AI-WEAR: TEXT READER GLASSES FOR VISUALLY IMPAIRED STUDENTS USING RASPBERRYPI WITH AUDIO-VISUAL CALL AND GOOGLE ASSISTANCE

Vincent D. Manlesis[1], Jose Andrei D. Manguiat[2]

[1] Department of Engineering and Computer Science, Laguna College of Business and Arts, Philippines

Abstract

This study proposes the Text Reader Glasses using Raspberry Pi to improve the learning status of visually impaired students. The device includes a text-to-speech feature for reading, Google assistance for online support, and video streaming via Jitsi Meet, which allows them to interact with their teachers. It is controlled in two ways: through voice command and user-friendly buttons with engraved Braille letters. The system combines the concept of Optical Character Recognition and Text to Speech Synthesizer (TTS) in Raspberry Pi. Text-to-Speech conversion scans and reads the extracted text using the OCR technique and then gives audio output through the headset. Furthermore, the system incorporates GSM/GPRS to access internet connection using cellular data when Wi-Fi is unavailable. The device produces an average 99.44% accuracy rate in font 16 of Courier New, and the device can run up to 8 hours of usage with all the system features. This project has been subjected to extensive testing and maintenance, which impacted its evaluation results. Subsequently, it is found out that the device functions well and has passed the standard level of ISO/IEC 25010. Although there are some adjustments to consider, the proposed device is a potential asset to provide innovative learning opportunities for the visually impaired in distance modalities.

Keywords: *Text Reader Glasses, Raspberry Pi, Visually Impaired, Audio-Visual Call, OCR, TTS.*

INTRODUCTION

Limited assistive technologies in the Philippines remain a constraint for visually impaired students' right to education. Although the government and non-government institutions support Special Education (SPED) Centers, a steady decline in enrolment persists nationwide. Worsen by the COVID-19 pandemic, learners with visual impairment struggle to obtain an inclusive and high-quality education. Today, their main challenge is the inability to read digital text information since Braille is limited to printed materials. Although advanced assistive devices exist, the costs are prohibitive and lack features that enable communication between the students and teachers in virtual learning. Until now, their access to low-cost and innovative reading tools is insufficient for interactive education. Thus, it is critical to address this issue to comprehend and develop solutions, providing the needs of visually impaired students while meeting the new demands of the evolving education system.

According to the UNICEF Philippines survey (2021), access to education and lack of assistive technologies are the main concerns of students with disabilities amidst the pandemic. During the transition of the education system into distant modalities, the Department of Education (DepEd) employed alternative learning methods for over 400,000 SPED students to continue their education. DepEd provided Self-Learning Modules with alternative learning modalities to address the accessibility of education for every learner with special needs. However, Save the Children (2020) argued that educational disruptions such as lack of government support, access to the internet, and accessible software learning tools still struggle the learners. Relatively the

Department of Education's 2021 National Expenditure Program did not provide allocation for SPED assistive equipment.

In relation, this study aimed to develop a low-cost, efficient, and user-friendly assistive device that improves the learning status of visually impaired students. The proponents intended to create an AI-Wear: Text Reader Glasses using Raspberry Pi with Audio-Visual Call and Google assistance. This device has text recognition for reading, real-time video streaming for online help, Google assistance for online information, and user-friendly control buttons with engraved Braille letters. This assistive device could elevate both teaching and learning methods in special education while contributing to the importance of technological advancement today. This study was conducted with the professionals, guardians and visually impaired students of Calamba Integrated School SPED Center, within the school year 2021-2022.

LITERATURE REVIEW

Wahab et al. (2020) emphasized that the development of assistive tools is necessary to accommodate visually impaired students' problems in education. Several studies demonstrate the help of assistive devices to visually impaired learners, increasing their communicative abilities and independence in learning. Indeed, blind and visually impaired people are potential subjects of new technologies. Internet and new assistive devices serve as opportunities for them to live with convenience and expand their capabilities. Thus, devising a system to elevate their learning methods and access to the digital world is highly significant (Sarkar et al., 2021).

