International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Speech and Text based Assistive Application for Special Needs Individuals

Dr. Ashwini Rao
Assistant Professor, MPSTME, IT Dept, NMIMS University
Mumbai (M.H.), India
ashwini.rao@nmims.edu

Srijamya Ranjan
Student, MPSTME, IT Dept, NMIMS University

Archit Anand
Student, MPSTME, IT Dept, NMIMS University

Rashil Rambia
Student, MPSTME, IT Dept, NMIMS University

Abstract— As per the survey report released last year on disability by National Statistics Office, it was seen that people with some kind of mental or physical disability is around 2.2% of India's total population. The project, "Saksham" that suggests be independent, aims to eliminate the need for human assistance and to provide equal opportunities and a more normal way of life to those with language or other disabilities. In the direction of building individual strength and also get great improvement in the academic performance of adults and kids with special needs, Assistive technology is now being used as an innovative tool. The entire application have been developed keeping in mind that it needs to provide all our users with instantaneous access to selected features specially catered to help them in completing their daily tasks.

Keywords Assistive Technology; Speech Recognition; Optical Character Recognition; Dyslexia

I. INTRODUCTION

Many of us are lucky enough to live our daily life without problems arising when carrying out small, seemingly mundane tasks like reading a report, appearing for a presentation, or getting any of our own official paperwork handled; in fact, we probably don't even give it a second thought. However, not everyone can say the same and there are so many individuals who might have to put in some extra effort from their side for these same jobs.

Individuals who have difficulty in hearing or are deaf, or those with speech and/or language disabilities require assistance in daily tasks. The technology that is now popularly being used for assisting, adapting, and creating rehabilitative devices for helping elderly people or people with disabilities is Assistive technology. It is often seen that people with disabilities find it difficult to even carry out their day-to-day tasks on their own and in some cases not even with assistance. For helping physically challenged people, we now have computer programs called assistive/adaptive software that are designed for specialized hardware applications which can be easily used by people with special needs.

In many situations, especially at school or work, they may have trouble reading documents or communicating with peers. The paper proposes a mobile application, which will solve this purpose by providing them with easily usable speech and text conversion features which they can use on their own with no assistance, even while traveling. These features will help them overcome social barriers and help them interact with ease, leading to a relatively normal life.

Dyslexia is one of the most common types of language disabilities. People with this disability find it difficult to read, write, spell words, create sentence structure and in some cases even lack communication and math which are all skills mainly needed for understanding language. As per the research conducted by Horn et al. [1], it was observed that around 15-20% of the population may have symptoms of Dyslexia and one in ten school-aged children are diagnosed with dyslexia. Yet many others around the world are not properly diagnosed or have partial symptoms, which can result in them not getting proper attention and the help they require.

It was stated in a report by World Health Organization that in India around 63 mil-lion people suffer from either partial or complete deafness, and of these the number of children would be at least 50 lakhs. A nice innovative tool that is being used to improve academic performance and build a good amount of self confidence among special needs adults and kids is Assistive technology. Our aim with this application is to enable these individuals to live their life with ease and make them feel like an equal by giving them a sense of independence.

Primarily people who are deaf, hard of hearing, or who have speech and/or language disabilities use assistive technology involving voice communication. It is also used to a lesser extent by people with visual or motor disabilities. The application has features that will provide them with the capability to bridge the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

gap they face by eliminating/simplifying some of the daily issues that they may encounter.

II. Author's Contribution

Our application "Saksham", proposed in the paper will provide assistance to differently abled users such as deaf-mute individuals or those with learning disabilities like dyslexia, dysgraphia in carrying out some of their everyday tasks. The project aims to eliminate the need for human assistance and to provide those with disabilities with equal opportunities and a more normal way of life.

III. Related Work

Assistive Technology are services or devices that enable people with disabilities to accomplish daily living tasks; assist them in communication, education, work, or recreation activities; and ultimately, help them achieve greater independence and enhance their quality of life. Comprehensive search strategy to select and filter two quantitative studies and three qualitative studies, was conducted in the United States, Malaysia, and Spain, on use of Assistive Technology (AT) by researchers Horn et al. [1]. Authors reviewed and compared the use of assistive technology by Dyslexic students, such as Screen Reader and Livescribe Pen. In most cases, writing scores increased with the use of word prediction and speech-to-text features, while a good improvement in reading was also seen with the text-to-speech functionality. The eye tracking studies conducted lend support to the claim that individuals with dyslexia benefit from visualization, larger font sizes and spacing's improve readability. The font sizes of 18, 22, and 26 points on a 17-inch computer screen was found to be preferable among dyslexics for better readability. Also, when compared with other colors, a white font on a black background or a black font on a white background was more readable. Spacing ranging from zero, +seven, to 14% was more readable. The most preferred typeface was found to be Arial, Courier, Verdana, or Helvetica and popular Font style that should be set was Roman or Sans Serif. It was proved from the experiments that appropriate parameter settings for online text makes it more readable and hence could benefit dyslexic students when they read the online content.

