Human computer interaction(HCI) based Smart Voice Email (Vmail) Application - Assistant for Visually Impaired Users (VIU)

Sherly Noel

Assistant Professor, Department of Computer Science & Engineering,
CMR Institute of Technology,
Bengaluru, India
sherly.n@cmrit.ac.in

ABSTRACT— **Communication development is creating a revolution in the current digital era. A formal or casual communication is now sent through email. The growth in digital technology has given immense opportunity to visually impaired person. This application is developed to ease the process of email writing not only for visually impaired person but for everyone. Now human voices can also be given as input instead of typing on the keyboard. So, additional skill required for typewriting will not be necessary anymore. This application recognizes user voice and performs comparisons with pre-sample voice stored in the database and executes the voice command. Common day-to-day spoken words are used as command language. It focuses on reducing the load incurred in human memory. The proposed work aims to develop a mechanism which converts Speech To Text (STT) for email composing and also converts Text To Speech (TTS) for reading emails. Google web kit API (Application Programming Interface) is used in this application for speech recognition. Investigations made on the application proves its effectiveness by delivering a better performance when compared with the various parameters like audible distance, accent, pace, words per minute (WPM), accuracy, and homophone words. The graphical analysis depicts the accuracy in terms of word recognition.**

Keywords—voice email, speech to text, text to speech, speech recognition, human computer interaction(HCI)

I. INTRODUCTION

Email has become a common application in professional and personal life. There are more than 250 billion [1] emails sent per day globally by 4 billion users around the globe. This indicates over half of the population using email to communicate letter, message, note, memo, etc. in just few minutes. Around the world more than 200 million people [2] are visually impaired according to the information from WHO (World Health Organization), 2019. Information available in World Wide Web (WWW) is primarily in the form of visual medium. To access and navigate information from WWW needs mouse and keyboard access [3]. Braille keyboards may not be affordable for many visually impaired users [4]. Users who lack in motor skills find troublesome in such activities.

To such users, speech recognition [5] offers a potential solution to access WWW in a comfortable and convenient way.

Sending email [6] might be little challenging for these people as it requires certain additional training and skill. Voice Email [7], [8], [9] will overcome this difficulty and it will assist the user to send email using digital technology with less or almost zero training. This application listens and recognizes human voice and converts them into readable text content. The user can completely write and edit email using their voice without any effort. Interactive Voice Response (IVR) [7] feature add benefit in navigation to the HTML page. Simple voice commands will execute the whole voice to text process enabling smoother and friendly experience to the user.

II. RELATED WORK

Anushka Dixit et al., [10] proposed an application which accepts voice command as input and with the help of voice to text converter, the speech is converted to a command to the application. Based on the command, the corresponding action will be performed. The output is converted from text to audio as speech through the text to speech convertor. This application has 3 modules which are Inbox, Sent mail and, Trash. Inbox contains all received mails. Sent mail is to compose a mail through which speech is converted to text. Trash is to keep a track of all deleted mails. Taslima Binte Hossaina et al., [11] proposed a voice mail android application, which is used to send email through voice. This application has several processing steps which involve processing the voice using API, converting to text, processing the text using API, converting to voice and connecting to the server using GMail API. This application uses a set of Google API for text to speech and speech to text conversion. Naziya Pathan et al., [12] and Pranjal Inglein et al.,[13] uses an Interactive Voice Response (IVR) between the user and the system. This mechanism interacts with the Email host system in which it directs the user through enquiries. IVR system works with the pre-recorded audio voice which assists the

users to proceed the application. This system allows the user to send and receive mails through the voice assistant. This system works in 4 phases. Speech to Text [14], Text to Speech, IVR and speech recognition. To recognize the speech Rudimentary speech recognition software is used to recognize speech which has a vocabulary of phrases and words. Milan Badigar et al.,[15] proposed and designed an android application, which process the voice to text and vice versa to send, read, forward and reply email. This system uses the user fingerprint to authenticate the user. Java Mail application platform is used to build mail and messaging applications. Navigation commands like compose, receive emails, sent mails and go back are used to perform certain actions. Database is stored in SQLite software library which is provided by Android.