The concept of the study has two categories: (1) user management and (2) module to operate the system. The module will serve as the main component of software and hardware enabling system operation. The data will process from the input module such as the camera, microphone, and GSM/GPRS for the internet connectivity. In this study, every hardware module designates to its process in the system.

User management's functionality depends on internet availability. Text recognition will produce a text-to-speech output using the output module such as speaker/earphone. The system accessibility requires an image processing library such as OpenCV and Tesseract OCR, used for output accuracy. The text-to-speech output is accessible with or without an internet connection. In contrast, the feature needed for the internet connection is the API system. Google assistance will provide the necessary online information for the user, and at the same time, Jitsi meet provides video chat to their professor if some problem occurs. For the system's efficiency, it provides a power source with 10000 mAh capacity. When the Internet Service Provider is not available, an alternative is the GSM/GPRS module. It can connect the system to the internet and have the features working together through configuration and integration.

RESEARCH METHOD

I. SYSTEM DESIGN

Figure 1 illustrates the block diagram of the proposed method. The framework of the proposed project is the Raspberry Pi board.

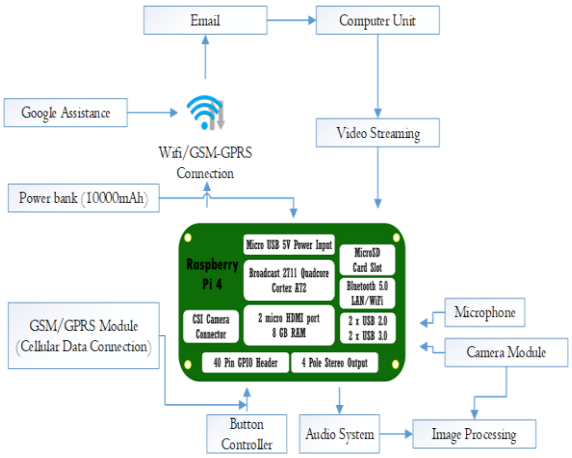

Figure 1. System Block Diagram

The components of the system have undergone individual coding according to their specific task. Raspberry Pi 4 Model B microcomputer served as the data holder and services provider of the Text Reader Glasses with Audio-Visual Call and Google assistance. It is the system's core wherein other components are integrated such as a camera connected to the USB port, GSM/GPRS module, an earphone/mini speaker, a microphone, and a power bank. A camera is necessary to capture the image of text, whereas earphones and mini speakers serve as the channel for the user to hear the spoken text. Moreover, GSM/GPRS module will be used to provide internet through data connection; the power bank will serve as the power source. Linux–Raspbian OS processes the conversions. The audio output is taken from the audio jack of the raspberry pi. The converted speech output is amplified using an audio amplifier. The Internet is connected through the Ethernet port in raspberry pi. The page to be read is placed on a base and the camera is focused to capture the image. The captured image is processed by the OCR software installed on raspberry pi. The captured image is converted to text by the software. The text is converted into speech by the TTS engine. The final output is given to the audio amplifier from which it is connected to the speaker.

This system is controlled through different buttons that are linked to the GPIO, which is the central operator of the system. Each button represents the target functionalities of the system, specifically the camera, text-to-speech, Google assistance, and video conferencing. Proponents undergo in-depth research activities to meet such requirements for the proposed project.

A. RPI SYSTEM

A software framework has been developed to operate the features of the system. Figure 2 illustrates the flow of the RPI system.

Using the controller button, the user can access the different features of the system by manually pressing it. The visually impaired can detect what button to press based on the Braille letters engraved on them. The camera button enables the system to capture and save the image file. The text-to-speech button processes the captured image into text and then converts it into voice output. It also allows the user to replay the text in the current image captured by the system. As for the video conference button, the system delivers video streaming through Jitsi Meet.

However, the system cannot run the programs simultaneously; it requires finishing the first executable program before selecting others. Based on the user's needs, once a button is pressed, the system executes that program and processes each output needed. Once the system finishes executing the user's request, the program will test the system if the Google assistance is terminated, it will allow the user to start it again. Whereas if the user does not want to use Google assistance, the system will return to the choices button. Google assistance will be temporarily unavailable until the user starts it again.

