionics, vol. 4, no. 1, pp. 274-284, Feb. 2022, doi: 10.1109/TMRB.2022.3146743.
[3] R. Iijima et al., "Audio Hotspot Attack: An Attack on Voice Assistance Systems Using Directional Sound Beams and its Feasibility," in IEEE Transactions on Emerging Topics in Computing, vol. 9, no. 4, pp. 2004-2018, 1 Oct.-Dec. 2021, doi: 10.1109/TETC.2019.2953041.
[4] R. Islam, M. Tarique and E. Abdel -Raheem, "A Survey on Signal Processing Based Pathological Voice Detection Techniques," in IEEE Access, vol. 8, pp. 66749 -66776, 2020, doi: 10.1109/ACCESS.2020.2985280.

[5] J. -W. Kim, H. Yoon and H. -Y. Jung, "Linguistic -Coupled Age -to -Age Voice Translation to Improve Speech Recognition Performance in Real Environments," in IEEE Access, vol. 9, pp. 136476 -136486, 2021, doi: 10.1109/ACCESS.2021.3115608.
[6] T. Tuncer, S. Dogan, F. Özyurt, S. B. Belhaouari and H. Bensmail, "Novel Multi Center and Threshold Ternary Pattern Based Method for Disease Detection Method Using Voice," in IEEE Access, vol. 8, pp. 84532 -84540, 2020, doi: 10.1109/ACCESS.2020.2992641. [7] M. Zhang, Y. Zhou, L. Zhao and H. Li, "Transfer Learning From Speech Synthesis to Voice Conversion With Non -Parallel Training Data," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1290 -1302, 2021, doi: 10.1109/TASLP.2021.3066047.
[8] Y. Hono, K. Hashimoto, K. Oura, Y. Nankaku and K. Tokuda, "Sinsy: A Deep Neural Network -Based Singing Voice Synthesis System," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 280 3 -2815, 2021, doi: 10.1109/TASLP.2021.3104165.
[9] R. Çolak and R. Akdeniz, "A Novel Voice Activity Detection for Multi -Channel Noise Reduction," in IEEE Access, vol. 9, pp. 91017 91026, 2021, doi: 10.1109/ACCESS.2021.3086364.
[10] Y. Hu, Q. Song, H. Peng and Q. Xiao, "Improvement of Noise Stability of Michelson Optical Fiber Voice Monitoring System," in IEEE Access, vol. 9, pp. 60482 -60488, 2021, doi: 10.1109/ACCESS.2021.3073953.
[11] L. Verde, G. De Pietro, A. Ghoneim, M. Alrashoud, K. N. Al -Mutib and G. Sannino, "Exploring the Use of Artificial Intelligence Techniques to Detect the Presence of Coronavirus Covid -19 Through Speech and Voice Analysis," in IEEE Access, vol. 9, pp. 65750 65757, 2021, doi: 10.1109/ACCESS.2021.3075571.
[12] W. Li, Y. Chen, H. Hu and C. Tang, "Using Granule to Search Privacy Preserving Voice in Home IoT Systems," in IEEE Access, vol. 8, pp. 31957 -31969, 2020, doi: 10.1109/ACCESS.2020.2972975.