Computers are the most widely used device. Other devices that are used in schools include iPads, projectors, and to a lesser extent, television sets. Another survey conducted by Sameer et al. [2] composed of 30 students and 20 parents depicts the use of different computer-assisted technologies aids to support the learning of children.

Optical character recognition or optical character reader is the electronic or mechanical conversion of images of typed, handwritten, or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposed on an image. There are several tools available for text extraction from images such as, Tesseract OCR, OpenCV, Google Firebase ML Kit, Amazon Textractor & many more.

The approach proposed by S. J. Ha et al. [4] uses integral image and design filters that are proper to detect text regions on the integral image. After the filtering, the center points in the regions are discovered by cascade text region verification followed by non-maximum suppression. Finally, text lines are extracted by grouping the points on the same line.

The authors Revathi et al. [3] discuss the growing requirements of text extraction from images and the tools available for this purpose. OpenCV and Tesseract are among the most popular and widely used tools. Image processing is a software-focused domain, it has a wide range of applications. Tesseract is an OCR Software used for the conversion of an image into text. If the quality of an image is not pristine, this software is prone to errors. So, the image needs processing using OpenCV and then processed using Tesseract for better results. Using OpenCV and Tesseract OCR the authors have suggested 3 methods:

- Using Manual Trackbar
- Trackbar in Autonomous mode
- Using Image Processing

The results of the three methods are tabulated in Table I below.

TABLE I. Accuracy report for proposed methods in [3]

	Without Processing	Trackbar in Manual Mode	Trackbar in autonomous Mode	Image Processing
Total Characters	180	180	180	180
Extracted Characters	104	174	158	165
Accuracy	57.78%	96.67%	87.78%	91.67%

During the literature survey of Text-to-Speech and Speech-to-Text conversion, it was found that many techniques were being used on different ways to promote the accuracy of English speech recognition based on an improved Hidden Markov model (HMM), as opposed to traditional methods such as Dynamic Time Warping and Artificial Neural Networks based methods that have many disadvantages. The Hidden Markov model can segment sentences into smaller units, such as characters, furthermore it is able to seamlessly combine and exploit language models in the de-coding process. The English speech recognition problem is described as seeking the most suitable word sequence given a segment of English voice.

The proposed English speech recognition system by Cuiling et al. [5] is made up of four parts: 1) Voice acquisition, 2) Speech model, 3) Speech recognition and 4) Speech recognition results. Since the performance of the traditional Hidden Markov model is not satisfactory, it is improved through adding a hidden layer to represent the evolution of the state transition. To test the performance of the proposed algorithm, the Aurora 2 English language database is used to conduct an experiment. In particular, the English speech recognition experiments are conducted with 1001 utterances of speech, and four types of noisy environments are contained: 1) subway, 2) babble, 3) car, and 4) exhibition hall.

Furthermore, in each experiment, six different settings were utilized, that is, SNR was set in the range of [-5, 0, 5, 10, 15, 20]. The results are shown in Table 2.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

TABLE II. AVERAGE RECOGNITION ACCURACY FOR DIFFERENT ENVIRONMENTS FOR HMMM

Environment	Average Recognition accuracy (%)	
	Standard HMMM	The improved HMMM
Subway	44.87	48.98
Babble	45.72	49.25
Car	52.79	57.95
Exhibition Hall	55.27	59.84

It is observed that the proposed method was able to solve the English speech recognition problem in different noisy environments and can effectively enhance the accuracy of the English speech recognition process.

IV. MARKET SURVEY AND PROTOTYPE DESIGN

Existing applications in the market do not serve the specific purpose of helping various types of disabilities; they do not cater to their problems and are not readily usable by such individuals. Also, these applications are complicated and don't really help the users much. Further, they are too expensive to use for the general public and only provide singular features at best.