Additionally, the study uses different Application Programming interfaces (API) for the technical assistance of the visually impaired student. API such as Google assistance and Jitsi meet are integrated into the system for online help. This kind of operation requires internet connectivity, which the system also provides through GSM/GPRS data connection and Wi-Fi.

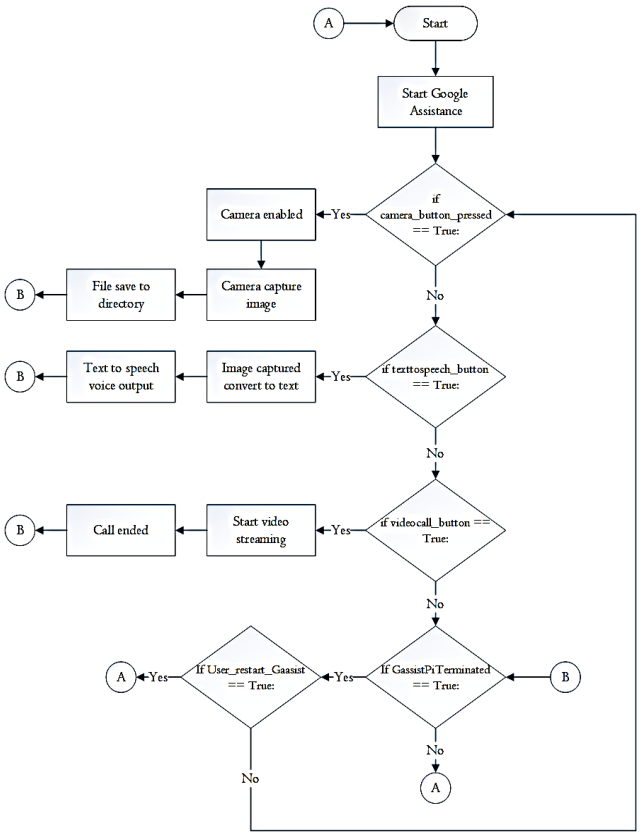

Figure 2. RPI System Flow Chart

B. NETWORK SYSTEM

As shown in Figure 3, the network system flow chart indicates the user's options for internet connectivity to provide the system's features. In this system flow, functionalities such as Google assistance and video streaming require an internet connection to use. Meanwhile, the text-to-speech conversion system is accessible even without an internet linkage. The system equips two sources of internet connectivity. It uses GSM/GPRS module that allows the user to have an internet

connection using a data connection. Also, it uses an automatic Wi-Fi connection based on the saved networks in the system.

With these, the user can maximize the use of the device. The visually impaired will be able to read text anytime and anywhere since the text-to-speech module of the system does not require an internet connection. Through GSM/GPRS module and saved networks in the system, the user can search information over the internet, current time updates, location, weather updates, and YouTube playing music that will enable the visually impaired to adopt the digital world independently. On the other hand, the video conferencing feature of the system is indeed helpful, promoting interactive learning even in distant modalities. With the existence of this functionality, professors can accommodate the issues and concerns of the visually impaired.

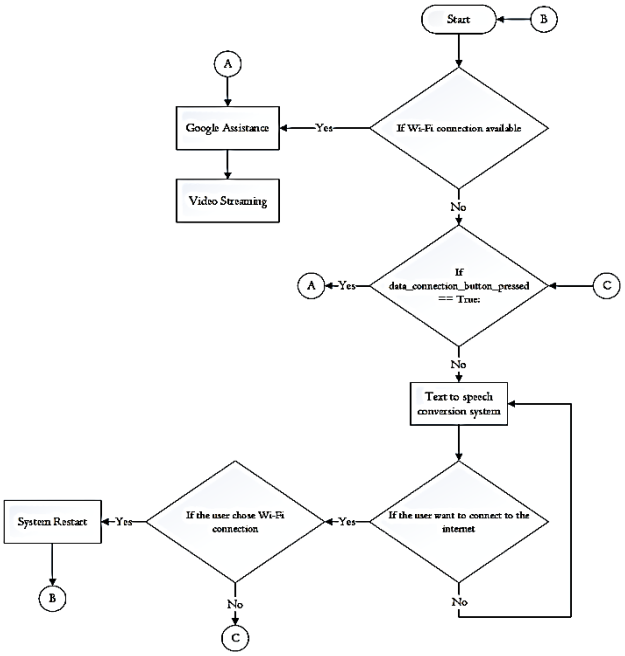

Figure 3. Network System Flow Chart

The RPI System and Network System Flow Chart can function separately; however, both are interconnected to enhance the system's capability in assisting visually impaired students' learning.