III. PROPOSED METHODOLOGY

Most of the email applications come with latest Graphical User Interface (GUI) which is least useful for visually impaired users (VIU). The development in GUI focuses mainly on normal users and not helping VIU. So, there is a necessity of technological development for such population.

The proposed methodology focuses on application development for visually impaired users to send emails which converts speech to text and text to speech. Speech and audio-based emails are the most suitable application for visually impaired users. This application is not only for visually impaired users but also can be well used by the normal users. The application works through voice commands which are simple and user friendly. Once the email is composed using text to speech, the user can verify the dictated content through text to speech conversion. The proposed method will benefit the visually impaired users not to stop them in writing an email.

A. Email Composer

The email composer is shown in the fig. 1. which comprises of five modules. They are processing user voice input, comparing voice with pre-recorded sample, text writing, verifying the output text and send email. The input is given by the user through his/her voice. The application receives the input through suitable hardware and processes the voice. During processing, the voice input received is compared with the sample voice stored in the database. The Google WebKit API recognizes and converts voice to text. The text output gets typed in the designated text box from the voice command received. The user can verify the output text through speech by the headphone connected and edit the content, as necessary. Upon confirmation, the application sends the email from outbox.

The application has got user sign-in page. User must register basic information which are name, email-id, and, password for authentication. The sign-in page is voice enabled which support VIU to self-register. The user must register his

domain address (e.g. @abc.com/org) to the application and this act as sender address. Voice recognition module recognizes user voice through various voice command build in the application and does authenticate and recognize the user. The application is built on Google WebKit API for speech recognition [16]. Voice processing module receives voice, compare the voice with pre-recorded sample which is stored in database, converts voice to text and composes the email. While composing email, actions like stop, pause, edit, delete etc were processed by the speech synthesizer web speech API. The application design for sending email is shown in fig. 2.
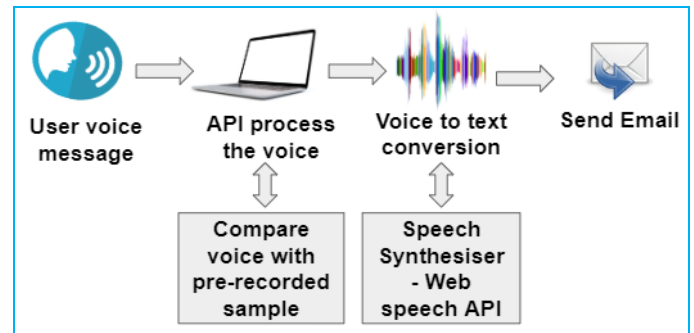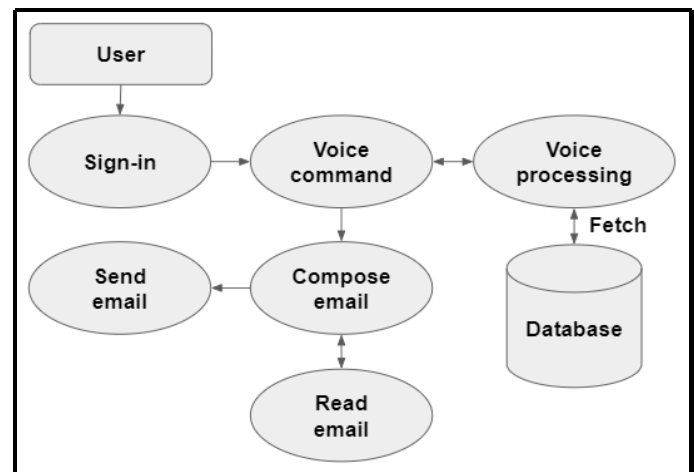