The main focus of an application for a special target audience should be to have a simple UI where all the functionality of the app are accessible easily from the home screen. This will not only make the job quicker, but also less cumbersome for the disabled individuals.

Existing Applications were researched are:

- Voice4U
- Proloquo2Go
- Braina
- Speak It!
- Dragon Anywhere

Voice4U: Is an augmentative and alternative communication app that helps students with communicative disorders express their ideas, thoughts, feelings, etc.

Braina: Braina is an intelligent personal assistant application for Microsoft Windows marketed by Brainasoft. Braina uses natural language interface and speech recognition to interact with its users and allows users to use natural language sentences to perform various tasks on their computer.

To better understand the needs of our users, we have conducted a contextual inquiry (CI) with sample target users to note down areas of importance that we need to focus on while designing the UI of the application.

Target Audience: Users of this application are the people who are differently abled such as deaf-mute individuals, partially blind individuals, or those with learning disabilities like dyslexia, dysgraphia.

Focus of CI:

- Are people able to understand what you try to convey to them? (Mute Users)
- Are you comfortable with using a smartphone?
- What problems do you face in your daily tasks?
- Have you used any such app before? If yes, did you find it useful?

Mode of Conducting survey: Physical and Virtual

Interviewed: School friends/batchmates with dyslexia [for deaf-mute we have contacted few NGOs, we have checked information from forums for disabled people and their family members/caretakers.

Interview Questions:

- Are you comfortable with using smartphones on a daily basis without any help?
- Further, are you familiar with camera scanner and voice recorder features? Have you faced any prior issues regarding these?
- What specific problems do you face in your daily life?
- In terms of hearing or vocal disabilities: what issues have you faced in regular work or school conversations?
- For learning disabilities: what kind of problems do you have in reading and understanding long texts?
- Have you used anything similar in the past? If yes, how was your experience?

Based on the contextual inquiry (CI) conducted, we have used the Figma software to create a prototype of our entire application and simulate the working environment. Figma is a vector graphics editor and prototyping tool, which is primarily web-based, with additional offline features enabled by desktop applications for MacOS and Windows. The Figma mobile app for Android and iOS allows viewing and interacting with Figma prototypes in real-time mobile devices. Using this we were able to create and test the screens shown in Fig.1 to Fig 4.

We have split the Read Easy module into three basic functioning parts which are essentially: camera for capturing & scanning images, text extraction using OCR, and text to speech conversion for this extracted text. These will make sure that we can work on and test each part of this feature properly.

For the camera, the phone camera is successfully accessed according to the user's input action using the inbuilt flutter camera plugin. Additional camera permissions were required to be added for this.

Text extraction from the image works but will be further improved; we have tried multiple methods such as flutter mobile vision, tesseract OCR, Google ML kit, to be sure of the most efficient extraction technique in practice. We have also tried using python script for text extraction but finally decided with techniques mentioned earlier as these are known to have a higher accuracy.

Text to speech works as per the expectations, we have used flutter's text to speech package for this function. It is delivering the requirements specified for the function accurately. We worked on the sound, pitch and accent of the voice message generated using Flutter_TTS (Text to Speech) package and its parameters. Other options that we tried and compared were a python script using Gtts (Google Text-to Speech).

V. PROPOSED SYSTEM

After careful consideration and an extensive literature review, the three main features that we have defined for the users of our application are as described below:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

A. Read Easy Module

People with learning disabilities such as dyslexia and partially blind people will be the users of this feature. The feature will help the users in extracting text from images they capture and convert the extracted text to speech.

This feature will open to a scanner which will essentially capture written data and provide a text to speech conversion for users with learning disabilities. E.g., Students/adults with dyslexia can use the app to scan and help them with studies or office work, and in special cases, even partially visually impaired users may be able to utilize this.