C. TEXT IMAGE PROCESSING

The main input requirements need is knowledge of image processing with OpenCV, and Tesseract OCR libraries. OpenCV package is used to read an image and perform particular image processing techniques. Tesseract OCR is a wrapper for Google's Tesseract-OCR Engine used to recognize text from scanned images. In this study, the preferred OS used was the Linux–Raspbian. This operating system is commonly administered for Raspberry Pi-based studies. The system also requires Python language to create instructions to be interpreted and integrated into the hardware modules and perform functions. Figure 2 depicts the system's image processing, which converts the camera-captured image into speech. The system will start with image acquisition and will go through a series of processes before the results of audio output.

C.1 Image Capturing

The first step is in which the device is moved over the printed/digital page and the inbuilt camera captures the images of the text. The system locates the image directory and reads it with OpenCV. When the existence of the file is validated, it performs the pre-processing.

C.2 Pre-Processing

The system will detect the image contour as it contains excessive background noise for the Tesseract to recognize. Moreover, if the image surpasses the file parameter, the system will automatically perform Region of Interest Extraction (ROI). ROI defines the borders of an object under consideration. Upon performing image contour detection and extraction, the system will conduct grayscale and image binarization via thresholding. It will alter the pixels of an image to make it easier for the system to analyze. Finally, noise removal is done pixel by pixel to make the image more precise and recognizable by the Tesseract OCR.

C.3 Segmentation

After pre-processing, the image with no noise moves to the segmentation phase. It is a process that tries to break up a picture of a sequence of characters into a picture of each symbol (characters). Interline spaces are looked for in the binary image. The image is broken up into groups of paragraphs across the distance if there are spaces between lines. The spaces between the lines in the paragraphs are checked to see if they meet in the background. A curve of the actual image is used to find out how wide the horizontal lines are. The lines are then looked at vertically to see if there are any spaces where two lines meet. In this case, pixel intensities determine how wide the words are. Then, character width computation is used to break the words down into their characters.

C.4 Text Extraction

When the image has been enhanced, its features are extracted according to different attributes such as height and width of the character, number of circles present, and number of vertically and horizontally oriented arcs and pixels in the various regions. The text in the enhanced image is extracted through the Tesseract Optical Character Recognition (OCR) software installed in the Raspberry Pi, the speed of text extraction depends on the font style and font size of the characters.

II. IMPLEMENTATION AND UNIT TESTING

This phase tackles the implementation of the designed software architecture, including the software coding. Each component of the system was developed and tested separately, then integrated to produce the whole system. This section emphasized the module testing wherein individual coding of each functionality was examined and modified thoroughly.

III. INTEGRATION AND SYSTEM TESTING

Each unit of the system was combined to produce the final product. The compatibility of the components was examined and evaluated according to the specified functionalities. Also, the system was tested according to its overall capability.

As an individual module of the system successfully operates, this was coordinated into the system. Processes and components underlying in Text-to-speech, Google assistance, and video-conference features of the system were integrated into the device. Other requirements were also administered. Upon developing the system, the compatibility of the functions and components was tested. Modification and adjustment were done to make the system operate simultaneously; the system has undergone testing until it successfully embodies the target outcome of the study.

FINDINGS AND DISCUSSION

A. Prototype

Figure 4 presents the whole system prototype composed of different components integrated to form one device that can function for different tasks. All initialization, data processing, and configuration took place on it.

Figure 4. System Prototype

The figure below demonstrates the system controls of the system prototype. There are two ways to control the system it is either voice command or button control which has engraved letters based on the braille system. Initial setup ports are included at the top of the prototype. The battery level indicator is where it can tell how many hours remaining usage of the device is.