Fig. 1. Email composer

Fig. 2. Application design for sending email

B. Application Modules

The proposed application composed of the following modules:

a) User Voice: During registration process, the user has to store his voice as samples in the database. The voice samples are in the form of commands. This voice data will be used further to authenticate the user. The user provides voice command to the machine through microphone inbuilt or an external microphone attached. The microphone inputs the received voice command to the composer built in HTML. All the functions are carried out through simple voice commands. So, the user need not require to access keyboard keys and also not required to remember keyboard shortcuts.

b) Voice Processing: The application is built with google WebKit API which is used for voice recognition and voice processing. The input provided by microphone is fed into the API and it primarily performs two consecutive activities. The first activity is the conversion of voice to text and the second activity is to authenticate the user voice.

Web speech API is the speech synthesizer used in the proposed system. This speech synthesizer converts text to speech and voice to text. Interfaces like speech recognition interface and speech synthesis interface are used for voice processing execution. The voice is processed by recognizing its distance, pace, continuity and language. The processed voice result is stored in a string format into the database.

c) Comparing voice with pre-recorded sample: The interface authenticates voice command by comparing it with sample voice stored in the database. The sample voice acts like a password which will be stored in PostgreSQL during the sign in as shown in fig.3.
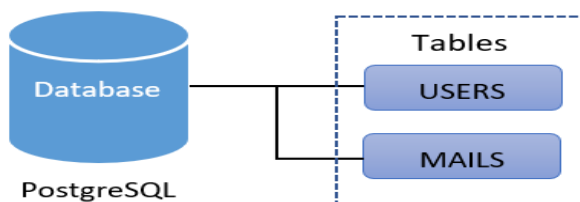

Fig. 3. PostgreSQL tables

d) Voice to text conversion: The speech synthesizer recognizes the following voice commands when commanded.

- "Compose Mail" - opens the email composer window
- "Edit To" – place cursor in To text box and type recipient email id
- "Edit Subject" – place cursor in subject text box and type the subject
- "Edit Body" - place cursor in email text area and type the dictated content
- "Mail Send" – send the email to the mentioned recipient
- "Mail Discard" – it discards the composed email
- "Read mail" – it reads the complete active mail content

The speech synthesis interface has inbuilt properties like pause, pending, speaking, resume, cancel, getVoice etc.

e) Send Email: The voice command "send email" sends the composed email. Users have an option to read the composed email before sending it. The application takes the user domain address (e.g. @abc.com/org) registered during sign-in as sender address.

f) Read email: The voice command "read email" reads the received email. The email can be heard through the speaker or headphone connected to the device.

IV. DESIGN FLOW OF COMMAND PROCESSING

The application starts with sign-in or user registration. The sign-in information once validated, it moves to user recognition. The application validates the user voice with the sample voice stored in database. If the application detects unauthorized user from internal process check, it skips to sign-in. The applications wait for the user voice command whether it is "compose email" or "read email".

If the command given by the user is "compose email", the application opens the email composer upon validation and the user start to compose email. Further voice commands "edit to", "edit subject" and "edit body" are required to compose the email. The STT interface process the speech to text and inputs the content to its intended component [17]. After composing the email, the user can read the composed email using the command "read email". The composed email is read to the user through text to speech interface. Finally, the composed email is sent to the mentioned recipient through "send email" voice command.

If the command given by the user is "read email", the interface reads the email to the user through text to speech recognition [18], [19]. The user can sign-out from the application. The flow chart of design flow of command processing is shown in the fig. 4.

V. NOISE ELIMINATION AND SECURITY

To enhance the accuracy of speech recognition process, noise elimination techniques are used. The input voice may contain noise or unwanted information. The web speech API has a built-in event handler which detects and eliminates the noise when it detects any sound. This application is tested for noise elimination with different environmental surroundings.

Security feature is enabled in this application by storing the user voice in the database. When the user provides voice command, the command is validated before processing. This feature prevents unauthorized user to access the application.

VI. EXPERIMENTAL RESULTS

This application is tested using the metrics distance, sound, speech and pronunciation [20]. The performance analysis is summarized in the below table.1.