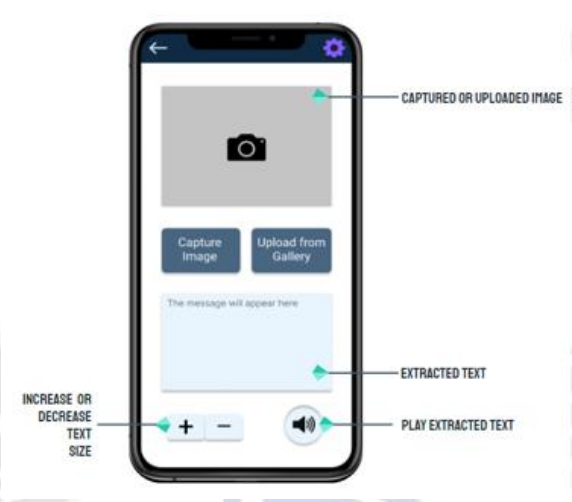

Figure. 1. Read Easy Prototype

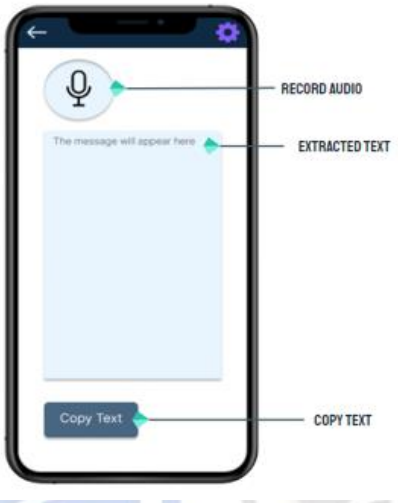

Figure. 2. Listen Freely Prototype

Figure. 3. Voice Bot Prototype

Figure. 4. Home Page Prototype

Components are:

1. Scan pictures from user device
2. Extract text from scanned picture using OCR
3. Convert text to voice using Text to Speech
4. Play audio message

B. Listen Freely Module

People with hearing disabilities will be the users of this feature. The feature will help the users to understand conversations spoken around them by recording and converting the audio to text format on their screens.

It will open to a voice recorder which will be able to record speech and provide voice to text conversion for users with hearing disabilities.

E.g., Deaf-mute individuals or partially deaf persons can use the app to easily participate in conversations around them.

Components are:

1. Record audio from user device
2. Convert audio to text
3. Display text message to user

C. Voice Bot Module

People with any form of speech disability will be the users of this feature and it will help them by synthesizing the text they type in the textbox to audio, which can then be played in conversations. This may substitute the use of sign language in situations where not everyone may be familiar with it. It will accept dialogue from users in the form of text and convert it to speech, which will help people with speech disability to interact with others.

E.g., A vocally challenged person can type out messages which will be read out as voice messages as everyone is not familiar with sign language.

Components are:

1. Accept text input from user
2. Convert text to speech
3. Play message

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

VI. Application Screen Shots And Test Cases

A. Screen Shots

Fig. 5 displays the Home Screen, prompting the user to select the feature that they may want to use.

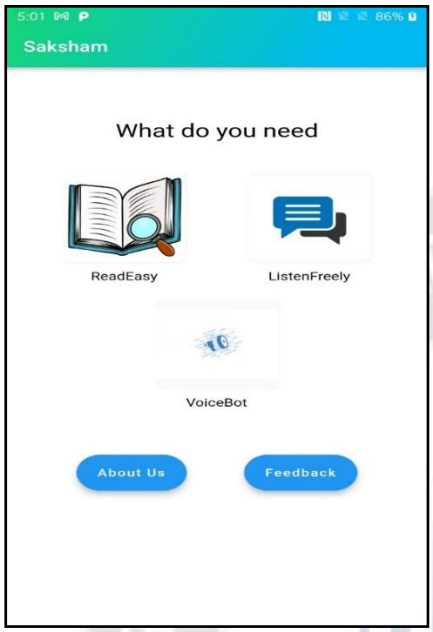

Figure. 5. Home Page

Fig. 6 and Fig. 7 displays the Read Easy module, that scans text the user supplies, converts it into speech and plays the audio message.

Fig 8 and Fig. 9 present the screen for Listen Freely module that prompts the user to speak few lines, converts this into text and displays the message.