Figure 5. System Start up

Figure 6 shows the initialization upon booting of the device. This is where its functions are initialized and also where confirmations of system requirements start- up. It is an LED and sound-based start-up confirmation.

Figure 6. Control Buttons

B. System Testing

The functions of the system including video conferencing and text recognition were tested. The results are as follows.

Figure 7 depicts the real-time viewing from the text reader glasses through a smartphone or computer. By this, the superuser can communicate in real-time to their professor. The system used the Jitsi Meet as a video conference app integrated into the website (https://aiwear.live). When the superuser activates the video-conferencing feature, an email will be sent to professor to notify them that the video chat has started.

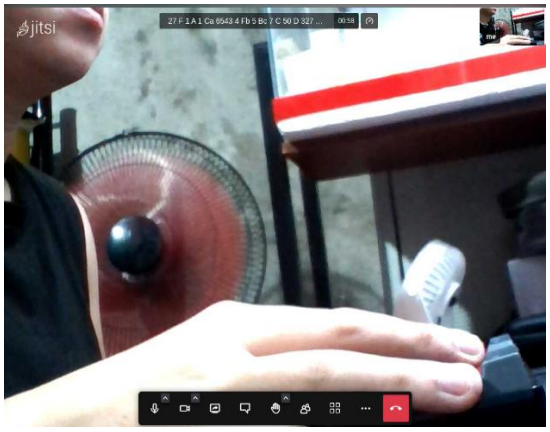

Figure 7. Real-time Video Viewing and Communication

Figure 8 depicts the thresholding and binarization of the captured image of the Pi camera using OpenCV image processing. A bounding box will serve as the pre-processing image marking for the Region of Interest (ROI) extraction. It will change the pixels of an image to make the image easier to analyze and that will be used by the system to find easily the image contour or edges for ROI (Region of Interest) extraction.

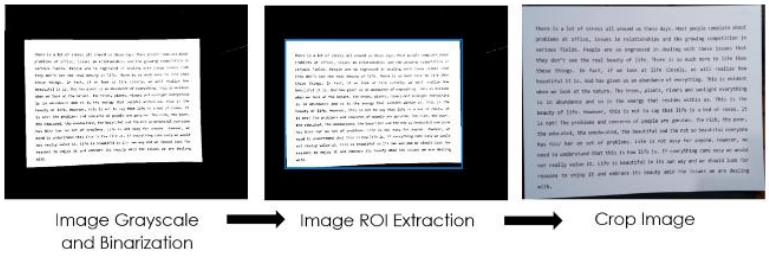

Figure 8. ROI Extraction

Figure 9 shows the pre-processing of the image after the extraction of ROI. After converting the image into grayscale and applying the thresholding it will convert the image into binarize image. Once the system receives the binarized image it will remove the noise from the image before the text extraction through dilation and erosion. The dilation procedure will look for similar pixels that overlap between the structuring element and the binary picture. If there is an overlap, the pixels under the structural element's center position will be set to 1 or black. In contrast, erosion is the

process that determines whether or not there is a complete overlap with the structural element. If there is no complete overlapping, the center pixel indicated by the structural element's center will be set to white or 0.

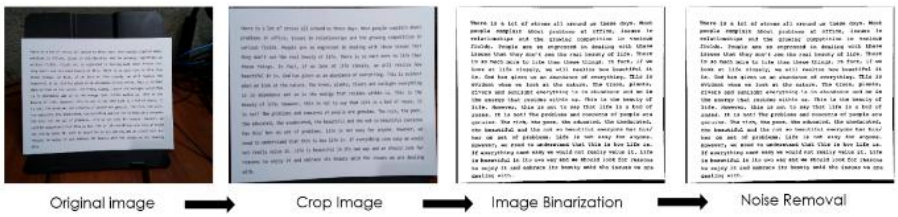

Figure 9. Pre-Processing

The figure below illustrates the most accurate distance during the trials when capturing an image including the height of the table and chair. This may vary depending on the height of the user. The distance of the camera from the text to capture is based on the specification of the camera. The angle used is the standard reading position in which the perspective view of the camera can capture.