TABLE I. PERFORMANCE ANALYSIS

| Test Case	Aim	Description	Result
1	Voice audible test	Ability of the model to receive voice in distance (near & far)	40 to 50 CM
2	Voice pitch	Ability of the model to receive voice with different tone	60 - 100 DB
3	Voice pace	Ability of the model to receive voice tone (slow & speed)	40 - 60 words per minute
4	Accent	Ability of the model to receive voice with different English accent	Pass (Indian Accent)
5	Noise elimination test	Ability to eliminate voice in different environments	Excellent
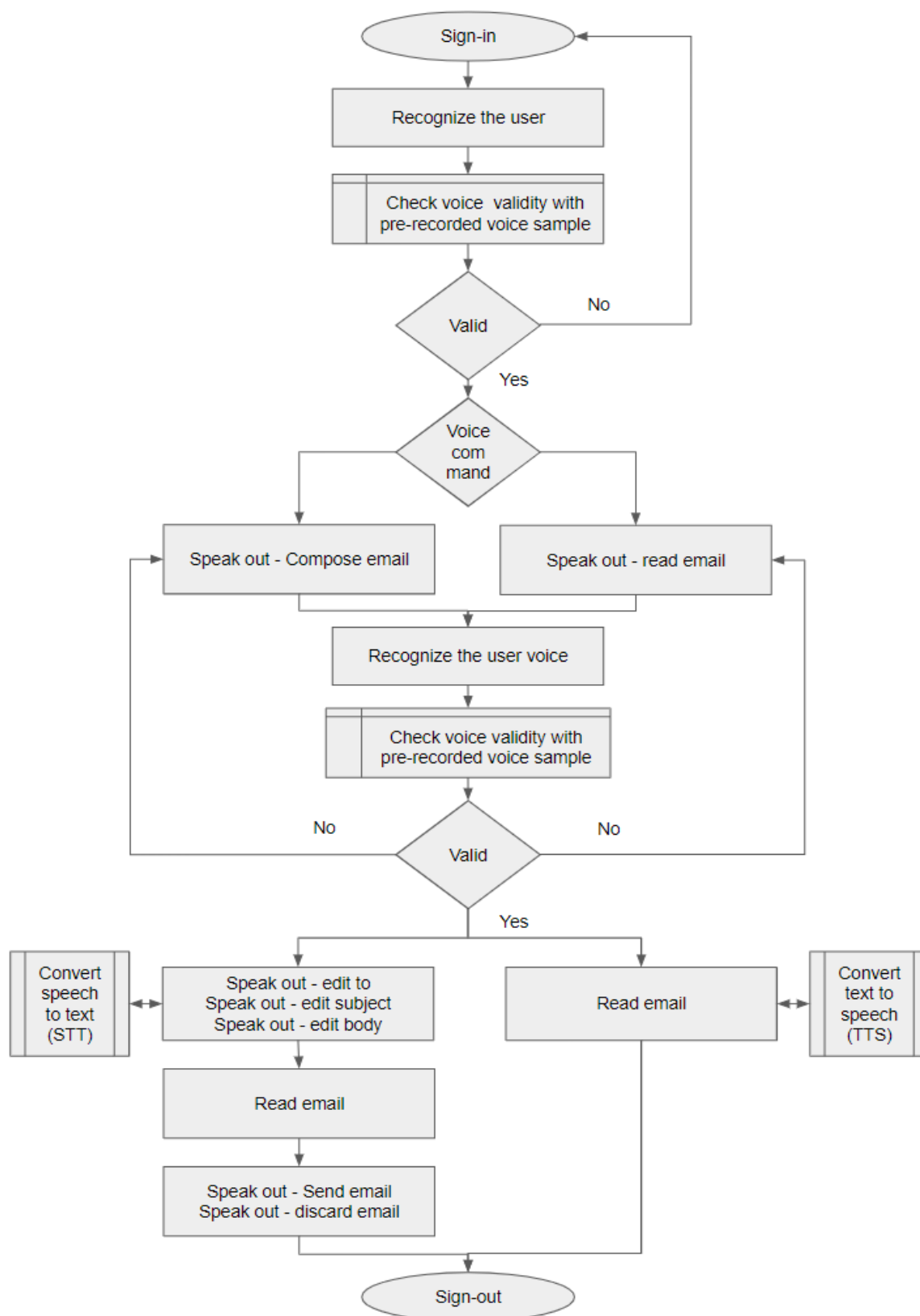
Fig. 4. Design flow of command processing