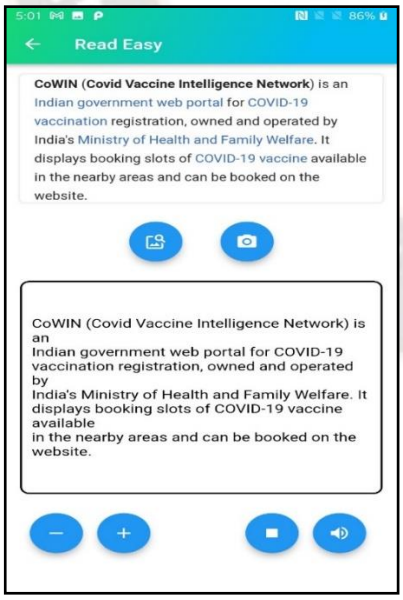

Figure. 6. ReadEasy Scanning Image from Gallery

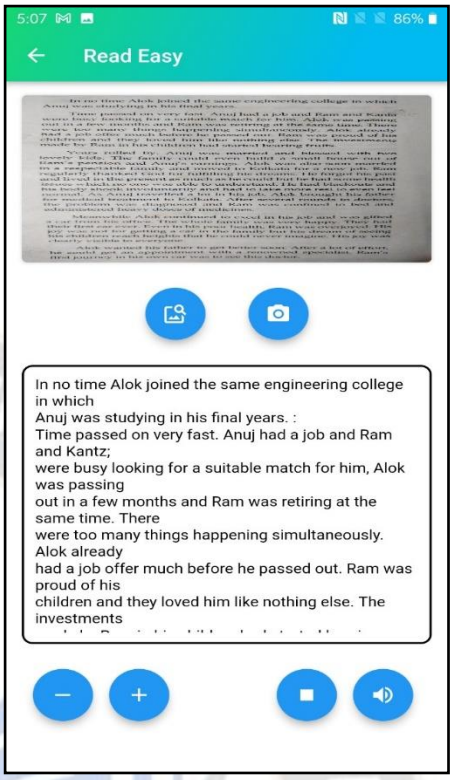

Figure. 7. ReadEasy Scanning Image using Camera

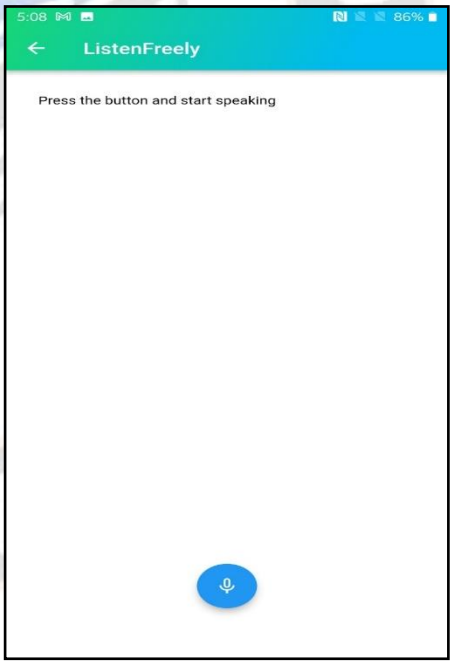

Figure. 8. Listen Freely Page

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

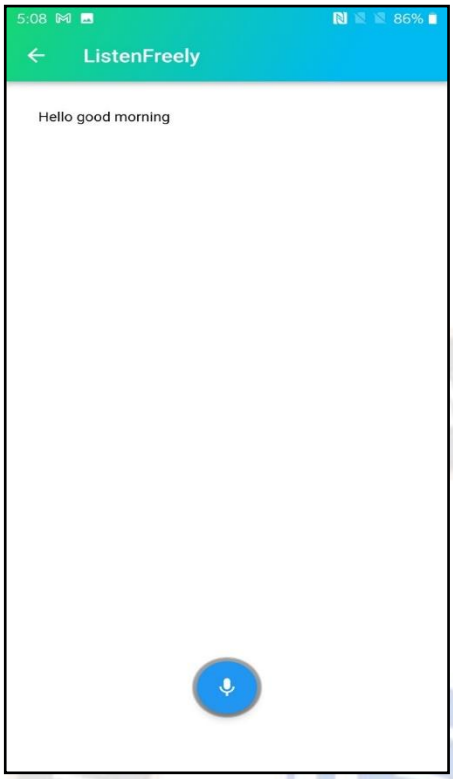

Figure. 9. Listen Freely page displaying captured text

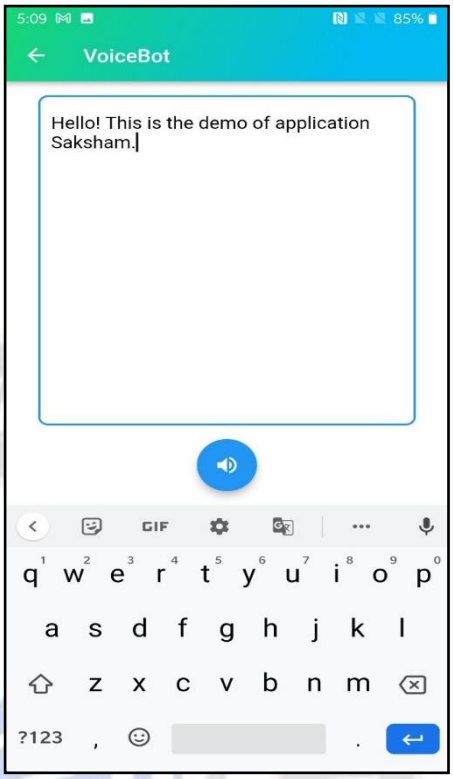

Figure. 11. VoiceBot Page after entering text

Fig. 10 and Fig. 11 displays the Voice Bot Screen that takes text input from the user and converts it into speech and then plays the audio message.

Fig. 12 displays the Feedback Screen that is used to capture users feedback and suggestion on "Saksham".

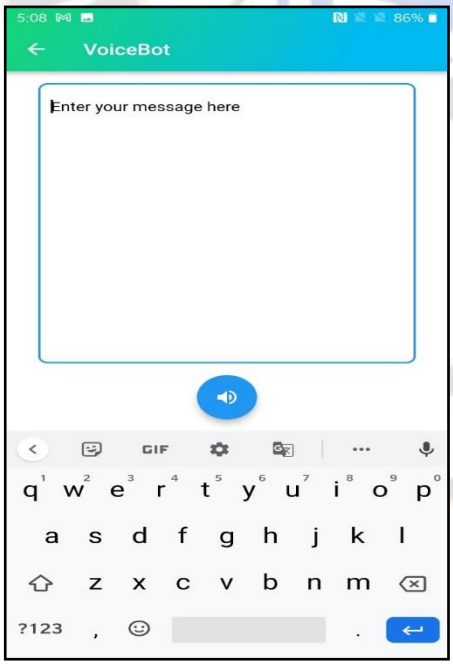

Figure. 10. VoiceBot Page

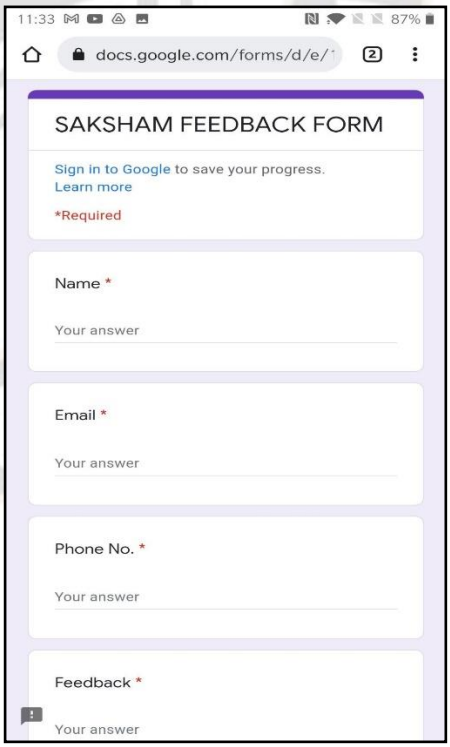

Figure. 12. Feedback Form

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

VII. TEST CASES

To test the application's accuracy and speed we divided the test cases into three special categories corresponding to each module.

A. Read Easy Module

The main component to be tested in the Read Easy module was the optical character recognition on images in different lighting conditions and proximity. For this we used a dataset of various images clicked by us under different conditions. These include:

- Images with a combination of text, objects & faces
- Images with varying light intensity
- Images at different distances

For measuring the accuracy of OCR, we decided to cross check the number of words in the original file against the number of words that are correctly extracted by the OCR model.
Given below are the results of the testing:
Case 1: Upload from camera

TABLE III. TESTCASES AND RESULT FOR READ EASY UPLOAD FROM CAMERA

Test Case No.	No. of words in the image	No. of words correctly identified	Accuracy
1	22	17	77.27%
2	11	7	63.64%
3	247	185	74.9%
4	28	22	78.57%
5	46	35	76.08%
6	31	24	77.41%
7	15	11	73.33%
8	109	90	82.56%
9	87	71	81.6%
10	68	54	79.41%

The average accuracy of the module on completion of testing is 76.447%.