B.1. Capturing Distance

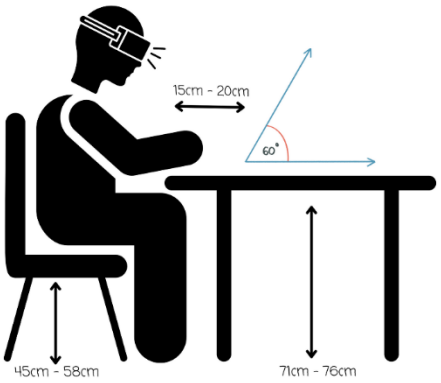

Figure 10. Image Capturing Distance

The data gathered for the accuracy test is based on the system captured and its font characteristics. Times New Roman font is short descenders to allow tight line pacing and a relatively condensed appearance. Arial font is the overall treatment of curves that is softer and fuller than in most industrial-style sans serif faces. Consolas font characters have the same width and also include hanging or lining numerals; slashed, dotted, and normal zeros; and alternative shapes for several lowercase letters. Courier new font has a higher line space than other fonts it is where the punctuation marks were reworked to make the dots and commas heavier. Calibri font has subtle rounded stems and corners, which are visible at larger sizes, creating a warm and soft character.

The accuracy rate shows a high percentage in Courier New which is a monospaced type. It shows an average of 99.44% accuracy rate in words and 99.83% in character. The lowest accuracy shows in Calibri font which average of 98.46% in word and 99.68% in character. In contrast with the average text extraction, Courier New shows a high text extraction process in an average of 7.24 seconds and the lowest text extraction is the Times New Roman with an average of 4.11 seconds. The text extraction process also affects the temperature of the system. The higher the temperature of the system the lower the response rate.

The accuracy rate in the word of the sample text is tested through its pronunciation, which means even if the result shows a capital letter and the actual shows a small letter it is considered

correct on the accuracy test. In line with this, error such as capitalization and punctuation in text extraction is recorded in the accuracy rate of character.

B.2. Run time of System

Table 1 demonstrates the amount of energy that can be extracted from a power bank depends on the battery's condition and the electronic circuitry. The energy efficiency of a power bank is the ratio of the amount of energy supplied at its output port to the amount of energy stored in its battery. The efficiency shows an average of 88.24% of the battery. Compared to the total run time test, the hours recorded are higher in terms of the efficiency rate of the power supply. The power efficiency analysis shows an average total run time of 8 hours and 48 minutes, which is still suitable for the time spent by visually impaired students in learning.

Table 1. Total Run Time of the System

Battery Capacity (mAh)	Battery Voltage (V)	Input Energy (Wh)	Output Energy (Wh)	Efficiency (%)	Run Time Hours Based on the Efficiency (Hrs:Min:Sec)	Total Run Time of the System (Hrs:Min:Sec)
10000	3.7	37	32.65	88.24	8:48	10:09:56

B.3. Decibel Microphone Test

The test was executed through different commands. It was tested using the categorization of without background sound and with background sound. Based on the findings, without background sound was much easier to dictate command since it ranges from 54dB to 57dB. It was equivalent to a normal conversation in a home. While with background sound, the tone of voice needs to increase to recognize the command since it ranges from 61dB to 64dB. This range of decibels is like the conversation in a restaurant. The test for getting the command using voice speech proves that the system is reliable. This will help the user to easily run the device and command the system for the information that the user wants.

C. Project Evaluation

The Text Reader Glasses for Visually Impaired Students using Raspberry Pi with Audio-Visual Call and Google Assistance has gone through several trials to determine the system's functionality suitability, reliability, portability, usability, performance efficiency, security, compatibility, and maintainability. Table 2 presents the summary of the result of the evaluation.

Table 2. Summary of the Result of the Evaluation Procedure

Particulars	Weighted mean	Interpretation
Functionality Suitability	4.07	Very Satisfactory
Reliability	4.33	Very Satisfactory
Portability	4.50	Very Satisfactory
Usability	4.63	Outstanding
Performance Efficiency	4.37	Very Satisfactory
Security	4.20	Very Satisfactory
Compatibility	4.60	Outstanding
Maintainability	4.36	Very Satisfactory
Group Mean	**4.38**	**Very Satisfactory**

CONCLUSION

There are several issues concerning education, particularly for students with special needs. One of these is the inability of the visually impaired to access quality learning opportunities. Since the establishment of traditional education, visually impaired students have faced significant challenges, affecting their self-esteem. Furthermore, the transition to distance learning has amplified their problems requiring them to navigate the digital world. Hence, this calls for attention and innovative solution.