As homophones are the words that sound similar, it creates an ambiguity for the model in speech recognition. Evaluations are also made using homophones in table.2 to understand the accuracy of the proposed model.

TABLE II.   WORDS RECOGINITION IN HOMOPHONE PHRASE

Sample Words		Correct Recognition	Incorrect Recognition
Word 1	Word 2		
praise	prays	20	0
eight	ate	19	1
two	to	18	2
peak	peek	20	0
by	buy	18	2

Fig.5 indicates the graphical representation with x-axis denoting sample words and y-axis denoting words recognition count with the tested model. Using the model, accuracy is measured with different input words. In order to check the accuracy, considerations are made using correctly recognized, incorrectly recognized, and not recognized words.
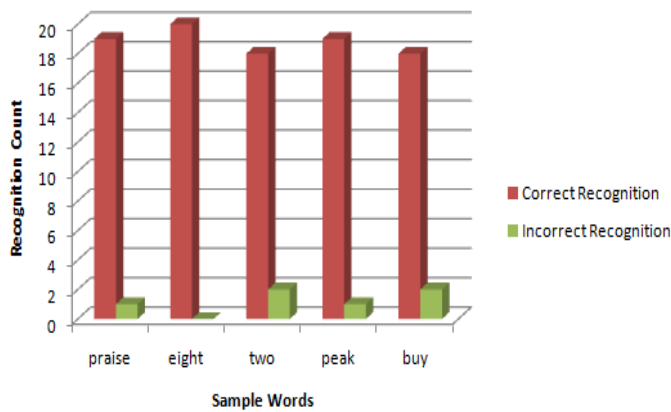

Fig. 5. Word Recognition

VII. IMPLEMENTATION

The email management application MailBox is used for user interface. This interface allows the user to register user name, Email id and password.

Fig. 6. Mailbox Registration page

The user interface registration page is shown in fig. 6.

Fig. 7. Mailbox login page

After registration, the user can sign-in using the login page which is shown in fig. 7. To login the page, the user needs to provide email id and password. If the login is successful with the registered email id and password, the inbox interface is displayed.

Fig. 8. Email composer

If the user needs to send a mail or to read the mail, using the voice command, the user can say "compose email". This voice command opens the email composer which is shown in the fig. 8.

Fig. 9. Mailbox – Inbox

If the user wants to read the inbox emails, the user voice command with "read email" and the emails from inbox are read. The inbox emails are shown in the fig. 9.

VIII. CONCLUSION AND FUTURE WORK

In today's digital world, communication through email have become day-to-day activity for most of us. The proposed application benefits the visually impaired users to send emails seamlessly through voice commands without any dependency. Also this application helps the normal user to send emails with less effort. The application has a limitless enhancement opportunity.

Instead of authenticating the user by using email ID and password, enhancements like fingerprint reader can be used. Additional voice commands like reply, forward, file attachment, starred etc., can be added to enhance the interface. Semantic suggestions and auto corrections can be added to the email composer to access better. More research and enhancements can be performed using NLP algorithms to recognize local languages. The stated enhancements boost user experience and it can be productive too.

REFERENCES

[1] Tharani K, Shalini R, Jeyanthi I, Deepalakshmi R "Voice based mail attachment for visually challenged people" International Journal of Scientific & Engineering Research (IJSER) Volume 8, Issue 5, May-2017.