Case 2: Upload from gallery

TABLE IV. TESTCASES AND RESULTS FOR READ EASY UPLOAD FROM GALLERY

Test Case No.	No. of words in the image	No. of words correctly identified	Accuracy
1	22	20	90.91%
2	11	9	81.81%
3	220	200	90.9%
4	28	28	100%
5	46	45	97.82%
6	31	28	90.32%
7	15	15	100%
8	109	98	89.9%
9	87	80	91.95%
10	68	67	97.05%

The average accuracy of the module on completion of testing is 93.066%.

B. Listen Freely Module

We needed to measure the accuracy of the STT package recognizing words in voice recording samples against the total number of words spoken.

TABLE V. LISTEN FREELY TEST CASES AND RESULTS

Test Case No.	No. of words	No. of words correctly identified	Accuracy
1	26	26	100%
2	4	4	100%
3	5	5	100%
4	5	5	100%
5	7	6	85.7%
6	7	5	71.4%
7	8	8	100%
8	16	15	93.75%
9	10	10	100%
10	15	15	100%
11	24	23	95.83%
12	24	21	87.5%
13	33	27	81.81%
14	12	12	100%
15	21	19	90.48%

The average accuracy of the module on completion of testing is 93.76%.

C. Voice Bot Module

This module requires testing using text input, and on running test scenarios the audio output seems to run satisfactorily. Hence, we can conclude no further testing is required for this module.

VIII. CONCLUSION AND FUTURE WORK

Saksham aims to provide individuals with language or other disabilities with equal opportunities and a more normal way of life. The various fresh features in our application have all been designed with the sole purpose of being able to help our users to single-handedly read, write and converse with others around them without missing a beat. The user interface makes it so simple that no person ever feels lost while using it and all new users can get the hang of the flow as quickly as possible.

The goal of this project Saksham has always been to bridge the gap and make everyone feel at ease no matter the environment, be it at the workplace or with friends and peers. This application has been designed with great care, with the sole purpose of fulfilling this mission so that our users feel that every single element has been designed uniquely for them and their needs.

Moving forward, we plan to conduct various user studies to receive feedback from each class of special needs individuals that Saksham aims to help so that we can get an appropriate idea of their response to both the application's current user interface as well as the functioning of each individual module in a real time setting. This will guide us in making the application better suited to the work environment it is aimed to be deployed in. This will also help us in including various other components like live GPS tracking and integration of IoT based hardware devices (example: smart blind stick) in future.

REFERENCES

[1] Horn, Tara Dawn, and Tonya Huber. "Assistive Technologies and Academic Success for Students with Dyslexia: A Literature Review." International Journal of Educational Technology and Learning 9.1 (2020): 52-59.
[2] Abuzandah, Sameer. "The Success of Using Assistive Technology with Disabled and Non-Disabled Students." Asian Journal of Sociological Research (2021): 15-18.
[3] Revathi, A. S., and Nishi A. Modi. "Comparative analysis of text extraction from color images using tesseract and opencv." 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2021.
[4] Ha, Seong Jong, Bora Jin, and Nam Ik Cho. "Fast text line extraction in document images." 2012 19th IEEE International Conference on Image Processing. IEEE, 2012.
[5] Cuiling, Lv. "English speech recognition method based on Hidden Markov model." 2016 International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE, 2016.
[6] Tsai, Chung-Yao, Chin-Kuan Kuo, Y. Wang, S. Chen, I-Bin Liao and C. Chiang. "Hierarchical prosody modeling of English speech and its application to TTS." 2014 17th Oriental Chapter of the International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (COCOSDA). IEEE, 2014.
[7] Mahanta, D., Sharma, B., Sarmah, P., & Prasanna, S. M. "Text to speech synthesis system in Indian English." 2016 IEEE Region 10 Conference (TENCON). IEEE, 2016.
[8] Mullah, Helal Uddin, Fidalizia Pyrtuh, and L. Joyprakash Singh. "Development of an HMM-based speech synthesis system for Indian English language." 2015 international symposium on advanced computing and communication (ISACC). IEEE, 2015.
[9] Venkateswarlu, S., Kamesh, D. B. K., Sastry, J. K. R., & Rani, R. "Text to speech conversion." Indian Journal of Science and Technology 9.38 (2016): 1-3.
[10] Yang, Y., & Pedersen, J.O., "A comparative study of feature selection in text categorization," In Proceedings ICML, pp. 412-420, 1997