The aforementioned issues were resolved with the introduction of AI-WEAR: Text Reader Glasses for Visually Impaired Students using Raspberry Pi with Audio-Visual Call and Google Assistance. The Text Reader Glasses system, which incorporates Jitsi Meet and Google Assistance API, is intended to provide a high-quality education for visually impaired students. Compared to the existing learning materials or gadgets for visually impaired students, such as the braille system, this device assists them in reading, learning, enjoying online, and interacting with people at a distance. Thus, it will increase their confidence in acquiring independent learning, allowing them to keep up with today's technologically demanding society.

The innovation in this study involves multiple trials and maintenance, which significantly impact the assessment results of the system. The device produces an average 99.44% accuracy rate in font 16 of Courier New which proves its reliability for the information dictated by the system. In contrast, it includes some factors to maintain its accuracy, such as proper lighting and distance of the camera in the text image. With all the system features, the device can run up to 10 hours of usage. Additionally, the test for getting the command using voice speech proves that the system is reliable. This will help the user efficiently run the device and command the system for the information the user wants. Therefore, the innovation of assistive devices in the form of Text Reader Glasses significantly impacts the current situation of visually impaired students and non-

students. It caters to their needs to access learning opportunities by providing features and functions intended to upgrade their educational method.

LIMITATION & FURTHER RESEARCH

The study was limited to Raspberry Pi as a microcomputer, which would be the system's data holder and services provider. The system did not provide object detection and navigation capabilities. An unstable connection can also render the device unusable as it could seek information over the internet, and the device could not operate effectively. Depending on the stability of the connection, the time accuracy of the system's response may vary. The system could only allow one email to be connected to the video conference system, which would receive the notifications. The text detection was limited to colors, font size, font style, and lighting provided by the surroundings.

Moreover, the system did not support the user is unsteady in capturing the text. It was needed to be stable when capturing an image to avoid blurring. Capturing images converted to text was manually operated through a button and a voice command when internet connection is available. In terms of google assistant, the information could be less accurate based on the student's needs, and the words were translated literally into the desired language. Furthermore, Google Assistant and YouTube API were limited to units/queries per day. Google Assistant API only provides 500 units/queries per day, equivalent to 500 times getting information over the internet. YouTube API provided 10000 units/queries that allowed users to play music without exceeding 200 songs.

The development of AI-WEAR: Text Reader Glasses for Visually Impaired Students using Raspberry Pi with Audio-Visual Call and Google Assistance the researchers highly recommend the following:

1. Improve the casing for the weight and size of the device.
2. Improve camera features such as:
    a. HD Lens for more detailed image quality and
    b. Wide Lens for a wide range of image
3. Use notification if the camera has the right angle and distance to capture.
4. Increase the number of word capabilities for text detection.
5. Increase the number of font styles it can recognize.
6. Add detection of a mathematical equation.
7. Use other voice synthesizers for much clearer results in both English and Filipino language.
8. Use other technology rather than GSM Module for cellular data connection.
9. Add features such as Braille system reader, and object and bills detection.

REFERENCES

[1] Sarkar, S. et al. (2021). Smart Reader for Visually Impaired Using Raspberry Pi. IOP Conference Series Materials Science and Engineering 1132(1).

[2] Save the Children Philippines (2021). Impact of COVID-19 to Children with Disabilities. Journal of Special Needs Education.

[3] UNICEF Philippines (2021). The situation of Children with Disabilities in the Context of COVID-19. Results of a rapid online survey in the Philippines. UNICEF Philippines.

[4] Wahab, S. et al. (2020). Text Reader for Visually Impaired Person. Journal of Physics. IOP Publishing. DOI:10.1088/17426596/1755/1/012055