[2] Siddhartha, B., Arun kumar P. Chavan, and B.V. Uma "An Electronic Smart Jacket for the Navigation of Visually Impaired Society." Materials Today: Proceedings -2018.

[3] Ummuhany sifa U., Nizar Banu P K, "Voice Based Search Engine and Web page Reader". In International Journal of Computational Engineering Research (IJCER). Pages 1-5, May - 2013

[4] Braile Writing System, Retrieved on May 01, 2019 from https://www.britannica.com/topic/Braille-writingsystem.

[5] Anchal Katyal, Amanpreet Kaur, Jasmeen Gill "Automatic Speech Recognition: A Review", International Journal of Engineering and Advanced Technology (IJEAT), Volume-3, Issue-3, February 2014.

[6] Jayachandran.K, Anbumani.P, "Voice Based Email for Blind People "International Journal of Advance Research, Ideas and Innovations in Technology. Volume3, Issue3, 2017.

[7] Rijwan Khan, Pawan Kumar Sharma, Sumit Raj, Sushil Kr. Verma, Sparsh Katiyar "Voice Based E-Mail System using Artificial Intelligence" International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-3, February 2020.

[8] M.J. Carmel Mary Belinda,N. Rupavathy, N.R. Mahalakshmi "A voice based text mail system for visually impaired" International Journal of Engineering &Technology,132-136 – 2018.

[9] Hashemi Hassan, Mehdi Khabazkhoob, Mohammad Saatchi, Hadi Ostad imoghaddam, Abbasali Yekta "Visual Impairment and Blindness in a Population-Based Study of Mashhad", Iran. Journal of Current Ophthalmology, 2018, pp. 161-168

[10] Anushka Solanki, Anushka Dixit, Anushri Saxena, Amit Khare, Kavita Namdev. "E- VOICE Mailing System for Blinds Using Android", International Research Journal of Modernization in Engineering Technology and Science, Vol 02, Issue: 04, April-2020.

[11] Taslima Binte Hossaina, Yeasmin Ara Akterb, Md. Ataur Rahmanc "Voice mail application for visually impaired persons", Recent Research in Science and Technology, 15-18 special issue on 6th International Conference on Natural Sciences and Technology Vol 12, 2020.

[12] Naziya Pathan, Nikita Bhoyar, Ushma Lakra, Dileshwari Lilhare, "V-Mail (Voice Based E-Mail Application)", International Research Journal of Engineering and Technology (IRJET), Volume: 06, Issue: 03, Mar 2019.

[13] Pranjal Ingle, Harshada Kanade, Arti Lanke "Voice based e-mail System for Blinds" International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) Volume 3, Issue 1, PP 25-30, 2016.

[14] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya SS "Text-to-Speech Conversion System using OCR", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 1, January 2015.

[15] Milan Badigar Nikita Dias Jemima Dias Mario Pinto "Voice Based Email Application for Visually Impaired", International Journal of Science Technology & Engineering (IJSTE) Volume 4, Issue 12, June 2018.

[16] Preeti Saini, Parneet Kaur "Automatic Speech Reorganization: A Review", International journal of Engineering Trends and Technology Volume 4 Issue 2-2013.

[17] Rijwan Khan, Pawan Kumar Sharma, Sumit Raj, Sushil Kr. Verma, Sparsh Katiyar, "Voice Based E-Mail System using Artificial Intelligence" International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-3, February, 2020.

[18] Ingle, Pranjal, Harshada Kanade, and Arti Lanke, "Voice based e-mail System for Blinds." International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) (2016): 25-30.

[19] Isewon, Itunuoluwa, O. J. Oyelade, and O. O. Oladipupo. "Design and implementation of text to speech conversion for visually impaired people." International Journal of Applied Information Systems 7.2 (2012): 26-30.

[20] Srija Rallabhandy and Sireesha Rodda "KeyBoard-Less Online Shopping for the Visually Impaired Using Natural Language Processing and Face Recognition Mechanism", Smart Intelligent Computing and Applications, Smart Innovation, Systems and Technologies, 160, 2020.